University of Southampton

Cancer Sciences

Faculty of Medicine

QUANTITATIVE PROTEOMICS PROFILING OF B-CELL CANCERS

Harvey Eugene Johnston

Thesis for the Degree of Doctor of Philosophy

September 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE
Cancer Sciences

Doctor of Philosophy

QUANTITATIVE PROTEOMICS PROFILING OF B-CELL CANCERS

Harvey Eugene Johnston

Approximately 800,000 leukaemia and lymphoma cases are diagnosed worldwide each year; with incidence spanning the extremes of both age and economic development. Burkitt's lymphoma (BL) is an aggressive B-cell lymphoid tumour, prevalent in pre-industrialised countries, with the primary sufferers aged between 4 and 7. Chronic lymphocytic leukaemia (CLL) is the most common adult leukaemia in the Western World; typically presenting with a leukaemic population of $CD5^+$ B cells in patients over 60 years of age. CLL is a heterogeneous disease with a variable clinical course which presents challenges in prognosis and treatment. A more comprehensive understanding of CLL and B-cell cancer biology and the identification of disease biomarkers and therapeutic targets has the potential to improve clinical outcomes.

Murine models offer a controlled and accessible means of studying B-cell cancers. Transgenic mice have been developed which use the immunoglobulin heavy chain gene enhancer (Eμ) to impose B-cell specific proto-oncogene expression. Eμ-*myc* and Eμ-*TCL1* mice model aspects of BL and CLL, respectively.

To achieve non-biased, quantitative proteomics profiling of B-cell cancers, this investigation has utilised isobaric labelling, two-dimensional liquid chromatography and mass spectrometry. The proteomes of whole cell lysates of human and mouse B-cell cancers were quantitated against non-cancer B-cell controls to determine global cancer-specific protein expression. A parallel analysis of pre-terminal and terminal mouse plasma was conducted using sub-proteome enrichment and size exclusion chromatography to identify potential disease biomarkers. A method was developed which more effectively utilises quantitative isobaric-labelled data to conclude differences in protein expression.

Proteomic characterisation quantitated 7391 proteins (FDR <1%) for Eμ-*myc* and Eμ-*TCL1* B-cell tumours relative to pre-tumour and WT controls, identifying over 2000 differentially expressed proteins, amongst which were anticipated findings such as myc and TCL1. Common and tumour-specific regulation of pathways, potential targets of inhibition and cell surface proteins were characterised; most notably of which was the interleukin 5 receptor in Eμ-*TCL1* tumours. Treatment with interleukin 5 induced proliferation and survival of Eμ-*TCL1* tumour cells, validating this novel finding. This constitutes one of the most comprehensive characterisations of B-cell cancers to date.

2095 proteins were profiled in Eμ-*myc* and Eμ-*TCL1* plasma identifying tumour lysis products as the major signature in terminal tumours. Additionally signatures of protein secretion, shedding and immune response were present in a tumour-specific manner. An early, pre-terminal signature of tumour development was also identified in the Eμ-*TCL1* model.

Profiling of CLL samples quantified 5956 proteins across 14 samples with findings agreeing with expected differential expression in CLL, relative to healthy B cells; the most comprehensive proteomics characterisation of CLL to date. The results suggested novel targets of immunotherapy and inhibitors, especially in the context of B-cell receptor signalling. Novel biology such as global spliceosome upregulation was also uncovered. CLL subtype-specific differences were identified, however, the strongest signature was that of a subtype-independent pattern of protein expression.

This investigation has provided unparalleled characterisations of B-cell neoplasms suggesting novel biological mechanisms and therapeutic targets, particularly in CLL. This supports the further characterisation of B-cell cancers, and other cancers, by the presented methodologies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

DECLARATION OF AUTHORSHIP

I, Harvey Eugene Johnston declare that this thesis entitled: "Quantitative Proteomics Profiling of B-Cell Cancers" and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. No parts of this work has been published before submission. The following submission has been made of part of this work:

SPIQuE: A Method for Quality-Weighted Isobaric Tag Quantitation
Harvey E. Johnston, Cory H. White, Yawwani Gunawardana, Oliver Bills, Christopher H. Woelk, Spiros D. Garbis *Journal of Proteome Research (Submitted)*

Signed: ……

Date: ……..

ACKNOWLEDGEMENTS

First and foremost, huge thanks are due to Mark Cragg for his supervision throughout this project, his rigorous proofreading of this thesis and for, alongside Paul Townsend, writing the Medical Research Council grant that funded this work. I would also like to thank Paul for securing the funding from Wessex Medical Research that purchased the Mass Spectrometer with which this work was performed. I would also like to thank Roger Allsopp for swimming the channel to help raise this money. I would like to thank Graham Packham and Christopher Sutton for their feedback in the examination of this thesis.

I would additionally like to thank Matt Carter for providing assistance and knowledge relating to the Eμ-*myc* and Eμ-*TCL1* models and for providing phenotype characterisations, Kerry Cox for her assistance with genotyping, all the members of the BRF staff for their assistance with animal work and everyone in Tenovus Research Laboratory for providing various help and good humour in the completion of this work. I am very grateful to Sam Larkin for her assistance and training related to Western blotting, Sarah Bailey for training with cell culture and Kath Woods-Townsend for her work with FFE that enabled this project to be funded. I must also thank Spiro Garbis, Theo Roumeliotis and Antigoni Manousopoulou for training and assistance with proteomics.

Appriciation is due to Cory White for writing the code enabling the implementation of SPIQuE, Oliver Bills for translating this into an online platform, Christopher Woelk for his guidance with the SPIQuE project, Yawwani Gunawardana for conducting a machine learning validation of the SPIQuE approach and Elena Vantaga for assistance with IRIDIS.

I am also very grateful to Marta Larrayoz Ilundain, David Oscier, Matthew Rose-Zerelli, Andy Steele, Jon Strefford and Renata Walewska for their helpful discussions relating to the design and results of the CLL proteomics; especially Renata for providing the CLL samples from Bournemouth Biobank.

I am also extremely grateful for the many people who have made this work possible indirectly, often by helping to maintain some level of sanity, or by preparing me for the test of said sanity that is a PhD. Firstly I would like to thank my Mum, Tina, for providing an eccentric and entertaining upbringing and ensuring that, in a good sense, I will never be normal. I thank her for inspiring tenacity and determination which have helped enormously in this work. Thanks are also due to Grace for putting up with an annoying, though luckily, mostly absent, big brother.

I'd like to say a huge thanks to, his lordship Lady Marcus Pennington, who I am indebted to for allowing me to adopt him, and for imparting graphical design wisdom that helped with the presentation of this thesis. I am very grateful for the support of Caroline and Roland, and Raoul and Pascal for simply being excellent cousins. I would also like to thank Annie and Freya for introducing me to drunken notes and always having a comfy sofa.

I would like to thank RMRK et al. for embracing my et al. status – and, without showing too much favouritism – I'd like to say a huge thanks to Ruth for her friendship, support and tolerance in addition to refraining from obviously falling asleep when I talk about mass spec. I'd also like to thank Ruth for her moral support and for the provision of snacks during the writing of this thesis.

Thanks are due to John 'Juan!' Ferdinand for being brilliantly British and introducing the concepts of reclettes, spatchcocking, overgenerous cloving and 4 am baking, without which, undoubtedly, this thesis would not quite be the same. I'd like to thank James for putting up with me as a housemate, listening to my rants, and for his coding of some of my less successful ideas, Kim; for providing many ridiculous, but excellent, conversations and Lizm; for being Lizm, for her persistent ducking and for being a superb brewery tour buddy. Other thanks must go to Ali Mac and his cheese sandwich story and Florence for his affinity for high velocity condiments. Additional acknowledgement is due to Mark for his almost impressive tolerance for Jägerbombs.

I also have to say a big thanks to all the office people for providing a brilliant atmosphere for the first half of my PhD especially Special (Sam), Mum (Sarah), Tranny/Tryer (Tom), Rage (Nola), Lost (Nunzia), Chav (Charlie), Fussy (Franco), Evil (Helen), Part-timer (Becky) and Captain Lemon. I will miss the days of wrapping every item at Special's desk in clingfilm, mocking c-chops, poking the bear and Tranny's persistence to end every night out in a dress.

Sofia, James and Diana definitely deserve plenty of credit for their good humour in the face of challenging situations. I would like to thank Mark Coldwell for helping to solidify my desire to go into research and David O'Connor for his kind advice and support during my BSc and for my PhD applications.

ABBREVIATIONS

2D	Two-dimensional
aa	Amino acid
ABC	ATP-binding cassette
ACN	Acetonitrile
AK1	Adenylate kinase 1
AKT	Protein kinase B
AP-1	Activator protein 1
APF	Antiproliferative factor
Arf	Alternative reading frame
ASS	Argininosuccinate synthase
ATM	Ataxia telangiectasia mutated
BCA	Bicinchoninic acid assay
BCL2	B-cell lymphoma 2 protein
BCR	B-cell receptor
BL	Burkitt's lymphoma
BLNK	B-cell linker protein
BSA	Bovine serum albumin
BTK	Bruton's tyrosine kinase
CAM	Cell adhesion molecule
Capg	Macrophage capping protein
CD	Cluster of differentiation
CDC	Cell division cycle protein
CDK	Cyclin-dependent kinase
CFSE	Carboxyfluorescein succinimidyl ester
CID	Collision-induced dissociation
CKAP4	Cytoskeleton-associated protein 4
CLEC17A	C-type lectin domain family 17, member A
CLL	Chronic lymphocytic leukaemia
CMIP	C-Maf-inducing protein
Cor1A	Coronin 1A
CXCR	Chemokine (C-X-C motif) receptor
DAVID	The database for annotation, visualisation and integrated discovery
DDA	Data-dependent acquision
DLBCL	Diffuse large B-cell lymphoma

DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulphoxide
DNMT	DNA methyltransferase
DOC	Deoxycholic acid
DTT	Dithiothreitol
Eμ	Enhancer of μ immunoglobulin
EBV	Epstein-Bar virus
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epithelial growth factor receptor
Eno1	Alpha-enolase
ERK	Extracellular-signal-regulated kinases
ESI	Electrospray ionisation
FA	Formic acid
FACS	Fluorescents assisted cell sorting
FBS	Foetal bovine serum
FBW7	F-box and WD repeat domain-containing 7
Fc	Ig fragment crystallizable region
FCRL	Fc receptor-like
FDR	False discovery rate
FFE	Free flow electrophoresis
FGF2	Fibroblast growth factor 2
FITC	Fluorescein isothiocyanate
FL	Follicular lymphoma
FL (1/2/3)	Fluorescents channel
FSC	Forward scatter
FTMS	Fourier Transform mass spectrometry
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GO	Gene ontology
GSK3β	Glycogen synthase kinase 3 beta
HCD	High-energy collisional dissociation
HD	Healthy donor B-cell
HDAC	Histone deacetylase
HDACi	Histone deacetylase inhibitors
HLA	Human leukocyte antigen
HMOX	Heme oxygenase

Hsp90	Heat shock protein 90
ICAT	Isotope-coded affinity tags
ICPL	Isotope-coded protein labels
ID	Inner diameter
Ig	Immunoglobulin
IgePalCA^{630}	Octylphenoxypolyethoxyethanol
IGHV	Ig heavy-chain variable-region gene
IL16	Interleukin 16
IL5	Interleukin 5
IL5RA	Interleukin 5 receptor alpha
INPP5	Inositol polyphosphate-5-phosphatase
IPA	Ingenuity pathway analysis
IQSEC1	IQ motif and SEC7 domain-containing protein 1
ITAM	Immunoreceptor tyrosine-based activation motifs
ITG	Integrin
ITIH	Inter-alpha-trypsin inhibitor
iTRAQ	Isobaric tags for relative and absolute quantitation
KCN	Potassium channel
KDM4B	Lysine-specific demethylase 4B
KIF	Kinesin motor proteins
LAX1	Lymphocyte transmembrane adapter 1
LC	Liquid chromatography
Lcp1	Plastin-2
LGALS1	Galectin-1
LIT	Linear ion trap
LY6D	Lymphocyte antigen 6D
Lyn	Lck-yes-related novel kinase
m/z	Mass to charge ratio
MACS	Magnetic assisted cell sorting
MAPK	Mitogen activated protein kinases
MBL	Monoclonal B-cell lymphocytosis
MCL	Mantel cell lymphoma
Mcl-1	Induced myeloid leukemia cell differentiation protein
M-CLL	Mutated CLL
MCM	Mini-chromosome maintenance

Mdh2	Malate dehydrogenase, mitochondrial
MDM2	Murine double minute
MHC	Major histocompatibility complex
miR	Micro RNA
MMTS	Methyl methanethiosulfonate
Moe	Moesin
MP	Mobile phase
MS	Mass spectrometry
MS/MS (MS2)	Tandem mass spectrometry
M_w	Molecular weight
MWCO	Molecular weight cut-off
myc	Cellular myelocytomatosis homologue
MYD88	Myeloid differentiation primary response gene 88
Myo9	Myosin-9
MZB1	Marginal zone B- and B1-cell-specific protein
NFAT	Nuclear factor of activated T cells
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NOTCH1	Neurogenic locus notch homolog protein 1
Npm	Nucleophosmin
O/N	Over night
OCT	Optimal cutting temperature
OD	Optical density
p300	300 kda Histone acetyl transferase
p53	Tumour protein 53 gene product
PAGE	Polyacrylamide gel electrophoresis
PARP	Poly [ADP-ribose] polymerase
PBMC	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PBST	PBS tween
PCR	Polymerase chain reaction
PDGF	Platelet-derived growth factor
PE	Phycoerythrin
PEP	Posterior error probability
PerCP	Peridinin chlorophyll
PI	Propidium iodide

PI3K	Phosphatidylinositol-3-kinase
PIGR	Polymeric Ig receptor
PKCβ	Protein kinase C beta
PRMT1	Arginine methyltransferase 1
PSM	Petide-spectrum match
PTEN	Phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase
PTPN	Tyrosine-protein phosphatase non-receptor
PVDF	Polyvinylidene fluoride
q	False discovery
RBC	Red blood cell
ROR1	Receptor tyrosine kinase–like orphan receptor 1
ROS	Reactive oxygen species
RP	Reverse-phase
RPMI	Roswell Park Memorial Institute medium
RRAS2	Ras-related protein R-Ras2
RRM	Ribonucleotide reductase M
RS	Regulation score
RT	Room temperature
RTK	Receptor tyrosine kinase
SATB1	Special AT-rich sequence-binding protein 1
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SF3B1	Splicing factor 3B subunit 1
SHM	Somatic hypermutation
SHP	Src homology region 2 domain-containing phosphatase
Shp1	Tyrosine-protein phosphatase non-receptor type 6
SILAC	Stable isotope labelling by amino acids in cell culture
SIRT	NAD-dependent protein deacylase sirtuin
SLC	Solute carrier
SPIQuE	Statistical Processing for Isobaric Quantitation Evaluation
SSC	Side scatter
STAT	Signal transducer and activator of transcription
SuPrE	Sub-proteome enrichment
TAE	TRIS, acetate, EDTA

TCB	Terminal cardiac bleed
TCEP	Tris(2-carboxyethyl)phosphine
TCL1	T-cell leukaemia/lymphoma 1
TCR	T-cell receptor
TE	Tris, edta
TEAB	Triethylammonium bicarbonate
TGF	Transforming growth factor
TLR	Toll-like receptor
TMT	Tandem mass tag
TP53	Tumour protein 53kda gene
TRIS	Tris(hydroxymethyl)aminomethane
TVCB	Terminal vena cava bleed
U-CLL	Unmutated CLL
VDAC	Voltage-dependent anion-selective channel
WB	Western blotting
WEE1	Wee1-like protein kinase
WT	Wildtype
ZAP70	Zeta-associated protein of 70 kda
ZIP	Zinc transporter Zrt- and Irt-like protein

1.0 INTRODUCTION

1.1 HALLMARKS OF CANCER

In complex multicellular organisms, it is vital that every cell functions to the benefit of the whole organism. Cells must proliferate when required and undergo cell death, should the cell's role become superfluous or detrimental. Cancer can be thought of as the reversal of these two obligations; when apoptosis is evaded and proliferation becomes continuous and unregulated. The result is cells that can be considered 'autonomous', where their function within the organism is no longer fulfilled and their presence can rapidly become harmful to other cells' functions, potentially leading to the death of the host.

Cancer can be defined by a distinct set of characteristics that contribute to cellular homeostasis dysregulation. These characteristics include cellular proliferation, immortalisation, immune evasion, tumour suppressor dysregulation, angiogenesis, metastasis and reduced apoptosis shown in detail in **Figure 1.1** [1, 2]. Each of these individual characteristics may be relatively harmless, however, the acquisition of multiple traits can cause the selective evolution of a clonal population of cells exhibiting all, or the majority, of these hallmarks [1, 2].

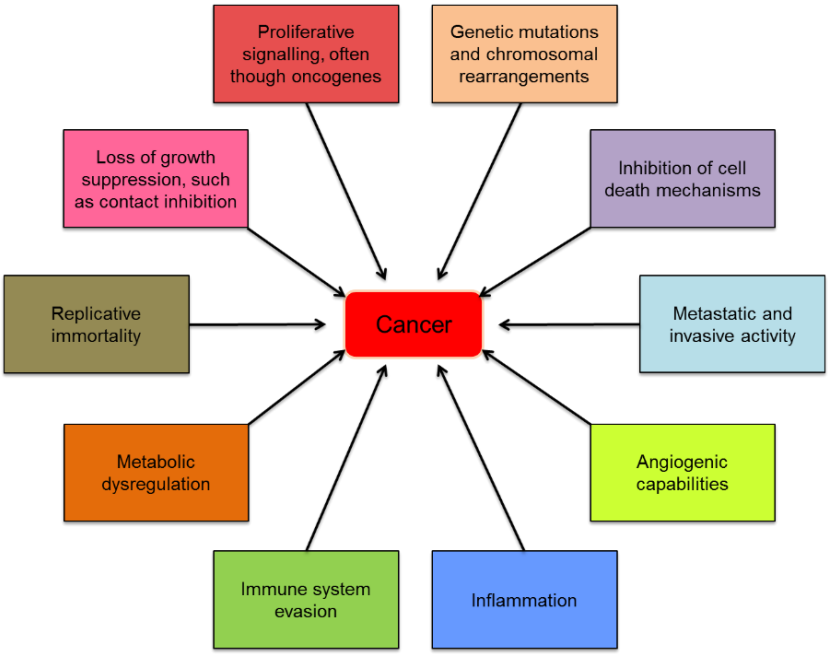


Figure 1.1. **Hallmarks of cancer.** Properties attributable to tumour development and growth; adapted from [1, 2].

1.1.1 THE MOLECULAR BASIS OF CANCER

The primary source of tumourigenesis is thought to be the accumulation of alterations at the genetic level. Aberrations to the genome influencing cancer can include single nucleotide changes, chromosomal translocations and gene duplications or deletions [3-5]. Such mutations can arise as a result of either the intrinsic infidelity of DNA replication and repair, or through environmental induction, such as chemicals, viruses and radiation [6]. Mutations can also be hereditary, passed from the parental genome to the offspring, presenting an increased risk factor [7].

The vast majority of mutations will have no effect on a cell's phenotype, but with cells and their progeny potentially living for many decades, an attribute exacerbated by longer life expectancy, the effect of acquired mutations is cumulative (**Figure 1.2**) [8]. Certain mutations can occur which affect the fidelity of DNA replication or stability of the genome and thus accelerate the acquisition of further mutations [9]. Aberrations to the epigenome also have the potential to harmfully alter gene expression, affecting chromatin structure, transcription factor activity and phenotype [10].

The emergence of cancer hallmarks can ultimately be attributed to the effects, directly or indirectly, of protein abundance or function [11]. These can be brought about through alterations to the protein coding sequence, enhancing or disrupting the physical and functional properties of a protein. Mutations in the regulatory region of genes can alter the levels of gene transcription,

and in turn, up- or down-regulate the translation of proteins. Aberrations to the pathways responsible for messenger RNA (mRNA) or protein degradation can also adversely influence protein abundance. The downstream effects of these proteomic changes ultimately determine the phenotypic characteristics, or hallmarks, of a cancerous or pre-cancerous cell [11].

The most prevalent mutations, which influence proteins involved in proliferation or survival, are most likely to induce avoidance of cell death and increase the cell number with such traits. This is the basis of clonal expansion and the evolution of cancer (**Figure 1.2**) [1, 2, 12, 13]. Proteins with such effects are typically derived from 'proto-oncogenes'.

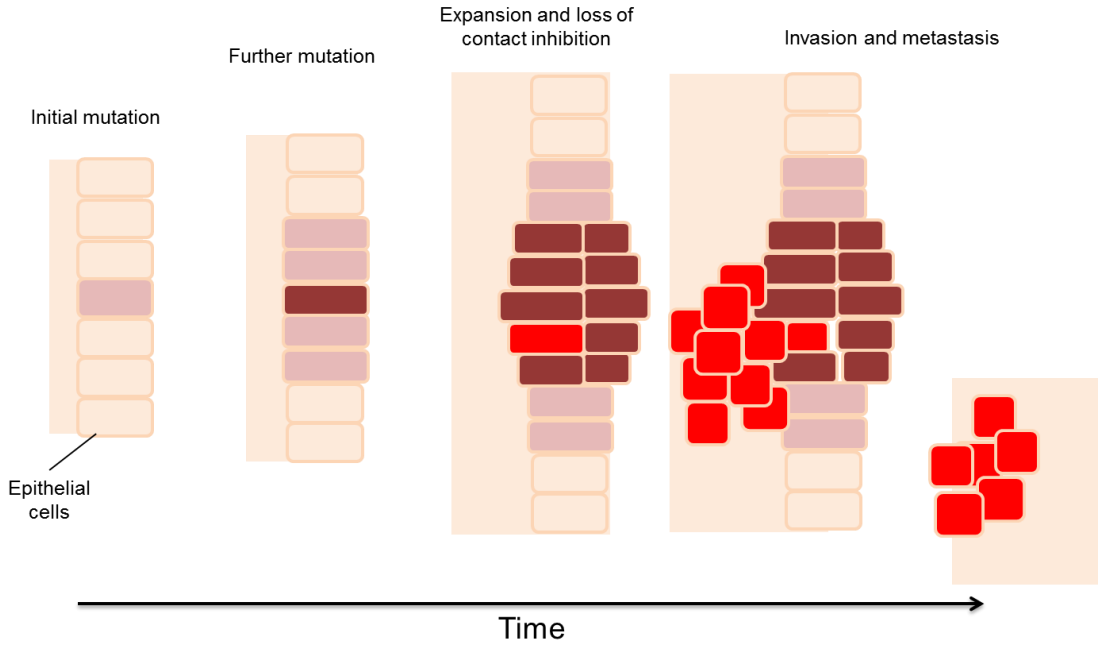


Figure 1.2. Acquisition of cancer traits. The acquisition of mutations and clonal expansion from a single epithelial cell to a metastatic and invasive cancer over time.

1.1.2 Oncogenes and Proliferation

The default reaction of a mammalian cell to complete isolation is to undergo apoptosis [14]. For a cell to maintain a typical state of function, whether proliferating or quiescent, signals from its environment are required to keep it alive. These signals can originate from the extracellular matrix, from adjacent cells or from soluble factors, such as those in the blood or in interstitial fluids. In many cases signalling from all three are required. Often, this relies on transmembrane proteins, which have specific interactions with survival or growth signals on the cell's exterior and transmit this information to the interior of the cell. This initiates downstream signalling which induces protein expression causing a cell to grow, proliferate or remain in a quiescent state [14].

Signals that induce cell growth and proliferation are termed ‘mitogenic’. Mitogenic signals act to upregulate gene expression involved in cell cycle progression and DNA replication so that cells can divide and proliferate. Aberrations to these mitogenic pathways form the basis of increased proliferation in cancerous cells; where pathways that function to promote healthy growth, become incorrectly constitutive. Genes with the potential to express proteins with such an effect are termed ‘proto-oncogenes’. Proto-oncogenes are a set of genes that have a functional role under typical conditions but upon dysregulation, are capable of inducing one or more neoplastic traits [1, 15]. Under such conditions, proto-oncogenes are termed ‘oncogenes’. For a proto-oncogene to induce a malignant state, mutations or dysregulation must occur which either result in an overexpression or alteration to the primary sequence of the proto-oncoprotein [1, 2]. Oncogene activity can be derived from transcription factors; such as myc (described in **Section 1.4**), protein kinases; such as Raf and Src, kinase co-activators; such as TCL1 (described in **Section 1.5**), signal transducers; such as Ras, mitogens; such as platelet-derived growth factor (PDGF) and mitogen receptors such as PDGF receptor or epidermal growth factor receptor [1, 2].

1.1.3 TUMOUR SUPPRESSION

It seems somewhat counter intuitive to the principles of evolution, both at cellular and organism level, that almost all of more than a trillion nucleated cells in the human body has an inherent probability of leading to death through cancer. While cancer is a leading cause of death, it most frequently arises in the elderly [16]. Murine models, for example, which overexpress oncogenes can live for months without tumour development [17]. Together, these observations demonstrate that oncogenesis is not entirely dictated by the action of aberrant oncogenes.

It is clear that mechanisms must exist to prevent single mutations from independently producing malignant neoplasms. The group of proteins responsible for these anti-cancer mechanisms are termed ‘tumour suppressors’. Tumour suppressors have numerous roles in detecting, reversing and compensating for the tumorigenic effects of oncogenes and oncogenic stresses [18, 19]. Many tumour suppressors have critical roles in providing negative feedback to cell proliferation signalling, or inducing senescence or apoptosis, should oncogenes activity or abundance become constitutive or excessive. Proteins such as p53 and ARF (described in **Section 4.1.2**), the retinoblastoma protein, Rb, and Phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase (PTEN), all demonstrate protective roles against the action of oncogene in the development of cancers. Loss of tumour suppressor function is a common component of the hallmarks of cancer [1].

1.1.4 TUMOUR ORIGINS AND SUBTYPES

Different cell types demonstrate different propensity for tumour development and the same oncogene or inactivated tumour suppressor can have vastly different effects on different tissues. This has been demonstrated by the use of knock-out mouse models which often demonstrate a specific tumour arising from the loss of each oncogene (detailed in **Section 1.3**). For example, p53 knock-out in mice results predominantly in sarcomas and lymphomas, whereas Rb knock-out induces brain and pituitary cancers [20, 21]. Additionally the same cell type, acquiring the activation of different oncogenes can result in different cancer phenotypes [22]. In healthy individuals, as opposed to mutant mouse models, cancer can be traced to a single cell acquiring an activating or inactivating mutation of a proto-oncogene or tumour suppressor, respectively (**Figure 1.2**).

A translocation event creating an oncogenic fusion protein, BCR/ABL, in chronic myeloid leukaemia, for example, was identified in a pool of non-leukaemic cells alongside their cancerous counterparts. This suggested that cancer arose from only a subset of these cells, indicating a clonal cancer origin [23]. This and other examples highlight an increased frequency of cells of origin being from undifferentiated progenitor cells, likely owing to their already activated capacity for self-renewal and proliferation [1, 2, 22].

One such cell type which presents a remarkable variation of tumour types is B cells, varying from rapidly proliferating solid tumours, to slow-developing chronic leukaemias.

1.2 B-CELL CANCERS

Cancers arising from B-cell have a high diversity of clinical presentation; ranging across a growing number of sub-classifications, derived from many cell types at many stages of differentiation. B-cell cancers can present as lymphomas, where the cancer localises to lymph nodes or a leukaemia where abnormal numbers of cancers cells enter the blood. At present, thirty different classifications exist for lymphoma alone, a number that has doubled in the last decade [24, 25]. The consensus from these classifications is that lymphoid neoplasia present an overlapping spectrum of phenotypes, oncogenic mechanisms and treatment responses [24, 25]. This presents a challenge to clinical approaches made to diagnose and treat patients of these diseases. B-cell-derived cancers make up the 95% of lymphoma cases and most frequently arise from B cells in germinal centres [26]. Common lymphoma subtypes include mantel cell lymphoma (MCL), follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL) and Burkitt's lymphoma (BL). B-cell cancers also make up a large proportion of leukaemia cases, predominantly in the form of chronic lymphocytic leukaemia (CLL).

1.2.1 B-CELL DEVELOPMENT AND SIGNALLING

'B' cells, termed due to their origination from the bursa of fabricus in birds, are a principal cell of the adaptive immune system which produce immunoglobulins (Ig) to recognise and ultimately eliminate non-self antigens [27]. In mammals, B cells differentiate from hematopoietic stem cells within the bone marrow (**Figure 1.3**) and are defined by cell surface expression of proteins such as CD19, CD20, CD23, CD45/B220, CD79a and CD79b [28]. The process of B-cell development involves the V(D)J (variable, diversity and joining) recombination of the regions coding for the Ig genes. This generates the vast diversity of Ig molecules required for the recognition, binding and elimination of previously unencountered pathogens. Each naïve B cell expresses a unique Ig at its surface in a complex with CD79a and CD79b, termed the B-cell receptor (BCR). The role of BCR signalling is diverse, with roles in proliferation, apoptosis and cell survival at different stages of B-cell development. During B-cell selection, for example, it is essential that any B cells with Ig capable of self-recognition are not allowed to develop. BCR signalling can therefore induce receptor editing, anergy and clonal deletion via apoptosis to prevent this. In mature B cells, BCR engagement with a cognate non-self antigen triggers BCR signalling thereby selectively inducing an expansion of the B cell encoding an Ig with antigen specificity. BCR signalling upon antigen engagement involves the recruitment of several tyrosine kinases, which activate downstream signalling (**Figure 1.4**). This involves pathways such as phosphatidylinositol-3-kinase (PI3K)/AKT, nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), Ras and mitogen activated protein kinases (MAPK). Signalling triggers cytoskeletal rearrangement, BCR internalisation, receptor editing and differentiation into memory B cells and plasma cells; which no longer express the BCR and instead secrete Ig [28].

The propensity for rapid expansion in B cells makes dysregulation of the BCR pathway, and other downstream effectors, such as the myc proto-oncogene, a potential sources of aberrant growth that may lead to cancer.

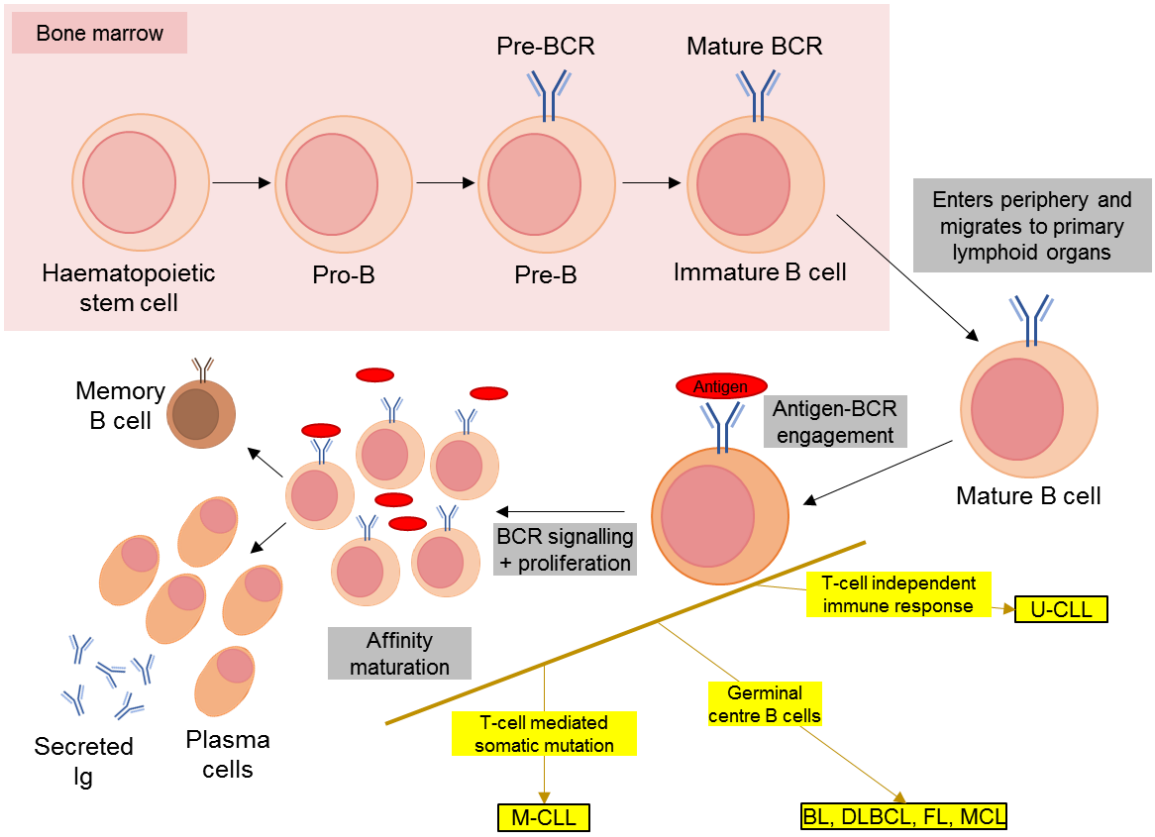


Figure 1.3. Schematic of B-cell development and tumourigenesis. B cells differentiate through several stages from haematopoietic stem cells defined by the V/(D)/J recombination status of the heavy and light Ig chains. Mature (naïve) B cells migrate to primary lymphoid organs and upon antigen stimulation proliferate and clonally expand while undergoing somatic mutation to evolutionarily increase the affinity of Ig to the non-self antigen. B cells presenting high-affinity epitopes differentiate to produce plasma cells capable of secreting the high-affinity Ig and memory B cells. Cancers can arise from many of these B-cell subtypes, but most commonly from B cells in germinal centre, especially lymphomas. CLL is thought to arise from a subset of mature B cells which express CD5. Two subtypes of CLL, mutated-CLL (M-CLL) and unmutated CLL (U-CLL) are thought to arise from $CD27^{+}$ and $CD27^{-}$ negative subsets of $CD5^{+}$ B cells, respectively. The involvement of T cells in the B cells immune response mediates somatic mutation of the *IGHV* gene [28, 29].

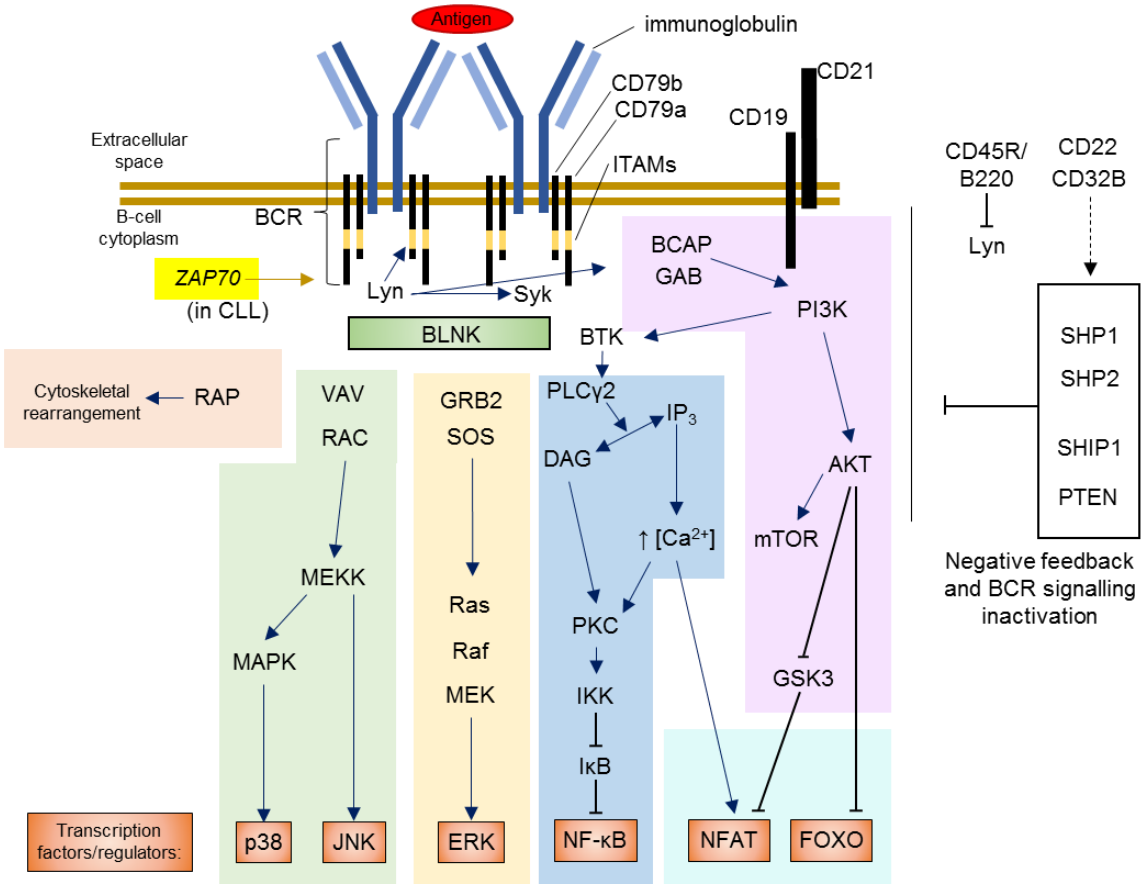


Figure 1.4. The B-cell receptor signalling pathways. The B-cell receptor (BCR) consists of three main components; CD79a, CD79b and an Ig molecule. Upon antigen binding, Lyn (Lck-yes-related novel kinase) phosphorylates Syk and the immunoreceptor tyrosine-based activation motifs (ITAMs) on CD79a and CD79B. Phospho-ITAMs play a role in the localisation of further proteins including the B-cell linker protein (BLNK), which upon phosphorylation localises further proteins which are subsequently activated and promote downstream signalling. B-cell survival and proliferation are driven by transcription factors and regulators via protein such as MAPK p38, extracellular-signal-regulated kinases (ERK), nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), nuclear factor of activated T cells (NFAT) and forkhead box O proteins. Proteins such as the Ras-related protein regulate alterations in cytoskeletal structure. Additional pathways exist which act to enable endocytosis of the antigen-BCR complex. Under cancerous conditions, zeta chain-associated kinase 70 (ZAP70) can be involved in BCR signalling. The BCR signalling pathway is also negatively regulated predominantly by phosphatases such as the Src homology region 2 domain-containing phosphatases (SHP); SHP1 and SHP2; and phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase (PTEN); reversing the activation of AKT induced by PI3K. Negative regulation is modulated by CD22, CD45/B220 and the inhibitory Fc receptor CD32B.

1.2.2 BURKITT'S LYMPHOMA

Despite being amongst the most proliferative human cancers, BL, is highly treatable and, in developed countries, has a cure rate of approximately 90% [30, 31]. BL is scarce in the western world, making up just 2% of lymphomas in the UK. In equatorial Africa, however, incidence is vastly higher, with BL accounting for 74% of all childhood cancers occurring most frequently between 4 and 7 years of age [32, 33].

Of the sporadic form of the disease, observed in developed countries, the causes and molecular mechanisms driving the disease are well characterised. BL results from a genetic abnormality in which genes are aberrantly rearranged between different chromosomes, termed chromosomal translocations). As many as three quarters of cases are attributable to a translocation of t(8;14) which brings the genetic enhancer element of the μ Ig heavy chain (Eμ) to drive the expression of the proto-oncogene *myc* [34, 35]. Other translocations observed in BL include t(2;8) (p12;q24) and t(8;22) (q24;q11); all involving *myc* gene expression amplification [34]. These translocations mimic the propensity for antibody production B cells in the expression of *myc* affecting predominantly either germinal centre B-cell or memory B-cell origin [36, 37]. When tumour suppressor mechanisms fail to dampen this excessive *myc* function, proliferation and neoplastic transformation occurs. This can result in large tumours with a cell replication rate of approximately 24 hours [31].

Another classification of BL is that of endemic BL, observed in regions coinciding with a prevalence of Epstein-Bar virus (EBV) infection [38]. EBV induces B-cell proliferation and mutation that substantially increase the chances of a translocation event triggering overexpression of myc [39]. EBV also acts to inhibit the tumour suppressive mechanisms that would otherwise see the apoptosis of the initial tumour cells [40]. A similar effect is seen with human immunodeficiency virus infections, termed HIV-associated BL and malaria [41, 42].

The treatment regimen for BL is defined as using multiple rounds of aggressive chemotherapy. The high proliferation rate of BL contributes to the efficacy of the treatments which interfere with DNA replication, giving a very high treatment response rate [31]. Recent studies have shown that rituximab therapy is also beneficial in BL therapy when used in combination with this chemotherapy regimen [33]. While BL responds very well to current treatments, there are still issues with treatment resistance and toxicity that suggests a greater understanding of the disease would potentially allow improved clinical approaches [33].

1.2.3 CHRONIC LYMPHOCYTIC LEUKAEMIA

Chronic lymphocytic leukaemia (CLL) is the most common adult leukaemia of the western world, constituting around 1% of all cancers diagnosed in the UK [16, 43]. It is a disease with two very different potential outcomes; an indolent form which can persist for more than a decade and is often not a direct cause of mortality, and a progressive form which can quickly develop into a life threatening disease within eighteen months [44, 45]. CLL predominantly effects the elderly and has a prevalence of 2:1 in male patients [43]. Given a lack of clear environmental risk factors, an increased risk of CLL amongst relatives of sufferers, and the higher incidence of CLL development in Caucasian populations [46], there is almost certainly a strong genetic component to CLL development, yet to be fully understood.

1.2.3.1 DIAGNOSIS OF CLL

CLL is defined as a leukaemic, lymphocytic lymphoma of the peripheral blood, lymph nodes, spleen and bone marrow, presenting with a persistent (>3 months) clonal, $CD5^+$, $CD19^+$ B-cell count, above that of $5x10^9$ cells per litre of blood [47, 48]. CD23 and CD20 are also typically expressed.

A possible precursor state, which does not always progress to CLL, is monoclonal B-cell lymphocytosis (MBL) which is defined as an asymptomatic presentation with CLL-like cells below 5000 cells/µl [50, 51]. MBL represents part of a continuous spectrum, which prognosticates CLL as a potential outcome [52, 53].

CLL or MBL are frequently diagnosed coincidentally when a blood sample is taken for other reasons, but can also present with symptoms such as fever, weight loss, night sweats and enlarged lymph nodes [49].

1.2.3.2 TREATMENT OF CLL

Counterintuitively, compared with most other cancers, MBL and early CLL are left untreated due to the high degree of uncertainty surrounding progression. Watchful waiting for clinical characteristics of advanced CLL, such as splenomegaly, lymphocyte doubling time and anaemia/thrombocytopenia, is currently used to assess the appropriate time to begin treatment [47], based upon the Rai/Binet staging system [44, 54]. Early treatment potentially risks selecting for the most aggressive CLL cells which can then occupy the therapy-induced niches.

Treatment of CLL remains predominantly based around cytotoxic chemotherapy and radiotherapy; acting to disrupt DNA synthesis in proliferating cells inducing apoptosis with drugs such as cyclophosphamide and fludarabine. Immunotherapy is widely regarded as a safer treatment strategy, as well as superior when used in combination with chemotherapy. The use of antibodies such as rituximab and alembtuzumab specifically target the B-cell antigens CD20 and

CD52, respectively, and induce a lethal immune response against the cells expressing these proteins.

CLL is frequently treated with rituximab, a monoclonal antibody targeted at the B-cell specific marker CD20, coupled with the chemotherapeutic adjuvants cyclophosphamide and fludarabine [55]. Survival rates are significantly impacted by refractory CLL, in which CLL cells become resistant to treatment, requiring further, more effective first-line treatments [56].

More recent therapies have focused on targeting more specific intracellular targets with small molecular inhibitors. The BCR, for example is known to drive CLL. Inhibiting components of BCR signalling have proven highly effective in CLL treatment such as the inhibition of Bruton's tyrosine kinase with ibrutinib [57] and the inhibition of PI3K δ with idelalisib [58]. BCL2, an anti-apoptotic protein commonly overexpressed in CLL and other B-cell malignancies, has also been the focus of inhibition therapy. BCL2 inhibitors, such as obatoclax, navitoclax/ABT263, ABT737and ABT199 [59-61], demonstrate a re-establishment and induction of apoptosis.

1.2.3.3 CLL BIOLOGY

CLL has previously been described as a disease of accumulation, rather than proliferation, with circulatory CLL cells predominantly in G_0 phase of the cell cycle, simultaneously overexpressing anti-apoptotic proteins [62-64]. An *in vivo* study measuring cell turnover with deuterium labelling, demonstrated a more dynamic picture of continuous cell production and cell death [65]. The microenvironment plays a critical role in these dynamics in CLL, with the vast majority of cell division occurring in the secondary lymphoid tissue, induced by chemokines, interleukins, integrin signalling and accessory cell contact [66-68].

1.2.3.4 PHENOTYPES AND ORIGIN OF CLL

CLL cells have an immunophenotype of the B-cell antigens CD19 and CD23 with the co-expression of CD5, a marker expressed predominantly on T cells [48]. A downregulated expression of CD20 and other B-cell traits are also characteristic [69]. Many variable characteristics exist amongst cases of CLL, such as the somatic mutational status of the Ig heavy-chain variable-region gene (IGHV). CLL with mutated IGHV (M-CLL) is typically indolent, whereas unmutated cases (U-CLL) often result in a progressive disease [70].

Several studies have attempted to conclude the precursor cell type from which CLL arises, producing conflicting evidence [71-74]. Early microarray characterisations suggested that gene expression signatures were most analogous to memory B cells [72]. Later suggestions were made that marginal zone B cells were a likely origin of CLL, on the basis of the type of BCRs observed in CLL [75-77]. An additional suggestion was that a B-cell subset expressing ZAP70, a

commonly dysregulated protein in CLL might represent a precursor to CLL [78]. Further investigation has indicated CLL development from two $CD5^+$ B-cell subtypes; $CD5^+CD27^+$ and $CD5^+CD27^-$ which may correlate with the mutated and unmutated status of the IGHV in CLL cases, respectively [74]. Given the expression of CD5 and comparative $CD5^+$ B cells gene expression profiles, it has also been proposed that this B-cell subset is the most likely cell of CLL origin [29].

1.2.3.5 MOLECULAR AETIOLOGY

Unlike with BL, where overexpression of myc is a clear fundamental cause of the disease, the molecular basis of CLL is multifaceted. While several subcategories have been proposed, defined by cytogenetic abnormalities, gene mutations, IGHV status and immunophenotypes, CLL can still develop in the absence of these features. Additionally, while many of these features confer an increased risk of progressive CLL, none can definitively predict it.

1.2.3.6 CYTOGENETIC ABNORMALITIES

CLL is observed with chromosome aberrations in more than 50% of cases [79-82]. These present opportunities to identify genes and subsequently proteins which confer a significant influence on the development or progression of CLL. The most frequently recurrent genetic lesion is the loss of 13q14, relates to the loss of the miRNAs miR15 and miR16, inhibitors of the expression of the anti-apoptotic protein BCL2 [83]. 17p contains the *TP53* gene which encodes the tumour suppressors, p53. Loss of this region is a strong predictor of short survival time [84]. 11q deletions are similarly associated with aggressive CLL, inducing the loss of genes encoding the DNA damage sensor ATM and the pleiotropic regulator BIRC3 [85]. The acquisition of a third copy of chromosome 12 termed ‘trisomy 12’ is perhaps the least understood chromosomal abnormality [86, 87]. Trisomy 12 status is related to an increased expression of integrins, suggested to be via epigenetic mechanisms, however, no regulators have been proposed [88, 89].

1.2.3.7 GENE MUTATIONS

Several genes are reported as frequently mutated in CLL, indicating the potential functional importance of the encoded proteins. Indeed, of the top mutated genes in CLL, most have a clear function in the pathogenesis of the disease [90]. Alongside the loss of *TP53* due to 17p deletion, point mutations of *TP53* is a common observation in CLL, which confers a similar, highly increased disease aggression [91]. *NOTCH1* encodes a transmembrane receptor and transcriptional regulator ‘neurogenic locus notch homolog protein 1’ (NOTCH1) that normally influences cell fate determination upon binding ligands such as Jagged1/2. Upon activation, its

effects are pleiotropic, regulating pathways involved in proliferation, apoptosis and angiogenesis [92]. Approximately 10% of CLL cases have *NOTCH1* mutations which consistently act to increase both the activity and abundance of NOTCH1 [93]. *NOTCH1* mutations appeared almost exclusively without *TP53* mutations and conferred a similar risk of progression [94]. The gene encoding splicing factor 3B subunit 1 (*SF3B1*) is mutated, again, in around 10% of CLL cases at first presentation [95, 96]. While the specific role of *SF3B1* mutation is elusive, the recurrence of mutations within certain 'HEAT' repeat domains suggests that functional aberration to the spliceosome influence CLL development [95]. Other frequently mutated genes include ATM [97] and BIRC3 [98] also observed deleted with the loss of 11q, and myeloid differentiation primary response gene 88 (MYD88) a critical component of Toll-like receptor signalling, also mutated in other B-cell cancers [99].

1.2.3.8 IGHV STATUS AND BCR SIGNALLING

The BCR is a complex of the transmembrane adaptors CD79a and CD79b, non-covalently bound to a surface Ig (**Figure 1.4**). In B cells, the engagement of antigen triggers downstream signalling which induces proliferation and survival [27].

The BCR has been shown to play a critical role in CLL pathogenesis, promoting cell survival and proliferation independently of normal, non-self antigen stimulation [100]. The aggression of CLL is often differentiated by the mutation status of IGHV, with CLL originating from B cells prior to somatic mutation, correlated with the most aggressive form of the disease [101]. Ig-mediated BCR cross-linking in U-CLL induced proliferation, whereas none was observed on M-CLL [102]. BCR signalling and CLL proliferation can be induced by cytoskeletal proteins exposed by apoptosis [103, 104], though autoantigenic signalling, where BCR signalling is induced by self-recognition [105] or potentially by non-self antigens from bacteria or viruses [106, 107]. The role of BCR signalling in CLL is supported by the clinical success of inhibitors of this pathway. It is further supported by knockout studies of the key BCR signalling protein; protein kinase C beta (PKCβ). Loss of PKCβ in models with a predisposition to a CLL-like disease failed to develop a malignancy [108].

The requirement for antigen engagement in BCR signalling in CLL has been questioned however [105], given the observations of low surface Ig expression and mutations and aberrations to BCR components [109, 110]. Several downstream components, such as the inhibition targets BTK and PI3K, have been studied to elucidate key molecules acting in the BCR pathway.

The protein tyrosine kinase, ZAP70 is typically observed regulating T-cell receptor signalling, however, ZAP70 has been shown to be interchangeable in BCR function, acting as an adaptor [111]. Expression in CLL cells correlates with both U-CLL status and enhanced BCR

signalling [112, 113], and has been shown to amplify downstream signalling such as that of NF-κB [114].

Nuclear factor of kappa light polypeptide gene enhancer in B cells (NF-κB) are a family of ubiquitous, pleiotropic transcription factors, which in B cells are activated by receptors such as CD40, inducing cell survival and proliferation. In CLL, NF-κB is induced by BCR signalling, PI3K and ZAP70 [115]. NF-κB appears to be active without, but is still enhanced by, CD40 stimulation [116, 117], as well as by cytokine signalling [118]. The role of NF-κB in CLL is strengthened by the observation that inhibition led to the induction of apoptosis [119].

1.2.3.9 IMMUNOPHENOTYPES

The heterogeneity of CLL is also apparent at the cell surface, with proteins such as CD38 [120] and CD49d [121], demonstrating differences in both expression and functional influence between cases.

Expression of ADP-ribosyl cyclase 1, the antigen CD38, has been shown to correlate with the proliferation rate of CLL and poor clinical outcome [122]. CD38 normally functions as an intracellular metabolic regulator and transducer and has intrinsic links with immune cell function and development including that of the BCR; features that appears to be amplified in cases of CLL overexpressing CD38 [123, 124].

CD49d, also known as integrin alpha 4 (ITGA4), pairs with either integrin beta 1 or 7 to act as receptor binding the extracellular matrix, vascular endothelial cells or mucosal tissue. In CLL it appears that α4β1 acts to mediate cell-cell, and cell-matrix interactions resulting in enhanced transendothelial migration, [125, 126] and enhanced microenvironment signalling conferring resistance to apoptosis [127]. CD49d expression correlates with trisomy 12 and has been shown to be regulated by DNA methylation [89]. A 3000 patient cohort defined CD49d as the single most accurate flow cytometry-based predictor of CLL aggression [121]. CD49d expression also has a significant correlation with other features of aggressive CLL including; CD38, ZAP70, and U-CLL [128].

1.2.3.10 OTHER PATHOGENIC MECHANISMS IN CLL

While CLL has many differential features, most appear to amplify a state that can already be defined as CLL. For instance, the role of BCR signalling in M-CLL appears to be less than that of U-CLL [102]. Several characteristic features, such as CD5 expression, have emerged, however, which demonstrate a consistent presence in the vast majority, if not all, cases of CLL.

The apoptosis regulator, BCL2, acts to inhibit apoptosis by sequestering pro-apoptotic proteins that induce the permeablisation of the mitochondria, an initial step that irreversibly

induces apoptosis. BCL2 is observed upregulated in CLL, as well as a number of other B-cell cancers, from which it derived its name; B-cell lymphoma 2 (BCL2) [129].

As mentioned, BCL2 can be upregulated either by loss of miRNAs by 13q14 deletions. Even in the absence of a 13q14 deletion, epigenetic silencing has been shown to downregulate miR15 and miR16 expression, allowing BCL2 overexpression, independently of any chromosomal aberration [130, 131]. A further mechanism has been described where BCL2 mRNA can be stabilised by the protein nucleolin which aberrantly translocates to the cytosol in CLL [132]. Other members of the BCL2 protein family have roles in the pathogenesis of CLL such as induced myeloid leukaemia cell differentiation protein (Mcl-1) and BCL2L1/BCLX [133, 134].

Receptor tyrosine kinase–like orphan receptor 1 (ROR1), a developmental embryonic protein, has recently been identified with frequent expression on CLL and other cancers [135, 136]. ROR1 appeared to induce NF-κB signalling and CLL survival in response to the developmental signalling protein WNT5a [137]. WNT5a was later shown to be expressed by CLL, self-inducing this pathway as well as increasing CLL cell motility [138]. The predominantly developmental expression of ROR1 has made it an attractive target for immunotherapy [139, 140].

The expression of the chemokine receptor CXCR4 in CLL mediates cell migration into tissues such as the bone marrow and lymph nodes [66, 141]. These provide a microenvironment in which CLL cells are stimulated by several other chemokines, interleukins and ligands, inducing cell survival, proliferation and therapy-resistance [142].

The proto-oncogene T-cell leukaemia/ lymphoma 1 (*TCL1)* has also been shown to contribute to CLL, in as many as 90% of cases [143]. The role of *TCL1* in CLL is supported by the observation of a CLL-like presentation of leukaemia in mice overexpressing *TCL1* in B cells [144, 145]. TCL1 function is described further in **Section 1.5**.

1.3 MODELLING CANCERS IN MICE

Mice provide an excellent pre-clinical platform for the controlled development of therapeutics and insights into cancer mechanisms, such as oncogenes. Cancers can be modelled in mice by the insertion of oncogenes into targeted cells by retroviral transduction, by xenografting human cancers, or by the engineering of transgenic mouse strains with a specific predisposition to develop spontaneous tumours, often induced by loss or amplification of tumour suppressors or oncogenes, respectively [146, 147].

Transgenic animal models provide an opportunity to observe polyclonal cancer development and evolution within the microenvironment of a whole organism. The temporal effects of the tumour-host interactions throughout tumourigenesis can also be studied. The effects of the immune system, availability of nutrients, microenvironment, circadian fluctuations and organism wide homeostatic mechanisms, which all likely influence tumour development, phenotype and evolution of cancers, can be captured, which may fail to be evident with cell culture or tumour xenografts [146, 148, 149]. Studying cancer in mice is also benefited by their short lifespan, small size and human comparability [148]. Mice have many physiological, metabolic and signalling mechanisms conserved though evolution, and cancers arising in humans and mice often share similarities in molecular nature and presentation.

Mice are, however, genetically and phenotypically very different from humans, with the last common ancestor existing around 80 million years ago, and have approximately a 3000-fold difference in physical size [150]. The metabolic rate of mice, which is 7 times that of humans, results in a far greater rate of DNA damage [151]. These differences have resulted in the evolution of, often, very different tumour suppressor mechanisms. For instance, loss of function of the *Brca2* gene is strongly associated with breast cancer development in humans, however, in mice conferred no tumour susceptibility [152]. Developmental and biological effects of an overexpressed or knocked-out gene can also mean that the sporadic evolution of tumours doesn't recapitulate human tumour evolution from a single cell [153].

Murine cancer models therefore have a significant role in the preliminary understanding of oncogenes and basic biology of cancers, that when translatable has proven beneficial to several cancer types [154]. They also provide the opportunity to investigate and manipulate cancers by means, which in humans might pose more ethically challenging questions.

1.4 THE *MYC* PROTO-ONCOGENE

Myc was first observed in tumours as the viral chimeric homologue *v-myc*, induced by the avian myelocytomatosis virus, due to genome integration of viral DNA into the *myc* gene [155-157]. This genetic insertion upregulated the *myc* gene and imposed a state of continuous proliferation and protein synthesis, resulting in a cell with ideal conditions for rapid viral replication, but also a dysregulation with the potential to drive malignancy [158, 159].

Myc encodes a helix-loop-helix leucine zipper transcription factor with affinity for E-box elements, the nucleotide sequence 'CAnnTG' (where 'n' denotes any nucleotide) associating with its dimerisation partner Max (shown in **Figure 1.5**) [160, 161]. E-box elements can direct the influence of the *myc* protein to approximately 15% of genes in the human genome [160, 162] . *Myc* is upregulated in response to mitogenic signals such as growth factors, hormones and

notch signalling (**Figure 1.6**) [163-165]. This leads to cell cycle progression and inhibits genes involved in cell cycle arrest or differentiation [159, 166]. *Myc* has other pivotal roles in regulating metabolism, cell adhesion and migration, angiogenesis, apoptosis sensitisation and pluripotency [167]. *Myc* is described as a pluripotency factor, owing to its involvement in the generation of induced pluripotent stem cells [168].

Myc exerts its effects through the recruitment of transcription factors and histone actetyltransferases (HATs) which induces chromatin dissociation and transcriptional upregulation [169-172]. *Myc* also has extensive roles in transcriptional repression, suggested to be just as numerous as its transcriptional activation capacities [157, 162].

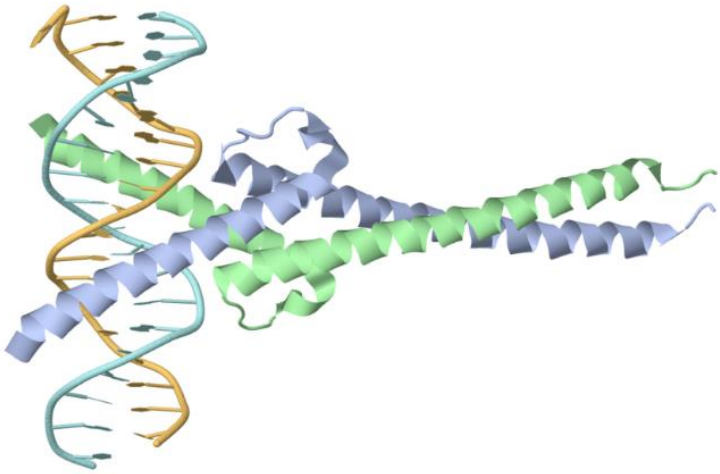


Figure 1.5. Crystal structure of myc. The myc-max heterodimer forming a helix-loop-helix leucine zipper and associating with an E-box element, 5'-CACGTG-3'. Adapted from the crystal structure defined by [161] using jmol [173].

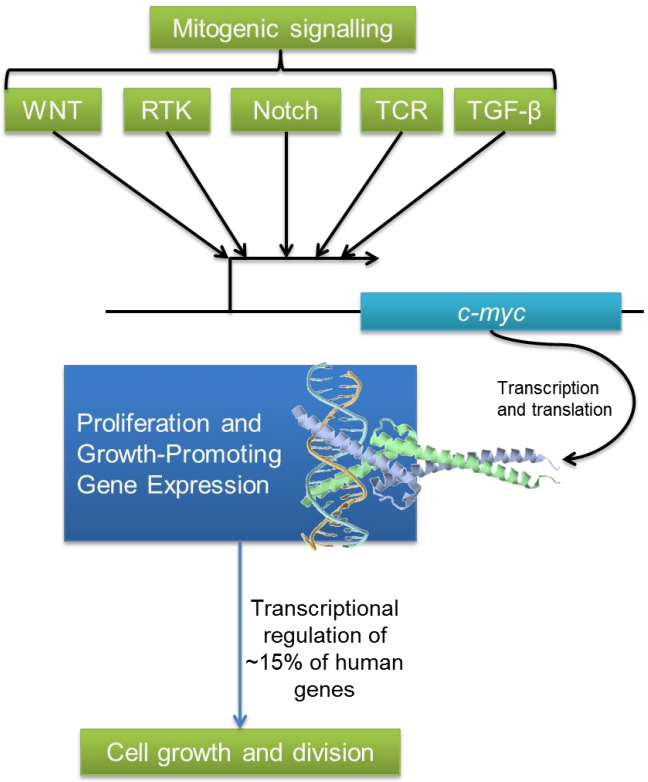


Figure 1.6. The function of myc in proliferation and cell growth in response to mitogenic signalling. Myc transcription is upregulated by many mitogenic signals. On translation, myc regulates many genes with the end effect being cell proliferation and growth. Adapted from [165]. RTK; Receptor tyrosine kinase, TCR; T-cell receptor, TGF; tumour growth factor.

1.4.1 THE *MYC* ONCOGENE

Ectopic *myc*-driven cell proliferation and tumourigenesis is seen in around 20% of cancers [174]. The intrinsic properties of *myc* make it particularly aggressive as an oncogene, with many of its effects contributions to neoplastic-like traits [1, 167]. It can be induced by various genetic aberrations, such as retroviral promoter insertions, chromosomal translocations or gene amplifications [35, 175-177]. *Myc* overexpression can also result from dysregulation of its related factors. Mutation or suppression of FBW7, for example, a protein responsible for *myc* degradation, allows an accumulation of *myc* protein [178].

1.4.2 *MYC* ONCOGENE REGULATION

The *myc* protein abundance is intrinsically regulated by multiple mechanisms to protect against the oncogenic potential of its overabundance (**Figure 1.7**) [165]. The alternative reading frame protein, p19ARF, for instance, is expressed in response to the transcriptional activities of *myc* [179]. If *myc* expression is high, ARF accumulates and inhibits the E3 ubiquitin ligase,

MDM2, preventing degradation of the tumour suppressor p53 [179, 180]. This allows an elevation above the normally very low endogenous levels of the tumour suppressor p53 and acts to trigger cell cycle arrest and apoptosis [181]. Studies have shown that mutations to this, and parallel pathways, are highly favourable to neoplastic development, specifically in the context of lymphomagenesis [21, 182]. When aberrations to components of these detection pathways occur, excessive *myc* expression no longer triggers p53 to initiate cell cycle arrest [182]. This loss of detection signifies a critical step in the development of cancer [165]. In B-cell lymphomagenesis, CD19 was shown to induce myc protein stability in two independent studies [183, 184] correlating strongly with poor survival.

Secondary to its effect on driving proliferation, *myc* appears to further accelerate tumourigenesis by the effect of its overexpression on genomic instability [185]. The downstream effects of *myc* increase the production of reactive oxygen species (ROS) which can directly induce DNA damage [175]. There is also evidence genetic damage can be induced by *myc* independent of ROS production [186]. These two distinct mechanisms are amplified by the ability of *myc* to circumvent cell cycle arrest, which normally occurs in response to genetic aberrations [175, 187].

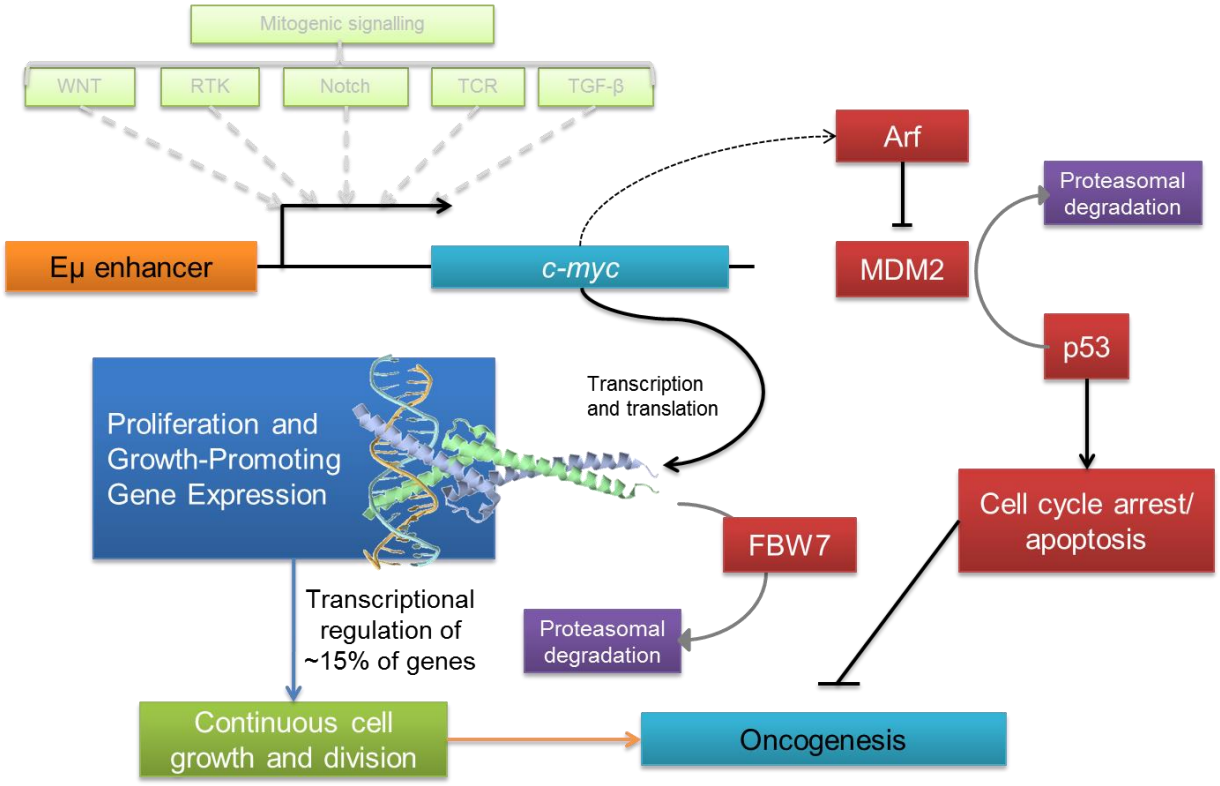


Figure 1.7. Tumour suppressors counteracting oncogenic myc overexpression. The effect of the immunoglobulin heavy chain enhancer region (Eμ) on the expression of myc in proliferation and cell growth in B cells. The translocation of Eμ to myc is observed in ~75% of BL cases. Myc overexpression induces the expression and accumulation of alternative reading frame protein (Arf). Arf inhibits murine double minute 2 protein (MDM2)-induced degradation of p53. Accumulation of p53 induces cell cycle arrest and cell death, a mechanism that has evolved to inhibit oncogenesis [165, 178-180]. FBW7 is responsible for the degradation of myc. Loss of any aspect of the tumour suppressor pathways on the right hand side can allow myc overexpression to go unchecked, allowing oncogenesis.

1.4.3 THE Eμ-*MYC* MURINE MODEL

BL is modelled in mice by Eμ-induced overexpression of human myc, which recapitulates the substantial overexpression of myc specifically in B cells [188]. Eμ-*myc* mice live for approximately 100 days, before succumbing to an aggressive lymphoma [176] (**Figure 1.8**) modelling both molecular and pathological aspects of BL [176, 188-190]. The most common presentations are splenomegaly, peripheral lymph node tumours, thymomas and mesenteric lymph node tumours [176]. As early as prenatally, Eμ-*myc* mice develop an enhanced population of polyclonal pre-B cells, with larger cell morphology and increased replicative potential [191] – typical of the effects of *myc* on cells; inducing a stem-like state, preventing progression towards terminal differentiation, accelerating mitosis and stimulating cell growth [167]. Eμ-*myc* mice therefore provide a model through which the BL translocation, and thus the biology of *myc*-driven tumours, can be studied [192].

Additionally, applications of the Eμ-*myc* model have been used to elucidate and understand aspects of myc biology. Prior to tumourigenesis, for instance, the Eμ-*myc* model demonstrates a substantial increase in myc transcript level, while only a marginal increase in myc protein [192]. This observation suggests that substantial feedback mechanisms occur supressing the overabundance of myc protein. The model has been used to understand aspects of myc biology such as myc-induced, p53- and ARF-dependent cell cycle arrest and apoptosis [182], and the cooperation of other oncogenes with myc such as Ras and Raf, accelerating tumourigenesis [193].

Some aspects of the Eμ-*myc* model do not accurately recapitulate BL, due to the presence of the genetic aberration to myc early in B-cell development. This is contrasted with BL where translocations often occur later into development. As a result of this, characteristics of DLBCL are also present [189].

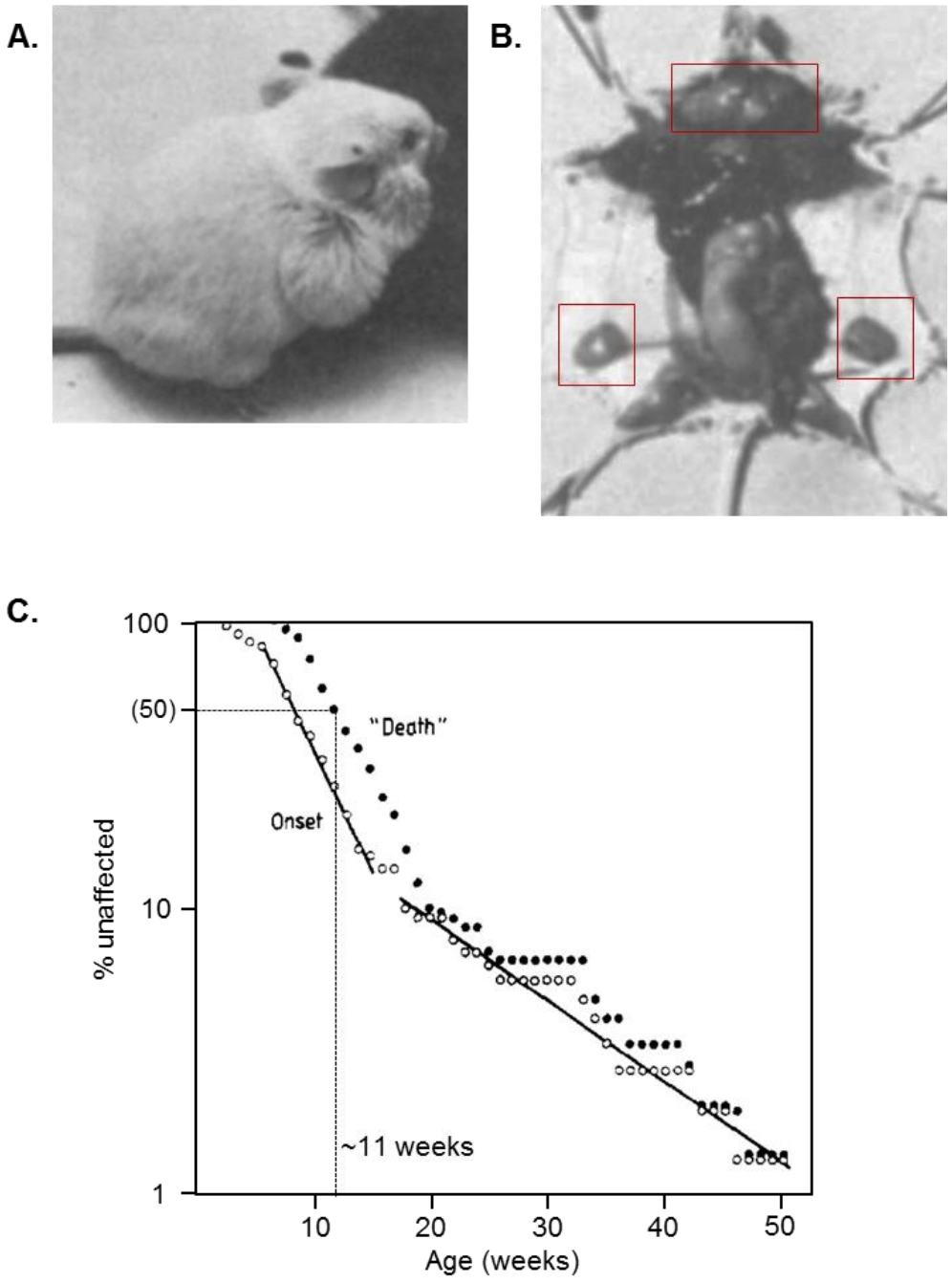


Figure 1.8. Characteristics of the Eμ-*myc* mouse model of Burkitt's lymphoma. A. An 11-week old Eμ-*myc* mouse demonstrating spontaneous lymph node tumours. **B.** A dissection of the mouse from A. highlighting tumours of the cervical (neck) and inguinal (abdominal) lymph nodes. **C.** The survival and emergence of such tumours in 134 Eμ-*myc* mice indicating a very high disease penetrance and a rapid progression from onset to lethality. The median time to death was approximately 11 weeks, with the vast majority succumbing to terminal tumours within 100 days. Adapted from [176] and [188].

1.5 THE *TCL1* PROTO-ONCOGENE

The TCL1 protein is a small 14 kDa homodimerising peptide which consists of an 8 strand beta barrel structure shown in **Figure 1.9** [194]**.** The proto-oncogene has been shown to be expressed in the spleen and lymph nodes, as well as in circulating B cells [195]. The expression of *TCL1* in B cells was shown to be in strong correlation with the differentiation of B cells as they develop, with naïve and pre-B cells expressing *TCL1* significantly more than germinal centre B cells or plasma cells [196]. The role of *TCL1* was extended to cover the development of T- and B cells, observed by reduced hematopoietic development in *TCL1*$^{-/-}$ mice [197]. *TCL1* has also been shown to have roles in early embryogenesis [198].

1.5.1 THE *TCL1* ONCOGENE

The *TCL1* gene was named 'T-cell leukaemia/lymphoma 1' due to its initial discovery, adjacent to a break point region on human chromosome 14 in a case of T-cell chronic lymphocytic leukaemia [199]. *TCL1* transcript expression was detected in T-cell leukaemias in 10 T-cell leukaemia cases [200]. This conclusion was later disputed by data suggesting that *TCL1* transcript could not be detected in 20 analysed T-cell lymphoma samples, suggesting that B-cell material expressing *TCL1* and may have been a contamination [201, 202]. It was concluded that *TCL1* expression did not have as great a role as originally thought in T-cell leukaemias and was actually strongly expressed in many B-cell tumours [195]. Expression across various cancers of hematopoietic cells was accessed by immunohistochemistry to further show the prevalence of *TCL1* expression in B-, but not T-, cell tumours, especially that of B-cell CLL and BL [203]. The negative correlation between B-cell differentiation state and *TCL1* expression was shown to be maintained upon tumourigenesis based on the B-cell subtype from which the tumour was derived [196]. The *TCL1* gene was confirmed as a proto-oncogene in transgenic mice where Lck promoter-regulated *TCL1* expression induced tumours of T-cell origin [204, 205].

1.5.2 *TCL1* REGULATION

While activation of *TCL1* is frequently attributed to translocations, it is also observed to be genotypically normal in many B-cell tumour types. Analysis reveals that hypomethylation of the *TCL1* promoter region is frequent in *TCL1*-expressing chronic lymphocytic leukaemia and BL [208]. There is strong evidence that the chaperone protein, heat shock protein 70 (HSP70), plays a role in enabling TCL1 function. Inhibition of HSP70 causes TCL1 degradation and also decreases the activation of AKT through TCL1 [209]. Recent work has shown that TCL1 overexpression is induced by the loss of miR-3676, a microRNA which strongly downregulates

TCL1 expression. miR-3676 is coded within the same region (17p13) as p53, and appears to be frequently co-deleted [210].

1.5.3 TCL1 AND AKT

Co-immunoprecipitation and yeast two-hybrid screening demonstrated a direct interaction between TCL1 and AKT1 (protein kinase B). This interaction increased the phosphorylation potential of AKT1, inducing its promotion of cell survival and proliferation [211, 212]. The interaction was also shown to enable co-translocation to the nucleus [211]. The role of TCL1 in AKT activation was later demonstrated to enable transphosphorylation between AKT1 and AKT3 [213]. This evidence suggested that the primary role of TCL1 was the activation of AKT, promoting kinase activity and upregulation of mitogenic and anti-apoptotic pathways [211]. However, the loss of the tumour suppressor PTEN, a protein which reverses the PI3K-dependent effects of AKT, while causing enhanced proliferation and survival, did not lead to tumourigenesis in B cells [214]. This observation led to the conclusion that the effects of TCL1 on AKT alone were not sufficient to cause malignancy in B cells, however does not discount an involvement of the AKT pathway in CLL development.

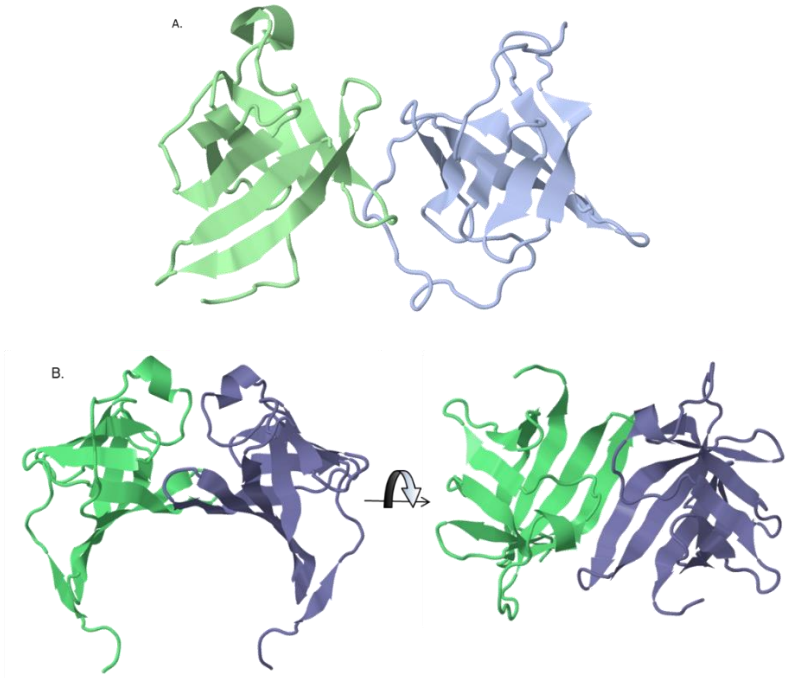


Figure 1.9. Crystal structures of the TCL1 protein. **A.** Murine TCL1 homodimer. The subunits form two beta-barrel structures. Adapted from the crystal structure defined by [206]. **B.** An alternative conformation seen of human TCL1 forming a single beta sheet across the homodimer's two subunits. The left view is rotated from the top, 90° out of the page to give the right view. Adapted from the crystal structure defined in [207] crystal structures visualised in Jmol [173].

1.5.4 TCL1 AND P300

The effect of TCL1 was examined on NF-κB, a proto-oncogene also observed to induce a CLL-like cancer in murine B cells [215]. While no direct interaction was observed, the transcriptional activator p300 was shown to be an intermediary in the activation of NF-κB [216]. The p300-TCL1 interaction was also shown to inhibit the activity of activator protein 1 (AP-1). AP-1 has numerous roles in cellular biology, but in the context of CLL, appears to have the potential to induce apoptosis via MEKK-1 [216, 217]. Thus, inhibition of AP-1 by TCL1 through p300 appears to reduce the expression of MEKK-1 and inhibit this apoptotic pathway in CLL [216]. TCL1 therefore appears to cause proliferation and survival through AKT and NF-κB while inhibiting apoptosis that may be induced by AP-1; all phenotypes that support cancer hallmarks [2]. These mechanisms are summarised in **Figure 1.10**.

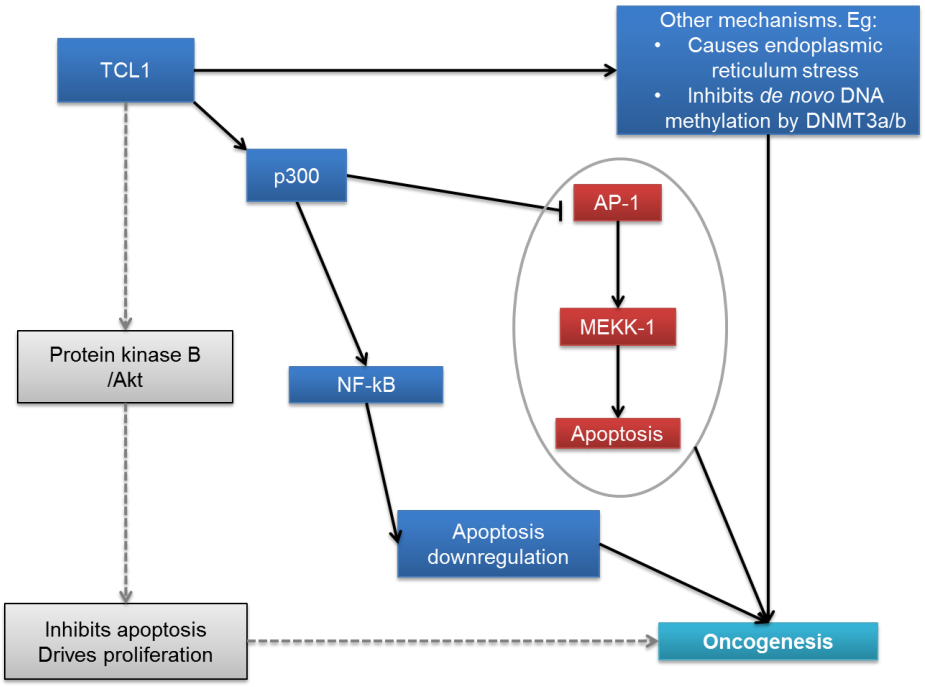


Figure 1.10. Potential Oncogenic mechanisms of TCL1 in CLL. TCL1 activates necrosis factor kappa-B (NF-κB) and inhibits activator protein 1 (AP1) through p300. This results in reduced apoptosis. The role of protein kinase B is unclear, and other mechanisms are emerging which implicate further mechanisms of TCL1 in cancer. [218-221].

1.5.5 THE Eμ-*TCL1* MURINE MODEL

To investigate the oncogenic potential of *TCL1*, specifically in B cells, the Eμ Ig enhancer was used to induce B-cell overexpression of the human *TCL1* gene in mice [144], like with the Eμ-*myc* model. The Eμ-*TCL1* mouse model is developmentally and phenotypically

normal, barring the observation of a higher *TCL1* expression in B cells. By 4 months, the majority of these mice present with a marginally raised $CD5^+$ population of B cells in the bone marrow, spleen and peritoneal cavity, but remain asymptomatic (**Figure 1.11A**). By 8 months, an expanded $CD5^+$ B-cell population is observable and the disease manifests as lethal at around 11 months of age with splenomegaly and leukaemia (**Figure 1.11B**) [144, 145]. The Eμ-*TCL1* mouse has been proposed as a model for studying the molecular basis of, primarily aggressive, CLL and for better understanding of the pathogenesis of the disease [144, 145]. Several clinical characteristics such as the relatively slow disease development, accumulation of a $CD5^+$ population in typical CLL-localised compartments, capability of CLL-like T-cell dysfunction induction [222-224] and organ infiltration are apparent in the model [145]. Another characteristic of the Eμ-*TCL1* mouse model which mimics that of CLL is the appearance of an aberrant epigenetic profile in B cells from as early as three months of age [225, 226], potentially attributable to TCL1's interactions with DNA methyltransferases [221]. The potential relevance of the model to CLL is further supported by the observation of increased TCL1 expression in 90% of CLL cases in addition to a correlation between TCL1 expression and poor prognosis [143, 227].

The Eμ-*TCL1* model has been used as a preclinical model, initially evaluating fludarabine demonstrating similar response to CLL [145]. To increase the throughput of such experiments, the use of adoptive transfer of an established leukaemic population is applied. This technique has been used to test the *in vivo* efficacy and tolerability of, for instance, deacetylase inhibitors [228], the mTOR inhibitor rapamycin [229] and a dual PI3K/mTOR inhibitor [230].

While other proposed transgenic mouse models of CLL exist, the Eμ-*TCL1* mouse has the highest disease penetrance [231]. This has made it one of the most widely used animal models of CLL. In addition to the evaluation of Eμ-*TCL1* tumour progression, a large number of studies have been conducted to evaluate the resulting phenotype arising from crossing Eμ-*TCL1* mice with other transgenic CLL mouse models and knockout models. Some noteworthy examples include crosses with; p53 knockout mice [232], B-cell activating factor (BAFF) transgenic mice [233], a proliferation-inducing ligand (APRIL) transgenic mice [234], ROR1 transgenic mice [235] and the miR-29 transgenic mice [236], each resulting in an accelerated disease progression.

As outlined in **Section 1.3**, the Eμ-*TCL1* mouse and other CLL mouse models are, however, limited by several factors common to mouse models of cancer and may not provide a completely accurate recapitulation of CLL.

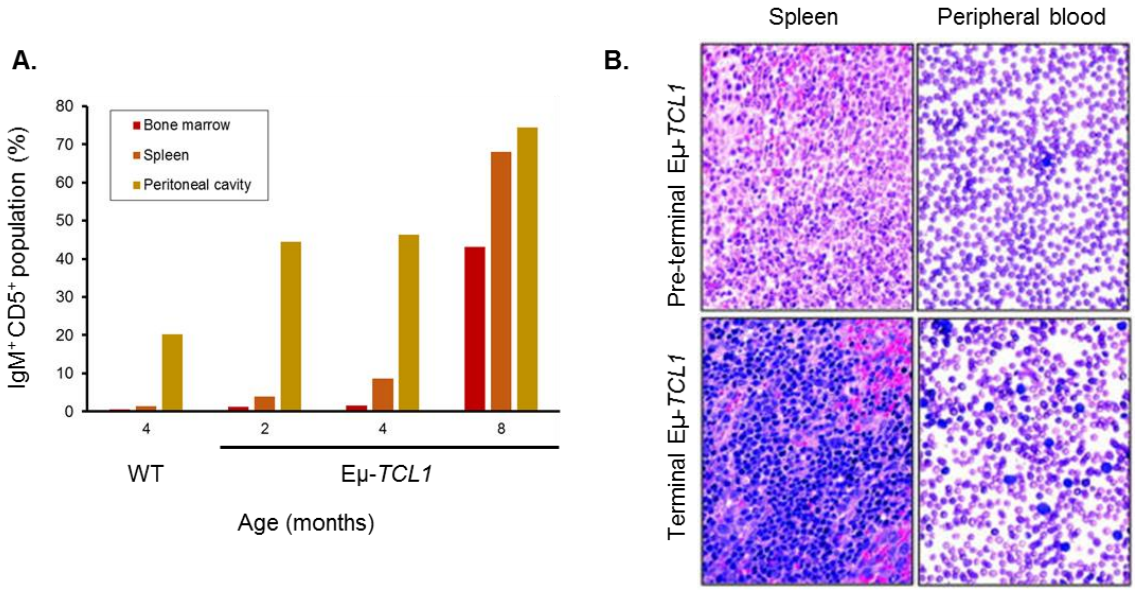


Figure 1.11. Characteristics of the Eμ-*TCL1* mouse model of CLL. A. The appearance of $CD5^+$ B cells in the spleen, blood marrow and peritoneum of the Eμ-*TCL1* mouse model (adapted from [144]). **B.** Haematoxylin (purple)/eosin (pink) stained samples from pre-terminal and terminal Eμ-*TCL1* mice demonstrating an expansion of lymphocytes in the spleen and blood in the terminal animals (adapted from [145]).

1.6 'OMICS

The term 'Omics' describes the systematic consideration of all the constituents among a class of related biological molecules. Genomics, epigenomics, transcriptomics, proteomics, metabolomics and lipidomics are all common examples of fields that aim to characterise the entirety of their given area in a comprehensive and unbiased way. These 'omics' approaches are based around the central biological dogma which describes the process from which DNA is transcribed to mRNA which is translated to proteins, which in turn affect almost every omic system [237-239]. In the context of cancer, omics looks to identify patterns in molecular aberrations that are characteristic of neoplasia. Understanding aberrations in this way is critical to identifying subtle differences between normal and cancerous cells, which can be exploited to target cancers or oncogenic mechanisms.

Genomics focuses on understanding cancer at the genetic level, with the origins of cancer often being directly attributable to genomic aberrations. A single nucleotide change can induce altered translation, and can cause dramatic changes to the function of a protein, which may have a critical role in tumour suppression. For example, *TP53*, has more than 2500 known

mutations identified in cancers which affect its primary sequence [240]. Genomics also attempts to appreciate the effects of mutation on the enhancers and promoter regions as well as other non-coding DNA.

Transcriptomics describes the characterisation of expression of the total RNA molecules present in cells or tissue at a point in time [239]. In cancer, it is frequently used as a means of identifying aberrant gene expression, relative to that of typical tissues from which the cancer has arisen [241] or for comparison between different cancers. The chemical and physical nature of mRNA makes the specific identification and quantification of transcripts a rapid and reproducible process. Microarray technology, such as Affymetrix GeneChip microarray enables large scale analysis of many thousands of transcripts simultaneously [241]. Deep sequencing approaches, such as the shotgun sequencing 'RNA-seq', present the opportunity to gain high-resolution total transcriptome characterisation [242].

Translation is a complex and tightly regulated process and numerous RNA binding factors, processing bodies, RNA secondary structures and ribosomal mechanisms dictate when and how much protein is translated from a transcript at a given time and condition [243]. Transcription has even greater complexity under the aberrant conditions of cancer [244]. Transcript expression is therefore rarely in direct correlation with protein expression, presenting a requirement for the direct measurement of relative protein expression [245]. To more accurately understand the phenotype of a cell, investigation of protein expression is critical.

1.7 PROTEOMICS

Proteomics is defined as the global approach to the characterisation of the protein complement expressed by cells, tissues or organisms under defined biological conditions [246, 247]. Proteins are the biomolecules responsible for the execution of the functional information defined by a genome, contributing to essentially every biological process. Protein study therefore presents a greater opportunity to characterise the molecular basis of phenotype. While transcriptomic characterisations are becoming comprehensive and relatively cost-effective, the correlation between protein and mRNA expression is low for the vast majority of proteins [248, 249].

The breadth of proteomics coverage lags behind genomics and transcriptomics characterisations due to the complex physicochemical properties of proteins [250]. Proteins present an enormous diversity due to many combinations of the 20 amino acids from which they are derived, amplified by many potential post-translation modifications. Where RNA and DNA detection can be facilitated through their logarithmic amplification by approaches such as the polymerase chain reaction, proteins lack such properties and their detection is determined by the

sensitivity of available proteomics methods and techniques [251]. The size of the proteome relative to that of the genome also presents a challenge, with potentially more than 5 times as many proteins than coding genes [252].

Approaches in proteomics are dominated by those employing mass spectrometry (MS), however some methods such as antibody microarrays have been described [253]. MS enables precise mass measurements of a protein or peptide to be determined allowing for confident conclusion of the proteins chemical identity [254, 255]. MS proteomics can be subdivided into two main approaches; top-down, which attempts to characterise intact proteins and bottom-up, which first employs digestions to fragment proteins into peptides. The application of bottom-up proteomics to complex samples has been described as 'shotgun proteomics' [256].

One of the most fundamental challenges in proteomics is the separation of complex protein mixtures for characterisation by MS. A consensus between the several available forms of MS characterisations is the need for low analyte complexity to enable sensitive and accurate analysis.

1.7.1 2D-GEL ELECTROPHORESIS

During its inception, proteomics was based upon two-dimensional (2D) polyamide gel electrophoresis, the earliest form of protein separation for proteomics. Using the size and isoelectric point of proteins as a means of protein differentiation allowed the 2D visualisation of the proteome; identifying protein 'spots' [257]. The presence or absence of spots in the same coordinates on two comparable 2D gel separations could therefore indicate a biological difference between samples. The combination of 2D gel electrophoresis with mass spectrometry (MS) provided a means of identifying the differentially expressed protein. 2D gel electrophoresis has demonstrated utility, for example, in classifying leukaemias [258] and cell surface marker identification across a range of cancers [259]. The method, however, is limited by sensitivity and low throughput, given that each spot is unknown until manually analysed [260].

1.7.2 LIQUID CHROMATOGRAPHY

Liquid chromatography (LC), much like that of 2D gels, takes advantage of the physicochemical differences that exist between proteins or peptides to provide a means of pre-separation prior to analysis by MS. LC is typically based on the use of columns within which different molecules have different levels of interaction, described as a retention. LC consists of two major components; a mobile phase, the continuously flowing solvent transporting the analytes, and the stationary phase, the material with which the analytes can interact.

The most widely used form of LC in proteomics, reverse-phase (RP) LC, enables the separation of peptides on the basis of hydrophobicity [261, 262]. RP LC uses a stationary phase

of hydrophobic resin of long hydrophobic alkane molecules, typically 8 or 18 carbon atoms in length, termed C8 and C18, respectively. Peptides are introduced to the stationary phase while dissolved in a hydrophilic mobile phase, which causes a partitioning of the peptides into the stationary phase, resulting in retention. Charged species such as salts do not partition, resulting in rapid elution – giving RP LC a 'desalting' capability. The hydrophobicity of the mobile phase is then steadily raised, described as a 'gradient', which allows peptides to partition back into the mobile phase. The retention time of each peptide is therefore proportional to the hydrophobicity [262].

1.7.3 SHOTGUN PROTEOMICS

Shotgun proteomics approaches, favouring the digestion of proteins into peptides prior to analysis has several advantages over intact protein characterisation. Trypsin, a digestive enzyme, is by far the most popular means of proteolysis, cleaving at the C-terminus of lysine and arginine residues [263]. Prior to proteolysis, proteins are reduced and unbridged cysteines are alkylated, permanently linearising the protein molecules. Proteolysis reduces the dynamic range of several physicochemical properties of the analytes, such as charge, mass and hydrophobicity, dramatically simplifying LC and MS. Trypsin proteolysis disrupts protein-protein interactions and reduces the approximate mass of analytes to between around 500 and 5000 Da. Another effect of trypsinisation is the distribution of the positively charged lysine and arginine residues to, most frequently, one instance per peptide. Although proteolysis results in an increase in sample complexity and identification redundancy, the resulting low-charged and low mass peptides are highly favourable for separation using LC and characterisation by MS, far outweighing these disadvantages [264].

The advent of modern MS proteomics was facilitated by the development of electrospray ionisation (ESI). ESI involves the application of an electrical current between the solvent and the aperture through which ions can enter the MS [265]. As evaporation occurs, the accumulated charge increases until the repulsive forces exceed that of the surface tension, resulting in Coulombic fission. Coulombic fission describes the cascading explosion of charged droplets which provides a means by which peptides are converted into a gaseous phase [266]. Once in the gaseous phase, and positively charged, it is then possible for the molecule to be detected by MS.

ESI allowed the interfacing of liquid chromatography (LC) with MS, termed LC-MS [265]. Coupling LC and MS enabled an increase in throughput as well as improved peptide separation and identification [267]. A recent advancement to ESI has been in the addition of 5%

dimethyl sulphoxide (DMSO) to the mobile phase, which substantially improves ESI-facilitated peptide detection [268].

The combination of prefractionation and subsequent MS-coupled fractionation, termed 2D-LC MS, allowed the extensive fractionation and characterisation of peptides, in a pipeline termed ‘multidimensional protein identification for shotgun proteomics (MudPIT)’ [256].

1.7.4 MASS SPECTROMETRY

MS characterisation of peptides can be conducted in several ways, depending on the MS instrumentation and means of ionisation. In all cases, the peptide molecules must be converted to the gaseous phase with a charge, focused with lenses, captured with a mass analyser, such as; quadrupole, ion trap, time of flight or Orbitrap, and the mass to charge ratio (m/z) determined with a detector.

For the Orbitrap Elite (**Figure 1.12**) and ESI, for example, the gaseous peptides pass through an aperture and are initially focused with a source lens and filtered to remove neutral species a beam blocker. Charged ions are guided using electrostatic charges and are focused by further lenses to be delivered to collision cells or mass analysers. Captured ions are also filtered based on the m/z by a quadrupole mass filter [269].

The Orbitrap Elite has two mass analysers; a linear ion trap (LIT) analyser and a Fourier Transform MS (FTMS) analyser. These offer complementarity in their means of detection, with the LIT giving a low resolution but high sensitivity and the FTMS with an ultra-high resolution and a relatively low sensitivity [269]. For the determination of peptide masses with a full spectrum scan, a resolution of up to 240,000 can be achieved at 400 m/z [269, 270]. The LIT captures and analyses ions using an electromagnetic field that allows linear oscillation and highly sensitive detection by photomultiplier. FTMS captures ions in an orbital oscillation, hence the name Orbitrap, which uses the resonant frequency of an orbital ion to determine its m/z with a high degree of accuracy [271].

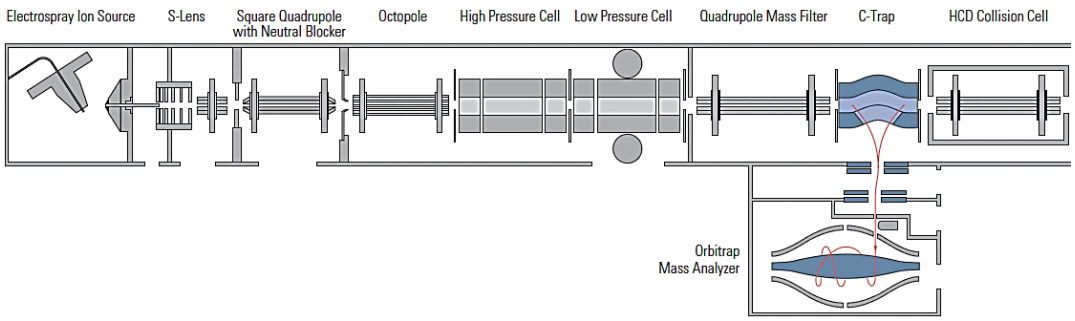


Figure 1.12. Schematic detailing the components of the Orbitrap Elite. Gaseous phase ions from the ESI source move from left to right. Adapted from [269].

1.7.5 Data-Dependent Tandem Mass Spectrometry

Determination of a peptide's mass is not sufficient to confidently identify its sequence. Using mass filtering, it is possible to isolate a mass window around a peptide, termed a precursor, and subject the precursor to fragmentation in isolation [264]. This process, termed MS/MS, MS^2 or tandem MS, generates product ions which are representative of fragmentation around the peptide bonds. These fragments can subsequently be used to attempt to confidently identify the peptide, in a process known as peptide mass fingerprinting.

Data dependent acquisition (DDA) is the process by which the full spectrum of ionised peptide masses are first determined with MS and are then subjected to MS/MS in order of intensity, starting with the most abundant. This process can also be described as 'top N', where N is the number of peptides characterised before another full spectrum scan is performed and the process is repeated. Common practise is to use dynamic exclusion which ignores the peptides which were characterised in the previous scans, to focus on the lower abundant and newly eluting peptides from the LC [264].

Two common forms of peptide fragmentation include collision-induced dissociation (CID) and high-energy collisional dissociation (HCD). CID causes peptide fragmentation with the use of helium gas, which, with an atomic mass of just 4 Daltons, has relatively low transfer of energy upon collision. This results in typically large fragments of the peptide being generated [272]. HCD uses nitrogen gas capable of transferring far more energy with a molecular mass of 28 Da, producing far smaller peptide fragments [273]. Both gasses are additionally inert and non-reactive.

Both CID and HCD induce peptide fission at the C-N bond of the O=C-N-H peptide bond. For each fragmentation event, two fragments are produced; the N-terminus fragment is termed a 'b' ion and the C-terminus fragment is termed a 'y' ion. Under different fragmentation methods, fission can occur between the α-carbon of an amino acid and the carbon of the peptide bond generating 'a' and 'x' ions or between the nitrogen of the peptide bond and the α-carbon amino acid to give 'c' and 'z' ions [274].

1.7.6 Target-Decoy Peptide Searching

Shotgun proteomics experiments typically produce many thousands of MS/MS spectra. In order to identify the peptides to which the precursor masses and respective fragment masses relate, a process termed target-decoy searching is used. This first analyses the potential matches based on theoretical spectra. Theoretical spectra are derived from a proteome by the *in silico* digestion of all the protein sequences in a proteome using a rule that relates to the means of proteolysis used for the shotgun proteomics experiment. For trypsin, this generates peptides

based upon an anticipated cleavage at the C-terminus of arginine and lysine residues. Another consideration with trypsin is the proline rule which prevents proteolysis when located on the C-terminus side of arginine and lysine. Theoretical precursor masses and fragment masses can then be determined for the resulting theoretical peptides calculated on the known masses of the constituent atoms of the peptide and fragments. The result of the 'target' search conducted between the theoretical spectra and MS/MS-derived spectra is termed a peptide spectrum match (PSM) [256, 264].

Decoy searching was introduced due to the high frequency of false discovery using target searching for PSMs. With both the spectra and theoretical spectra numbering many thousands, the odds of matches occurring by chance is very high. To adjust for this false discovery, a second search is used which reverses the sequences of the target proteome and conducts a second search. This 'decoy' search defines the probabilistic score of those spectra matching by chance. Comparison of the probabilistic scores of target and decoy matches presents an overlap which allows the unlikely PSMs to be disregarded. The threshold at which PSMs are disregarded is frequently set allowing for a 1% overlap, termed a false discovery rate (FDR) of 1%. This can also be expressed as a 'q-value', and for a <1% FDR equates to $q<0.01$. Several approaches exist which implement the process of target decoy searching. The most commonly adopted being Andromeda, SEQUEST and MASCOT. [274-277]. Although the specific methods used by these engines vary, they follow the basic target-decoy process described, assigning each PSM with a false discovery rate. Again, several methods of estimating the FDR of PSMs exist. A frequently employed, powerful method is that of semi-supervised machine learning, using an algorithm termed 'Percolator'. Percolator considerers approximately 20+ features of the target and decoy PSMs providing a far more accurate and efficient means of separating the true matches in the data [278].

1.7.7 PROTEIN INFERENCE

The result of a shotgun proteomics experiment is often the identification of several thousand peptides. After proteolysis, however, all information regarding the protein of origin of a peptide is lost. Strategies must therefore be implemented to infer a peptides origin on the basis of protein sequences. Many peptides simply map uniquely to a single protein, however, this process is frequently complicated by protein isoforms where several or all of the identified peptides match to two distinct protein sequences. Disregarding these data as redundant, especially for proteomes of higher eukaryotic organisms, where redundancy is high, impacts the quantity of usable data. A process of grouping proteins on the basis of a high degree of similarity is therefore frequently employed [279]. Protein grouping attempts to use as much data as possible to explain as many of the peptides with the smallest number of protein groups [280]. This process results in a list of

proteins where only those proteins with a high probability of having been present in an analysed sample are presented based on all the information available from both the identified peptides and the aligned proteome. Protein grouping is of particular importance when attempting to quantify proteins, as peptide quantitations are only of use if they are proteotypic to a protein group or unique protein. Redundant peptides can have ambiguous origins and their quantitations are often disregarded.

1.8 QUANTITATIVE PROTEOMICS

The qualitative analysis of a particular protein present in a single sample is a useful and interpretable result. However, the appreciation of such a result within the context of a complex biological matrix is challenging. From the perspective of cancer drug target and biomarker discovery, quantitative proteomics approaches allow the characterisation of disease samples quantitatively, relative to healthy, reference, control proteomes. These controls form a point of comparison from which biological and, potentially, clinical hypotheses can be drawn. To address this requirement, techniques have emerged which allow quantitative comparisons between two or more samples within large scale proteomics experiments. These approaches can be subdivided into label-based and label-free methods. Label-free approaches compare between data derived from separate 2D-LC MS/MS experiments, whereas label-based approaches focus on combining quantitation into a single 2D-LC MS/MS workflow [281].

1.8.1 LABEL-FREE QUANTITATION

The earliest and most basic form of quantitative proteomics, compared the relative sizes of protein spots on 2D gels. However, as described in **Section 1.7.1** this is a labour-intensive process. Shotgun proteomics dramatically increased this throughput, with the quantitative analyses comparing protein abundances using a method termed peptide, or spectral, counting. Spectral counting works on the principle that there is a correlation between the number of PSMs matching to a protein and the protein's abundance [282]. Another label-free approach uses the relative intensities determined for comparable precursors, on the basis of mass and retention time [283]. Label-free approaches are regarded as the least accurate means of MS quantitation, predominantly due to inter-experimental variation [284].

1.8.2 LABEL-BASED QUANTITATION

The two major types of label-based proteomics are those which covalently modify peptides as part of the proteomics workflow and those which use metabolic labelling to incorporate stable isotopes into the protein during its synthesis in a cell. Peptide labelling can further be subdivided into approaches which perform relative quantitation using the intensity of precursors with MS1 and methods which use isobaric tags to quantify sample-specific

abundance with MS^2. The primary aim of labelling strategies is to enable simultaneous sample processing eliminating the variability seen when chromatography is performed iteratively. The incorporation of sample-specific stable isotopes allows samples to be pooled and while maintaining a means of deriving relative quantitation using MS.

Stable isotope labelling by amino acids in cell culture (SILAC) allows cell lines to be grown under identical conditions, varying only in supplementation with amino acids incorporating stable isotopes, typically arginine and lysine. Upon MS analysis, the relative spectral intensities of stable isotope labelled and unlabelled peptides allow direct quantitation between the up to three cell cultures [285-287]. This principle has also been extended to whole organisms, such as mice [288]. Metabolic labelling is costly and limiting however, especially in the context of primary human samples.

Several methods attempt to overcome the challenges of metabolic labelling by preforming the stable isotope labelling at the point of sample processing. ^{18}O incorporation, for instance can be induced during proteolysis using $H_2{}^{18}O$ in the solvent [289, 290]. The first use of tag-based covalent peptide labelling was that of isotope-coded affinity tags (ICAT) [291]. ICAT was able to covalently modify cysteine sulphydryl groups with either ^1H or ^2H-containing tags inducing a mass-shift of 8 Da. These tags were, however, limited by the scarcity of cysteine residues in the proteome and the propensity of deuterium to alter chromatographic retention [292]. Several other isotopic mass tags capable of facilitating MS1 quantitation have since emerged including; isotope-coded protein labels (ICPL) [293] and dimethyl labelling [294]. Quantitation with MS1 using SILAC, ^{18}O labelling, ICPL and dimethyl labelling is limited to a comparison of no more than three samples per experiment. This is predominantly due to the potential overlap resulting from natural isotopic variability.

1.8.3 Isobaric Labelling

Isotope-coded tags emerged which used stable-isotope labelling in a manner which ensured each sample-specific label imposed the same physicochemical changes upon peptides. These tags, which used a reporter group and a balancing group to maintain identical mass, were termed isobaric labels. By maintaining identical peptide masses and properties with different labels, it is possible to maintain sample complexity, maintain chromatographic retention and allow the co-fragmentation of chemically identical peptides as a single precursor from several samples. This enabled the generation of a single MS/MS spectrum providing peptide identification and relative quantitation. The two predominant types of isobaric labels are isobaric tags for relative and absolute quantitation (iTRAQ) and tandem mass tags (TMT). Both iTRAQ and TMT contain an amine reactive group capable of modifying primary amines at either lysine

residues or the N-terminus of peptides. Upon reaction, the peptide becomes covalent modified with a reporter group and a balancing group described in **Figure 1.13**. These tags avoided the use of deuterium, differentially incorporating ^{15}N and ^{13}C into the mass tag and balancing group.

TMT were the first isobaric tags described, which allowed for the relative quantification of two samples [295]. Using MS/MS, peptides could be quantified with two peaks at 126 and 127 Da, termed reporter ions. The first variant of iTRAQ offered the simultaneous quantitation of 4 samples, termed a 4-plex, with quantitation occurring with MS^2 based on the relative intensities of reporter ions at 114, 115, 116 and 117 Da [296]. The location of these measurements is termed the reporter region. A later variant introduced 4 further labels; 113, 118, 119 and 121 Da, expanding iTRAQ capabilities to an 8-plex [297]. A 6-plex version of TMT was introduced which was recently developed to incorporate a further four labels using the reporter region 126 to 131 Da [298, 299]. These additional labels utilised the differential masses of the neutron in ^{15}N compared with a neutron in ^{13}C to create mass differences of less than 10 mDa detectable by high resolution instruments. This also proposed an expansion to a 10-plex system which was later described [300].

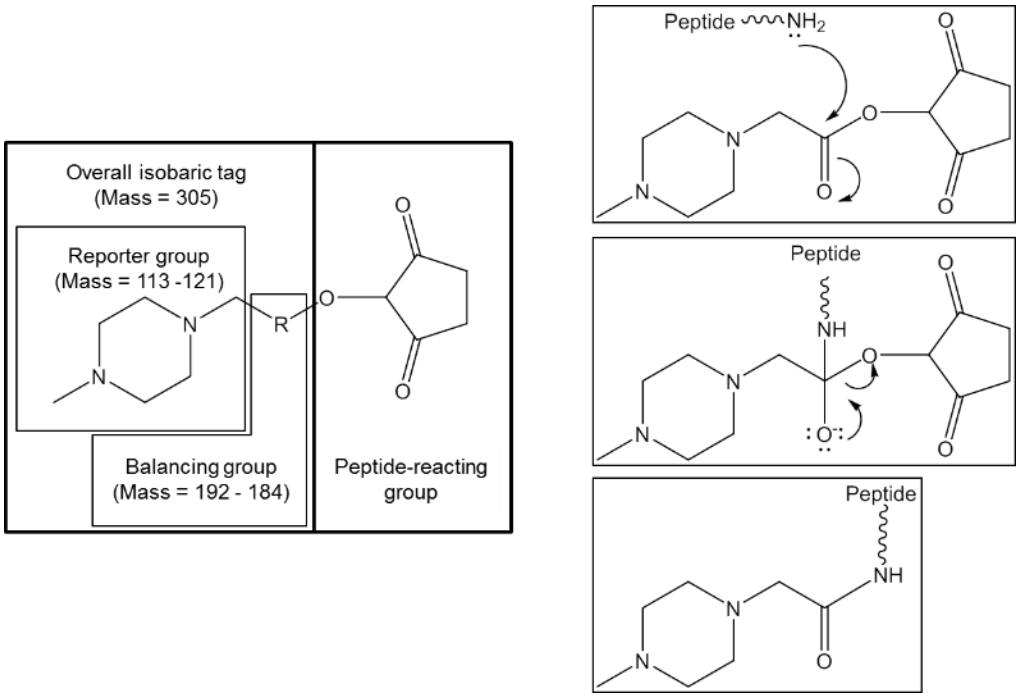


Figure 1.13. The chemical basis of iTRAQ and TMT peptide labelling. The left hand box describes the chemical groups responsible for each property of the iTRAQ and TMT reagents, specifically detailing the masses of the iTRAQ 8-plex labels. The right hand boxes describe the process of reductive amination, or amide coupling, in which peptides are covalently modified by iTRAQ or TMT labelling groups. This reaction also occurs at lysine residues, and is frequently neutralised by the use of hydroxylamine (H_2NOH).

1.8.4 ISOBARIC LABELLING LIMITATIONS

Under optimum conditions, TMT and iTRAQ workflows enable the effective qualitative and relative quantitative analysis of many thousands of proteolytic peptides derived from complex biological extracts. However, relative quantification may be impeded by isotopic impurities, labelling inefficiencies, intrinsic detection noise, and the typical limitations of bottom-up proteomics approaches [250]. Arguably, the greatest limitation to accurate relative quantitation when using multiplex isobaric labelling workflows lies in reporter ion ratio compression due to precursor ion co-isolation (**Figure 1.14**) [301, 302].

In principle, bottom-up proteomics attempts to describe the entire spectrum of proteins comprising highly complex biospecimens by proteolytic peptide characterisation. As a result of this complexity, peptides with similar mass-to-charge ratios and chromatographic retention times frequently co-elute and tandem mass spectrometry often lacks the fidelity required to discretely isolate each precursor ion. Co-isolation and co-fragmentation are, therefore, extremely common, causing distorted ratios due to the undifferentiable summation of the co-isolated precursors' reporter ions.

As only a minority percentage of proteins are modulated in many biological conditions, there is a high probability of differentially expressed peptides co-isolating with non-differentially expressed peptides; resulting in a compression of the detected relative abundance. When deriving protein expression ratios, equal averaging of all PSM reporter ion ratios further exacerbates compression.

A number of advanced MS-based approaches have been proposed to eliminate or minimise the effects of co-isolation. Approaches include the multi-notch MS^3 fragmentation, gas-phase purification, traveling wave ion mobility separation, extensive chromatographic separation and optimised data dependent acquisition settings [303-308]. Alternatively, post-data acquisition approaches to reducing the co-isolation effect from mainstream bottom-up proteomic workflows have been proposed. These are based upon either weighting or filtering their spectral outputs. One such approach attempts to filter out spectral data with less than 30% isolation interference [309]. Calculation of the signal-to-interference gives an improved indication of compression rates caused by co-eluting peptides [310]. The use of interference-defined cut-off values for differentially expressed peptide assignment also holds promise in improving quantitation accuracy [311]. Weighting correction has been applied to the intensity of reporter ion signals, given the reduced quantitation accuracy at low intensities [312].

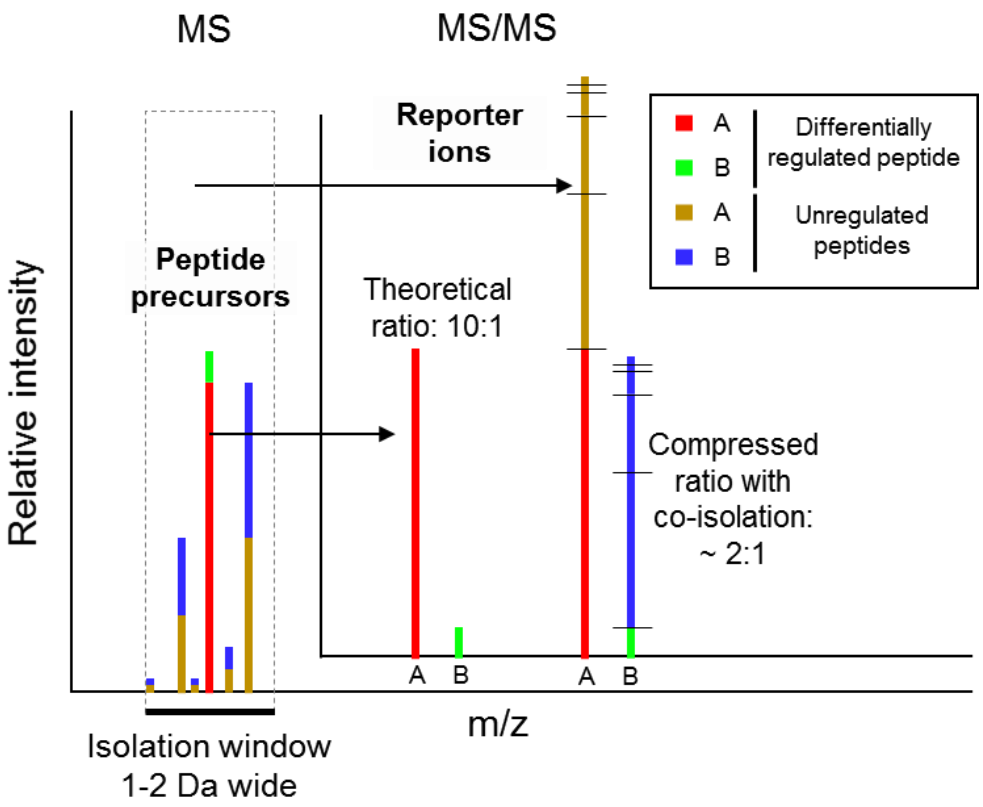


Figure 1.14. Ratio compression induced by peptide co-isolation. Co-isolation of additional peptide precursors (left) within the MS isolation window results in a compression to the detected reporter ion ratio at the MS/MS level.

1.9 PLASMA PROTEOMICS

Minimally invasive tests of blood fluids such as plasma and serum hold great potential for the early diagnosis of cancers, as well as many other disease states. As little as a millilitre of blood can provide insight into the many hundreds of homeostatic mechanisms operating inside the body [313]. Few blood-based protein biomarkers exist, however, which successfully allow the diagnosis of cancers [314, 315]. Cancer biomarkers offer the most promise when the protein can be traced to the cancer itself, offering a specificity to disease detection, and potentially correlating with progression. Such proteins, however often have a very low abundance in the blood, relative to the typical blood proteins [316]. Biomarkers can also be proteins or molecules that are upregulated in response to the presence of a cancer, such as an immune response and inflammation.

Albumin, along with around 20 other proteins, such as Ig and macroglobulins, make up over 99% of the protein content of plasma and serum [317]. The remaining ~1% is therefore difficult, to isolate and analyse. Shotgun proteomics approaches, due the use of DDA

characterisation which survey the most abundant peptides (top N), are severely impeded by this dynamic range of concentrations [317]. Plasma proteomics is often favourable over serum proteomics as the use of an anticoagulant can reduce the loss of proteins during clotting and can expedite sample processing [313].

To reduce this dynamic range and to enable efficient bottom-up characterisation of the lower abundant plasma/serum proteins, an approached termed 'immunodepletion' is frequently adopted. Immunodepletion uses antibodies specific to the most abundant proteins in serum/plasma to selectively remove those proteins which cause interference with shotgun proteomics [318]. While the principle is sound, this methodology suffers from limitations such as the removal of many proteins non-specifically and having a high degree of variability [319].

Another approach to reducing the dynamic range of proteins involves the use of size exclusion chromatography (SEC) [320]. SEC allows for the prefractionation of whole serum or plasma thereby isolating fractions dominated by certain proteins, which excludes these proteins from the other fractions. Under harsh denaturing, but non-reducing, conditions, SEC of whole plasma or serum produces 4 fractions; a high molecular weight fraction, an Ig-dominant fraction, an albumin-dominant fraction and a low molecular weight fraction. The isolation of these fractions, termed sub-proteome enrichment by SEC (SuPrE-SEC) enabled the identification of almost 2000 proteins in serum. This method was later developed to enable quantitative analysis with isobaric tags demonstrating reproducibility between technical replicates [321].

1.10 B-CELL CANCER PROTEOMICS

Many studies have taken a global approach to characterising the molecular nature of B-cell malignancies. Genomics [322, 323], microRNA expression [324-326] and transcriptomics [327, 328] have all been interrogated in an attempt to better understand the nature of B-cell cancers. Proteomic characterisations of B-cell cancers are less common due to relatively lower throughput and results yield [329, 330].

A large number of studies have utilised 2D gel electrophoresis in the proteomic analysis of B-cell malignancies [331-340]. Protein microarray have also been employed [341]. B-cell cancer proteomics experiments have predominantly focused on the analysis of material derived from cell lines [329, 337, 339, 340, 342-349], although a number of studies have considered primary material [333-335, 347, 350].

The specific focus of studies into B-cell malignancies has varied greatly with characterisation focusing on the effects of therapeutics on B-cell cancers [347, 351], the elucidation of biological mechanisms [342] and the identification of proteome relating to more

or less aggressive versions of a cancer type [333, 334]. Sub-proteomes of B-cell cancers, investigating proteomic expression amongst different cellular compartments has also been studied [338], with the most interrogated compartment being that of the cell surface proteome, with the most potential for therapeutic targeting with immunotherapy [346, 348, 350]. Consideration has also been given to the posttranslational modification of proteins in B-cell cancers using phosphoproteomics [343, 344] and glycoproteomics [346].

One of the most noteworthy of these studies was the MS characterisation of cell lines derived from B-cell lymphomas using SILAC [345]. This revealed over 7,500 proteins and 55 differential markers for each lymphoma subtype.

1.10.1 CLL PROTEOMICS

11 published studies have attempted to use proteomic methodologies to characterise CLL to date. The most recent and comprehensive published study attempted to elucidate biological signatures differentiating 9 M-CLL from 9 U-CLL samples using three iTRAQ 8-plexes. This study fully quantitated 2024 proteins, and while not identifying any discrete markers of IGHV mutation status, the data suggested a role for differential adhesion and migration pathways between U- and M-CLL [352]. Another noteworthy study characterised 12 CLL samples with isobaric-tagged shotgun proteomics, quantifying 655 proteins. A number of proteins were suggested as markers of poor prognosis, including TCL1 [353]. One study which compared healthy B cells to CLL with iTRAQ quantitation identified 536 proteins [354]. Proteomics was conducted to characterise the effects of B-cell response to BCR stimulation in primary CLL, identifying upregulated protein responses [355].

In addition to iTRAQ proteomics, other methods have been employed to characterise CLL, such as phosphoproteomics, characterising signalling induced by chemokines [356], membrane proteomics [357], ICAT characterisation of membrane and cytosolic proteins [358] and label-free 2D-LC MS/MS [359]. 2D gel electrophoresis was also conducted in several studies characterising CLL [360-362], of which one also focused on the characterisation of membrane proteins [363].

1.11 AIMS AND OBJECTIVES

The overall aim of this investigations was to implement the latest advances in quantitative proteomics to characterise, as comprehensively as possible, a selection of B-cell malignancies. While several published studies to date have described proteomic characterisations of B-cell cancers, none have successfully reached the full potential of quantitative proteomics in primary samples.

Eμ-*myc* and Eμ-*TCL1* represent two well-studied murine models of oncogene-driven B-cell tumourigenesis; in many aspects, providing good representations of the human cancers BL and CLL, respectively. These two models provide the opportunity to study the full developmental course of two differing B-cell cancer phenotypes. While BL and the Eμ-*myc* model present with solid, rapidly proliferating tumours, predominantly at a young age, CLL and the Eμ-*TCL1* model are characterised by a far slower expansion of B cells in several compartments in addition to displaying a blood-borne leukaemia which presents at an older age. Modelling these cancers in mice using Eμ-driven oncogenes, while partially artificial, allows predictable development under spontaneous conditions in the space of months rather than decades. It additionally allows for the characterisation of the effects of tumourigenesis on the model over its time course, through the signatures identifiable in the plasma, potentially revealing insight into how a whole organism progressively responds to pre-terminal and terminal tumours. The contrast between these two models poses interesting questions as to the similarities and differences in protein expression responsible for each cancer phenotype. Any commonalities between these contrasting cancers may have implications ranging further than just CLL and BL, especially in the context of other B-cell cancers. Findings such as targets of small molecular inhibitors and cell surface targets of immunotherapy promise additional, clinical potential. To date, no published studies have used primary samples, murine or otherwise, to quantitatively characterise these B-cell cancers or their respective plasma proteomes using MS proteomics.

Evaluation of primary human CLL samples also holds substantial potential in providing greater insight into the nature of the disease. Given the leukaemic population which emerges in CLL, the acquisition of purified cancer cell samples is achievable from a non-invasive blood sample without many of the contaminating factors and confounding issues which are present in solid tumour samples. Despite many years of research into CLL, prognostics and diagnostics have failed to offer consistently effective clinical benefits to patients. This leaves a clear requirement for a greater understanding of CLL and the heterogeneous nature of the disease. Characterisation of protein expression of CLL holds potential to identify novel immunotherapy targets, inhibition targets and subtype specific differences with wide ranging clinical implications. To date, the published studies performed using whole cell shotgun proteomics

comparing CLL, to healthy B cells, have achieved the characterisation of less than 1000 proteins. Other characterisations, not incorporating healthy B-cell controls have quantified around 2000 proteins in CLL.

Quantitative proteomics has the potential to provide quantitations for thousands of proteins simultaneously and non-biasedly for a set of up to 10 biological samples. Such near-comprehensive characterisations of protein, rather than mRNA, expression has far greater applicability to understanding the phenotype of cells. Proteins are also the molecules which are targeted by the majority of therapeutics, given their functional roles in promoting the hallmarks of cancer. Isobaric tags have emerged as one of the most widely adopted approaches in quantitative proteomics, however still remain limited by the effects of precursor co-isolation. This is an area which has several solutions, but few which are readily implementable to the majority of proteomics workflows. To most effectively utilise quantitative MS data, it was considered it might be possible to predict and non-biasedly select the most reliable peptide quantitations.

This investigation therefore aimed to:

- Evaluate and validate the success of preliminary quantitative proteomics data which characterised tumours of the Eμ-*myc* model
    - To concluded the validity of the isobaric tag proteomics quantitations derived from a 2D-LC MS/MS analysis
- Develop an approach to reducing the effects of co-isolation induced ratio compression on isobaric-labelled quantitations
- Use 2D LC-MS/MS proteomics to quantitatively characterise the tumour B-cell proteomes in the Eμ-*myc* and Eμ-*TCL1* models relative to that of WT and pre-cancerous Eμ-*myc* and Eμ-*TCL1* B cells
    - To improve upon current understanding of the B-cell cancers in these models
    - To identify cell surface protein expression suggestive of immunotherapy targets
    - To identify intracellular pathways and proteins which might offer targets of small molecular inhibitors
- Use SuPrE-SEC in combination with 2D LC-MS/MS proteomics to quantitatively characterise the blood plasma proteomes of the Eμ-*myc* and Eμ-*TCL1* model at pre-cancerous, early cancerous and terminally cancerous stages, relative to WT plasma proteomes

    - To identify proteins emerging in the plasma as a result of each tumour type
    - To understand the mechanisms by which these proteins appear in the plasma
    - To identify a protein signature in plasma prior to terminal tumour development
- Use 2D LC-MS/MS proteomics to quantitatively characterise proteomes of CLL, relative to healthy donor B cells
    - To improve upon the current understanding of CLL biology
    - To identify novel CLL cell surface markers and immunotherapy targets
    - To identify intracellular pathways and proteins which might offer targets of small molecular inhibitors
    - To identify subtype-specific differences that might explain the heterogeneity of CLL

2.0 MATERIALS AND METHODS

2.1 MATERIALS

Tris(hydroxymethyl)aminomethane (TRIS), sodium dodecyl sulphate (SDS), $MgCl_2$, disodium ethylenediaminetetraacetic acid (Na_2EDTA), NH_4Cl, $NaHCO_3$, agarose, NaCl, sodium deoxycholate (DOC), sodium dodecyl sulphate (SDS), guanidine hydrochloride, glycine, HPLC and LC-MS grade acetonitrile (ACN) and formic acid (FA) and 100 µm cell sieves were purchased from Fisher Scientific.

Tween20 (tween), sodium heparin, asparagine, 2-mercaptoethanol, sodium azide, propidium iodide (PI), sodium citrate, octylphenoxypolyethoxyethanol ($IgePalCA^{630}$), triton x-100, protease inhibitors, ponceau S, acetic acid, carboxyfluorescein succinimidyl ester (CFSE), methyl methanethiosulfonate (MMTS), tris(2-carboxyethyl)phosphine (TCEP), triethylammonium bicarbonate (TEAB), DMSO, hydroxylamine and ammonium hydroxide (NH_4OH) were purchased from Sigma.

Dulbecco's modified Eagle's medium (DMEM), Roswell Park Memorial Institute medium (RPMI), glutamine, pyruvate, penicillin and streptomycin were purchased from Life Technologies.

Coulter Isoton III dilutent and zapoglobin red blood cell (RBC) lytic reagent were purchased from Beckman Coulter. Mouse B-cell isolation kits and magnetic cell sorting columns were purchased from Miltenyi Biotech, Germany. *E. coli* JM109 competent cells, proteinase K, 2 x Green GoTaq buffer were purchased from Promega. 23 gauge needles, 1 ml syringes and FACS flow were purchased from BD Biosciences.

Foetal bovine serum (FBS) was purchased from Lonza, Lidocaine cream from Teva, 100 µl capillary tubes from Blaubrand, erythrolyse from Serotec, lyphoprep medium from Axis-Shield, interleukin 5 recombinant protein from Peprotech, CoolCell freezing regulator from Biocision, gel red from Biotium, O'Gene Ruler (0.1 µg/µl) 1 kbp ladder from Fermentas

3x red loading buffer and 1.25 M (30x) dithiothreitol (DTT) were purchased from Cell Signaling Technology. Precision Plus Protein All Blue Standard 10-250 kDa and AnykD gels were purchased from BioRad. Immobilon 0.45 µm pore polyvinylidene fluoride (PVDF) membrane was purchased from Millipore and non-fat milk from Marvel.

TMT 10-plex isobaric labelling reagents and 2 kDa M_w cut-off (MWCO) Slide-A-Lyzer dialysis cassettes were purchased from Thermo Scientific. Proteomics grade trypsin was purchased from Roche. iTRAQ 8-plex isobaric labelling reagents were purchased from ABSciex.

2.2 ANIMALS

Mice were bred and maintained in-house. They were housed with 12 hour light-dark cycles, with carefully monitored ambient temperature and humidity and were fed standard chow and water ad libitum. All procedures were carried out in accordance with home office licences PPL30/2450 and 30/2970 and PIL30/9925. Eµ-*myc* [C57BL/6J-TgN(Igh*Myc*)22Bri/J] hemizygous and Eµ-*TCL1* [C57BL/6J-TgN(Igh*TCL1*)22Bri/J] hemizygous mice were used. Eµ-*myc* mice and Eµ-*TCL1* mice were used as models of BL and CLL, respectively. These mice were observed daily for signs of lymphoid tumours, hunching, general signs of discomfort or shortness of breath indicative of tumour development. Eµ-*TCL1* mice were screened monthly for blood count and percentage of $B220^+$ $CD5^+$ B cells in the blood. Once the percentage of leukaemic $B220^+$ $CD5^+$ in the peripheral blood reached 50% of total white blood cells, mice were palpated weekly to detect splenomegaly. Eµ-*TCL1* mice were considered terminal when the blood percentage reached 80% leukaemic cells or when splenomegaly progressed to 3 cm in

length. Mice were then euthanised by exsanguination by terminal bleeding under isoflurane anaesthesia. Littermates and transgenic controls were housed and samples collected under identical conditions.

For the purposes of deriving B cells which could be compared between malignant, premalignant and healthy WT conditions, spleens were collected. Spleens were removed and placed in magnetic assisted cell sorting (MACS) buffer (PBS, 2 mM Na_2EDTA, 0.5% v/v FBS) on ice to provide biological material for analysis and cell culture.

2.3 Isolation of Human Peripheral Blood Mononuclear Cells

Ethical approval for the use of human samples was granted under REC reference 228/02/t. Peripheral blood mononuclear cells (PBMCs) were derived from healthy donors and CLL patients by Southampton Blood Services or Bournemouth Tissue Bank, respectively. PBMCs were isolated from whole blood or blood cones by density gradient centrifugation at 600 g for 20 minutes using lyphoprep medium and were washed with 40 ml of PBS. PBMCs were frozen in 10^7 cells/ml in FBS containing 10% DMSO (**Section 2.11.1**).

2.4 Genotyping

Ear punches were used as a means of identifying individual animals. Each ear punch was digested in 100 µl of DNA isolation buffer (50 mM TRIS (pH 8.9), 12.5 mM $MgCl_2$, 0.5% v/v Tween20, 0.5 µg/µl Proteinase K) overnight at 55°C. Samples were then heated to 95°C for 5 minutes and centrifuged at 15,000 g for 5 minutes. The supernatant contained the genetic material for PCR. The Eµ-*myc* and Eµ-*TCL1* transgene were detected with PCR and agarose gel electrophoresis to detect the PCR product (**Section 2.15** and **Section 2.16**). For Eµ-*myc*, the primers (annealing temperature; 55°C): 5'- CAG CTG GCG TAA TAG CGA AGA G -3' and 5'- CTG TGA CTG GTG AGT ACT CAA CC -3' generated a product of ~900 bp for the Eµ-*myc* transgene. For Eµ-*TCL1*, the primers (annealing temperature; 58°C): 5'- GCC GAG TGC CCG ACA CTC -3' and 5'- CAT CTG GCA GCA GCT CGA -3' generated a product of ~250 bp for DNA positive for the Eµ-*TCL1* transgene.

2.5 Screening for Eµ-*TCL1* Tumour Progression

Mice were warmed at 37°C for 10 minutes to increase vasodilation. Lidocaine cream was applied to the tail tip to produce local anaesthesia. Mice were restrained and less than 1 mm of tail tip was removed by scalpel blade. Petroleum jelly was applied to tail bases and massaged towards the tip. Blood samples were collected from tail tips with 100 µl capillary tubes containing 5 µl of heparin solution (50 µg/ml sodium heparin in PBS) and was placed on ice.

30-100 µl of heparinised blood was sampled monthly from Eµ-*TCL1* mice by tail bleeding. 10 µl of blood was immunostained with anti-CD5 FITC and anti-B220 PercP. Cells were washed as described in **Section 2.14.1**, treated with 1 ml of erythrolyse RBC lysis buffer and washed again. Flow cytometry analysis was used to determine the percentage of $CD5^+$ $B220^+$ lymphocytes in the peripheral blood which was considered as the leukaemia population. Lymphocyte count was determined by measuring the cellularity **(Section 2.7)** of blood diluted 1 in 5 with PBS.

2.6 SPLENOCYTE SUSPENSION

Spleens from Eµ-*myc*, Eµ-*TCL1* or wildtype littermates were washed with ~30 ml of MACS buffer and passed through 100 µm cell sieves, using 1 ml syringe plungers into 10 ml of MACS buffer. Splenocyte suspensions were centrifuged at 300 g for 5 minutes at 4°C, supernatant aspirated and the cell pellet resuspended in 40 ml of MACS buffer. Suspensions were passed through a second sieve and a 100 µl aliquot taken for cellularity determination (**Section 2.7**). Suspensions were then centrifuged and resuspended in the appropriate medium and volume as required.

2.7 CELLULARITY

Cellularities of splenocytes, PBMCs and cell lines were determined with a Coulter Industrial D Cell Counter (Beckman Coulter, UK) following manufacturer's instructions. 20 µl of cell suspensions were diluted in 10 ml of Coulter Isoton III dilutent with zapoglobin RBC lytic reagent if required.

2.8 B-CELL ISOLATION

For the isolation of B cells for both mouse and human samples, negative isolation was conducted using antibody cocktails which enabled magnetic depletion of non-B cells. Eluting without physical labelling of the B cells prevented any antibody binding to cell surface proteins in B cells which may cause downstream signalling altering protein expression. B cells were analysed by flow cytometry immunostaining with CD19 and CD3 to determine $CD19^+CD3^-$ B-cell purity. Viability of cells was assessed by propidium iodide staining (**Section 2.14.2**).

2.8.1 MOUSE SPLENIC B-CELL ISOLATION

Mouse B cells were isolated from whole splenocytes using the mouse B-cell isolation kit (Miltenyi Biotech) according to the manufacturer's instructions. This method isolated B cells by negative selection, using biotinylated antibodies targeted against all non-B cells; CD4, CD43 and Ter119. Streptavidin magnetic beads bind these antibodies. Using a MACS column coupled to a

strong magnet, all labelled non-B cells from a splenocyte suspension are retained, eluting enriched B cells. For proteomics experiments, mouse B-cell fractions were centrifuged at 300 g for 10 minutes at 4°C and fully aspirated. B-cell pellets were washed once in 20 ml of 1x red blood cell lysis buffer (155 mM NH_4CL, 10 mM $NaHCO_3$, 0.1 mM Na_2EDTA (pH 7.4)) before further washing in PBS as described in **Section 2.10**.

For tumours, where the cellularity was far higher than a normal spleen due to the expansion of the tumour cells, the use of MACS cocktails proportional to the cell number of non-B cells would have been inefficient. Optimisation demonstrated that in the presence of tumour cells, 100 µl of antibody cocktail was sufficient to deplete non-tumour cells from a single tumour-bearing spleen. Based on this, 50 µl of antibody cocktail would have been sufficient, as 1/5 of a tumour-bearing spleen was efficiently isolated with just 10 µl. However, given the variable cellularity within tumour-bearing spleens, an excess of 100 µl was concluded to be sufficient to consistently ensure high sample purity.

2.8.2 HUMAN PERIPHERAL BLOOD B-CELL ISOLATION

Human B cells were isolated from PBMCs by negative B-cell enrichment using the EasySep human B-cell enrichment kit without CD43 depletion. CD43 depletion was avoided, as CLL has been shown to have CD43 expression in some cases. As with MACS, all non-B cells were labelled with antibodies specific to CD2, CD3, CD14, CD16 and CD56 and glycophorin A, which were subsequently bound by magnetic beads. Rather than a column, cells were placed within a large magnet in a 5 ml tube. When cells were poured off, non-B cells remained in the magnet. The non-magnetically retained cell fraction containing enriched B cells was washed with PBS as described in **Section 2.10**.

2.9 CELL CULTURE

Cell lines derived from Eµ-*myc* mouse tumours were grown in 'Eµ-*myc* media' (DMEM supplemented with 2 mM glutamine, 1 mM pyruvate, 45 units/ml penicillin, 45 µl/ml streptomycin, 200 µM asparagine, 50 µM 2-mercaptoethanol and 15% FBS. Initially, splenocytes in a single cell suspension from a tumour-bearing Eµ-*myc* animal were cultured at a density of $5x10^7$ cells/ml in Eµ-*myc* media supplemented with 15% FBS. After approximately 2 passages, or when the cell culture demonstrated steady proliferation and minimal cell death, cells were adapted to Eµ-*myc* media supplemented with 10% FBS. Cell cultures were maintained for 10-15 passages, with cells typically being harvested for analysis between 5 and 10 passages. Eµ-*TCL1* splenocyte suspensions were cultured in Eµ-*myc* media at a density of $5x10^6$ cells/ml. These were cultured at 37°C in 10% CO_2 in six well plates. Cells were cultured for 48 hours

prior to any treatment to acclimatise cells. Cells were cultured for a maximum of 7 days due to Eμ-*TCL1* splenocytes not readily proliferating in culture.

HeLa and MCF7 cells were cultured in RPMI and DMEM, respectively, supplemented with 10% FBS at 37°C, 5% CO_2. Cultures for proteomics were passaged and harvested by scraping at 70% confluence to avoid protein digestion that would result from trypsin-based passage. *E. coli* JM109 competent cells were cultured in lysogeny broth at 37°C for 24 hours with continuous agitation.

2.10 Cell Washing For MS Samples

Cells were washed at least 3 times in 40 ml of sterile PBS, centrifuging at 300 g for 5 minutes and fully aspirating and resuspending with each wash. Mouse cells were washed with ice cold PBS and centrifuged at 4°C, human samples were maintained at room temperature and *E. coli* cells were centrifuged at 5000 g. Washing in PBS was repeated to remove contaminants that may be bound to the cells, predominantly FBS proteins, as well as cell debris.

2.11 Cell Freezing

2.11.1 Freezing Cells to Maintain Viability

To maintain cell viability during freezing, 10^7 cells/ml were resuspended in FBS with 10% v/v DMSO and transferred to a cryovial. Cells were frozen using a CoolCell freezing regulator to lower the temperature at -1°C/minute to minimise ice crystal formation. These were placed in -80°C for one week and transferred to liquid nitrogen for long term storage.

2.11.2 Snap Freezing for Cell Lysates

For the purposes of cell lysate generation, cell pellets were rapidly frozen in liquid nitrogen. Cells were suspended in PBS and transferred to a 1-2 ml tube and centrifuged at 2000 g for 5 minutes. Supernatants were fully discarded and tubes were snap frozen in liquid nitrogen.

2.12 Terminal Bleeding

Blood collection by terminal vena cava bleeding (TVCB), adapted from [364] was used to avoid issues with the standard terminal cardiac bleeding, including; red blood cell lysis, cardiac displacement by thymomas and other tumours and potential contamination from the puncturing of several layers of tissue to reach the heart. Mice were anaesthetised with isoflurane with O_2 until unconscious. The isoflurane supply mask was secured to a dissection board. Complete anaesthesia was assessed by applying strong pressure to all feet and observing reactions. Once unresponsive, the mice were pinned by their limbs and sprayed liberally with 70% ethanol to ensure the sterility and removal of contaminants. A midline incision was used

from the base of the peritoneum to the base of the sternum, and laterally across each skin flap at around half way up the peritoneum to avoid major blood vessels. Organs were moved to one side, while avoiding unwanted haemorrhage, revealing the inferior vena cava located to the left of the vertebral column, behind the liver. Ethanol was again sprayed liberally. 50 µg/ml sodium heparin in PBS was drawn with a 23 gauge needle attached to a 1 ml syringe and all the air expelled, so that heparin filled the dead space. Needles were placed by bevel-down venepuncture of the inferior vena cava just below the junctions of the hepatic and renal veins. Blood samples were drawn slowly so as not to cause vessel collapse, taking around 60-90 seconds to allow refilling. Needles were withdrawn from the vein and the mouse second-killed by neck dislocation. Blood was ejected gently from the syringe (after needle detachment) and placed on ice.

2.13 PLASMA ISOLATION

Heparinised blood samples were stored on ice prior to centrifugation at 2000 g for 15 minutes at 4°C. Plasma supernatant was taken and subjected to a further centrifugation at 2000 g for 5 minutes at 4°C and repeated if red blood cell contamination remained. Samples were rejected if red blood cell lysis was visible by eye. Plasma was stored in liquid nitrogen.

2.14 FLOW CYTOMETRY

2.14.1 IMMUNOPHENOTYPING

2 x 10^5 cells were resuspended in 100 µl with MACS buffer and either the manufacturer's recommended concentration, or 10 µg/ml, of antibody (**Table 2.1**) were added and incubated for 30 minutes in the dark. Cells were washed with 3 ml of FACS washing buffer (PBS, 1% w/v BSA and 10 mM sodium azide), centrifuged at 180 g, the supernatant discarded and cells resuspended in 100 µl of FACS flow for flow cytometry analysis by a FACScan or FACScalibur (BD). Unstained cells were used to adjust the voltage of the forward scatter (FSC) and side scatter (SSC) and to adjust the voltages of the FL1, FL2 or FL3 laser intensities, so that unstained events fell below 10^1 on a logarithmic plot. Cells stained with each antibody were individually used for compensation to ensure that each fluorescent dye was detected discretely in its intended FL channel. Typically 10,000 or 50,000 events were captured. The results were analysed by Cyflogic (CyFlo Ltd, Finland).

2.14.2 CELL VIABILITY ANALYSIS (PI EXCLUSION)

2 x 10^5 cells were suspended in 100 µl with MACS buffer, PI added to a final concentration of 3.6 µg/ml and incubated at room temperature for 1 minute. PI is excluded from live cells by the integrity of the plasma membrane and the polar charge of the PI molecule. PI

penetrates dead cells with no membrane integrity, binding DNA with high affinity. Binding DNA results in an approximately 100-fold increase in fluorescent intensity detectable in FL2/FL3.

2.14.3 Cell Cycle Analysis (Hypotonic PI)

5 x 10^5 cells were suspended in 0.5 ml of hypotonic PI solution (50 µg/ml PI, 0.1% w/v sodium citrate, 0.1% w/v triton x-100), vigorously vortexed and incubated at 4°C for 15 minutes. This simultaneously causes cell lysis while leaving the nuclei intact and allowing the PI to stain the DNA. The resulting cell nuclei were analysed by flow cytometry (FL2 on linear scale) by a FACScan at the lowest flow rate.

2.14.4 Cell Division Analysis (CFSE Staining)

Cells were resuspended in PBS at a concentration of 10^7 per ml. Carboxyfluorescein succinimidyl ester (CFSE) was added at a final concentration of 5 µM by rapid mixing to give an equal distribution of the label across all cells. After a 15 minute incubation at room temperature, an equal volume of FBS was used to neutralise the CFSE. After treatment, cells were then exposed to various assay conditions, after which cell fluorescence was measured using flow cytometry with CFSE detectable by FL1. Cells undergoing division distribute half of their stained contents into each daughter cell and therefore demonstrate a 50% reduction in fluorescence.

2.15 PCR

1 µl of supernatant containing isolated DNA was mixed with 1 µl (10 pM) of each primer, (diluted in TE (10 mM TRIS pH 8, 1 mM EDTA), 10 µl of 2 x Green GoTaq buffer and 7 µl of ultrapure water. The PCR was performed using a C1000 thermal cycler (Biorad) as follows: **1.** Initial denaturing; 95°C for 5 minutes, **2.** Denaturing; 95°C for 1 minute, **3.** Annealing**;** primer-specific annealing temperature for 1 minute, **4.** Elongation; 74°C for 2 minutes, Steps 2-4 repeated 30 times then **5.** 70°C for 10 minutes and; **6.** 4°C indefinitely.

2.16 Agarose Electrophoresis

DNA samples were subjected to electrophoresis through 1% w/v agarose gels made in 40 mM tris, 20 mM acetic acid, 1 mM EDTA (TAE) buffer containing 0.005% v/v gel red at 150 volts for 30 minutes, alongside 7 µl O'Gene Ruler (0.1 µg/µl) 1 kbp ladder, and known positive and negative control samples for the transgene being evaluated.

2.17 SDS-PAGE AND WESTERN BLOTTING

2.17.1 CELL LYSIS

Cells pellets were resuspended in 30µl of radioimmunoprecipitation assay (RIPA) cell lysis buffer (0.15 M NaCl, 1% v/v octylphenoxypolyethoxyethanol (IgePalCA^{630}), 0.5% w/v sodium deoxycholate (DOC), 0.1% w/v sodium dodecyl sulfate (SDS), 0.05 M TRIS (pH 8) and 1% v/v protease inhibitor (Sigma)) per million cells and incubated on ice for thirty minutes. Lysates were triturated 20 times with a 23 gauge needle and centrifuged at 16,000 g for 5 minutes at 4°C and the supernatant collected. The centrifugation step was repeated if the sample remained cloudy. Lysates were stored at -20°C or -80°C long term.

2.17.2 PROTEIN QUANTIFICATION

The protein concentration of cell lysates was determined by bicinchoninic acid assay (BCA) protein quantification kit (Thermo Scientific) according to the manufacturer's instructions. The kit was used to estimate the protein concentration of samples relative to a standard curve of bovine serum albumin (BSA). The working reagent reacted with protein to induce an absorbance change with proportionality to the protein present in a sample. This colour change was measured at 562 nm with a Varioskan plate reader (Thermo Scientific). The BSA standard curve covered a concentration range of 0 – 2000 µg/ml and lysates were diluted 1 in 10 with cell lysis buffer to consistently fall within this range. Standards were diluted in cell lysis buffer to maintain consistency with cell lysates.

20 µl of each standard and each 1 in 10 diluted lysate were loaded into a 96-well plate in triplicate with 200 µl of the working reagent, agitated for 5 minutes and incubated at 37°C for 30 minutes to ensure the reaction was complete. The equation of optical density relative to protein concentration of the BSA standard was inverted, making protein concentration the subject, and applied to the unknown lysates to determine the protein concentration from optical density at 562 nm.

2.17.3 SDS-PAGE

Lysates were diluted based upon their concentration to derive a consistent protein content and volume. Lysates were mixed with 3 x red loading buffer, 30 x DTT (final concentration 41.7 mM), and heated at 95°C for 5 minutes. Denatured lysates and 10 µl of Precision Plus Protein All Blue Standard 10-250 kDa were resolved with AnykD gels at 200 volts for 40 minutes in SDS-PAGE running buffer (25 mM TRIS, 192 mM glycine, 0.1% w/v SDS).

2.17.4 WESTERN TRANSFER

Proteins were transferred by 'wet transfer' to Immobilon 0.45 µm pore PVDF membrane at 100 volts for 90 minutes in transfer buffer (SDS-PAGE running buffer with 20% v/v methanol). The PVDF membrane was first pre-activated for 30 seconds in 100% methanol, and then placed between two pieces of filter paper and two sponges arranged as follows: Positive electrode, sponge, filter paper, membrane, gel, filter paper, sponge, negative electrode.

Ponceau staining was used to demonstrate successful protein transference and assisted in cutting membranes if required. After transfer, membranes were washed in methanol, agitated in 0.1% (w/v) ponceau S in 5% (v/v) acetic acid for 5 minutes, agitated in distilled water for 5 minutes and imaged. Ponceau stain was removed by washing in PBS with 0.05% (v/v) tween (PBST) for 5 minutes.

Membranes were then placed in 5% (w/v) non-fat milk in PBST for 1 hour to neutralise the hydrophobicity of the PVDF and prevent further, unwanted protein association.

2.17.5 IMMUNODETECTION BY WESTERN BLOTTING

Western blotting (WB) was conducted with primary antibodies diluted in blocking buffer (5% (w/v) non-fat milk in PBST (PBS with 0.05% (v/v) tween). **Table 2.2** identifies the manufacturer, specific dilutions, incubation conditions used and the animal in which the antibody was derived. Membranes were washed 4 times for 5 minutes with PBST.

Secondary antibodies used were based on the animal in which the primary antibody was derived; described in **Table 2.2**. All were diluted in 2.5% (w/v) non-fat milk in PBS with 0.05% (v/v) Tween and 0.01% (w/v) SDS. Secondary antibodies were conjugated to one of two fluorescent dyes either fluorescing at 680 nm (red) or 800 nm (green). Membranes were washed as before but were also washed twice briefly with PBS due to a low level of fluorescence possessed by the tween detergent.

Membranes were imaged using an Odyssey Imager (Licor) which scans for the two specific wavelengths of light corresponding to the emission spectra at 680 nm (red) or 800 nm (green). These have no overlapping emission spectra so give highly discrete signal detection of protein bands when using secondary antibodies conjugated to the two dyes. Scans were set to 50 nm to give a high resolution for the best quality image for relative quantification. Scanning intensities were set to 11.0 for 800 nm and 8.0 for 680 nm. All other settings used were as default.

Mouse anti-GAPDH antibody was used as a loading control to ensure that the levels of protein loaded for each sample was consistent.

2.17.6 WB Expression Analysis

Relative quantification of band fluorescence intensity was derived using Image Studio 2.0 software (Licor). Each protein band intensity was divided by the band intensity of GAPDH for each sample to gain a normalised value for protein expression.

Normalised values for relative protein expression in each lysate were dividing by the normalised value for WT B-cell lysate. This provided a ratio of expression relative to the WT B cells.

A value describing the concordance between the WB-derived relative expressions with MS-derived relative expression was determined. This was determined by dividing the $\log_2$ (WB-derived ratio) by the respective $\log_2$ (MS-derived ratio).

2.18 Plasma Sub-proteome Enrichment

2.18.1 Size Exclusion Chromatography

120 µl of plasma was diluted with 380 µl of 6 M guanidine hydrochloride in 10% methanol and centrifuged at 16,000 g for 5 minutes to remove any insoluble precipitates or particles which could block the filter. The supernatant was loaded onto a 500 µl loading loop and resolved on the basis of protein size by three KW804 size exclusion chromatography (SEC) columns (300 mm length x 8 mm inner diameter (ID) x 7 µm particle size) (Shodex, Japan) with 6 M guanidine hydrochloride in 10% methanol at 1 ml/min flow rate with a LC-20AD high performance liquid chromatography (HPLC) system (Shimadzu) maintained at 30°C. Eluting protein was detected over 50 minutes at a wavelength of 280 nm, which reproducibly identified six peaks; the first three peaks eluting between 22 and 31 minutes were treated as a high molecular weight (M_w) sub-proteome, the peak at 31-34, as the Ig-dominant sub-proteome, 34-42, the albumin sub-proteome and the material eluting after 42 minutes as the low M_w sub-proteome.

2.18.2 Dialysis and Protein Purification

Each low M_w sub-proteome, solubilised in 6 M guanidine, was dialysed using 12 ml 2 kDa M_w cut-off (MWCO) Slide-A-Lyzer dialysis cassettes with five 5 L 18.2 MΩ/cm water exchanges at 8 hour intervals. Soluble protein was extracted from cassettes and additionally each cassette rinsed with 3 ml of 0.1 M TEAB to extract any precipitated protein. Protein was lyophilised using a vacuum concentrator at room temperature (Thermo Scientific). The resulting protein was solubilised in 0.5 M TEAB with 0.05% SDS and processed and analysed by 2D-LC MS/MS as described in **Section 2.19**. Due to the low protein yield of the low M_w sub-proteome,

only 30 µg of protein material (rather than the typical 100 µg) was digested and labelled for MS analysis.

2.19 QUANTITATIVE MASS SPECTROMETRY PROTEOMICS

2.19.1 LYSATE PREPARATION

Snap frozen cell pellets were initially lysed on ice by trituration with a 23 gauge needle in 10 µl per million cells of 0.5 M TEAB with 0.05% SDS. If the lysate was too viscous to triturate, a further 50% buffer was added. Disrupted cells were then sonicated by an Ultrasonic Processor XL (Qsonica) at 20% amplitude for 2 x 30 seconds on ice, with a 30 second interval. Trituration allowed a greater degree of control over the initial cell lysis and ensured a good degree of disruption to cells prior to sonication providing maximum and consistent protein extraction. Lysates were cleared at 16,000 g for 10 minutes at 4°C and transferred to a fresh tube. Clearing was repeated if the lysate remained cloudy.

2.19.2 PEPTIDE PREPARATION

Lysate protein concentration was determined by a Direct Detect spectrometer (Millipore) analysing 2 µl of lysate for amide bond absorbance between wavenumbers 1702 and 1602 cm^{-1} or by BCA assay. 100 µg of lysate for each samples was reduced with 50 mM TCEP for 60 minutes at 60°C and were alkylated with 200 mM MMTS for 15 minutes at room temperature. Denatured protein material was digested overnight at room temperature in the dark, using 30:1 proteomics grade trypsin.

2.19.3 ISOBARIC TAG PEPTIDE LABELLING

Peptides were incubated with either iTRAQ 8-plex or TMT 10-plex isobaric tags according to the manufacturer's instructions. Reactions were allowed to proceed for 2 hours at room temperature, and were quenched with 16 µl of 5% v/v hydroxylamine for 15 minutes at room temperature. Hydroxylamine (H_2N-OH) contains a primary amine which readily reacts with the mass tag peptide reacting group preventing any labelling of the incorrect peptides once pooled. Reactions were individually lyophilised at room temperature in a vacuum concentrator (Thermo Scientific). Lyophilised, labelled peptides were serially reconstituted in a single volume of 100 µl of 2% v/v ACN, 0.1% v/v NH_4OH.

2.19.4 PEPTIDE PRE-FRACTIONATION

Peptides were first resolved using high-pH (0.1% v/v NH_4OH) RP C8 chromatography (150 mm x 3 mm ID x 3.5 µm particle, XBridge, Waters) at 300 µl/min with a LC-20AD HPLC system (Shimadzu) maintained at 30°C, using the mobile phases (MP); A – 99.9% H_2O, 0.1% NH_4OH, B – 99.9% ACN, 0.1% NH_4OH. The typical two hour gradient was as follows; 0

minutes; 2% B, 10 minutes; 2% B, 75 minutes; 30% B, 105 minutes; 85% B, 120 minutes; 2% B. Prior to peptide injection, the column was first run through a full gradient without sample to remove all contaminating material and equilibrated with 2% MP B for 30 minutes. Fractions were collected in a peak-dependent manner and were lyophilised. High pH chromatography was used to orthogonally complement the low pH separation subsequently conducted for each peptide fraction.

2.19.5 Peptide Fraction Resolution and Characterisation by LC-MS

Lyophilized peptide fractions were individually reconstituted in 30 µl of 'loading' MP (2% ACN, 0.1% FA) and 10 µl loaded by a Dionex Ultimate 3000 (Thermo Scientific) at 20 µl/minute for 4 minutes onto a C18 PepMap100 trapping cartridge (5 mm × 500 µm ID, 5 µm particle) (Thermo Scientific) in 'loading' MP. After peptide loading, the trapping cartridge was brought in line with an Acclaim PepMap 100 column (50 cm × 75 µm ID, 3 µm particle) at a flow rate of 300 nl/minute with 2% MP B (94.9% ACN, 5% DMSO, 0.1% FA) in MP A (2% ACN, 5% DMSO, 0.1% FA). Several reverse phase elution gradient lengths were used, proportionally extrapolated from the 100 minute gradient: 0 minutes; 2% B, 5 minutes; 2% B, 50 minutes; 22% B, 75 minutes; 45% B, 85 minutes; 85% B, 90 minutes; 85% B, 95 minutes; 2% B.

Peptide elution was directly coupled to electrospray ionisation at 2.4 kV using a PicoTip nESI emitter (New Objective), and were characterised with an Orbitrap Elite Velos Pro mass spectrometer (Thermo Scientific). MS characterisation of eluting peptides was conducted between 350 and 1900 m/z at 120,000 mass resolution. The top 12 +2 and +3 precursor ions per MS scan (minimum intensity 1000) were characterised by tandem MS with high-energy collisional dissociation (HCD) (30,000 mass resolution, 1.2 Da isolation window, 40 keV normalised collision energy) and collision induced dissociation (CID) (ion trap MS, 2 Da isolation window, 35 keV) with a dynamic exclusion (±5 ppm) of 60 seconds. A maximum of 200 ms ion injection time was allowed. Additionally the DMSO ion at 401.922718 was used as a MS lockmass and, where applicable, the TMT-H+ ion at 230.170422 was used as an internal MS/MS lockmass.

2.19.6 MS Data Processing

Target-decoy searching of raw spectra data was performed with Proteome Discoverer version 1.4.1.14 software (Thermo Scientific). HCD and CID spectra were first extracted from raw files due to requiring different search parameters. Spectra were subject to a two stage search, both using SequestHT, with Percolator used to estimate FDR with a threshold of $q \leq 0.01$. The

first allowed only a single missed cleavage, minimum peptide length of 7, precursor mass tolerance of 5 ppm, no variable modifications and searched against the human/mouse UniProt Swissprot database (downloaded 01/15). The second search used only spectra with q > 0.01 from the first search, allowed 2 missed cleavages, minimum peptide length of 6, searched against the human/mouse UniProt trembl database (downloaded 01/15), precursor mass tolerance of 10 ppm and a maximum of 2 variable (1 equal) modifications of; TMT or iTRAQ (Y), oxidation (M), deamidation (N,Q) and phospho (S,T,Y). In both cases, fragment ion mass tolerances of 0.02 Da for the HCD spectra and 0.5 Da for the CID spectra and fixed modifications of Methythio (C), TMT or iTRAQ (K and N-terminus) were used. Reporter ion intensities were extracted from non-redundant PSMs with a tolerance of 20 ppm.

2.19.7 Quantitative Analysis of MS Data

Peptide spectrum match data was exported from Proteome Discoverer with the associated features: isolation interferences, ion injection times, posterior error probability (PEP) scores, charge states, precursor intensities and the reporter ion intensity values. Data was submitted to the Statistical Processing for Isobaric Quantitation Evaluation (SPIQuE) (detailed in **Section 2.20** and **Chapter 4**) and processed with the default '-10:-9:-5:-5:+3' weighting, with median normalisation and T-test. For the purposes of comparing differential protein expression between two experiments for example, in the case of the CLL characterisation, ratios for each CLL sample to each of the three healthy 'bridging' controls were generated and of these, to reduce the effect of noise and to conserve only the most consistent observations, only the least deviated of the three $\log_2$ (ratios) were used for further analyses. For the mouse data, ratios were generated describing each tumour pool relative to the two WT controls.

The resulting $\log_2$ (ratios) were then used to define the 'regulation score' (RS), a value that described both the magnitude and consistency of the considered ratios. Where previous analyses have relied on filtering by the average and the standard deviation of technical or biological replicates [321, 365] this has a tendency to exclude proteins with a high standard deviation resulting from high fold change values. Averaging also allows single outlier $\log_2$ (ratios) to be over-represented giving a false impression of regulation. To normalise against this, RS was defined as that of the average of the $\log_2$ (ratios) divided by 1 + the standard deviation. A high absolute RS was therefore indicative of the proteins with the highest average and lowest standard deviation, highlighting the most consistently up or downregulated proteins. For unregulated or inconsistently regulated proteins the RS tended towards 0.

The thresholds at which a RS could be considered as indicative of differential protein expression were extrapolated from the principles described by Levin, 2011 [366] based on the

number of available sample comparisons and total variation. Variation was considered for each protein by the incorporation of SD as a denominator into the RS.

For the analysis of mouse samples where 4 biological/technical replicates were available, a threshold of ±0.5 was used. Using this threshold, the lowest detected fold change was 1.44 with an average variation of just 5.4%

For the analysis of human samples, 13 biological replicates were analysed and an approach was taken considering only the minimum of ratios relative to the three controls. This provided a higher statistical power allowing for a lower threshold of ±0.25. The lowest fold change concluded as differentially expressed was 1.21 with a standard deviation (SD) of 0.08.

2.19.8 Statistical Analysis of MS Data

Statistically significant differential expression was derived from the $\log_2$ (ratios) of biological and technical replicates by a one-sample (when describing regulation versus the ratios' denominator) or two sample, FDR-corrected t-test. For multiple test correction, p-values were ordered from smallest to largest and each FDR-corrected p-value determined as the minimum value of either; the uncorrected p-value multiplied by the ratio of the total number of tests and the p-value's ranking in the list; or the corrected p-value immediately after this value. In all cases, a p-value of <0.05 was considered significant. In combination with the RS, statistics gave a complementary measure of differential protein expression, and in all cases differential proteins expression was defined as those proteins for a given comparison reaching an RS threshold as well as a p-value of <0.05. Ratio statistics derived by SPIQuE were not used for the determination of these statistics.

2.20 SPIQuE Development

2.20.1 *E. coli* Spike-in Design

Peptides from *E. coli*, HeLa and MCF7 cells were generated by methods described in **Section 2.19.** Peptides were assigned to 0.8mg of each TMT 10-plex label as follows: **126**- 100 µg HeLa, **127N**-50 µg HeLa, 0.25 µg *E. coli*, **127C**-20 µg HeLa, 1.25 µg *E. coli*, **128N**- 2 µg HeLa, 5 µg *E. coli*, **128C**- 25 µg *E. coli*, **129N**-1.4 µg MCF7, 12.5 µg *E. coli*, **129C**-14 µg MCF7, 2.5 µg *E. coli*, **130N**-35 µg MCF7, 0.5 µg *E. coli*, **130N**-70 µg MCF7, **131**-no peptide material added (buffer only). The labelled peptides were subjected to a 2D-LC MS/MS workflow. After fractionation peptides were combined into 13 pools for LC MS/MS analysis with HCD and CID fragmentation.

2.20.2 SPIKE-IN DATA ANALYSIS

Peptide spectrum match data was generated and exported from Proteome Discoverer with all associated features including the reporter ion intensity values and precursor intensity (not displayed by default). To minimize the overlap of peptides common to both species, a minimum peptide length of 7 was used. PEP was calculated by Proteome Discoverer, defined as the probability of an incorrect PSM assignment. Other measures of peptide match probability will likely provide an equivalency when used in place of the PEP score. Isolation interference, also known as precursor isolation purity, was calculated by Proteome Discoverer by the formula:

$$Isolation\ interference = \left(1 - \left(\frac{\sum Isolated\ precursor\ peaks}{\sum All\ isolated\ peaks}\right)\right) x\ 100$$

2.20.3 DEFINING THE CO-ISOLATION SCORE AND MULTI-FEATURE WEIGHTING MODEL

E. coli PSM data was first filtered to exclude any PSM reporter ion quantitations that were not able to derive a real number score from a ratio of 128C to 127N or 130N. Three 'co-isolation scores' were then calculated for each PSM, giving the following theoretical, uncompressed ratios:

$$log_2\left(\frac{(128\mathrm{C}\ +\ 129\mathrm{N})}{(126\ +\ 127\mathrm{N}\ +\ 130\mathrm{N}\ +\ 130\mathrm{C})}\right) \cong\ log_2(50{:}1)\ \cong\ 5.64$$

$$log_2\left(\frac{128\mathrm{C}}{(126\ +\ 127\mathrm{N}\ +\ 129\mathrm{C}\ +\ 130\mathrm{N},\mathrm{C})}\right) \cong\ log_2(5.6{:}1)\ \cong\ 2.49$$

$$log_2\left(\frac{128\mathrm{C}}{(126\ +\ 127\mathrm{N},+\ 129\mathrm{N},+\ 130\mathrm{N},\mathrm{C})}\right) \cong\ log_2(1.5{:}1)\ \cong\ 0.58$$

For each of the PSM features, comparable scores were calculated using percentile ranking (PERCENTRANK function in Microsoft Excel 2013). Scores for features correlating negatively with co-isolation scores, including; isolation interference, ion injection time and PEP were calculated using 1-percent rank. Using these percentile ranked scores, correlations of single or multiple features with the co-isolation scores could then be comparatively evaluated using linear regression. The machine learning validation of this approach (courtesy of Yawwani Gunawardana) is detailed in **Appendix A1**.

2.20.4 ONLINE IMPLEMENTATION OF AN R SCRIPT APPLYING THE MULTI-FEATURE WEIGHTING MODEL

A script was written in R statistical programming language (courtesy of Cory White) to enable the automated implementation of weighting to ratios and statistical calculation. This script was designed to take an input file in .CSV format containing reporter ion intensities and features for each PSM and use an experimental design file specifying the nature of the input file.

Lines of code were then used to specify the type of analysis and which ratios to be determined. The output file was generated based on this user input.

This script was incorporated into an online interface (courtesy of Oliver Bills) hosted at spiquetool.com. This involved the upload of a .CSV file and the generation of the experimental design file on the basis of the uploaded .CSV headers and user specified weighting factors. The user could then specify the ratios, statistical test and normalisation parameters, all through a simplistic user interface with dropdown menus, recommended defaults and instructions. These options then generated the lines of code creating custom script automatically which was then submitted by the tool to R for processing. The output was then provided as a compressed file for download. The full user guide for SPIQuE is presented in **Appendix A2**.

2.20.5 STATISTICAL TESTING OF HELA:MCF7 AND PRIDE DATA WITH SPIQUE

Data was exported from Proteome Discoverer with the quantitation setting adjusted to show raw quant values, no minimum threshold, no missing value replacement, apply quan value corrections, maximum fold change of 100 (but allowing usage above this), and co-isolation exclusion off (set to 100). Ungrouped peptides (PSMs) were filtered by quantitation usage. Field chooser was then used to select the following columns: protein group accession, reporter ion raw values, PEP, isolation interference, ion injection time, intensity and charge. The .xlsx file was converted to comma-delimited .csv format and submitted to spiquetool.com. Weightings and correlation directions of: isolation interference (10 -), ion injection time (9 -), PEP (5 -), charge (5 -) and intensity (3 +) were used. Normalisation was set to median, statistical test to T-test, missing values were not replaced and where weighting was required, a final exponential weighting of 1 was used, otherwise, this was set to 0. To align the SPIQuE-derived results with the additional proteome information (protein name, number of peptides etc.), the 'MATCH' function in excel was used to match and sort based on protein accession numbers.

2.21 BIOINFORMATICS

Several approaches were applied to attempt to identify topological and mechanistic patterns in each data set. Different approaches were taken to each data set on the basis of data complexity and each biological question being evaluated.

2.21.1 HIERARCHICAL CLUSTERING ANALYSIS

Protein lists with $\log_2$ (ratios) were submitted to Cluster 3.0 (University of Tokyo, Human Genome Centre). Hierarchical clustering was performed for conditions using a similarity metric of Euclidian distance and complete linkage. Java TreeView (version 1.1.6r2) (http://jtreeview.sourceforge.net/) was used to visualise the clustering.

2.21.2 Principal Component Analysis

Principal component analysis (PCA) was conducted using Qlucore Omics Explorer 3.1 (64-bit version) (Qlucore). The $\log_2$ (ratios) for all fully quantitated proteins were imported into Qlucore and the PCA option used to visualise the separation of the data on the basis of all protein quantitation events using the default settings.

2.21.3 DAVID (The Database for Annotation, Visualization and Integrated Discovery)

Proteins reaching significant ($p<0.05$) differential expression were submitted to the database for annotation, visualisation and integrated discovery (DAVID) [367] **(**https://david.ncifcrf.gov/). DAVID was used to evaluate the enrichment of trends and patterns in each given list of proteins relative to a background defined protein set; in most cases the set of fully quantitated proteins within each experiment. Default settings were used and in all cases only enrichment p-values with a Benjamini-corrected p-value of less than 0.05 were used. Enrichment was assessed for several gene properties, including Gene Ontology (GO) terms, chromosomal enrichment and canonical pathway, as described below.

2.21.4 Gene Ontology Terms

GO term enrichment analysis was performed with DAVID using all quantitated proteins as a background list. GO term lists for biological processes, cellular components and molecular functions were exported with their corresponding Benjamini-corrected p-values. These were submitted to REVIGO, a tool for the reduction and visualisation of GO terms using default settings. Colouration and circle size were altered to correspond to GO term size and p-value significance, respectively.

2.21.5 Ingenuity Pathway Analysis

Full protein lists were submitted to Ingenuity Pathway Analysis (IPA) which were filtered on the basis of regulation score and p-value. Annotations were exported relating to the canonical localisations of each protein, its canonical functional category and any potential therapeutic targets. Canonical pathway analysis of the CLL proteomics was assessed using IPA core analysis, given the greater clinical relevance of this data. Analyses were conducted filtering for enrichment for the significantly upregulated proteins ($RS>0.25$, $p<0.05$), the significantly downregulated proteins ($RS<-0.25$, $p<0.05$) and for a combined enrichment considering all differentially regulated proteins ($RS>0.25/RS<-0.25$, $p<0.05$). The analyses were conducted using a ‘stringent’ search against the ingenuity knowledge base allowing for only direct relationships and experimentally observed findings in species being examined. The regulation

score was used for the analyses to give a robust measurement representative of the overall trend in expression in all samples.

2.21.6 Membrane Protein Analysis

While GO terms exist describing a protein as 'integral component of membrane' (GO:0016021) and 'intrinsic component of plasma membrane' (GO:0031226), closer interrogation of the proteins annotated with these terms revealed that the coverage of the proteome is incomplete, even in surface proteins confidently described with cluster of differentiation (CD) numbers. In some cases these were simply annotated with 'membrane' (GO:0016020). To ensure all proteins with a high confidence of surface expression were included in this analysis of membrane proteins, more than one source of annotation was considered. Firstly, all proteins annotated with GO terms indicative of relation to membranes (GO:0005886, GO:0005887, GO:0009986, GO:0031226, GO:0009897, GO:0016020 and GO:0016021) were selected and subsequently filtered for any proteins also annotated with localisation to the endoplasmic reticulum (GO:0005783), mitochondria (GO:0005739), or nucleus (GO:0005634). This list was then combined with those proteins described by the MS surface atlas [368]. This resulting list was used to filter for those proteins identified in **Chapter 5.0** and **Chapter 7.0** and of these further filtered against the surfaceome database **[369]**, a tool combining surface expression observations from several resources.

2.21.7 Chromosome Enrichment and Mapping

Chromosome enrichment was conducted by the submission of data to DAVID (**Section 2.21.3**) using default settings and Benjamini-corrected p-values.

The Ensembl genome browser was used to map proteins onto their respective coding gene locations within the human genome. Uniprot accession numbers for proteins lists were submitted to the 'features on karyotype' tool (http://www.ensembl.org/Homo_sapiens/Location/Genome).

2.22 Antibodies

Several antibodies were used for the immunophenotyping of cells by flow cytometry summarised in **Table 2.1** and immunodetection of proteins by Western blotting, summarised in **Table 2.2**.

Target	Species/isotype	Manufacturer/ Clone no.	Mono/ Polyclonal
Mouse CD3	Rat/IgG2a	In house (KT3)	Monoclonal
Mouse CD19	Rat/IgG2a	In house (1D3)	Monoclonal
Mouse CD5	Rat/IgG2a	BD (53-7.3)	Monoclonal
Mouse B220	Rat/IgG2a	BD (RA3-6B2)	Monoclonal
Mouse IL5RA	Rat/IgG1	BD (T21)	Monoclonal
Human CD3	Mouse/IgG2a	Biolegend (OKT3)	Monoclonal
Human CD19	Mouse/IgG1	Biolegend (HIB19)	Monoclonal
Human IL5RA	Mouse/IgG1	R&D systems (26815)	Monoclonal

Table 2.1. Antibodies used for flow cytometry detection of proteins. (BD; BD Biosciences).

Target	Species	Dilution	Incubation conditions	Manufacturer/ Clone no.	Mono/ Polyclonal	Target size (kDa)
Myc	Rabbit	1:500	O/N 4°C	Abcam (AB32072)	Monoclonal	57
Myo9	Rabbit	1:1000	O/N 4°C	CST (#3403)	Polyclonal	230
Cor1A	Goat	1:1000	O/N 4°C	Genetex (GTX89757)	Polyclonal	60
Moe	Rabbit	1:1000	O/N 4°C	CST (#8146)	Polyclonal	78
Lcp1	Rabbit	1:500	O/N 4°C	CST (#5350)	Polyclonal	70
Hsp90	Rabbit	1:1000	O/N 4°C	CST (#4874)	Polyclonal	90
Mdh2	Rabbit	1:500	O/N 4°C	CST (#8610)	Polyclonal	35
Capg	Rabbit	1:1000	O/N 4°C	PT (10194-1-AP)	Polyclonal	39
Eno1	Rabbit	1:1000	O/N 4°C	CST (#3810)	Polyclonal	47
Npm	Rabbit	1:1000	O/N 4°C	CST (#3542)	Polyclonal	38
Shp1	Rabbit	1:1000	O/N 4°C	CST (#3759)	Monoclonal	68
GAPDH	Mouse	1:2500	1 hour RT	Abcam (AB8245)	Monoclonal	36
Rabbit Ig (Green)	Goat	1:40000	1 hour RT	Licor	Polyclonal	N/A
Goat Ig (Green)	Donkey	1:30000	1 hour RT	Licor	Polyclonal	N/A
Mouse Ig (Red)	Goat	1:15000	1 hour RT	Licor	Polyclonal	N/A

Table 2.2. Antibodies used for immunoblotting detection of proteins. (O/N; overnight, RT; room temperature, CST; Cell Signaling Technology, PT; Protein Tech, N/A; not applicable).

3.0 ANALYSIS AND VALIDATION OF iTRAQ 2D LC-MS/MS Eμ-*MYC* PRECANCEROUS AND LYMPHOMA DATA

3.1 CHAPTER INTRODUCTION

Myc upregulation is proposed to drive proliferation in approximately 20% of tumours as well as having a functional role in tumourigenesis in around 70% of cancers. Multiple mechanisms have been described attributing *myc* dysregulation to oncogenesis. For example, *myc* has been observed to be among the most duplicated and amplified genes in cancer [370].

The translocation of an Ig enhancer and *myc* is well characterised as the initiating event of BL. The molecular events leading to tumour formation, however, are more complex. The overabundance of *myc* transcription and translation is not independently sufficient to induce tumour formation, with tumour suppressor mechanisms inhibiting *myc*-driven oncogenesis. Downregulation of myc degradation by loss or inhibition of FBW7, loss of the p53/ARF/MDM2 negative feedback axis and BCL2 overexpression have all been characterised as oncogenic mechanisms supporting *myc*-driven oncogenesis [178, 179, 181, 182, 192, 371]. Eμ-*myc* mice were generated to confirm the effects of the *myc* oncogene upon B-cell lymphomagenesis [188].

This model demonstrated that constitutive expression of *myc* resulted in the appearance of aggressive BL-like tumours at approximately 100 days of age [176, 188, 190].

Further understanding into the molecular events responsible for Eμ-*myc* and BL tumours offer insight not only into BL and myc-driven tumourigenesis, but potentially other insights into other B-cell cancers. Given the success of immunotherapy as an approach to treating B-cell cancers, methods were considered which would allow the efficient detection and discovery of novel cell surface proteins. One such approach that was considered was a method termed free-flow electrophoresis (FFE). FFE works on the principle that cells have variable surface charges which provide a potential means of prefractionation. By applying an electrical current perpendicularly to cells suspended in free flowing buffer, this surface charge causes differential attraction and repulsion from the electrodes. The resulting separation is suggested to therefore enrich for outlying sub-populations at the extremes of the isoelectric normal distribution [372]. In the context of 2D-LC MS/MS characterisation of cancer samples, it was thought that the characterisation of these rare sub-populations may enable the enrichment and detection of cell surface proteins and immunotherapy targets that would otherwise fall below the detection threshold of MS when analysed as part of the whole cell population.

An experimental design was considered which would compare proteomic expression in splenic B cells from Eμ-*myc* mice with advanced lymphoma (termed 'lymphoma'), compared with a wildtype (WT) splenic B-cell control. The characterisation of Eμ-*myc* splenic B cells, prior to any tumour formation, (termed 'Eμ-*myc*') was also considered with the potential to identify proteins regulated in the presence of constitutive *myc* expression prior to the emergence of a neoplastic phenotype.

The workflow used for this experiment is detailed in **Figure 3.1**, describing the isolation of WT, Eμ-*myc* and lymphoma splenic B cells, the sub-population enrichment of B cells using FFE and the pooling of these outliers sub-population. These three samples were then analysed by a 2D-LC MS/MS workflow utilising iTRAQ labelling of peptides as a means of relative quantitation enabling the inference of differential protein expression in three splenic B-cell phenotypes. Target-decoy searching of the spectra and protein inference and quantitation assignment was conducted to produce a quantitative proteomics output.

This work was performed prior to the commencement of this PhD project by Kath Woods-Townsend, who collected the samples and performed FFE and Spiro Garbis, who performed the 2D-LC MS/MS characterisation of the samples.

This chapter describes the analysis and Western blot validation of this quantitative proteome, attempting to conclude the validity of the shotgun proteomics workflow in the characterisation of B-cell cancers.

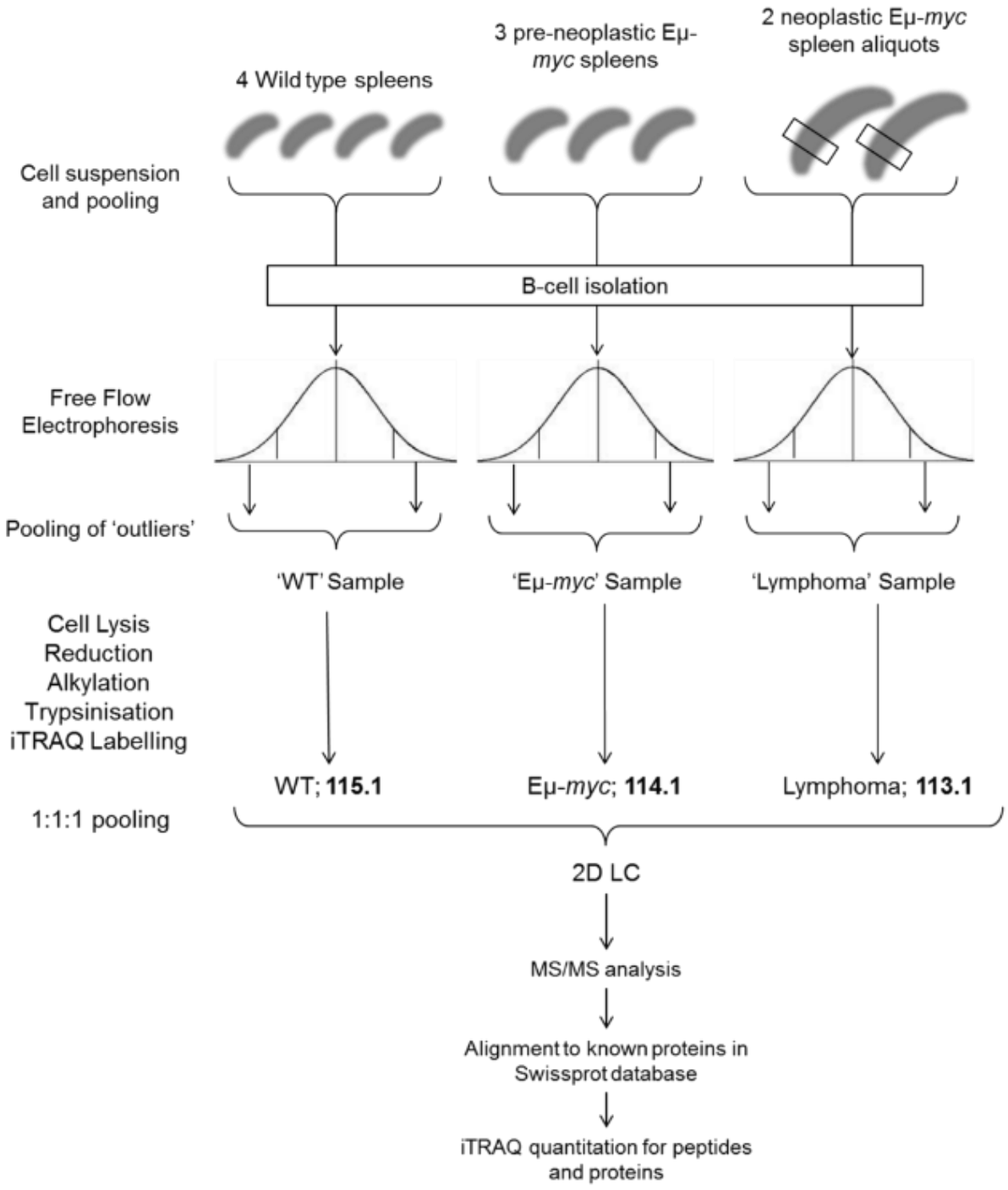

4 Wild type spleens
3 pre-neoplastic Eμ-myc spleens
2 neoplastic Eμ-myc spleen aliquots
Cell suspension and pooling
B-cell isolation
Free Flow Electrophoresis
Pooling of 'outliers'
'WT' Sample
'Eμ-myc' Sample
'Lymphoma' Sample
Cell Lysis
Reduction
Alkylation
Trypsinisation
iTRAQ Labelling
WT; 115.1
Eμ-myc; 114.1
Lymphoma; 113.1
1:1:1 pooling
2D LC
MS/MS analysis
Alignment to known proteins in Swissprot database
iTRAQ quantitation for peptides and proteins

Figure 3.1. Workflow detailing the sample procurement and 2D-LC MS/MS characterisation of Eμ-*myc* lymphoma. Spleens were collected and a single cell suspensions for each condition was pooled consisting of 4 WT spleens, 3 Eμ-*myc* spleens prior to any tumours and frozen aliquots of splenocytes derived from Eμ-*myc* mice with advanced tumours. Each pool was subject to B-cell isolation using negative selection and the resulting B cells subjected to FFE enrichment. The outlying populations were pooled for each condition and the resulting cells subjected to 2D-LC MS/MS quantitative proteomics characterisation. This involved the reduction, alkylation and trypsinisation of the cell lysates and iTRAQ labelling of the resulting peptides. The iTRAQ-labelled peptides were then prefractionated by liquid chromatography and each fraction subjected to chromatography in line with ESI and DDA of MS/MS. The work conducted up to the point of producing a quantitative proteomics output was conducted by Kath Woods-Townsend and Spiro Garbis.

3.2 ANALYSIS OF THE QUANTITATIVE PROTEOMICS DATA DESCRIBING Eμ-*MYC* AND LYMPHOMA

The resulting output of the 2D-LC MS/MS characterisation of WT, Eμ-*myc* and lymphoma B cells identified 372 proteins with a false discovery rate of less than 5%. 147 of these proteins were identified with at least three non-redundant, unique peptides. To visualise the distribution of the quantitations for the 372 proteins in these two conditions, relative to the WT B cells, a frequency distribution of the $\log_2$ (ratios) of protein expression was plotted (**Figure 3.2**). This described, using bins of 0.2, the number of proteins exhibiting varying extents of differential expression, approximately forming a normal distribution for each condition. Log_2 (ratios) were used on the basis that ratios and fold changes do not present a linear numerical scale suitable for mathematical analysis. Amongst these 372 proteins, 36 and 106 proteins were quantitated with more than a 2-fold ($\log_2$ (ratio) of ±1) differential expression in pre-lymphoma Eμ-*myc* and lymphoma B cells, respectively, highlighted by the number of proteins observed at the extremes of the plot. The standard deviations of $\log_2$ (ratios) considering all protein quantitations were 0.62 and 1.12 for Eμ-*myc* and lymphoma, respectively. This analysis demonstrated a greater variation in global protein expression in tumour, compared to pre-tumour samples.

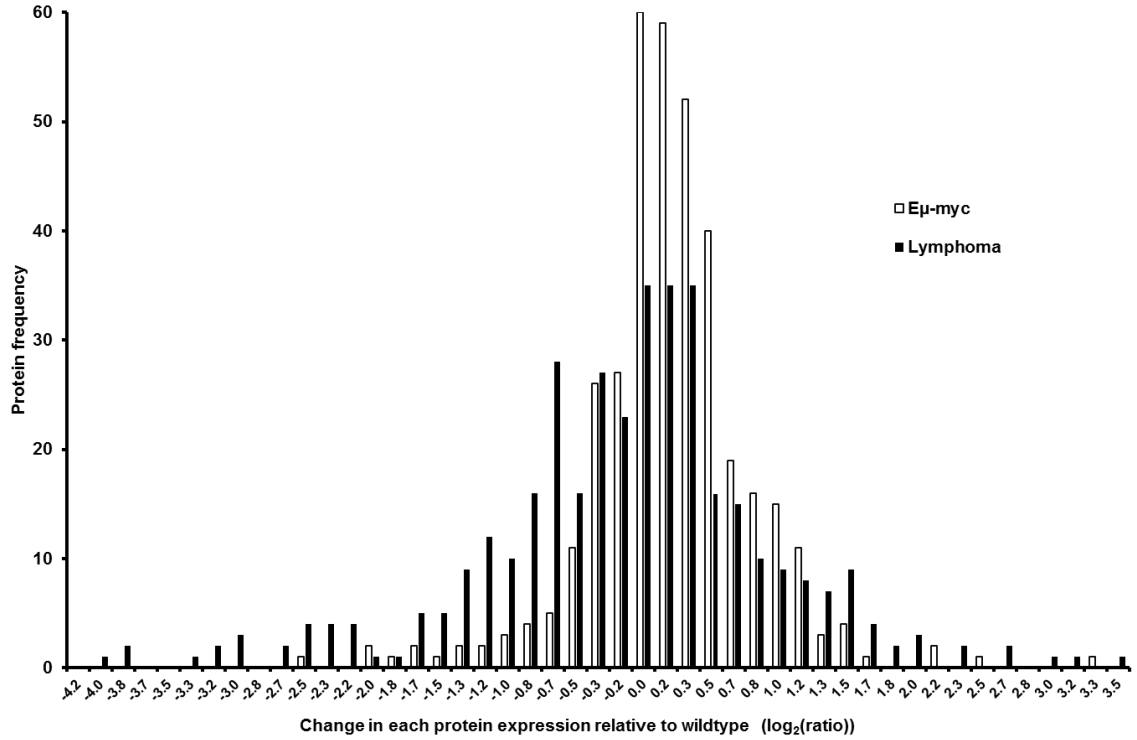


Figure 3.2. Detected differential protein expression in Eμ-*myc* lymphoma and pre-lymphoma, relative to WT B cells. Frequency distribution describing the number of proteins exhibiting $\log_2$ (ratios) in bins of 0.2 for Eμ-*myc* and lymphoma conditions relative to WT B cells. This summarises the regulation observed amongst the 372 proteins identified and quantitated by MS-iTRAQ analysis. This approximately forms a normal distribution with the greatest number of proteins with little or no regulation.

3.3 SELECTION OF VALIDATION CANDIDATES

In order to better understand the validity of the quantitative proteomices data, given a lack of positive controls such as myc protein expression, candidates were considered for quantitative validation using Western blotting.

There were many potential means of selecting candidates for validation and biological investigation. The MS-iTRAQ analysis generated a vast database of information pertaining to each peptide and therefore to each aligned protein. Factors for any protein such as the number of aligning peptides, sequence coverage, iTRAQ to background detection ratio, B- and Y- ion to background detection ratio, Eμ-*myc* ratio, lymphoma ratio, inter-peptide ratio variability and peptide confidence were considered. The majority of consideration, beyond these preliminary factors, was given to the expression ratios derived from the MS-iTRAQ analysis.

Proteins were selected for validation and further investigation using the pipeline shown in **Figure 3.3**. Candidates were filtered gradually to remove proteins that had no significant regulation observed in either Eμ-*myc* or lymphoma. Preference was given to candidates with a linear progression from WT to Eμ-*myc*, to lymphoma, to identify candidates with an expression correlated with tumour progression. This workflow also incorporated considerations for aspects that may influence the success of validation.

Figure 3.4 describes the shortlist of 18 proteins generated by steps 1-7 of **Figure 3.3**. The standard deviations of the pipeline-selected short lists were 1.08 and 2.59 for Eμ-*myc* and lymphoma, respectively.

Tyrosine-protein phosphatase non-receptor type 6 (SHP1), coronin 1A (Cor1A), macrophage capping protein (Capg), alpha-enolase (Eno1), nucleophosmin (Npm1), and malate dehydrogenase, mitochondrial (Mdh2), were selected from the shortlist in **Figure 3.4** for validation based upon the availability of validated antibodies. Plastin-2 (Lcp1), myosin-9 (Myo9), heat shock protein 90 (Hsp90) and moesin (Moe) were chosen based on availability of antibodies, peptides and biological interest.

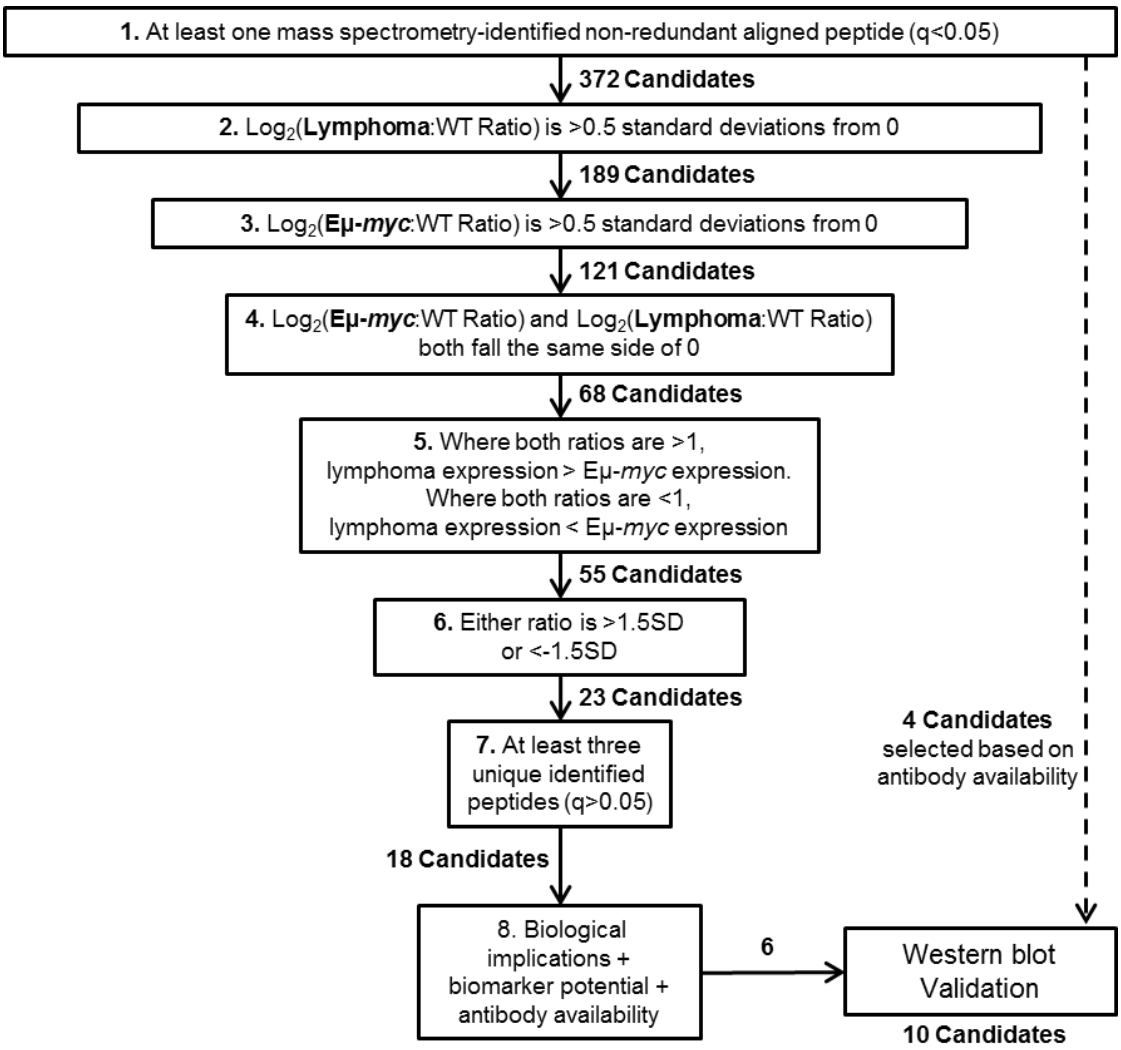


Figure 3.3. Candidate selection pipeline for the validation of quantitative MS proteomics. The pipeline employed to identify a selection of candidates for validation and investigation. This pipeline specifically tried to select candidates that show a directional progression from wildtype B cells to pre-neoplastic Eμ-myc B cells to neoplastic lymphoma cells, as well as strong peptide evidence suggesting a likely success of validation. 4 additional candidates were selected on the basis of antibody availability for confident candidate that did not fit criteria 4.

Name	Validated	UniProt Accession #	% Sequence Coverage (q>0.05)	Peptides (q>0.05)	Eμ-myc: WT	Lymphoma: WT
SAM domain and HD domain-containing protein 1 (**Samhd1**)		Q60710	5.1	4	**0.686**	**0.060**
Actin-related protein 2/3 complex subunit 4 (**Arpc4**)		P59999	11.3	4	**0.787**	**0.090**
Histone H2A type 2-B (**Hist2h2ab**)		Q64522	62.3	75	**0.373**	**0.156**
Core histone macro-H2A.1 (**H2afy**)		Q9QZQ8	27.4	28	**0.581**	**0.196**
Tyrosine-protein phosphatase non-receptor type 6 (**SHP1**)	*	P29351	7.4	4	**0.461**	**0.201**
Histone H4 (**Hist1h4a**)		P62806	71.8	37	**0.242**	**0.213**
Coronin-1A (**Cor1a**)	*	O89053	21.9	9	**0.745**	**0.244**
Poly [ADP-ribose] polymerase 1 (**Parp1**)		P11103	3.3	3	**0.731**	**0.291**
Macrophage-capping protein (**Capg**)	*	P24452	10.2	3	**0.773**	**0.322**
Histone H2A type 3 (**Hist3h2a**)		Q8BFU2	80	74	**0.488**	**0.360**
Macrophage migration inhibitory factor (**Mif**)		P34884	23.5	3	**1.445**	**2.805**
Alpha-enolase (**Eno1**)	*	P17182	35.5	10	**1.820**	**3.251**
Nucleolin (**Ncl**)		P09405	23.5	19	**1.614**	**3.342**
Splicing factor, arginine/serine-rich 2 (**Sfrs2**)		Q62093	9.5	3	**2.399**	**3.837**
Nucleophosmin (**Npm1**)	*	Q61937	29.1	7	**2.168**	**4.488**
Malate dehydrogenase, mitochondrial (**Mdh2**)	*	P08249	29.9	10	**1.528**	**4.786**
L-lactate dehydrogenase A chain (**Ldha**)		P06151	13	5	**3.076**	**5.754**
Serum albumin (**Alb**)		P07724	3.5	5	**2.992**	**29.648**
Plastin-2 (**Lcp1**)	*	Q61233	25.8	17	**1.380**	**0.177**
Myosin-9 (**Myo9**)	*	Q8VDD5	14.1	25	**0.920**	**0.286**
Heat shock protein HSP 90-alpha (**Hsp90aa1**)	*	P07901	11.9	9	**1.419**	**0.520**
Moesin (**Moe**)	*	P26041	26.7	20	**0.492**	**2.606**

Figure 3.4. Validation candidate shortlist. The 18 proteins selected after stage 7 of the selection pipeline in **Figure 3.3**, as well as four further proteins selected for validation based on availability of antibodies, MS-observed peptides and biological interest. The proteins shown with an asterisk were selected for WB validation based on a number of factors including; biological interest, number of peptides, availability of validated antibodies against the protein and the ratios observed in Eμ-*myc* and lymphoma.

3.4 DETECTING THE LINEAR PHASE OF QUANTITATIVE WESTERN BLOTTING

To determine the validity of the MS iTRAQ ratios, WB analysis was employed as an alternative method of evaluating relative protein expression in the three B-cell phenotypes. Prior to this analysis, it was essential to ensure that WB detection produced reliable, quantitative observations. The Licor Odyssey was used as a means of WB detection. This scans for the detection of fluorophore-conjugated secondary antibodies, thereby generating a uniform detection that gives a quantitative comparison between band intensities.

Ideally cells would first have been subjected to FFE prior to WB analysis to recapitulate the MS-analysed samples. However, given that no surface proteins were amongst the validation candidates, and that the cells at both ends of the FFE were analysed, the material analysed was assumed to be representative of differential expression of intracellular proteins. Additionally, FFE would have depleted cell material making WB detection unworkable. Therefore, for validation purposes, no FFE was implemented.

To evaluate the linear detection threshold for each of the 10 validation candidates described in **Figure 3.4**, WB titrations were first performed. WT B-cell lysates and cultured lymphoma B cells ('Eμ8') lysates were evaluated for quantitative detection using a titration of loading protein from 10-60 μg (**Figure 3.5, Appendix A3**). Analysis of Cor1A is shown as a model example (**Figure 3.5**). Cor1A demonstrated a linear phase and clear detection for both WT and lymphoma from 10 to 40 μg with an R^2 value of >0.98 for both lysates. It was concluded that at 30 μg of protein, a quantitative comparison of relative Cor1A expression was achievable between these B-cell phenotypes.

It was observed that when high concentrations of protein lysate were subjected to WB, protein detection became saturated. Above this point of saturation, there was no longer a proportional relationship between amount of protein loaded and the intensity of the resulting band. Below this threshold, protein detection consistently demonstrated a linear correlation to loading protein mass. Each of the 10 validation candidates subjected to this analysis demonstrated that 30 μg of loading protein was below the band saturation threshold for both WT and lymphoma B-cell lysates (**Appendix A3**).

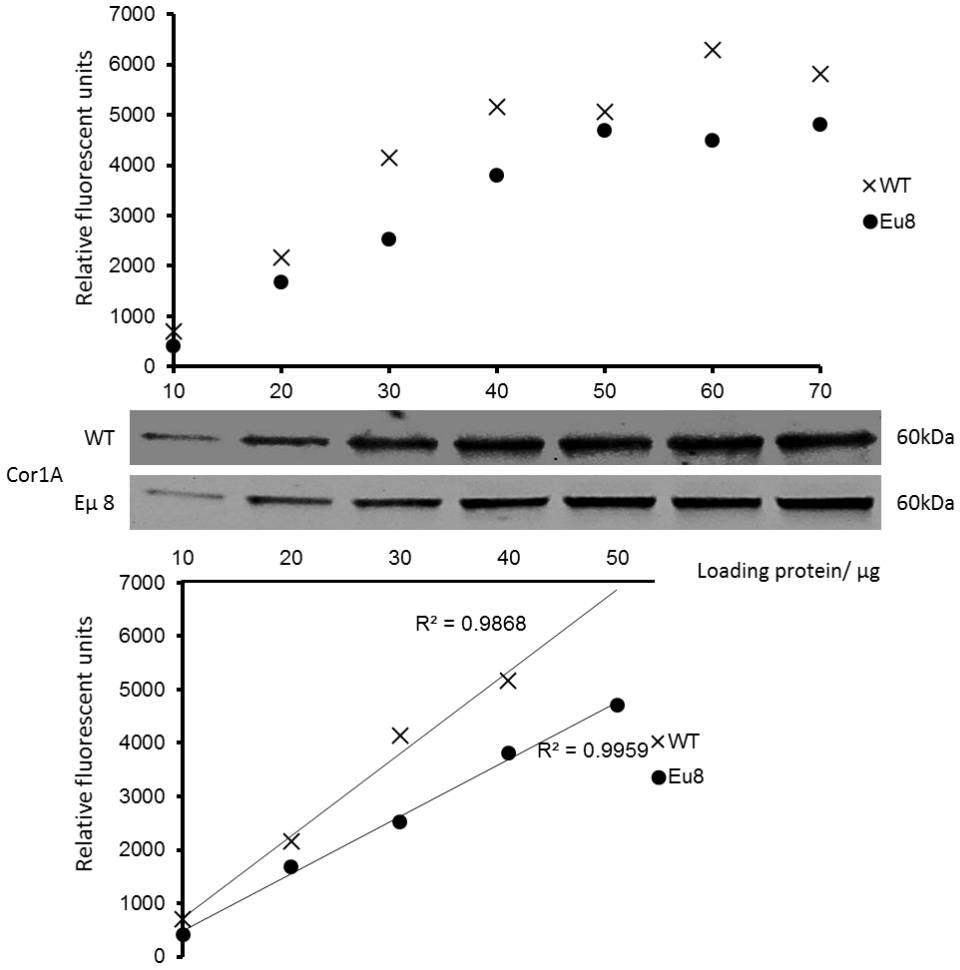


Figure 3.5. Determination of the linear phase of quantitative WB protein detection. WB detection was conducted for lysates of WT B cells and lymphoma (Eμ8 - Eμ-*myc*-derived lymphoma cell line) evaluating detection of 10-60 μg of protein loaded. The lower graph shows the linear portion of the upper graph with a trend line fitted and correlation coefficient described.

3.5 WB VALIDATION OF CANDIDATE EXPRESSIONS AND MYC EXPRESSION IN LYMPHOMA CELL LINES

The acquisition of fresh Eμ-*myc* lymphoma samples was challenging given the spontaneous nature of the disease development. Given that Eμ-*myc* lymphoma cells readily grow in culture (**Section 2.9**), it was concluded that early passage (<10) cultured Eμ-*myc* lymphoma cells would effectively recapitulate the characteristics of *in vivo* Eμ-*myc* lymphoma. This also provided a means of deriving sufficient protein to perform WB for all 10 candidates on the same tumour samples. Additionally, while different Eμ-*myc* lymphoma samples were used for validation, it was hoped that by assessing multiple tumours, the biological signature could be validated.

For quantitative WB analysis, a pool of splenic B cells from three WT mice (WT), a pool of splenic B cells from three Eμ-*myc* mice (Eμ-*myc*) and five individual Eμ-*myc*-derived lymphoma cell line lysates (Eμ6, Eμ8, Eμ13, Eμ14 and Eμ15) were selected (**Figure 3.6**). Five lymphoma cell line lysates were individually probed to assess the variation of protein between the lymphoma samples. This was based on the observation of a much greater SD of protein regulation for the lymphoma MS-iTRAQ ratios, as well as the know heterogeneity of cancers.

WB analysis for myc demonstrated a consistent and substantial upregulation of myc across the five lymphoma cell line lysates with an average of 18-fold increase from wildtype to lymphoma. Very little difference in myc abundance was observed between the WT and Eμ-*myc* B-cell lysates. No peptides for myc were identified by MS, so this served as qualitative biological validation.

The WB of Cor1A expression in WT, Eμ-*myc* and the 5 Eμ-*myc* lymphoma-derived cell lines is also shown in **Figure 3.6**. WB of Cor1A demonstrated lymphoma lysates with a consistent and substantial downregulation of Cor1A. A down regulation was also observed in the Eμ-*myc* B cells. The WB bands were quantitated and normalised to glyceraldehyde 3-phosphate dehydrogenase (GAPDH) expression and this normalised value was compared to that of WT expression to give a ratio. Quantitation and GAPDH normalisation of Cor1A expression exhibited an average abundance of 20% for lymphoma and 88% for Eμ-*myc* relative to that of wildtype B cells. This analysis was performed for all 10 proteins, shown in **Appendix A4**.

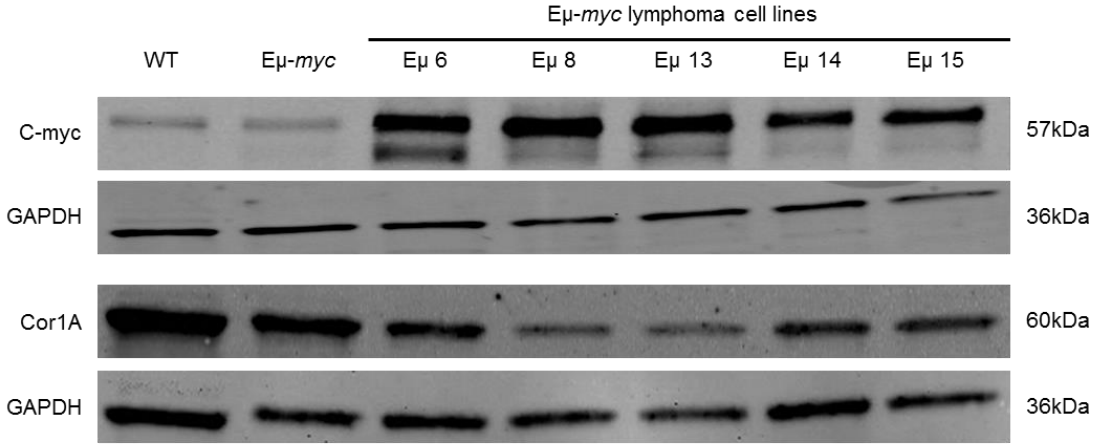


Figure 3.6. The expression of myc and Cor1A in Eμ-*myc* lymphoma and pre-lymphoma samples, relative to WT B cells. WB of lysates derived from a pool of splenic B cells from three WT mice (**WT**), a pool of splenic B cells from three Eμ-*myc* mice prior to any tumour development (**Eμ-*myc***) and five individual Eμ-*myc*-derived lymphoma cell line lysates (**Eμ6, Eμ8, Eμ13, Eμ14** and **Eμ15**). GAPDH serves as a loading control for each WB.

3.6 EVALUATING THE CORRELATION OF MYC AND COR1A EXPRESSION

WB of myc and Cor1A (**Figure 4.5**) demonstrated an inverse pattern of expression across the seven assessed samples. Where myc expression was low, in the non-cancerous lysates, Cor1A expression was high. In the lymphoma lysates, the inverse was observed. The observation of greater myc expression in Eμ 8 and 13 lysates was seen with a lesser expression of Cor1A. Quantitation of these relative expressions (**Section 2.17.6**) demonstrated an inverse correlation between the expression of myc and Cor1A across the seven probed B-cell states with an R^2 value of 0.7338 (**Figure 3.7**).

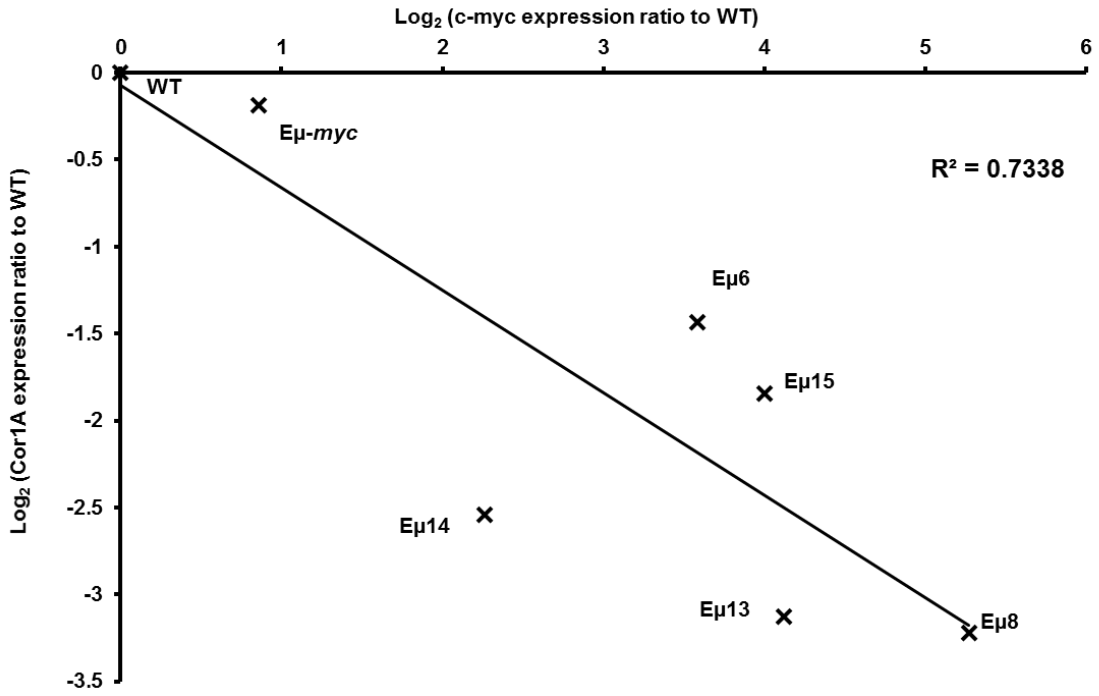


Figure 3.7. The inverse relationship between myc and Cor1A expression. The WB-determined $\log_2$ (ratios) of expression changes of myc and Cor1A in WT, to pre-neoplastic Eμ-*myc* and lymphoma (Eμ6, Eμ8, Eμ13, Eμ14 and Eμ15) B cells.

3.7 WB VALIDATION OF ITRAQ 2D LC-MS/MS DIFFERENTIAL EXPRESSION DATA

To generate a direct comparison between iTRAQ and WB-derived observations, WBs (**Appendix A4**) were quantitated, normalised to GAPDH, and ratios to WT B cells calculated. These ratios were plotted side by side to demonstrate the degree of agreement between MS and WB observation (**Appendix A5**).

Cor1A, again used as an example (**Figure 3.8**), demonstrated a consistent agreement between observations made by MS-iTRAQ and WB for the Eμ-*myc* and lymphoma samples relative to WT B cells. For Eμ-*myc* B cells, there was a marginal downregulation observed by

both iTRAQ and WB. For the 5 lymphoma cell lines, the average value of 20%, agreed closely with the 24% averaged determined by MS. While this value varied across the 5 lymphoma lysates, the downregulation was consistently below that of both the WT and Eμ-*myc* B cells.

This comparison was performed for all 10 validation candidates, shown in **Appendix A5**. Of these 10 candidates, for lymphoma lysates, 8 consistently agreed with the MS-iTRAQ-determined up- or down-regulation relative to WT. For Eμ-*myc* pre-lymphoma B-cell lysates, only two proteins, HSP90 and Cor1A, demonstrated an agreement between the MS- and WB-determined ratios to WT B cells.

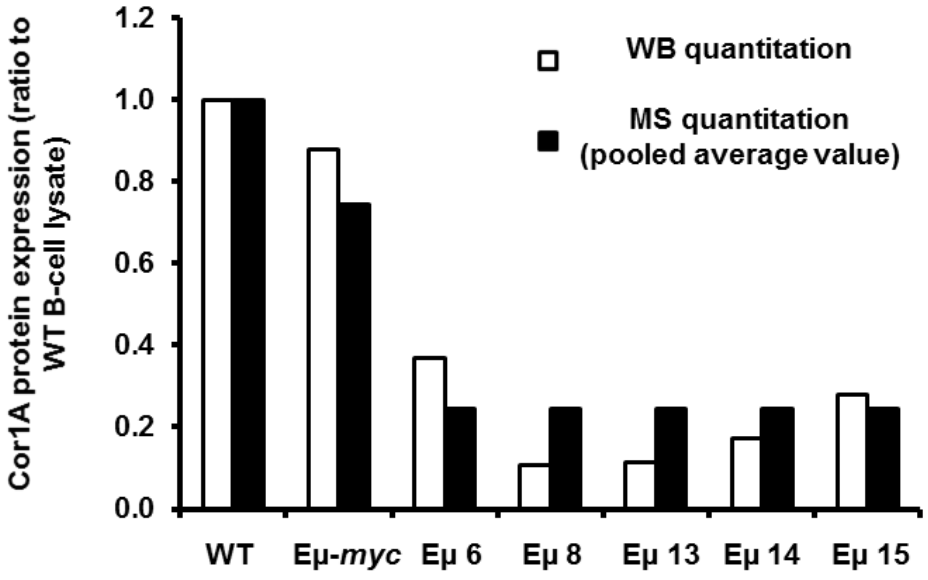


Figure 3.8. Comparison of the MS-determined and WB-determined ratios of Cor1A expression. Ratios of protein expression in splenic isolated B cells from pre-neoplastic Eμ-*myc* mice (**Eμ-*myc***) and Eμ-*myc*-derived cell lines from spontaneous splenic tumours (**Eμ 6-15**) were determined relative to wildtype B cells (**WT**) for WB- determined (open bars) and MS-determined (closed bars) quantitation methods.

3.8 DEFINING CONCORDANCE BETWEEN THE MS AND WB QUANTITATIVE DATA

To represent the overall agreement between relative protein expression determined by MS and WB, a ratio between the two values termed 'concordance' was defined. Concordance was calculated by dividing the WB-determined $\log_2$ (ratios) by the MS-determined $\log_2$ (ratios) thereby giving a single value indicative of the extent of ratio agreement for each WB observation. This was performed for both the ratios of Eμ-*myc* lymphoma (**Figure 3.9**) and Eμ-*myc* pre-lymphoma (**Figure 3.10**) relative to WT B-cell expression. For example, when the two

detection methods both agree exactly with, say, an upregulation of 4-fold ($\log_2$ (ratio) of 2), the concordance will equal 2/2 = 1. When the $\log_2$ (ratio) is larger, for WB than MS, the concordance value will be >1. When the $\log_2$ (ratio) is smaller for WB than MS, the concordance value will be less than 1 and when the regulation is in the opposite direction the concordance value will be less than 0.

Figure 3.9 demonstrated that 8 of the 10 evaluated candidates had consistently positive concordance for the Eμ-*myc* lymphoma B cells. Cor1A, Eno1, Mdh2 and Lcp1 all presented an approximate mean ratio of 1, demonstrating that in the pooled sample analysed by MS, the derived ratio has a good prediction of the average expression across several samples. Npm1 and Myo9 had concordances of <1 for all 5 Eμ-*myc* lymphoma B-cell samples which suggests the iTRAQ quantitation may have overestimated the extent of the up- and down-regulation of these proteins, respectively. Capg demonstrates the opposite, with all 5 Eμ-*myc* WB observations suggesting the MS analysis has underestimated Capg down-regulation. Moe and Hsp90 consistently demonstrate a concordance of <0, between the predicted expression changes from WT to lymphoma, with the MS predicting expression change directionality to be completely the opposite of that seen by WB.

The concordance of WB to MS showed the correct prediction of the direction of regulation in just 3 of the 10 proteins evaluated for Eμ-*myc* to WT ratios (**Figure 3.10**). Hsp90 demonstrated the only concordance of approximately 1 and Cor1A and Eno1 both suggested an overestimation of down- or up-regulation, respectively. SHP1, Capg, Npm1, Mdh2, Lcp1, Myo9 and Moe all had a negative concordance.

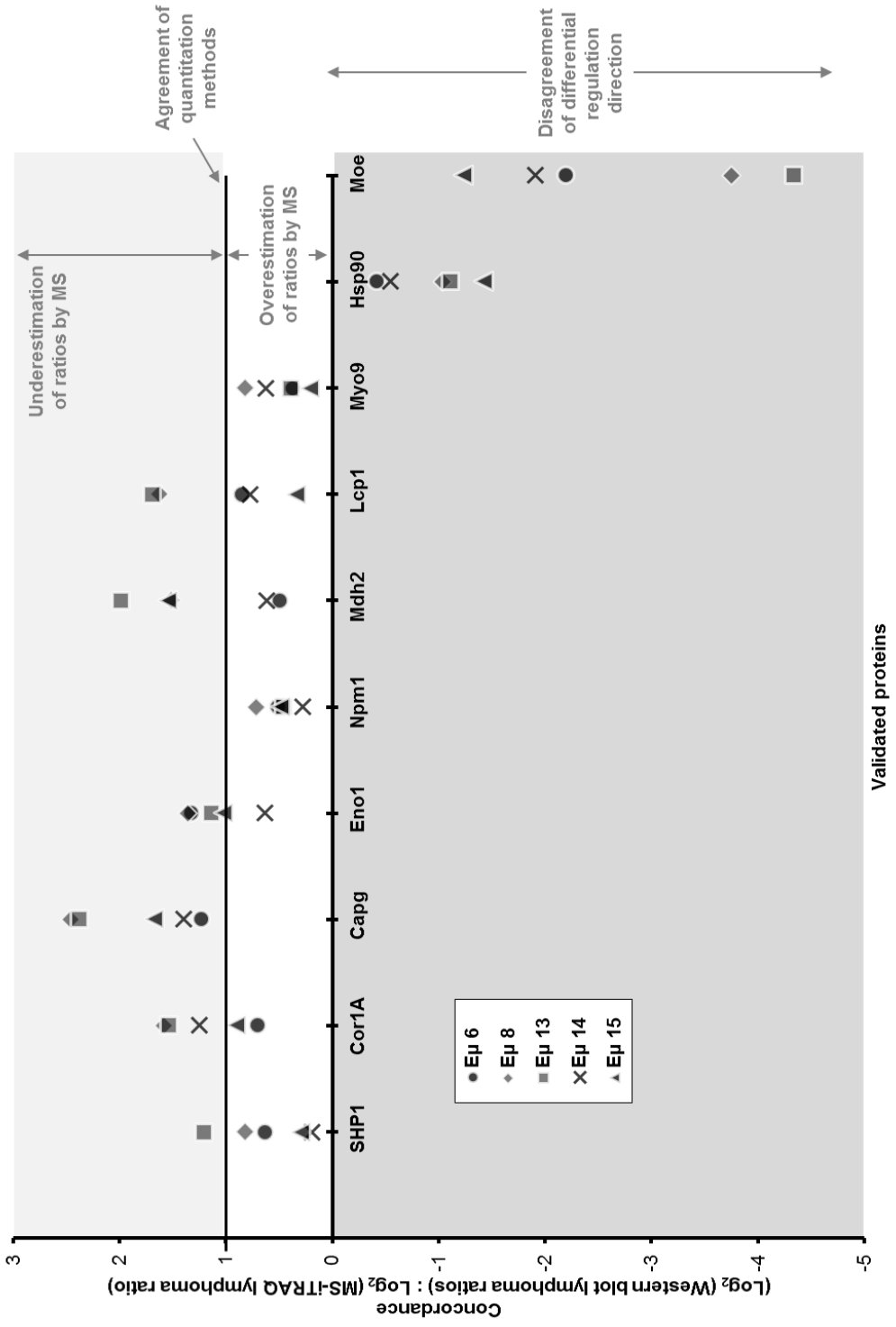


Figure 3.9. Eμ-*myc* lymphoma cell line concordances between log_2 (WB ratio) to log_2 (iTRAQ ratio). For each protein validated, this concordance ratio indicates how well the relative expression value produced by MS-iTRAQ analysis relates to the relative expression shown by quantitative WB data, both describing relative protein expression in Eμ-*myc* lymphoma cell lines relative to WT B cells.

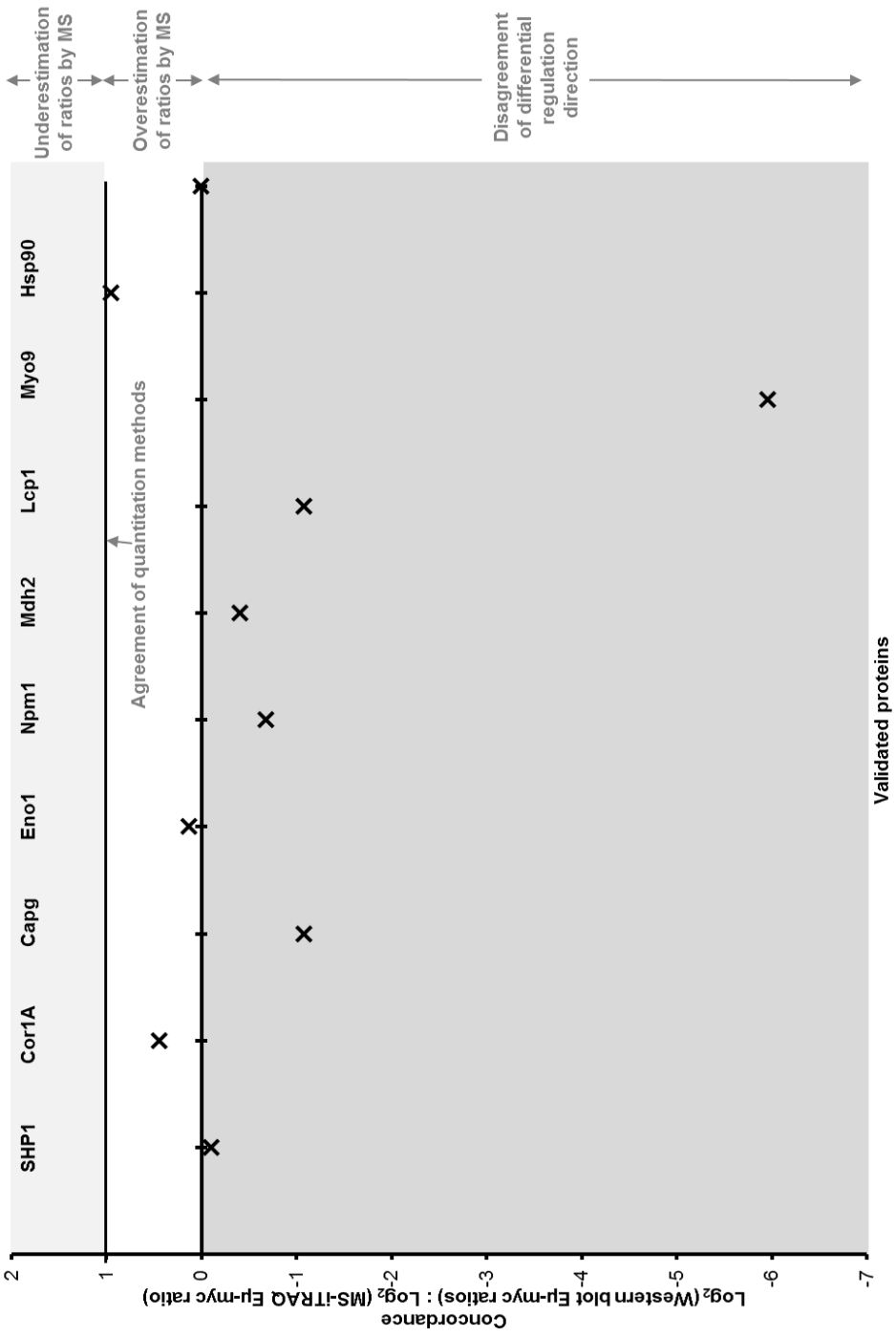


Figure 3.10. Eμ-myc pre-lymphoma concordances between log_2 (WB ratio) to log_2 (iTRAQ ratio). For each protein validated, this concordance ratio indicates how well the relative expression value produced by MS-iTRAQ analysis relates to the relative expression shown by quantitative WB data, both describing relative protein expression in Eμ-*myc* pre-lymphoma B cells relative to WT B cells.

3.9 EXAMINING THE EFFECT OF FOLD-CHANGE ON CALCULATED CONCORDANCES

All concordance values for Eμ-*myc* B cells and lymphoma cell lines were plotted against the fold change determined by WB (**Figure 3.11**). To demonstrate differential expression detection regardless of up- or down regulations, values were converted to absolute fold-changes. The concordance values for Moe and Hsp90 were excluded from this analysis. This demonstrated that at the lower fold-changes, candidate validation became increasingly disconcordant between the two detection methods. At fold changes of 2.5 to 12.5, the concordances observed were more accurate, consistently with a concordance ratio between 0.5 and 1.5. The concordance ratios for fold changes above 5 consistently demonstrate a concordance ratio of more than 1, indicating that the MS has underestimated the fold change in expression compared to that detected by WB.

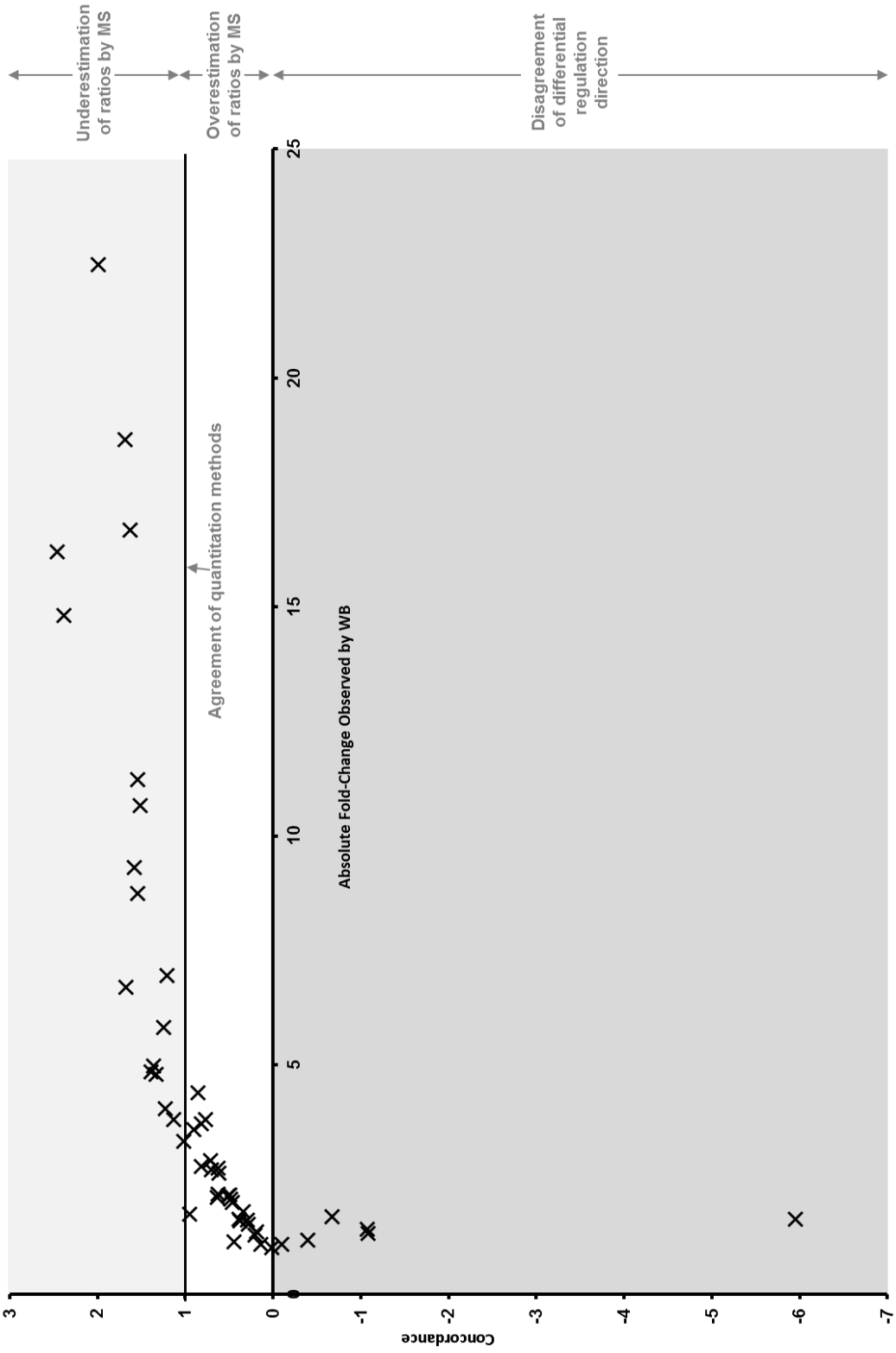


Figure 3.11. The effect of absolute fold-change on concordance. Values from the individually generated ratios for both pre-neoplastic Eμ-*myc* and lymphoma cell lines (excluding Hsp90 and Moe data) plotted against the absolute fold-change from WT determined by WB.

3.10 CHAPTER DISCUSSION

The overall aim of this initial investigation was to analyse and validate the success of 2D-LC MS/MS quantitative proteomics in the characterisation of B cells from Eμ-*myc* lymphomas and Eμ-*myc* transgenic B cells prior to lymphoma development relative to WT B cells using WB. While the experiment promised the potential of novel biological findings, the predominant focus was on the technical findings, given the uncertainty surrounding the ability of this emerging methodology.

It was hoped that protein dysregulation, such as the upregulation of pathways and oncogenes, loss of tumour suppressors and the global effects upon protein expression induced by the myc oncogene would be detected within the proteomics data. Such findings were, however, not effectively captured in the low number of identified proteins. The identification of 372 proteins was below what would be typically expected of a 2D-LC MS/MS characterisation of cancer cell lysates, with only around 5-30% of the potential number of confidently observed proteins detected [373, 374]**.** The isolation of the extremes of electrophoresed cells by FFE caused the cell number available for lysate production to be lower than ideal. This led to approximately only 30-40% of the ideal protein material being available for MS analysis. With MS analyses, detection of peptides is subject to various thresholds defined by detection limits and background noise. These limits are even more sensitive in the context of product ions which due to variable and imperfect fragmentation can result in more than a 10-fold reduction in sensitivity [281, 375]. Therefore, when half the mass of protein in a sample is available, substantially less than half of the peptides analysed will be above this threshold [281]. While this reduced the efficacy of this experiment, it reinforced the principle that further MS experimentation utilising a greater quantity of material promised a greater depth of analysis.

An impact of this reduced protein material was the lack of sensitivity to detect several key proteins, most notably peptides matching to myc – an observation that should have acted as a positive control for Eμ-*myc* lymphomas [188]. Other key proteins such as members of the p53 pathway also fell below the detection threshold, limiting the conclusions that could be drawn regarding pathway dysregulation and wider cell perturbations. Reduced sensitivity may also offer an explanation as to why no surface proteins were detected with at least three peptides. This lack of surface protein detection was a clear issue, as the original premise behind the FFE isolations was rare surface protein discovery. Overall, qualitatively, this lack of sensitivity limited the biological conclusions that could be drawn from the experiment. However, the higher-abundant differentially expressed proteins presented several opportunities to confirm the quantitative validity of the MS findings.

The overall trend in quantitative findings were approximately what might be expected comparing tumour, pre-tumour and healthy control samples. Greater protein dysregulation was anticipated in Eμ-*myc* lymphoma B cells, compared to that of pre-lymphoma Eμ-*myc* and WT B cells. This was best highlighted by the topological trend described in **Figure 3.2**, which presented a far wider normal distribution for lymphoma than pre-lymphoma Eμ-*myc* B cells. The greater number of differentially expressed proteins agrees with the notion of oncogenesis being a process requiring protein dysregulation [1, 2]. In order for a cell to become malignant, sufficient traits and autonomy must be acquired; phenotypically driven by up or downregulation of proteins.

For the pre-lymphoma Eμ-*myc* B cells the observation of lesser protein dysregulation was expected, relative to that of B cells which had progressed to lymphoma. Given the pleiotropic nature of the aggressive myc oncogene however, it might have been expected to observe a greater extent of protein dysregulation, more than the 36 proteins observed with >2-fold differential expression compared to WT B cells.

The reproduction of quantitative MS findings by WB offers a means of confirming the validity of the MS experiment and the methodology used, especially when no internal positive controls, such as the quantitation of myc were available. Validation of the MS findings presented several challenges, however. The ideal scenario for validation would have been aliquots of the same protein lysates used for the MS characterisation, however no such samples were available in light of the FFE isolations providing less than the optimum protein content. The implementation of FFE enrichment to a second set of validation samples would have been more suitable, however would have suffered from the same low yields; impractical for WB. While frozen tumour splenocyte aliquots were available they would not have provided sufficient material required for several WBs. Isolation of fresh tumour material from tumours of Eμ-*myc* mice was considered, however, this presented challenges based around the spontaneous tumour development and again, limited material. Overall it was concluded that given the high-abundance of the detected proteins, based on the number of peptides identified, that these proteins should be representative of a robust biological trend. Cultured Eμ-*myc* lymphoma cell lysates were therefore evaluated as they provided the most practical means of overcoming these issues. While distinct differences are known to exist between cells in culture [376]and within the tumour microenvironment, lymphoma B cells were used at the earliest possible passage, rarely in culture for more than 21 days (**Section 2.9**).

WB provides an opportunity to observe differential expression of proteins, however this requires the use of a mass of protein within a linear dynamic range of detection. Too little protein and the band may not be detected, too much and band saturation can occur, making

quantitative conclusions inaccurate. Prior to the WB analysis of the 10 validation candidates, it was essential to identify the ideal loading protein mass required for maximum band intensity, while remaining within the linear dynamic range of WB band quantitation. **Figure 3.5** and **Appendix A3** details these observations demonstrating that band saturation is a potential issue with regard to accurate relative quantitation for the majority of the proteins assessed. This is an expected observation on the basis that either no more protein can bind the membrane, no further primary or secondary antibody can bind, or the scanning becomes saturated. In fact, on this basis, actin was avoided as a loading control as saturation occurred at just 20 µg of protein (data not shown) while GAPDH provided a linear response at 30 µg. Care was taken to use identical conditions for these optimisation blots as several factors such as loading volume, gel type, lane width and secondary antibody could all have influenced the potential saturation point.

The observation of high upregulation of myc in all 5 Eµ-*myc* lymphoma cell lysates (**Figure 3.6**) supports the premise that Eµ-driven *myc* overexpression drives tumourigenesis [188]. It also confirms that the WB-analysed tumours were almost certainly derived as a result of Eµ-*myc* model, rather than any other tumour type that could have arisen spontaneously. The observed expression of myc in the pre-neoplastic Eµ-*myc* B cells was lower than might be expected. This observation is in agreement with literature suggesting that myc protein expression is only marginally upregulated in Eµ-*myc* B cells prior to lymphomagenesis, by 30-50% [377]. Increased transcription of the *myc* gene, driven by the Eµ enhancer [343], would lead to the conclusion that myc protein levels should also be elevated above that of wildtype, however no elevation was observed. This lack of myc upregulation could be due to negative feedback though protein degradation pathways or inhibited translation.

WB analysis of Cor1A (**Figure 3.6**) confirmed the MS-derived observation of downregulation in Eµ-*myc* and lymphoma B cells detailed in **Figure 3.4**. The consistent observation across the five lymphoma cell line lysates suggests that the downregulation may be seen in all Eµ-*myc*-derived tumours. WB also confirmed the intermediate expression of Cor1A observed in Eµ-myc B cells. These observations of Cor1A regulation in both conditions, either suggested a role in lymphomagenesis and lymphoma survival, or as a bystander of these processes, implicating Cor1A downregulation as a possible marker of tumour progression.

The relative expressions of myc and Cor1A across the seven probed cell lysates were plotted to examine the relationship between the two proteins (**Figure 3.7**), which appeared inverse by eye (**Figure 3.6**). The trend that emerged was that as myc expression increased, Cor1A expression, overall, decreased. While not highly correlative or conclusive, this suggested a general trend that directly or indirectly implicated myc in the downregulation of Cor1A

expression. This was of interest, as myc has not been implicated in targeting or suppressing the Cor1A gene directly [162].

Cor1A has been implicated in cytoskeletal restructuring, required for cell motility through invagination and cell protrusions [378, 379]. It has also been shown to be essential for T-cell immune function and survival through the increase of calcium ion release [380]. Cor1A$^{-/-}$ mice demonstrate a decreased formation of f-actin in T cells [381]. Combined with the observation that a key component of actin filament networks, actin-related protein 2/3 complex subunit 4 (p20) (**Figure 3.4**), is significantly downregulated in lymphoma, possible conclusions could be drawn as to the suppression of f-actin formation either as part of, or because of, lymphomagenesis [382]. Interestingly, a Cor1A mutation that was observed to retard T-cell function, showed no effect on B cells [383]. This led to the possible conclusion that while Cor1A was downregulated in lymphoma its loss or loss of function may not have had a critical role in tumourigenesis [383]. Cor1A was identified as a downregulated protein in 2 thymomas of p53$^{-/-}$ mice to an average of 26% of that of WT [384]. This consistent observation with the work presented here, implies that Cor1A downregulation may not be exclusive to Eμ-*myc* lymphomas. It would be of interest to see the effects of Cor1A knockout on the progression or presentation of Eμ-*myc* and other tumours. An earlier Eμ-*myc* tumour presentation in the absence of Cor1A would indicate a putative tumour suppressive role and would support a premise that cytoskeletal protein dysregulation plays a role in B-cell tumour development.

Concordance, described in **Section 3.8**, provides a measure for concluding the validity of the MS-iTRAQ-derived ratios for Eμ-*myc* lymphoma and pre-lymphoma, when comparing to WB ratios. The concordance values comparing 70 observations made by WB (**Appendix A4**) and the 30 MS quantitations (**Appendix A5**), are summarised in two graphs (**Figure 3.9** and **Figure 3.10**). The concordance values for lymphoma (**Figure 3.9**) demonstrated that the quantitative MS analysis had successfully identified the up or downregulation of 8 out of 10 of the chosen candidates and for 5 of these, the concordance indicated a close agreement between MS and WB. This indicates a good degree of success with regards to the determination of the directionality of differential protein expression. This demonstrated that the cultured lymphoma cells, without the application of FFE enrichment provided, on the whole, a suitable means of validating the MS. It also demonstrated that, for these evaluated, high-abundance intracellular proteins, the cells at the extremes of the normal distribution of FFE enrichment were comparable to the population as a whole.

The concordance analysis also highlights some potential limitations of this validation analysis, with two proteins, Hsp90 and Moe, regulated in the opposite direction between the two quantitation approaches. It is difficult to conclude which method is correct, given that the MS is

derived from at least 9 observations of peptides for each protein, while the antibodies employed in WB have been used in several publications. One potential source of the disparity is that of isoforms, which can be distinguished by MS, but not necessarily by antibodies, and sometimes *vice versa*. HSP90, for example has 3 major isoforms with several other minor isoforms with overlapping peptide sequence homology [385]. While HSP90 isoform variant AA1 is described and quantitated in **Figure 3.4**, this may have been constituted and quantitated by peptides matching several isoforms with potentially vastly different regulation. In order to more confidently quantify the specific isoform a far greater number of peptides are required for protein inference [279, 280]. Another source of error is potentially that derived from the analysis of cultured lymphoma cells which could have been either selected for, or induced, by the act of culturing these cells *in vitro*. The lack of FFE enrichment or sample handling differences could also be attributable to differences. However, the predominant trend of protein expression does appear to agree between MS and WB, despite these limitation. This suggests that, overall, both the WB analyses and MS analysis have provided a suitable and successful validation.

Concordance values of the Eμ-*myc* pre-lymphoma regulation directionality, relative to WT, was positive in only three of the ten instances (**Figure 3.10**). Only Hsp90, a protein which did not validate successfully for lymphoma, shows a concordance of more than 0.5. It was therefore concluded that validation for this condition was unsuccessful. While the same limitations discussed above were potentially true for the Eμ-*myc* pre-lymphoma B-cell samples, the observed fold changes, overall, were far smaller than that of lymphoma B cells. The detection of smaller ratios is more challenging by both MS and WB quantitation methods, given an intrinsic background variability in both approaches. Far more samples with technical replicates would be needed to conclude the majority of smaller fold changes represented by Eμ-*myc* pre-lymphoma, outside the scope of this analysis.

The theory that less, or negative, concordance occurs at lower ratio values was tested by plotting all of the concordance values (with the exception of those for Moe and Hsp90) against the absolute fold-changes derived by WB (**Figure 3.11**). This graph confirms that quantitative MS analysis of smaller fold-changes in expression are less accurately validated, demonstrating that any fold-changes below 2 are unlikely to validate successfully. **Figure 3.11** also indicates that larger fold changes have less accurate concordance, with iTRAQ underestimating the extent of the fold change. This trend is consistent with previous studies describing isobaric tag quantitation having a substantial ratio compression from the co-isolation of precursors [301, 302, 304, 386, 387], outlined in **Section 1.8.4**.

The validation and analysis described in this chapter suggest that the proteomics results have fallen short of fully harnessing the potential of MS technology. The lack of initial material

for the MS-iTRAQ analysis, observed proteins, myc peptides and surface proteins reduced the qualitative confidence in the data set as a whole. However, the study demonstrated novel methods of sample preparation and MS analysis and highlighted a number of validated candidates of biological interest. This validation was successful, despite the biological and technical variability in sample analysis and preparation. From a biological perspective, the ability of the technology to predict candidates that are consistently up or down regulated is vital to biomarker and target discovery, which appears to have been captured, at least in part by this proteomics approach. The observations of these 8 candidates with a consistent regulation across the original MS-iTRAQ analysed samples and the 5 lymphoma cell lines potentially demonstrated that sample pooling may be of benefit in this type of discovery, a generally vital approach given the limited capacity of iTRAQ and TMT multiplexes.

Retrospectively, the use of FFE was not ideal in the context of this MS experiment. While the principle of attempting to enrich for the detection of surface proteins using FFE was sound, it did not take into consideration the substantial challenges behind membrane protein solubilisation, digestion and characterisation using the standard 2D-LC MS/MS workflow [264], which exacerbated the depletion of cell material. Due to the highly selective nature of the FFE process, there was a substantial risk that the cells selected for analysis did not accurately represent the whole cell populations. The FFE selection generated bias based upon the phenotype of cell surface charge, and it was difficult to speculate as to the intracellular mechanisms that may have had some contribution to this phenotype. Possibly one of the most significant factors in the surface charge of cells was the effects of apoptosis altering the surface lipid composition, such as phosphatidylserine presentation during apoptosis [388]. Apoptosis is sensitised in Eμ-*myc* and lymphoma B cells as a result of myc overexpression [377, 389] and was likely increased upon the handling of cells outside of their preferred conditions and microenvironment.

Overall, this investigation has established that there is significant potential behind MS iTRAQ proteomics in the context of the Eμ-*myc* BL model, which may be applicable to many other types of cancer. It can be concluded that a more rigorous application of quantitative MS analysis has the potential to improve upon this initial study. Current techniques and technology, analysis of the whole populations of cells, rather than FFE isolation, and more rigorous experimental design may prove to generate higher numbers of proteins detected and quantified.

4.0 DEVELOPMENT OF AN OPTIMISED APPROACH TO PROTEOME QUANTIFICATION WITH ISOBARIC TAGS

4.1 CHAPTER INTRODUCTION

The generally successful validation of preliminary data in **Chapter 3.0** supported the implementation of further isobaric-labelled MS analyses. Several methodological advances also became available after this initial study, which potentially offered a far greater depth of analysis. Improvements to available technology also promised an increased number of protein identifications and quantifications.

While limitations of this experiment, such as the number of proteins and peptides identified, could likely be improved upon with advancements in sensitivity and mass accuracy of mass spectrometers, such as that of the Orbitrap Elite, one issue that remained was that of ratio compression induced by the co-isolation of precursor ions [301, 302, 304, 386, 387]; an effect highlighted by **Figure 3.11** and described in **Figure 1.14**.

Ratio compression arises within a complex proteome because differential expression is typically in the minority, while the majority of proteins do not differ in abundance. When a peptide from a differentially expressed protein is isolated during a shotgun proteomics

experiments the probability that any co-isolating peptides are without regulation is very high. The end result is therefore an averaging of quantitations, diluting that of the regulated protein.

While many options to avoid the effects of co-isolation exist, none were deemed optimal for implementation. The most common method applies a filter which disregards quantitations from spectra where more than 30% of the isolated ions are not attributable to the selected precursor, termed isolation interference [309]. The main limitation to this approach is that a large number of quantitations are disregarded, many of which still have valuable information which can support a finding of a differentially expressed protein. Another flaw to the approach is that equal consideration is given to all remaining quantitations which may still introduce ratio compression.

This chapter describes the development of a method for minimising the effects of co-isolation-induced ratio compression in isobaric-tagged proteomics. A 'spike-in' experiment was designed using *E. coli* and human cell lysates, providing non-redundant peptides with the aim of better understanding the effects of co-isolation. Using these data, a strategy was designed that implements weighting to ratio calculation and also incorporates statistical analysis. This experiment also provided an opportunity to evaluate the utility of the newly available TMT 10-plex reagents for potential use in later experiments.

The methods described in this chapter have been developed into an online tool termed Statistical Processing for Isobaric Quantitation Evaluation (SPIQuE) available at http://spiquetool.com.

4.2 PROTEOMICS DESIGN OF AN *E. COLI* SPIKE-IN EXPERIMENT

For the purposes of defining the extent of co-isolation, it was considered that the properties of a differentially expressed proteome should be simulated, while being fully traceable for data analysis purposes. Protein extracts from *E. coli* and human cell lines provided such an opportunity, with very minimal peptide redundancy between the two proteomes. This lack of redundancy and the availability of reference proteomes meant that, upon target-decoy searching, the origin of each peptide precursor could be easily concluded. The workflow and experimental design are shown in **Figure 4.1.**

E. coli and human cell lines were grown and harvested, using extensive cell washing to remove any potential contamination from growth media. Cells were lysed using trituration and sonication, and lysate protein concentration determined. Peptides were generated using trypsinisation of reduced and alkylated lysates. *E. coli* peptides were spiked into a TMT 10-plex experiment as a complex and minority, proteome at defined dilutions. The 10-plex otherwise

contained peptides comparing between titrated quantities of HeLa and MCF7 cell line-derived tryptic peptides **Figure 4.1**. This experimental design simulated the complexity, and low abundance, of a differentially regulated proteome encountered within clinical and biological specimens. Peptides were prefractionated using RP, high-pH liquid chromatography and non-adjacent fractions were pooled into a total of 13 fractions for analysis. These were each subjected to 6 hour LC-MS/MS analysis using DDA and HCD and CID peptide fragmentation.

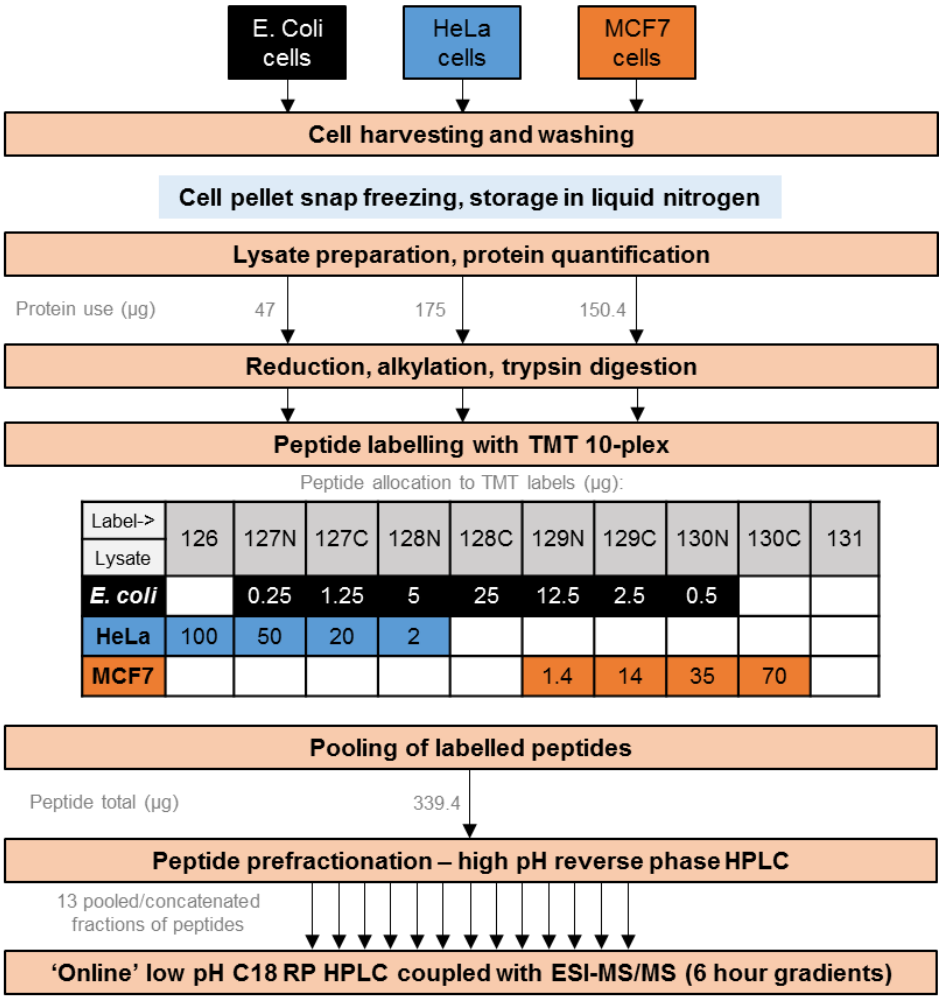


Figure 4.1. The experimental design and workflow for the TMT 10-plex containing spiked-in *E. coli* peptides. A workflow describing, in descending order: The 3 biological materials collected for MS characterisation, extensive washing, snap freezing and lysis of these samples and the production of peptides from each reduced and alkylated lysate. Peptides were then assigned to differential labelling using TMT 10-plex, which were pooled and subjected to prefractionation chromatography to reduce sample complexity. These fractions were pooled to give 13 fractions for analysis. Each peptide fraction was subject to a second round of chromatography, in line with electrospray ionisation, MS detection and data-dependent MS/MS characterisation.

4.3 DATA ANALYSIS OF THE *E. COLI* SPIKE-IN EXPERIMENT

Spectra produced by the MS analysis were analysed using target decoy searching as described in **Figure 4.2**. CID and HCD spectra were subjected to separate searches due to the differences in detection methods. Each target-decoy search was performed with Proteome Discoverer software against a proteome consisting of the combination of protein sequences for *E. coli* and humans. To facilitate data processing, protein accession numbers were modified to contain the suffix Ec_ for *E. coli* proteins and Hs_ for human proteins. The analysis of the combined, labelled peptide extracts resulted in the identification of a total of 8096 protein groups (<1% peptide FDR). Specifically, a total of 6499 human and 1597 E. coli proteins were observed with approximately 20% of the PSMs matching to the E. coli proteome.

To gain an impression of the extent of peptide co-isolation and the effects this had on the relative quantitations, the quartiles and percentiles for ratios derived from the reporter ions for human and *E. coli* PSMs were summarised (**Figure 4.3**). For the human PSMs the extent of suppression was most apparent in the '2%' ratios for which the median PSM quantitation ratios were indicative of 7.1% and 11% for HeLa and MCF7, respectively (**Figure 4.3A**). For *E. coli*, the minority proteome, suppression was far more apparent. The ratios defining 1% and 2% *E. coli* peptide content, relative to 100% (128C), were indicative of a median of 8.6% and 9.1% respectively, with many PSMs exhibiting substantial co-isolation in these channels (**Figure 4.3B**). Overall, extensive peptide co-isolation was observed for the E. coli PSM quantitations, with >80% of ratios compressed by >50%.

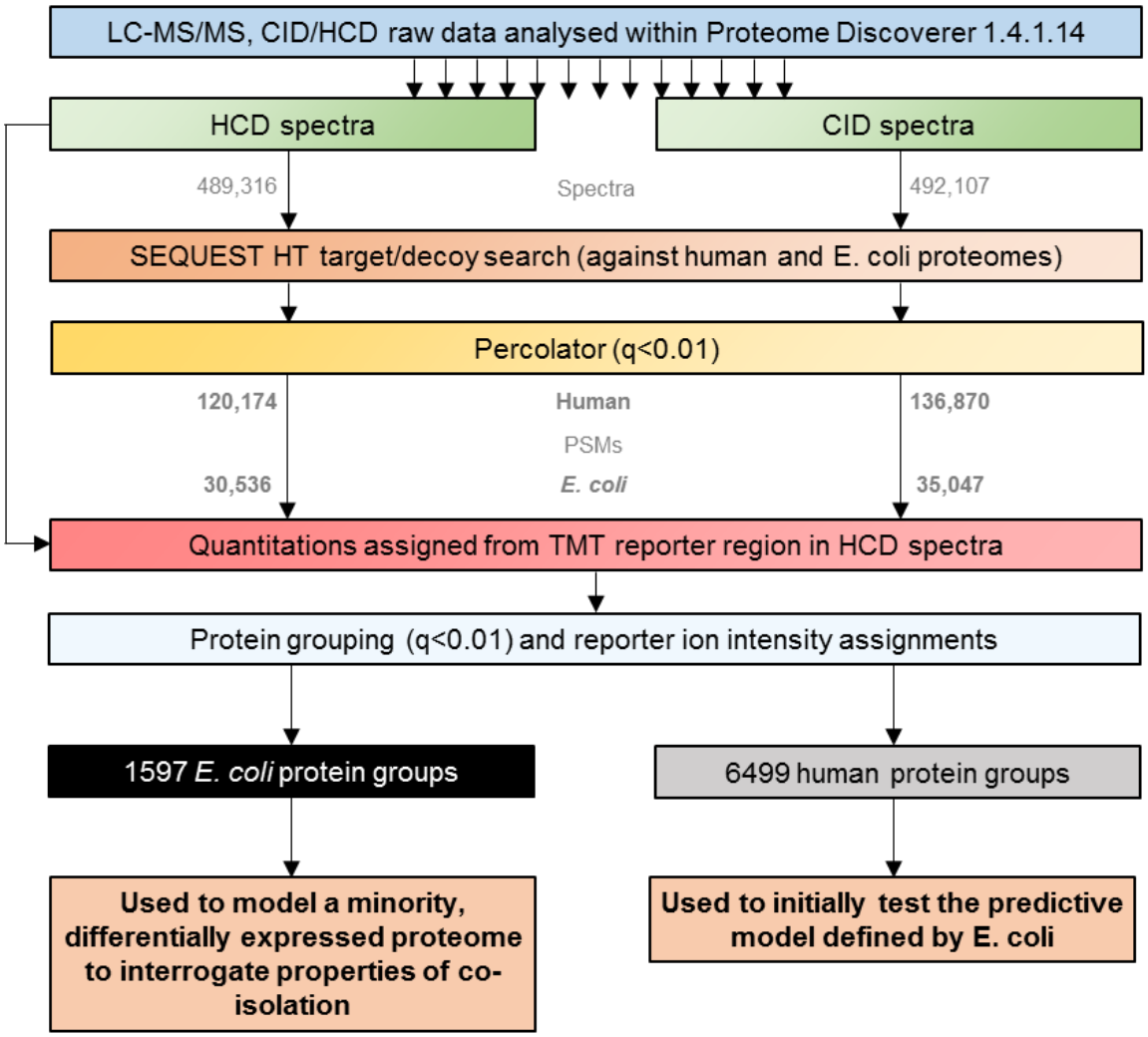


Figure 4.2. Identification and segregation of identified human and *E. coli* proteins. LC-MS/MS data were processed using Proteome Discoverer, searching spectra separately based upon CID and HCD fragmentation against a single proteome containing both the human and E. coli UniProtKB proteomes. The percolator algorithm was used to estimate FDR and a threshold of q<0.01 was used. Proteome Discoverer was also used to extract quantitation from reporter regions and group proteins. The data pertaining to each organism were separated for analysis.

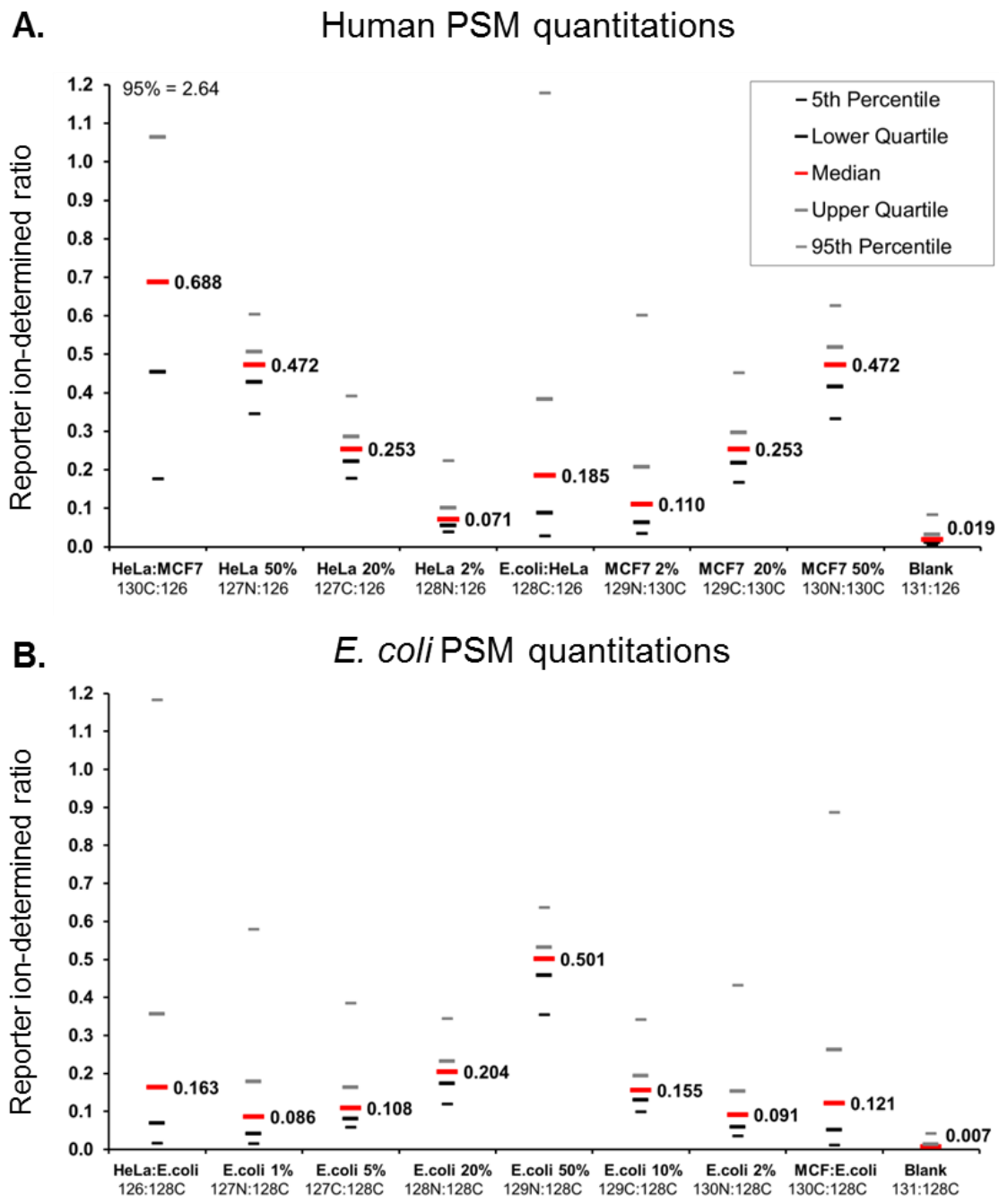


Figure 4.3. Q**uantification of human and *E. coli* PSM ratios and co-isolation.** PSM reporter ions were extracted from Proteome Discoverer and individual PSM ratios were calculated for those matching human (**A.**) and E. coli (**B.**) proteins to demonstrate the proteome-wide extent of ratio suppression and variance among observations. The median and quartiles were plotted to indicate the approximate distribution of derived ratios.

4.4 QUANTIFYING CO-ISOLATION TO ASSIGN A 'CO-ISOLATION SCORE'

With 35,793 PSM quantitations exhibiting a known extent of ratio suppression, visualised in the schematic in **Figure 4.4A**, it was considered that this information may have potential to identify trends that could be used to better understand the properties of co-isolation. The features of PSMs, such as the isolation interference, the MS-determined precursor intensity and the ion injection time used to derive the product ions of a spectrum were therefore evaluated for correlation.

To first determine the extent of peptide co-isolation as a quantifiable and measurable value, for all PSMs matching to *E. coli* proteins, 'co-isolation scores' were calculated from MS/MS-observed TMT reporter ion intensities. These were defined as a ratio of HeLa- and MCF7-dominant reporter ions, to *E. coli*-dominant reporter ions, as shown in **Figure 4.4B**. The theoretical values for a purely isolated *E. coli* peptide for these three models were 50:1, 5.6:1 and 1.5:1 (log_2 (ratios) of 0.58, 2.49, 5.64). Given the substantially greater abundance of human peptides, the probability of co-isolation for *E. coli* PSMs was weighted towards that of human peptide contamination. In total, 35,793 *E. coli* PSM quantitations were scored by this approach.

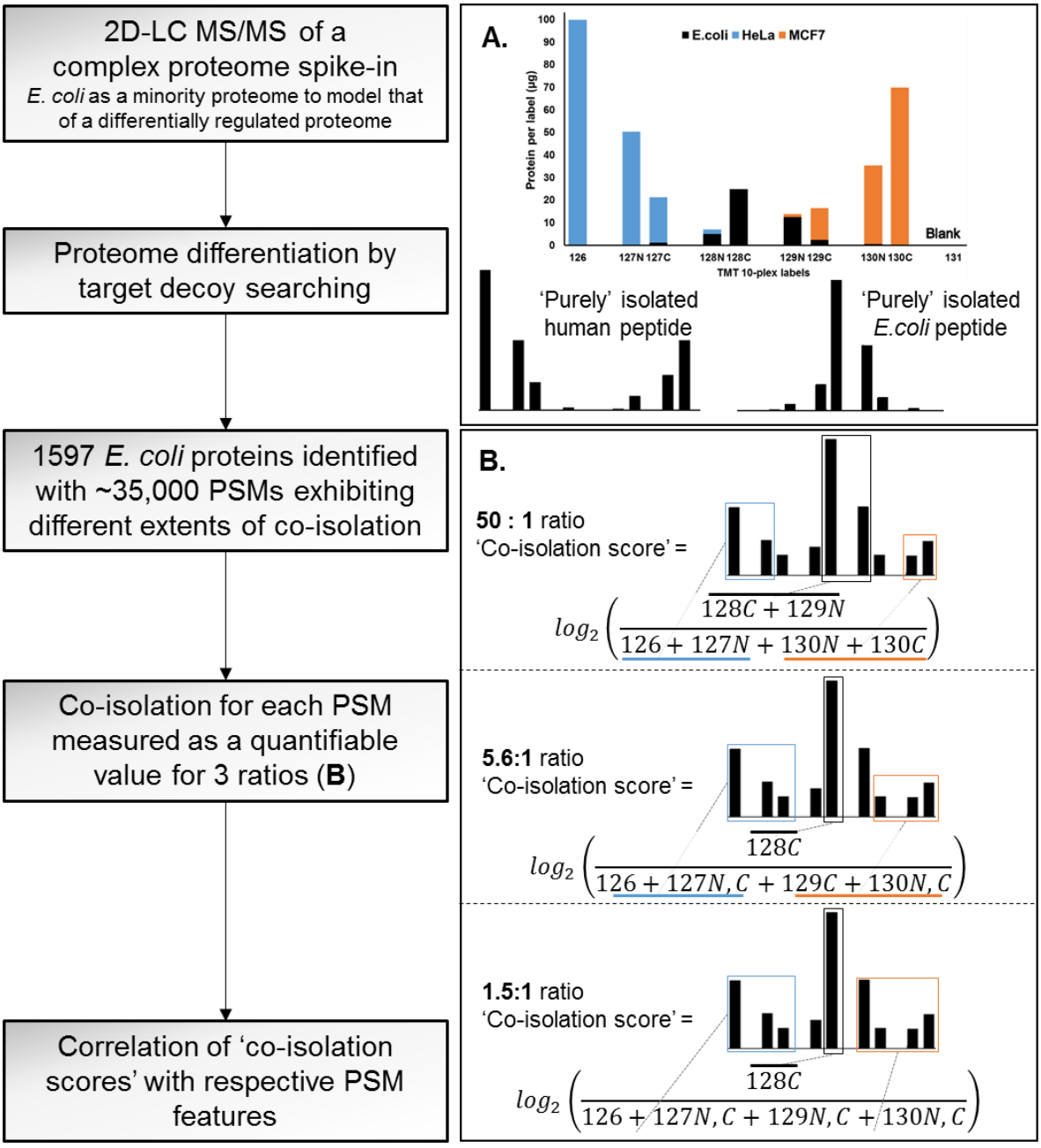


Figure 4.4. Using a complex-proteome spike-in experiment to estimate the correlation of co-isolation with PSM features. A. A representation of the *E. coli* peptide minority proteome 'spike-in' between two dilution series of peptides from HeLa and MCF7 human cell line lysates within a TMT 10-plex experiment. **B.** Calculation of the 'co-isolation score' defined for three ratios with theoretical values (as pipetted) of 1.5:1, 5.6:1 and 50:1. Compression of these ratios was used as a quantitative measurement of co-isolation.

4.5 CORRELATING THE 'CO-ISOLATION SCORES' WITH PSM FEATURES

Based on preliminary data describing high-abundant, regulated proteins, correlations were observed between PSM features and the extent of their ratio compression. To evaluate this trend on a larger, more comprehensive scale, the co-isolation scores were plotted against spectral features of the PSMs (**Figure 4.5, Appendix A6**). To normalise to the data and produce comparable scores, the features were first transformed using percentile ranking.

As expected, isolation interference had the strongest correlation with the extent of TMT-measured co-isolation (**Figure 4.5**). However, this was a poor determinant of compression to the reporter ions with a maximum R^2 value of 0.222. Ion injection time and precursor ion intensity values also demonstrated some degree of correlation with the co-isolation score of up to 0.175 and 0.177, respectively. This was consistent with the notion that higher precursor ion abundance levels would allow more selective isolation, and subsequently more accurate TMT reporter ion quantitations. The remaining features demonstrated very poor or no correlation with the co-isolation score. A notable exception was for peptide precursors with a charge of +2, that gave a statistically significant lower co-isolation rate than +3 charged peptide precursors for all 3 tested ratios ($p<0.0001$, T-test) (**Appendix A6**).

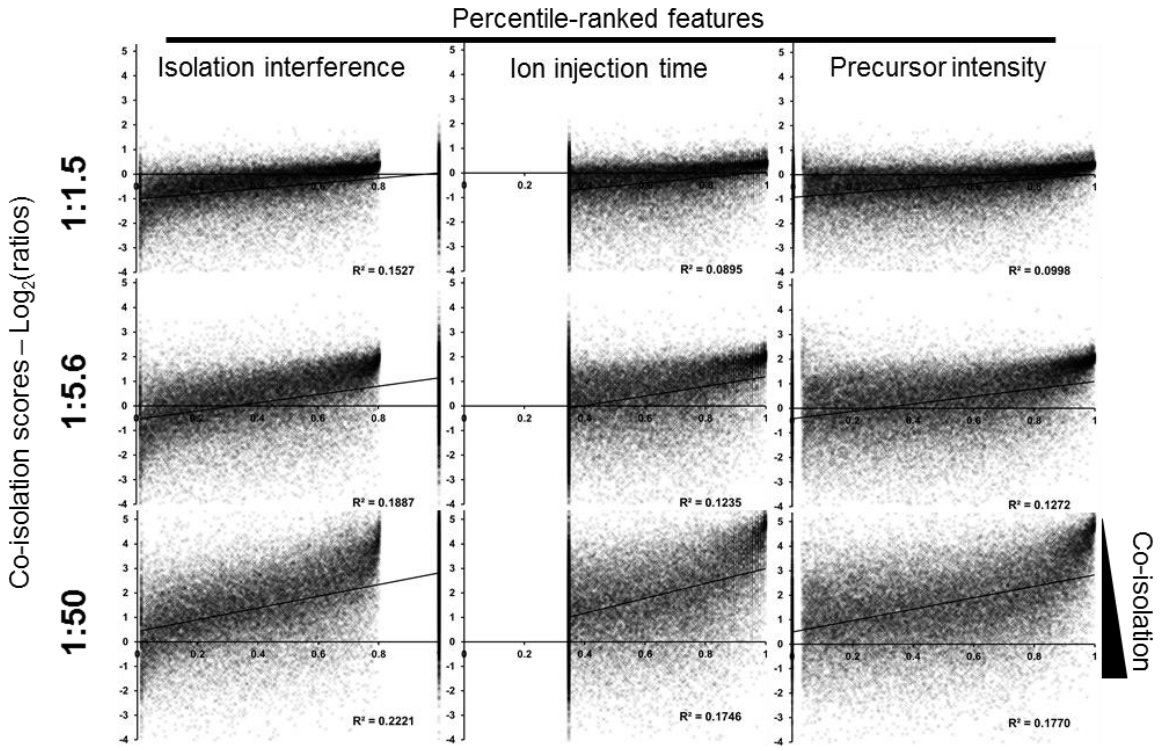


Figure 4.5. Correlating 'co-isolation scores' with PSM features. Correlation of the 'co-isolation scores' of the 1.5:1, 5.6:1 and 50:1 ratios (theoretical unsuppressed ratios, as pipetted) with percentile ranked PSM features.

4.6 CORRELATING THE 'CO-ISOLATION SCORES' WITH MULTIPLE PSM FEATURES

An approach using a combination of multiple features to determine correlation with ratio compression was considered. The hypothesis was that the properties of multiple features may improve upon the correlations seen with individual features. A resulting quality-defined score, would therefore estimate the probability of co-isolation and extent of reporter ion ratio compression. By extension, when applied to further data, this estimation could provide a means of emphasising reporter ion ratios with minimal compression based on PSM features alone.

Given the varying dynamic scales of PSM features, the process of scoring was simplified by percentile ranking to reduce extremes and to give comparable and cumulative scores between features (applied in **Figure 4.5**). Percentile ranked features also demonstrated marginally improved linear correlation coefficients with the 'co-isolation scores'.

Initially, iterative combinations of isolation interference, ion injection time and precursor intensity demonstrated utility in tandem. The strongest correlation (R^2=0.312) was observed when combining isolation interference and precursor intensity percentile-ranked values (**Figure 4.6A**). The impact of PEP score and precursor charge on the correlation coefficient were measured (giving a value of 1, multiplied by the relative contribution, to the cumulative score for +2 peptides). The PEP score was chosen as it is a measure of the statistical likelihood of a true peptide match, with most search engines providing an equivalent score. Both features improved the model (R^2=0.331, for both) and gave further improvement in combination (R^2=0.358). As precursor intensity and ion injection time are related physical parameters, ion injection time was separately tested in the place of precursor intensity and showed further improvement (R^2=0.362). Each of the three ratio models, 50:1, 5.6:1 and 1.5:1 were then individually assessed for the best possible correlation coefficient by iterative adjustment of these cumulative contributing factors.

Each of these individual models and the final model weighting that was generated as an average of these, gave R^2 values of 0.263, 0.315 and 0.370 for the 1.5:1, 5.6:1 and 50:1 ratios, respectively. The final, averaged, cumulative scoring of isolation interference, ion injection time and precursor intensity, PEP and charge of 10, 9, 3, 5 and 5, respectively, was plotted against the 'co-isolation scores' for each of the three ratios to illustrate the correlations of these models (**Figure 4.6B**). This percentile-ranked scoring system was termed a 'multi-feature weighted' model.

Finally, to confirm the efficiency of this linear regression model, machine learning was employed (conducted by Yawwani Gunawardana) which identified that the 10:9:3:5:5 model was highly correlative with the optimal approach (R^2 =0.9622) with machine learning-derived weighting factors of 0.343 : 0.257 : 0.136 : 0.104 : 0.232. When applied to data analysis

downstream the machine learning values gave a correlation of R^2>0.99 with the final 10:9:3:5:5 model. For simplicity, the 10:9:3:5:5 model was therefore applied to further data analysis.

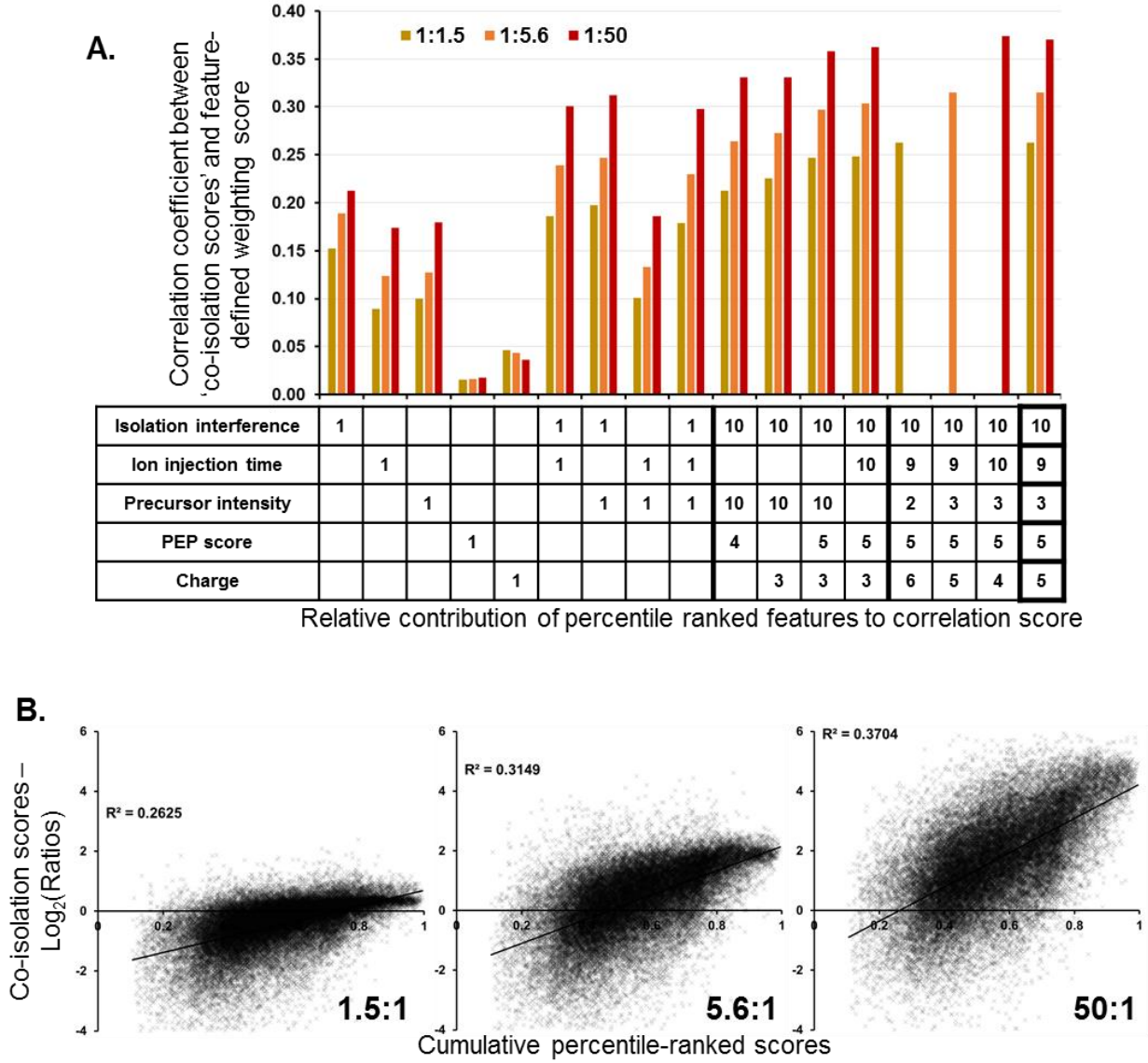


Figure 4.6. Using 'co-isolation scores' to define optimal multi-feature weighting. A. Evaluation of the correlation coefficient of individual and cumulative percentile-ranked features with the 'co-isolation scores' of PSMs. The optimal weighting was individually determined by iterative adjustments for the 1.5:1, 5.6:1 and 50:1 ratios and averaged to give the final multi-feature model. **B.** The final correlations between the cumulative percentile-ranked feature scoring and the 'co-isolation score' for the three co-isolation predicting ratios.

4.7 DEVELOPMENT OF AN ONLINE TOOL FOR THE APPLICATION OF MULTI-FEATURE WEIGHTING

A controlled means of investigating any potential value to the multi-feature weighted model was required. Ideally this would have taken the form of a complex isobaric labelled proteome containing several (20+) non-redundant or ^{13}C-labelled recombinant proteins spiked-in with known concentrations mimicking differential expression. Such an experiment, however, would have been extremely costly and was outside the scope of this analysis. Instead, the comparison of the relative expression between HeLa and MCF7 proteins also provided an opportunity to evaluate co-isolation, given that differential expression exists between these two cell lines. A general assumption can be made that for a given protein the quantitations of the surrogate peptides should, on the whole, be indicative of the same ratio of differential expression.

As well as evaluating the predictive capabilities of the multi-feature weighting model on individual proteins. The dilution series also offered the potential to observe the effects of co-isolation within the lower content peptide labels. 128N and 129N, contained just a fiftieth of the peptide content labelled by 126 and 130C and were therefore highly sensitive to ratio compression induced by E. coli peptide co-isolation. The same was true, but to a lesser extent for 127C and 129C labels. Four ratios were defined as 126:130C, 127N:130N, 127C:129C and 128N:129N termed the '100%', '50%', '20%' and '2%' ratios, respectively, on the basis of the relative masses of peptide material assigned to these TMT labels (**Figure 4.7A**).

In order to implement multi-feature weighting for the determination of protein ratios, requiring several complex calculations, a custom R script was written (courtesy of Cory H. White) which allowed the automation of this process. The script was designed to collectively analyse the reporter ion ratios for all PSMs matching each protein. Either all PSM ratios were combined equally to give an average ratio for each protein or, optionally, weighting could be applied to each PSM quantitation on the basis of PSM features. Secondarily to ratio calculation, the R script performed a statistical test of each ratio based on the individual quantitations of the PSMs. This test provided a multiple-test corrected p-value determined by a two-tailed, paired t-test. While not entirely suitable for this form of analysis, due to a general absence of normal distributions within the data, overall analysis showed that the ratios tended towards a normal distribution. The t-test was considered sufficiently robust in these situations however, as it was used predominantly as a means of filtering data to define instances where a clear difference was apparent in the individual ratios.

This script was developed into an online tool (courtesy of Oliver Bills) called SPIQuE which dramatically simplified the process of R script usage, providing an online interface with simplified options, data upload, automated processing and output download (available at SPIQuEtool.com).

4.8 Modelling Co-isolation-Induced Ratio Compression with Differential Protein Expression in Cell Lines

The SPIQuE tool was employed to normalise, statistically test and determine the protein ratios of each of the four ratio comparisons ('100%', '50%', '20%' and '2%'), independently of any weighting (**Figure 4.7**). Differential expression was defined as the observation of a >2-fold up or downregulation with a multiple-test corrected p-value of less than 0.05. The number of proteins reaching this threshold for each of the four ratios was demonstrated by Venn diagram (**Figure 4.7B**). Only a minor deterioration was seen for the '50%' ratios and statistics, with 12% of the '100%' proteins failing to reach significant differential expression ($p<0.05$, >2-fold). For the '20%' ratios, the impact was more apparent, with 28% of the original '100%' proteins determined as significantly, differentially expressed. The '2%' ratios demonstrated a substantial collapse of both ratios and statistics, with only 182 of the original 1007 '100%' proteins reaching significance and at least a 2-fold change.

Volcano plots were also used to detail the ratio compression and loss of statistical significance in the lower ratios; plotting the $\log_2$(ratios) against the $-\log_{10}$(p-values) (**Figure 4.7C**). The distributions for the volcano plots shows the normalisation efficiency in the '100%' and '50%' ratios. The effects of the *E. coli*-derived TMT reporters co-isolation in the '20%' and '2%' ratios are apparent, with the higher mass 129N and C channels skewing the results towards identifying an MCF7-dominant expression. The $\log_2$(ratios) and the $-\log_{10}$(p-values) for each sub-optimal comparison ('50%', '20%' and '2%') were plotted against the '100%'-derived values (**Figure 4.7D**) This further demonstrated the ratio compression and reduced statistical power for the low-percentage, high-interference channels.

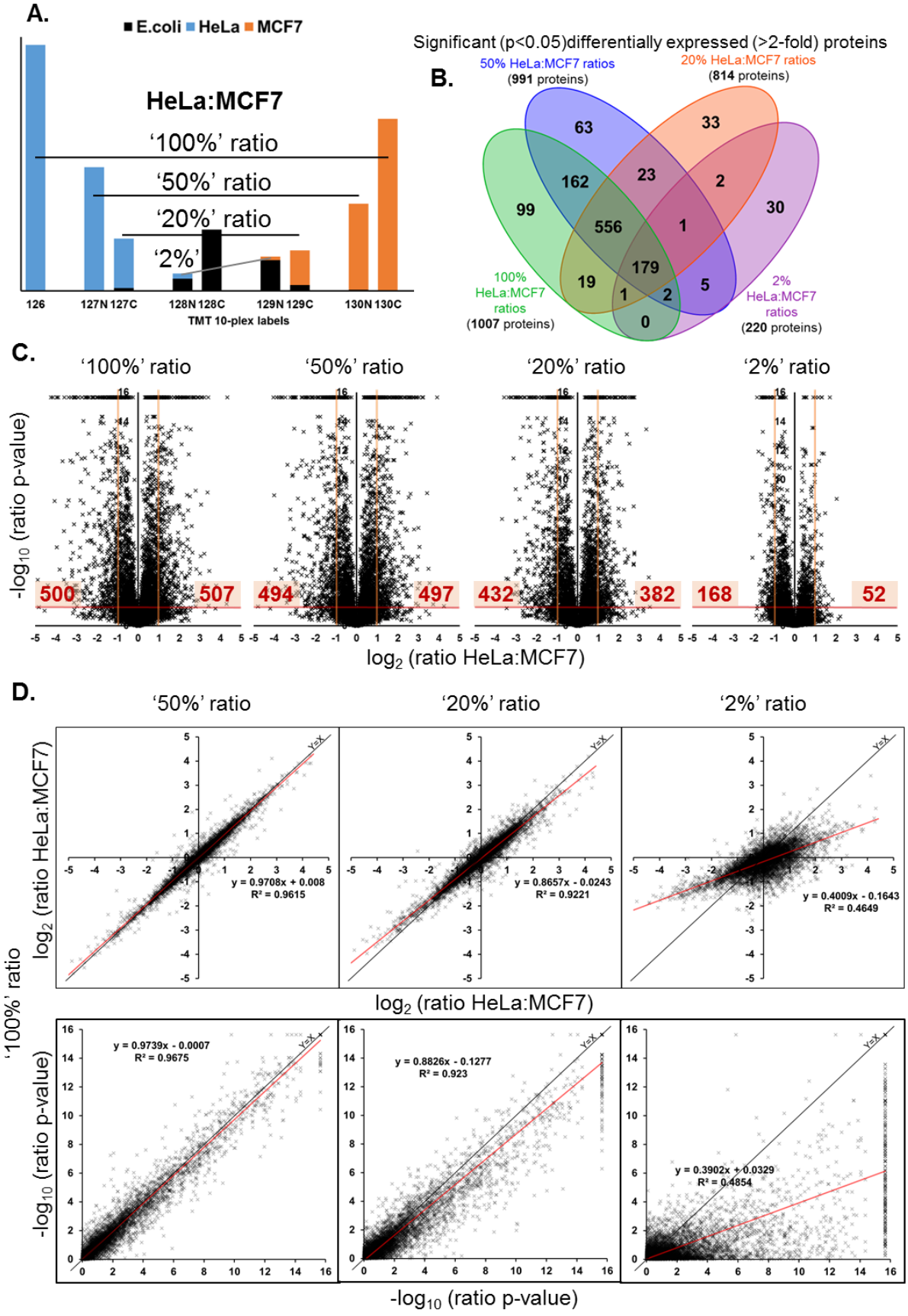
As anticipated, these results confirmed that, with a lower percentage of peptide mass per TMT label, and a higher probability of interference from co-isolated *E. coli* peptide-derived reporter ions, an increased rate of ratio compression was observed, negatively impacting both the ratios and statistics. It was concluded that the '100%', '50%', '20%' and '2%' ratios presented a suitable test case for evaluating the ability of the multi-feature weighting model to relieve the observed effects of ratio compression, from both human peptide co-isolation and *E. coli* peptide-derived co-isolation. Finally, it was concluded that the SPIQuE script was an effective and reproducible means of generating normalised ratios and statistics.

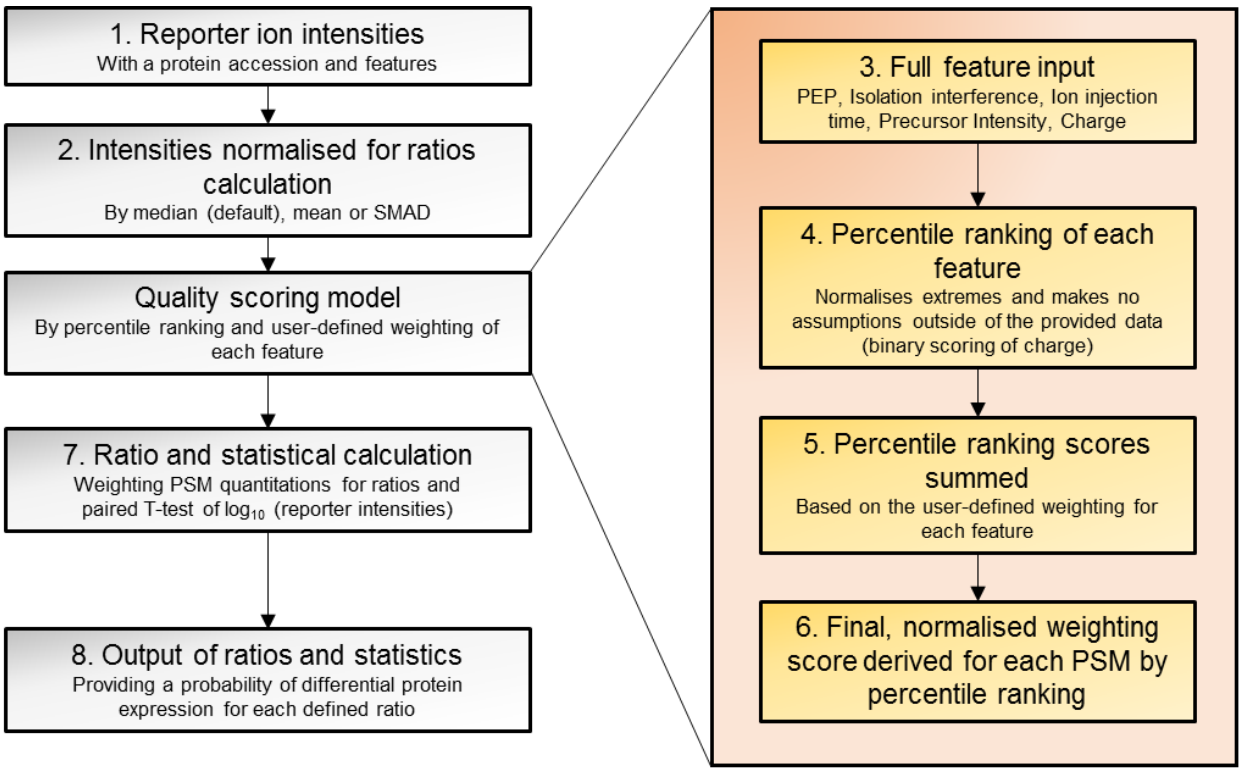
Figure 4.8 describes the steps taken and a worked example of applying the multi-feature weighting model to PSM-level data. Three proteins which demonstrated unweighted, statistical significance for all 4 HeLa:MCF7 ratios with between 8 and 12 PSM quantitations were chosen for this example. In these three cases, this highlights how an improvement can be derived from multi-feature weighting observed in both ratios and statistics. To visualise the effect of the PSM-weightings on these values for each protein, the individual weighting scores were plotted against their respective PSM-derived $\log_2$ (ratios) for the '100%' HeLa:MCF7 quantitation. This demonstrated a trend towards reduced weightings given to PSMs which exhibited greater relative ratio compression.

Multi-feature weighting was applied to the ratios of the HeLa:MCF7 proteome as a whole, for all 4 comparisons ('100%', '50%', '20%' and '2%') (**Figure 4.9**). When the resulting ratios were plotted against the unweighted ratios a trend emerged suggesting that multi-feature weighting provided a proteome-wide effect on ratio decompression. The vast majority of protein ratios were indicative of a greater or unchanged fold-change in protein expression after multi-feature weighing.

This trend was predominantly apparent for the '100%', '50%' and '20%' ratios, while even for the '2%' ratios, the few proteins demonstrating more than a 2-fold change, the majority of these had larger ratios with multi-feature weighing. Overall, for these proteins, multi-feature weighing provided a decompression effect averaging a 25% increase in the fold changes determined, relative to the unweighted analysis. The same evaluation was performed for proteins demonstrating an upregulation specifically in MCF7 cells which demonstrated an identical trend. While this demonstrated a successful decompression of ratios in proteins at the extremes of differential expression, further evaluation was required to determine the efficacy of the approach on more subtle ratios.

Figure 4.7. Peptide dilution series with spiked-in peptides models ratio compression from co-isolation. A. Defining the ratios used to model the effects of co-isolation on the human PSMs. **B.** The number of proteins with differential expression with both a fold-change of >2 and a multiple test-corrected p-value of <0.05 for each of the 4 defined ratios. Paired t-testing of the $\log_{10}$ (reporter ion intensities) evaluating all individual PSM ratios uniquely matching each protein. **C.** Volcano plots for each of the defined ratios plotting the $\log_2$ (ratios) of HeLa:MCF7 protein expression against the corresponding multiple test-corrected p-values (plotted as $-\log_{10}$ (p-values)) for each calculated protein expression change. **D.** The '50%', '20%' and '2'%' ratios and p-values from C. plotted against the corresponding '100%' values.


A.
E.coli HeLa MCF7
HeLa:MCF7
'100%' ratio
'50%' ratio
'20%' ratio
'2%'
126
127N 127C
128N 128C
129N 129C
130N 130C
TMT 10-plex labels
B.
Significant (p<0.05)differentially expressed (>2-fold) proteins
50% HeLa:MCF7 ratios
(991 proteins)
20% HeLa:MCF7 ratios
(814 proteins)
100%
HeLa:MCF7
ratios
(1007 proteins)
2%
HeLa:MCF7
ratios
(220 proteins)
63
33
162
23
2
99
30
556
1
179
19
1
2
5
0
C.
'100%' ratio
'50%' ratio
'20%' ratio
'2%' ratio
-log10 (ratio p-value)
500
507
494
497
432
382
168
52
log2 (ratio HeLa:MCF7)
D.
'50%' ratio
'20%' ratio
'2%' ratio
log2 (ratio HeLa:MCF7)
'100%' ratio
y = 0.9708x + 0.008
R² = 0.9615
y = 0.8657x - 0.0243
R² = 0.9221
y = 0.4009x - 0.1643
R² = 0.4649
y = 0.9739x - 0.0007
R² = 0.9675
y = 0.8826x - 0.1277
R² = 0.923
y = 0.3902x + 0.0329
R² = 0.4854
Y=X
-log10 (ratio p-value)



1. 2. 3. + 4. 5. 6. 7.

Protein Group Accessions	126	130C	Normalised Log2 (ratio 126:130C)	PEP	PEP Percentile-ranked score (5)	Isolation Interference [%]	Isolation Interference [%] Percentile-ranked score (10)	Ion Inject Time [ms]	Ion Inject Time [ms] Percentile-ranked score (9)	Intensity	Intensity Percentile-ranked score (3)	Charge	Charge Discrete score (1 for +2) (5)	Cumulative score (32 max)	Percentile-ranking of cumulative score	Protein-specific weighting
B1AVQ7	19430	184700	-3.78	9.96E-06	0.84	6	0.66	14	0.91	1429710	0.82	3	0	21.47	0.74	1.00
B1AVQ7	8435	65050	-3.48	4.16E-02	0.09	4	0.71	70	0.64	837642	0.72	2	1	20.55	0.69	0.93
B1AVQ7	11080	47380	-2.63	6.79E-02	0.04	4	0.71	104	0.54	560131	0.63	2	1	19.07	0.61	0.82
B1AVQ7	16580	68540	-2.58	9.94E-06	0.84	18	0.39	32	0.81	661617	0.67	3	0	17.41	0.50	0.68
B1AVQ7	8600	26460	-2.16	1.23E-05	0.81	15	0.45	98	0.55	347113	0.50	3	0	15.00	0.34	0.46
B1AVQ7	14260	36580	-1.89	2.80E-02	0.12	20	0.36	173	0.39	336711	0.49	2	1	14.17	0.29	0.39
B1AVQ7	11400	38950	-2.31	9.64E-02	0.02	20	0.36	161	0.41	361098	0.51	2	1	13.90	0.27	0.37
B1AVQ7	10050	20480	-1.56	1.03E-01	0.02	27	0.26	200	0.35	185052	0.33	2	1	11.82	0.15	0.20
C9J6N5	122800	31000	1.45	1.38E-05	0.80	0	1.00	63	0.67	1058441	0.77	2	1	27.35	0.96	1.00
C9J6N5	61220	17260	1.29	1.41E-05	0.80	7	0.63	103	0.54	663858	0.67	2	1	22.18	0.78	0.81
C9J6N5	122600	18490	2.19	2.33E-06	0.95	2	0.76	34	0.80	635893	0.66	3	0	21.55	0.75	0.78
C9J6N5	56270	13670	1.51	1.73E-05	0.77	12	0.51	159	0.41	376459	0.52	2	1	19.23	0.62	0.64
C9J6N5	195800	26450	2.35	1.19E-04	0.54	8	0.61	32	0.81	706526	0.68	3	0	18.15	0.55	0.57
C9J6N5	184200	29570	2.10	1.28E-02	0.18	1	0.78	41	0.77	533367	0.61	3	0	17.45	0.51	0.53
C9J6N5	136000	23080	2.02	1.24E-02	0.19	7	0.63	47	0.74	688333	0.68	3	0	15.95	0.41	0.42
C9J6N5	114200	30860	1.35	2.36E-02	0.14	31	0.21	128	0.48	469711	0.58	2	1	13.83	0.27	0.28
C9J6N5	111900	33730	1.19	7.29E-02	0.04	34	0.18	143	0.44	420669	0.55	2	1	12.68	0.20	0.20
C9J6N5	48900	18700	0.85	6.81E-03	0.23	37	0.16	200	0.35	189557	0.34	2	1	11.90	0.15	0.16
C9J6N5	46070	10160	1.65	5.58E-05	0.63	20	0.36	147	0.44	118015	0.23	3	0	11.35	0.13	0.13
C9J6N5	65430	23000	0.97	1.12E-01	0.01	55	0.05	200	0.35	271883	0.43	2	1	10.04	0.08	0.08
O14732	183700	11310	3.49	5.42E-03	0.25	0	1.00	67	0.66	919829	0.74	2	1	24.38	0.89	1.00
O14732	85530	7397	3.00	5.49E-03	0.25	0	1.00	117	0.50	528179	0.61	2	1	22.59	0.80	0.91
O14732	89110	8393	2.87	1.50E-05	0.79	4	0.71	116	0.50	570240	0.63	2	1	22.52	0.80	0.90
O14732	42140	5681	2.36	1.99E-02	0.15	0	1.00	200	0.35	164527	0.30	2	1	19.81	0.65	0.73
O14732	155000	10460	3.35	5.98E-03	0.24	9	0.58	78	0.62	796056	0.71	2	1	19.70	0.64	0.73
O14732	120800	12670	2.72	9.39E-03	0.21	9	0.58	84	0.60	736287	0.69	2	1	19.30	0.62	0.70
O14732	39820	4602	2.58	1.90E-05	0.76	35	0.17	200	0.35	198544	0.35	2	1	14.74	0.33	0.37
O14732	36950	16220	0.65	7.29E-05	0.60	55	0.05	200	0.35	162344	0.30	2	1	12.58	0.19	0.22
O14732	53460	11990	1.62	8.69E-02	0.03	30	0.22	184	0.37	336266	0.49	2	1	12.20	0.17	0.19

8.

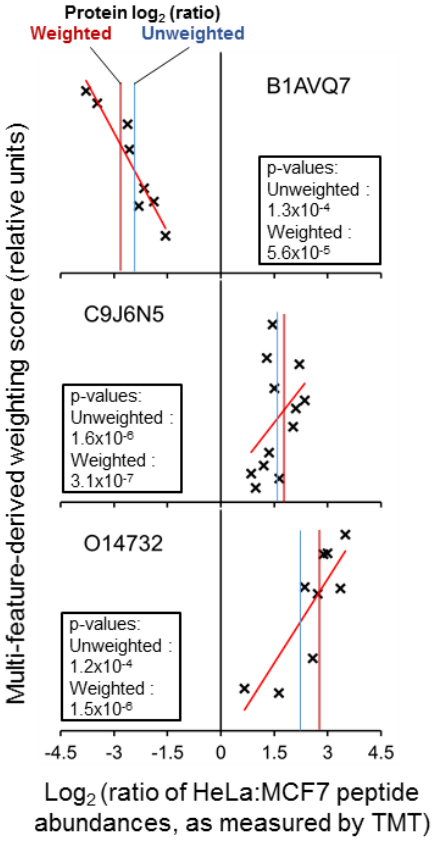


Figure 4.8. Workflow and worked examples of weighted ratio and p-value calculation by SPIQuE. The table demonstrate sets of PSMs specific to three differentially expressed proteins and details; the percentile ranking of PSM features, the cumulative and weighted summation of these ranks and the calculation of the final weighting factors used in the weighted calculation of the protein ratios and their statistics. The weightings are plotted against their respective $\log_2$ (ratios) to demonstrate correlation with ratio compression for the individual PSMs. The three proteins with differential expression were mucin-1 (MUC1) (B1AVQ7), FAM107B (C9J6N5) and inositol monophosphatase 2 (IMPA2) (O14732).

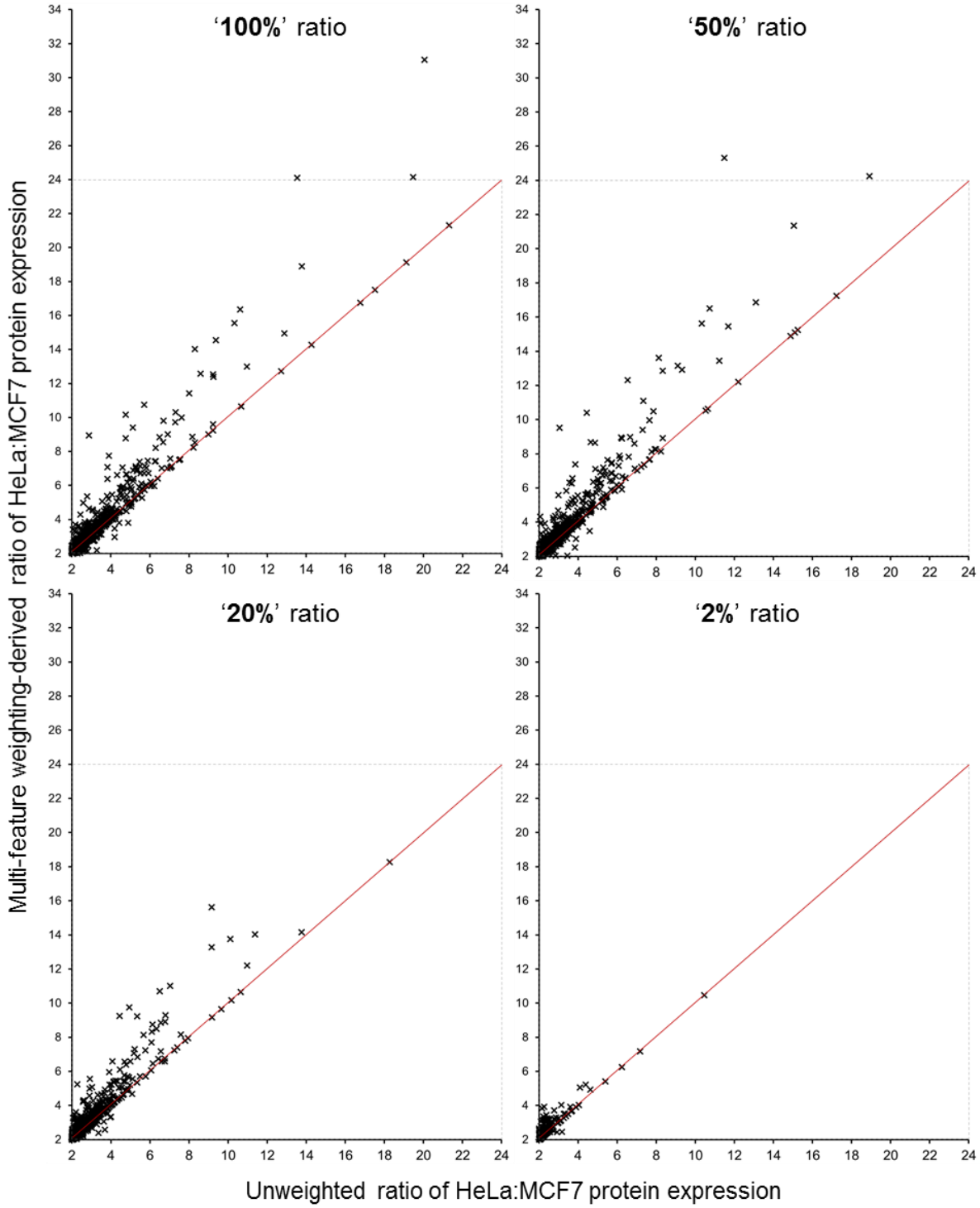


Figure 4.9. Ratio decompression in HeLa:MCF7 ratios from multi-feature weighting.
Ratios determined by multi-feature weighting plotted against ratios determined by the unweighted analysis.

4.9 RECOVERING DIFFERENTIALLY EXPRESSED PROTEINS FROM COMPRESSED RATIOS

To test the accuracy of the multi-feature weighting approach on a proteome-wide scale, the 1007 proteins identified for the '100%' ratios by the unweighted approach as significant ($p<0.05$) with a differential expression of at least 2-fold were used as a reference point. These proteins were considered the most confidently, differentially expressed, given that they were determined from the most confident data ('100%' ratios) with the most stringent, unweighted, approach. It was therefore concluded that determination of these proteins as differentially expressed ($p<0.05$, >2-fold) in the '20%' and '2%' ratios could provide a measure of the success of different approaches to ratio determination.

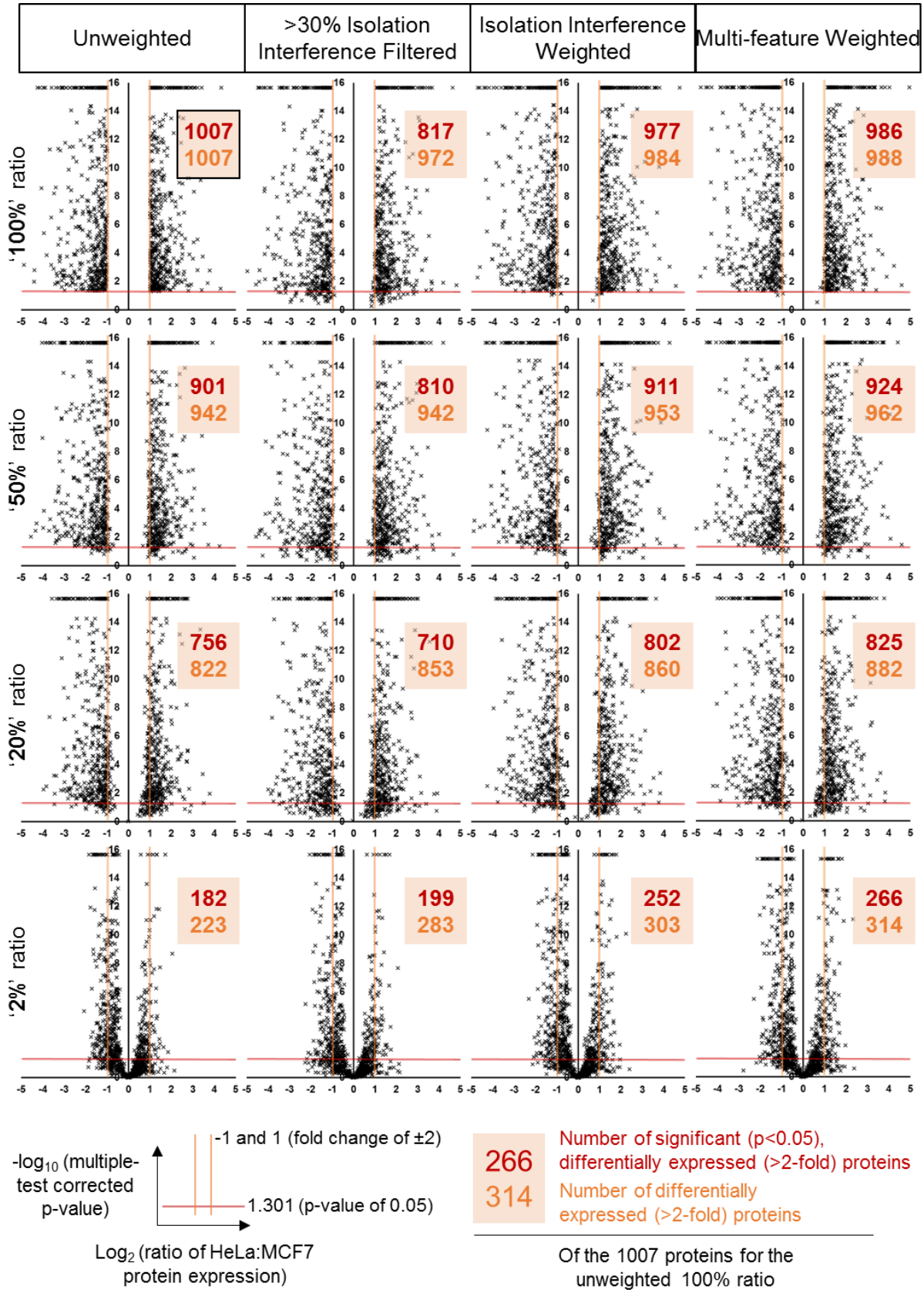
Each of the four ratios were subjected to analysis by four alternate means of ratio and statistical calculation: Firstly by the unweighted analysis described in **Figure 4.7**; secondly by the commonly adopted approach of filtering out spectra with >30% isolation interference; thirdly by weighting the relative contribution of PSM quantitations using the percentile-ranked isolation interference; and finally, by the '10:9:3:5:5' multi-feature weighting model. These analyses were then filtered to observe the determination rates of the 1007 confidently differentially expressed proteins. The $\log_2$ (ratios) and $-\log_{10}$ (p-values) for these 1007 proteins were plotted as volcano plots for these 16 analyses (**Figure 4.10**). Filtering out the PSMs with >30% isolation interference demonstrated some benefit only to the compressed '20%' and '2%' ratios with an additional 31 and 60 proteins observed with a >2-fold differential expression compared to unweighted, respectively. Filtering caused a loss of significance for many of the proteins, observable as a noticeable downward shift in the volcano plots, for all but the '2%' ratios. The reduced significance caused by the removal of these quantitations demonstrated that despite the high level of observed interference, they still had value in the calculation of protein expression statistics.

For the isolation interference-only and multi-feature weighting approaches similar effects were observed on the number of proteins reaching a >2-fold differential expression, but at a consistently greater degree than that of filtering. Multi-feature weighting was, however, consistently more effective than both isolation interference-centric analyses for ratio calculation and correctly determined 41% more proteins with >2-fold differential expression. Statistically, both the isolation interference and multi-feature weighting approaches failed to identify just 3% and 2% proteins as significant for the '100%' ratios, respectively; likely as an artefact of this alternative approach on marginally significant proteins. For the '50%', '20%' and '2%' ratios, however, a substantially larger number of the 1007 proteins were identified as significantly, differentially expressed ($p<0.05$, >2-fold) compared to the unweighted and filtered analyses. 'Multiple features' consistently outperformed 'isolation interference' as a statistical weighting

approach. The most pronounced improvement was observed for the '20%' and '2%' ratios, with 2.9% and 5.6% of the 1007 proteins recovering significance in both cases, respectively.

Overall, the multi-feature weighting approach consistently provided the most effective means of determining differential protein expression in compressed data, facilitating the strongest recovery of proteins which were impacted by ratio compression in the '20%' and '2%' ratios. Additionally multi-feature weighting provided up to 46% more significantly differentially expressed proteins ($p<0.05$, >2-fold) than the unweighted analysis.

Figure 4.10. Comparing the efficiency of approaches in the recovery of differential protein expression from compressed ratios. Differential protein expression for the 100%, 50%, 20% and 2% ratios was determined by: 1. unweighted analysis of all PSM quantitations; 2. only quantitations from spectra with <30% isolation interference; 3. quantitations weighted by their isolation interference alone and; 4. quantitations weighted by multiple PSM features (isolation interference, ion injection time, precursor intensity, PEP score and charge with relative of weightings of 10:9:3:5:5, respectively). The 1007 proteins identified as significantly, differentially expressed ($p<0.05$, >2-fold) by the unweighted, '100%' (126:130C) ratios were used as a reference point as the most confident differentially expressed proteins. Volcano plots demonstrate the $\log_2$ (ratios) and $-\log_{10}$ (p-values) determined by each approach for each of the ratios. The numbers describe the quantity of proteins in agreement with those 1007 'confident' differentially expressed for: 1. the number of proteins with both a p-value of <0.05 and a differential expression of >2 fold; and 2. the number of proteins with a differential expression of >2 fold.


Unweighted
>30% Isolation Interference Filtered
Isolation Interference Weighted
Multi-feature Weighted
'100%' ratio
'50%' ratio
'20%' ratio
'2%' ratio
1007
1007
817
972
977
984
986
988
901
942
810
942
911
953
924
962
756
822
710
853
802
860
825
882
182
223
199
283
252
303
266
314
-log10 (multiple-test corrected p-value)
-1 and 1 (fold change of ±2)
1.301 (p-value of 0.05)
Log2 (ratio of HeLa:MCF7 protein expression)
266
Number of significant (p<0.05), differentially expressed (>2-fold) proteins
314
Number of differentially expressed (>2-fold) proteins
Of the 1007 proteins for the unweighted 100% ratio

4.10 EVALUATING THE ADDITIONAL DIFFERENTIALLY EXPRESSED PROTEINS DETERMINED BY MULTI-FEATURE WEIGHTING

While the previous section demonstrated an ability to recover the determination of differentially expressed proteins by multi-feature weighting, it did not consider the potential false discovery that may have been caused by this novel approach. It was therefore considered important to interrogate any proteins identified reaching significant differential expression ($p<0.05$, >2-fold) uniquely as a result of multi-feature weighting. **Figure 4.11** evaluates the proteins determined by multi-feature weighting and dissects out those proteins which were not considered differentially regulated by established approaches.

Proteins reaching significant differential expression ($p<0.05$, >2 fold) by the unweighted, isolation interference filtered, isolation interference weighted and multi-feature weighted '100%' ratio analyses were plotted as a Venn diagram (**Figure 4.11A**). Of the 1120 significant and differentially expressed proteins determined by the unweighted and isolation interference-filtered approaches, 61 (5.4%) were not determined by multi-feature weighting. By contrast, multi-feature weighting, determined 134 additional proteins reaching significance for the unweighted or filtered analyses, the majority of these (70) were also present for the isolation interference weighting.

Figure 4.11B examines the number of additional proteins determined as significant, differentially expressed ($p<0.05$, >2 fold) for the '50%', '20%' and '2%' ratios. Whilst the majority of these were determined at '100%', 190 proteins reached significance uniquely to the other sub-optimal ratios. The proteins for the '100%' ratios, described in this Venn diagram were the same as those compared to the other approaches in **Figure 4.11A.**

The 70, 64 and 190 proteins from **Figure 4.11A** and **Figure 4.11B** uniquely determined as differentially expressed a result of multi-feature weighting were examined for significance ($p<0.05$) in any of the '100%', '50%' '20%' or '2%' ratios (**Figure 4.11C**). This was done to identify those substantial outliers which exhibited significance uniquely as a result of the multi-feature weighting. A total of 42 of the 134 '100%' multi-feature weighting-unique proteins were newly determined as significant. Of the 190 sub-optimal ratio-determined proteins, only 19 were not observed as significant for their respective '50%', '20%' and '2%' unweighted ratio analyses. This was potentially an artefact of the skewing effects of *E. coli* interference observed in **Figure 4.7C**.

These 42 and 19 proteins uniquely determined as significantly, differential expressed ($p<0.05$, >2 fold) with multi-feature weighting were evaluated by plotting the $-\log_{10}$(p-values) of the weighted versus the unweighted results (**Figure 4.11D**). This plot illustrates the greatly

variable effects multi-feature weighting had on the effects on p-values, relative to the unweighted approach. Some proteins reached significance though a marginal improvement, while others, such as the protein denoted with '1' had its p-value reduced from 0.711 to 0.047. Further investigation of these proteins exhibiting a sizeable shift in p-values by multi-feature weighting was therefore conducted.

To demonstrate the impact of the multi-feature weighting on the least-likely differentially expressed proteins, PSM weightings of the 20 least significantly regulated proteins, as determined without weighting, were plotted against their individual PSM $\log_2$ (ratios) (**Figure 4.11E**). For the majority of these proteins, a trend was apparent with weightings for each protein selectively favouring those PSMs with the greatest observed fold change. Given the assumption that all the PSMs for each protein should exhibit the same, or very similar fold changes, the multi-feature weighting model appears to be correctly predicting an approximate probability, for each of the PSMs, of co-isolation and ratio compression. Those cases with the greatest shift in significance, (plots 1-6) contain individual PSMs ratios indicative of the opposite direction of differential expression. Based purely on the PSM features, these were determined as 'poor' quantitation events and subsequently had a minimised contribution to the statistics and ratio calculations. While such a trend is absent from plot 11, it is clear that all the PSM quantitations are indicative of truly regulated proteins – the outlier in this case had 96% isolation interference.

By evaluating the individual PSM quantitations of these 20 proteins, it was clear that the multi-feature weighting model enabled efficient prediction of precursor co-isolation and ratio compression within spectra on the basis of PSM feature alone. Within the context of this controlled dataset, the multi-feature weighting model resulted in an improvement of statistics and reduced ratio compression that improved the determination of differentially expressed proteins with greater accuracy and in greater number. The question remained, however, if this approach would translate to other, less controlled datasets.

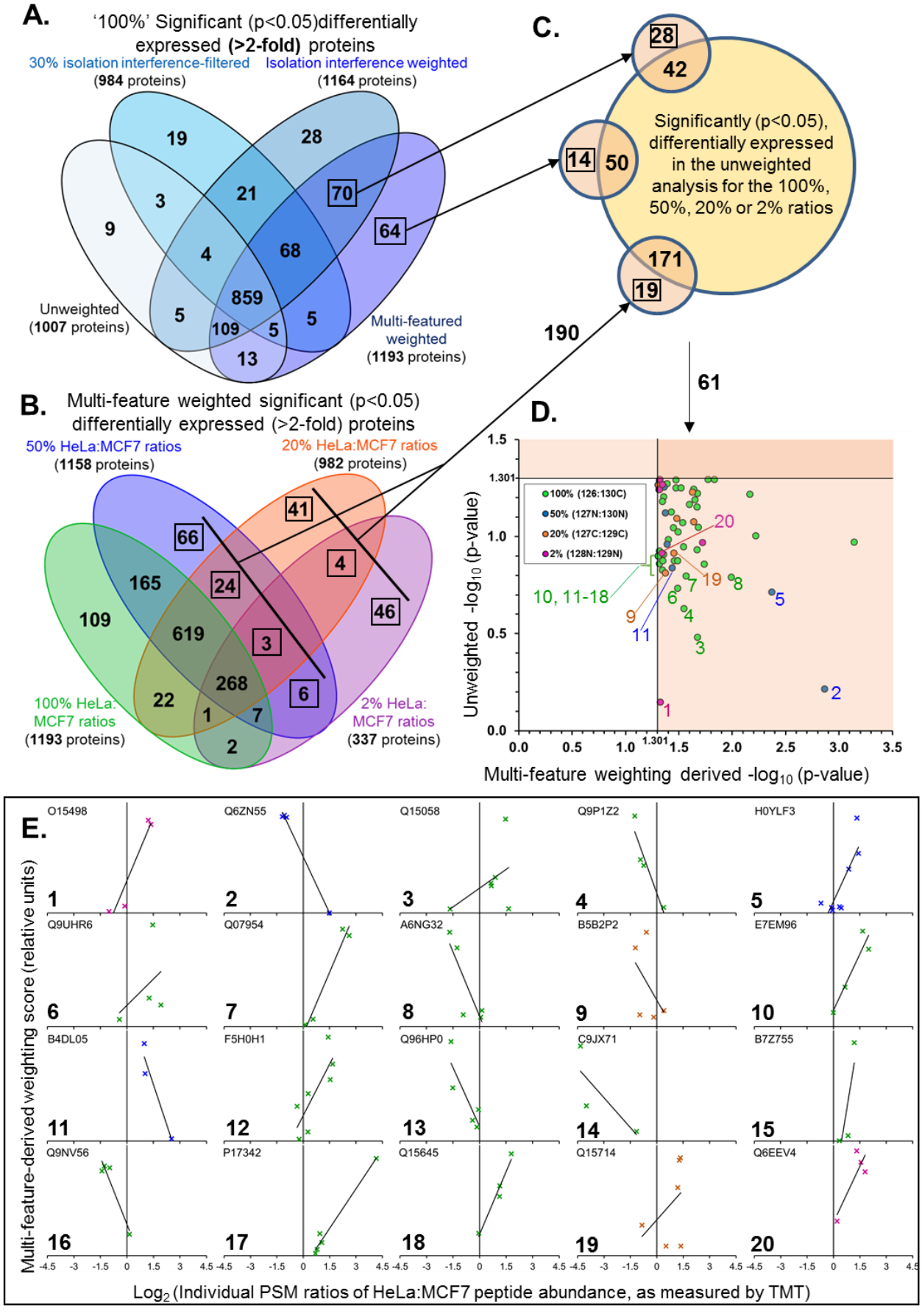

A.
'100%' Significant (p<0.05)differentially expressed (>2-fold) proteins
30% isolation interference-filtered (984 proteins)
Isolation interference weighted (1164 proteins)
Unweighted (1007 proteins)
Multi-featured weighted (1193 proteins)
19
28
3
21
70
9
64
4
68
859
5
109
5
5
13
B.
Multi-feature weighted significant (p<0.05) differentially expressed (>2-fold) proteins
50% HeLa:MCF7 ratios (1158 proteins)
20% HeLa:MCF7 ratios (982 proteins)
100% HeLa: MCF7 ratios (1193 proteins)
2% HeLa: MCF7 ratios (337 proteins)
66
41
165
24
4
109
46
619
3
268
6
22
1
7
2
C.
28
42
14
50
Significantly (p<0.05), differentially expressed in the unweighted analysis for the 100%, 50%, 20% or 2% ratios
171
19
190
61
D.
100% (126:130C)
50% (127N:130N)
20% (127C:129C)
2% (128N:129N)
Unweighted -log10 (p-value)
Multi-feature weighting derived -log10 (p-value)
E.
Multi-feature-derived weighting score (relative units)
Log2 (Individual PSM ratios of HeLa:MCF7 peptide abundance, as measured by TMT)
O15498
Q6ZN55
Q15058
Q9P1Z2
H0YLF3
Q9UHR6
Q07954
A6NG32
B5B2P2
E7EM96
B4DL05
F5H0H1
Q96HP0
C9JX71
B7Z755
Q9NV56
P17342
Q15645
Q15714
Q6EEV4

Figure 4.11. Evaluating the additional, significantly differentially expressed proteins determined by multi-feature weighting. A. The number of proteins identified from the '100%' (126:130C) ratios as significantly differentially expressed ($p<0.05$, >2-fold) by: 1. unweighted analysis of all PSM quantitations; 2. only quantitations from spectra with <30% isolation interference; 3. quantitations weighted by their isolation interference alone and; 4. quantitations weighted by multiple PSM features (isolation interference, ion injection time, precursor intensity, PEP score and charge with relative of weightings of 10:9:3:5:5, respectively) were examined for overlap using a Venn diagram **B.** Number of proteins identified as significantly, differentially expressed ($p<0.05$, >2-fold) by multi-feature weighting of the '100%', '50%', '20%' and '2%' ratios were examined for overlap using a Venn diagram. **C.** Proteins identified as significantly, differentially expressed ($p<0.05$, >2-fold) uniquely by the multi-feature weighting approach were compared and filtered to select only those which had gained significance by the multi-feature weighted approach using a Venn diagram. **D.** The $-\log_{10}$ (p-values) for proteins only reaching significance by multi-feature weighting were plotted against their respective unweighted p-values. **E.** For proteins reaching significance only by multi-feature weighting (D.), the individual PSM quantitations and their respective multi-feature-derived weighting scores of the 20 lowest unweighted p-values were plotted.

4.11 EVALUATING MULTI-FEATURE WEIGHTING IN UNCONTROLLED DATASETS

With the demonstration of ratio decompression, ratio recovery and the identification of additional, potentially differentially regulated proteins in the HeLa/MCF7 data, it was considered that this could be a data-set-specific phenomenon – given that the multi-feature weighting was defined using the spike-in within the same 2D-LC MS/MS analysis.

To demonstrate its utility under different experimental conditions, the multi-feature weighting approach was applied to an iTRAQ 8-plex data set acquired with an LTQ Orbitrap Velos (Johansson et al.) [390]. This quantitative proteome included biological replicates, which provided the opportunity to assess the reproducibility of significantly, differentially expressed ($p<0.05$, >2 fold) determination. The search report was downloaded from PRIDE **(**PXD000281), consisting of an analysis of biological replicates (controls; 113 and 114, treated; 119 and 121) of therapeutic response to fulvestrant in a human breast cancer cell line LCC2. The data was acquired by an LTQ Orbitrap Velos with HCD and CID fragmentation and an isolation width of 2 m/z, quantitating approximately 7000 proteins from approximately 190000 PSM quantitation events. Using SPIQuE, the two independent treated:control ratios of 119:113 and 121:114 were subject to unweighted and '10:9:3:5:5' multi-feature weighted analyses. **Figure 4.12A** describes the number of proteins determined as significantly, differentially expressed ($p<0.05$, >2 fold) as

a result of these two different approaches. The majority of such proteins were determined independently of weighting across the two ratios (90) but, when applied, weighting yielded an additional 44 (49%) proteins. Of these 44 proteins, 20 were commonly determined for both ratios. 17 proteins, determined for only one ratio, were confirmed as significantly, differentially expressed in the other ratio upon the implementation of weighting.

Plotting the weighting-derived $-\log_{10}$(p-values) for both ratios for each of the 44 weighting-determined significantly, differentially expressed ($p<0.05$, >2 fold) proteins highlighted a reproducible trend shifted in favour of the multi-feature weighting approach (**Figure 4.12B**). On closer examination, 12 proteins were determined to be uniquely significant as a result of multi-feature weighting – demonstrating a similar, wide distribution of p-value improvements to that seen in **Figure 4.11D**. The weighting-derived absolute $\log_2$(ratios) of these 44 proteins were also plotted against their respective unweighted values (**Figure 4.12C**). This demonstrated that for those 44 proteins with significance for both weighted and unweighted calculations, the multi-feature weighting-derived ratio improvements, also played a role in the additional determination of differentially expressed proteins ($p<0.05$, >2 fold).

The 12 proteins specifically acquiring significance as a result of multi-feature weighting; three for 119:113 (proteins 1-3), five for 121:114 (proteins 4-8) and four for both ratios (proteins 9-12), were evaluated by plotting individual PSM quantitations against respective weighting scores (**Figure 4.12D**). Weightings were once again indicative of a trend towards the selective emphasis of uncompressed ratios, again explaining the derived improvements to significance and ratios.

Two additional experiments were analysed by multi-feature weighting (**Figure 4.13A**), with data also downloaded from PRIDE (PDX001125 and PDX000413). The $-\log_{10}$ (p-values) for all quantified proteins were plotted to compare the weighted and unweighted analyses (**Figure 4.13B**). The number of proteins reaching significance ($p<0.05$) by each approach was shown by Venn diagram. These were plotted alongside the equivalent analyses of the HeLa:MCF7 ratios and the LCC2 ratios. When multi-feature weighting was applied, all eight ratios revealed a consistent pattern of statistical improvement. Only a small percentage of $-\log_{10}$ (p-values) decreased, and the vast majority increased as a result of the multi-feature weighting. The Venn diagrams demonstrated that the multi-feature weighting approach yielded ~10% more proteins reaching the threshold of significance and only ~3% losing significance. The ratios generated with and without weighting were evaluated using fold change cut-offs of 2 and 1.5. The number of proteins determined to reach these thresholds across the 8 ratios are illustrated in **Figure 4.13C**. For the 2-fold threshold, all 8 weighted analyses provided between 10 and 30% more proteins with, at most, a loss of 4%. The 1.5-fold cut-off demonstrated marginally less

advantage to weighting with between 8 and 26% more proteins and a loss of up to 7%. However, across all ratios, for all experiments, a consistent trend of improvements was observed as a result of multi-feature weighting.

The multi-feature weighting strategy demonstrated similar improvements across eight ratios in four independently acquired experimental datasets. Overall, this suggests that this approach has generic utility for improving the determination of significantly, differentially expressed proteins from isobaric tag labelled data.

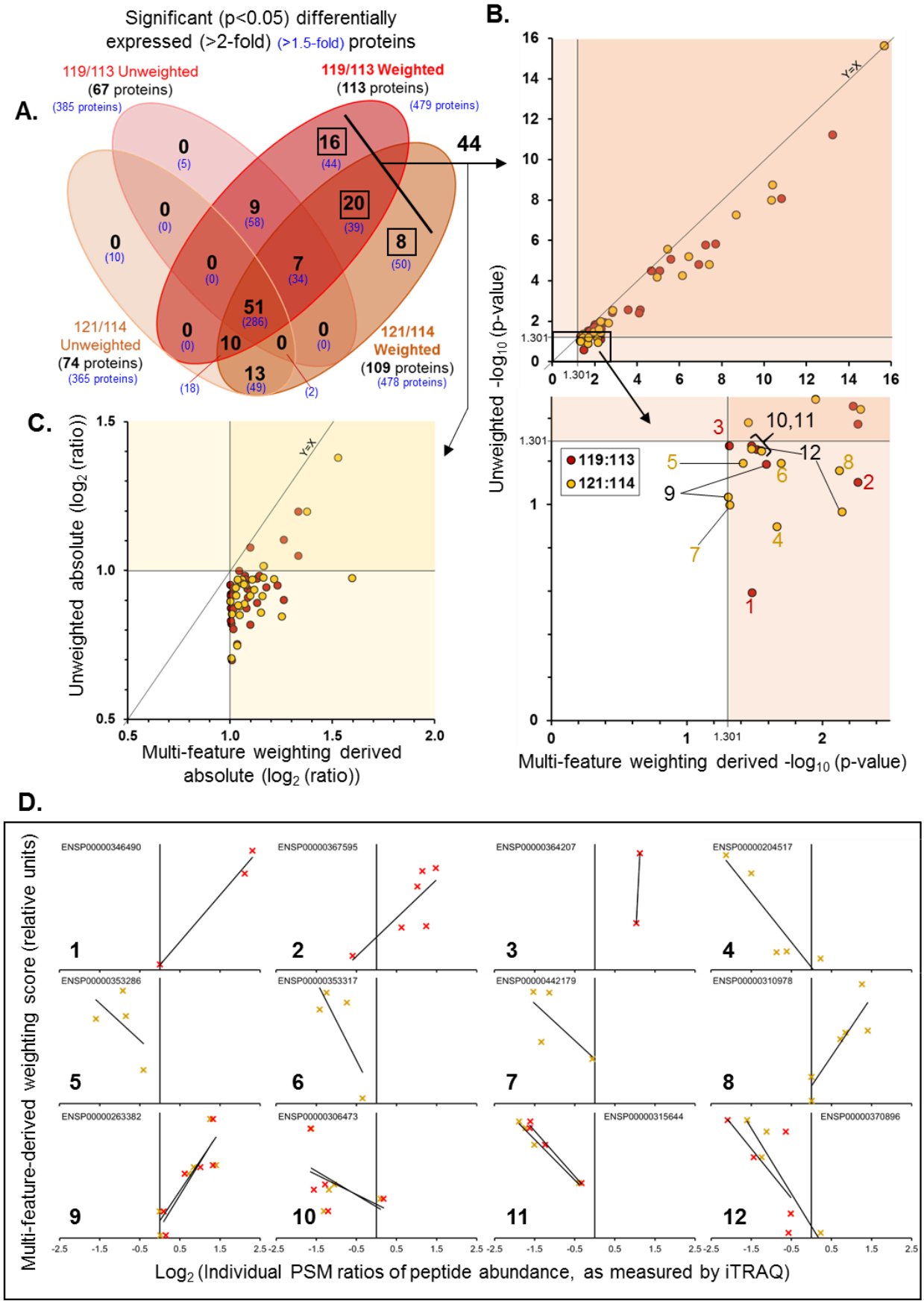
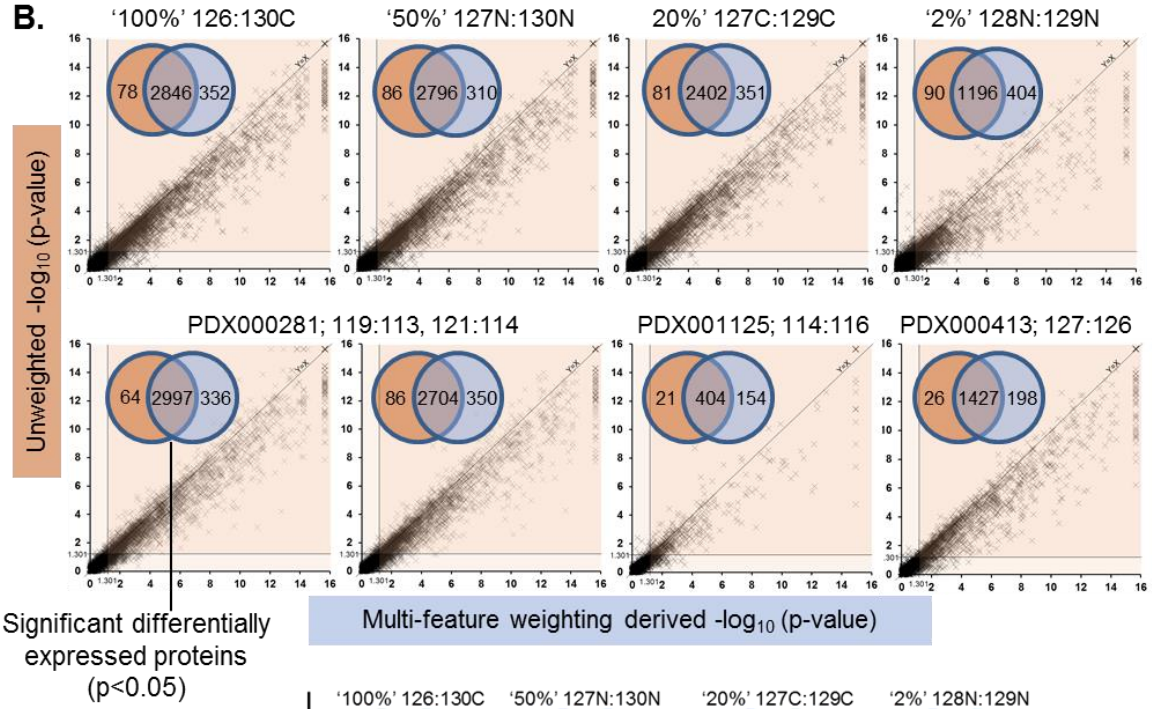

A.
Significant (p<0.05) differentially expressed (>2-fold) (>1.5-fold) proteins
119/113 Unweighted (67 proteins)
(385 proteins)
119/113 Weighted (113 proteins)
(479 proteins)
121/114 Unweighted (74 proteins)
(365 proteins)
121/114 Weighted (109 proteins)
(478 proteins)
0 (5)
16 (44)
44
0 (0)
9 (58)
20 (39)
8 (50)
0 (10)
0 (0)
7 (34)
51 (286)
0 (0)
10
0
0 (0)
13 (49)
(18)
(2)
B.
Unweighted -log10 (p-value)
Multi-feature weighting derived -log10 (p-value)
Y=X
1.301
119:113
121:114
10,11
12
C.
Unweighted absolute (log2 (ratio))
Multi-feature weighting derived absolute (log2 (ratio))
D.
Multi-feature-derived weighting score (relative units)
Log2 (Individual PSM ratios of peptide abundance, as measured by iTRAQ)
ENSP00000346490
ENSP00000367595
ENSP00000364207
ENSP00000204517
ENSP00000353286
ENSP00000353317
ENSP00000442179
ENSP00000310978
ENSP00000263382
ENSP00000306473
ENSP00000315644
ENSP00000370896

Figure 4.12. Multi-feature weighting analysis of differential protein expression in uncontrolled data. Multi-feature weighting analysis of an external dataset downloaded from PRIDE (project identifier PXD000281)**.** In brief, these data consisted of an analysis of biological replicates (controls; 113 and 114, treated; 119 and 121) of therapeutic response to fulvestrant in a human breast cancer cell line LCC2. Data was acquired by an LTQ Orbitrap Velos with HCD and CID fragmentation and an isolation width of 2 m/z. iTRAQ 8-plex quantitated approximately 7000 proteins from approximately 190000 PSM quantitation events. **A.** Multi-feature weighted and unweighted ratios and p-values were calculated using SPIQuE for the biological replicates 119:113 and 121:114. PSM quantitations and features were subjected to the SPIQuE workflow. The number of proteins determined as significantly differentially expressed ($p<0.05$, >2-fold) for each analysis was displayed as a Venn diagram. **B.** Comparison of the weighted and unweighted $-\log_{10}$ (p-values) for those proteins uniquely determined by the multi-feature weighting approach for the 119:113 and 121:114 ratios. Shading defined the regions of significance ($p<0.05$). **C.** The absolute ($\log_2$ (ratios)) for proteins identified as significantly, differentially expressed uniquely by the multi-feature weighting approach were plotted. Shading highlights the regions of a >2-fold change. **D.** For proteins reaching significance only by multi-feature weighting, the individual PSM quantitations and their respective multi-feature-derived weighting scores were plotted for 119:113 (1-3, red) and 121:114 (4-8, orange). For plots 9-12, both ratios gained significance.

A.

Identifier	Protein source	Quantitation method	Instrument	MS/MS	Isolation window	File	Peptide probability score
PXD001388	HeLa, MCF7 cell lines	TMT 10-Plex	Orbitrap Elite	HCD/ CID	1 m/z 2 m/z	'All'	PEP
PDX000281	LCC2 breast cancer cell line	iTRAQ 8-plex	LTQ Orbitrap Velos	HCD/ CID	2 m/z	'LCC2'	PEP
PDX001125	Maize seedlings root cells	iTRAQ 8-plex	Q-Exactive	HCD	2 m/z	'P12311a'	PEP
PDX000413	Mouse embryonic fibroblasts	TMT 6-plex	LTQ Orbitrap Velos	HCD	1.2 m/z	'Mix1'	Exp Value

B.



C.

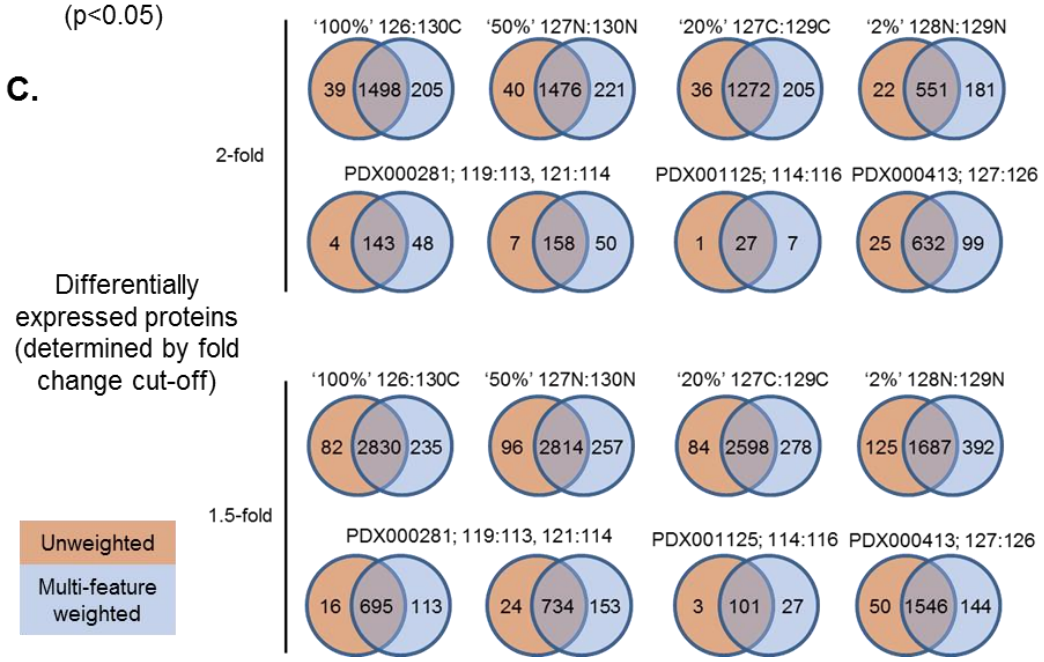


Figure 4.13. Weighted versus unweighted-determined differential expression across multiple datasets. A. The details of the evaluated experiments; from the present study, the LCC2 data (evaluated in **Figure 4.12**) and two additional isobaric tag-labelled proteomics datasets downloaded from PRIDE. **B.** PSM quantitations and features from these studies were subject to the SPIQuE workflow by both the multi-feature weighted and unweighted approaches to determine the respective ratios and significance values resulting from each method. The resulting $-\log_{10}$ (p-values) for each of the 8 ratios were plotted for weighted versus unweighted. The number of proteins reaching a significance threshold of $p<0.05$ by each method was demonstrated by Venn diagram (unweighted; left, weighted; right). **C.** As for B, Venn diagrams were used to demonstrate the number of proteins reaching a fold change threshold of 2 and 1.5 for each of the 8 ratios when the weighted and unweighted approaches were applied.

4.12 DISCUSSION, LIMITATIONS AND CONCLUSIONS

This chapter has evaluated the properties of PSMs in an attempt to better understand ways of predicting the extent of peptide co-isolation and ratio compression. Isobaric tags, such as TMT and iTRAQ, have become widely adopted approaches for relative quantitative proteomics, but their utility remains limited, primarily due to ratio compression resulting from co-isolation. Although approaches exist to minimise the effects of co-isolation, the techniques are not always widely accessible. The proposed multi-feature weighting approach, carefully defined using an extensive, complex spike-in proteome, has demonstrated an ability to predict the relative ratio compression of PSMs in both controlled and uncontrolled isobaric tag-labelled data. In combination with the use of SPIQuE this has demonstrated an improvement to both the calculation of ratios and the statistical evaluation of differential protein expression.

The correlation of features described in **Figure 4.5**, followed the expected principles of MS. The intrinsically related features of precursor intensity and ion injection time both correlated with ratio compression presumably on the basis that the higher the intensity of a precursor, the lower the probability of an equal or greater intensity precursor imposing ratio compression through co-isolation. The converse is true for peptides with a low intensity, with the additional effect of precursors being isolated which fall below the threshold of detection, but which still contribute to the total reporter ions when summated. The effect of co-isolation of sub-detection threshold precursors may be a component causing low correlation between isolation interference and the determined co-isolation scores. Additionally, some correlation will have been lost by co-isolation of two or more *E. coli* peptides, however given the minority status of these within the proteome any impact will have been minimised.

Peptides with a 2+ charge demonstrated significantly lower co-isolation and improved the co-isolation score. This may primarily be an effect of higher charged peptides being isolated in the low m/z detection region, which typically contains a greater number of peptides. The improvement to the multi-feature weighting model demonstrated by the use of the peptide probability measure, PEP score, possibly relates to the more effective fragmentation and intensities seen by peptides with better PEP scores giving a more intense and less noisy reporter region.

While the SPIQuE tool and multi-feature weighting provide an overall improvement to ratios and statistics for isobaric tag-labelled data, it remains limited by the low R^2 value of the multi-feature model (0.37). This is compensated for by the trend that emerges when a larger number of quantitations are available, as the model predicts the probability of ratio compression. When lower number of PSM quantitations are available for a protein, the probability of an incorrect p-value evaluation resulting from outlying weighting assignments is raised. As more accurate measures and features of PSM-specific co-isolation become readily available, such as a precise measure of peaks in the MS^2 spectra not derived from either the target peptide fragmentation [310] or common label-derived artefacts [391], the weighted statistics approach promises even more accurate improvements to the reduction of ratio compression.

The model's greatest utility is in guiding the shortlisting process of those proteins most likely to be differentially expressed in a proteome. The observation that more than 80% of the *E. coli*-modelled differentially expressed proteome's PSMs had more than 50% ratio compression is a stark demonstration of how many protein ratios are adversely effected by co-isolation. SPIQuE offers a means of automatically selecting the majority of these PSMs with a high probability of ratio compression and minimising their effect on the resulting protein ratios. In this capacity, SPIQuE provides a superior means of post-acquisition iTRAQ or TMT data analysis, allowing the user to more accurately consider the quality of their PSMs in their large scale quantitative analyses. When querying data for differential protein expression, the use of this tool as a secondary analysis may yield novel findings compared to the primary analysis, offering a simple, automated means of stimulating further hypotheses. However, the manual interrogation of these candidates' raw data is still important in drawing biological conclusions, regardless of the approach taken to ratio calculation.

Statistical analysis by t-test is limited in the context of iTRAQ and TMT experiments where proteins are quantified with quantitations with very consistent but minimal ratios, eg. <1.1. This can be accounted for by implementing a fold change cut-off for analyses.

The datasets used as test cases for SPIQuE, were from Orbitrap-derived analyses, due to availability, but the principles suggest that an isolation interference- and ion injection time-dominant weighting approach will be applicable to iTRAQ and TMT data from other types of MS analysis. Additionally, the use of alternative peptide identification probabilities in the weighting model, where the PEP score is not available, should be universally applicable, due the minor influence this has on the model and the comparable nature of these scores.

The focus of SPIQuE's development has been to provide a simple tool that can be used by any researcher, regardless of knowledge of scripting languages. The website spiquetool.com was created with this aim in mind.

In conclusion, SPIQuE and multi-feature weighting promise to assist users in the post-acquisition analysis of isobaric tag-labelled data, facilitating a more accurate interpretation of relative protein expression.

5.0 QUANTITATIVE PROTEOMIC CHARACTERISATION OF Eμ-*MYC* AND Eμ-*TCL1* B-CELL CANCERS

5.1 CHAPTER INTRODUCTION

Building on the characterisation of the Eμ-*myc* model described in **Chapter 3**, the further application of the latest proteomic methodologies was considered. Several areas for improvement upon this previous study design were identified. The isolation of fresh tumour B cells, rather than from frozen splenocyte aliquots, would ensure maximum cell viability ensuring minimum deviation from *in vivo* protein expression and identical handling compared with controls. The characterisation of the whole cell populations, rather than FFE enrichment, was used for the reasons outlined in **Section 3.10**, such as ensuring sufficient protein material and minimising cell handling**.** Consideration was given to utilising the full capacity of an iTRAQ 8-plex isobaric tag labelling set allowing the inclusion of more controls and biological replicates. It was also decided that increasing the number of samples per control pool to 6 would make the comparison to controls more robust, minimising any variation from single outlying samples.

In addition to these design improvements, several technical advancements had become readily available that promised a substantial improvement in the depth of proteome coverage.

Access to an Orbitrap Elite MS platform offered the potential to dramatically increase the depth of proteome coverage, offering ultra-high resolution Orbitrap analysis, 12 Hz LIT MS/MS acquisition speed and a non-linear ion path reducing background noise [269]. A means of increasing peptide ESI efficiency using DMSO also offered the potential to increase proteomic coverage with stronger peptide signals [268]. Access to software implementing the target-decoy validation tool 'Percolator' also became readily available; which uses semi-supervised learning to more accurately and efficiently determine peptide identifications [278]. The SPIQuE tool, described in **Chapter 4** also promised to increase the accuracy of protein ratios, which also acted to demonstrate the substantial depth of proteome coverage which could be generated using these approaches.

In addition to a more comprehensive characterisation of the Eμ-*myc* model of BL, the Eμ-*TCL1* mouse was chosen as a second B-cell cancer model. Eμ-*TCL1* mice develop a CLL-like B-cell tumour induced by the overexpression of the oncogene, *TCL1*, by the same Ig enhancer, Eμ, as the Eμ-*myc* model [144]. The incorporation of this model into the iTRAQ 8-plex experimental design allowed a simultaneous cross comparison between two contrasting B-cell cancer phenotypes.

The Eμ-*myc* model takes approximately 100 days to present with a terminal tumour, at which stage they develop aggressive lymphomas frequently presenting with splenomegaly, thymomas, mesenteric tumours and peripheral lymph node tumours. The time course from first presentation to lethality is rarely more than a week [176, 188]. In contrast, Eμ-*TCL1* mice present with an accumulation of blood-borne $CD5^+$ $B220^+$ leukaemic cells, at around 3-5 months of age [144]. This population accumulates over the period of approximately a year, when mice become terminal with splenomegaly with white pulp hyperplasia and other organs demonstrating lymphocytes infiltration [145].

The *TCL1* oncogene effects are less clearly defined than those for *myc*, with multiple pathways contributing to *TCL1*-driven oncogenesis. The pathways engaged by TCL1, such as the BCR, AKT, p300 and NF-κB activation, have been shown to contribute to cancer development in the Eμ-*TCL1* model [218-221]. The extent of these pathways' contributions to tumourigenesis and comprehensiveness of this list, however, remains uncertain. In contrast, for the Eμ-*myc* model, the pleiotropic, proliferative and stem-like-promoting effects of myc overexpression are well-characterised mechanism of cancer development [17, 165, 178-180].

Proteomics offers an unbiased, systematic means of identifying proteins differentially expressed in the Eμ-*myc* and Eμ-*TCL1* B-cell cancer models that may help to identify proteins

involved in tumourigenesis, cancer proliferation and potentially means of inhibiting growth in these tumours.

This chapter describes the characterisation of the Eμ-*myc* and Eμ-*TCL1* B-cell cancer models by quantitative iTRAQ proteomics, applying the 2D-LC MS/MS methods described in **Section 2.19** and the SPIQuE approach to quantitation detailed in **Chapter 4**.

5.2 Eμ-*MYC* and Eμ-*TCL1* Tumour Presentation

To better understand the Eμ-*myc* and Eμ-*TCL1* models for defining the biological conditions for proteomics characterisation, the natural history of the disease was observed in 27 Eμ-*myc* mice and 14 Eμ-*TCL1* mice. **Figure 5.1** demonstrates that the majority of Eμ-*myc* tumours appeared as terminal between 50 and 150 days. The median age of terminal presentation was 106 days and the SD of time to presentation was 69 days. Eμ-*myc* mice presented with splenomegaly in the majority of cases, as well as with frequent thymomas, lymph node tumours and mesenteric tumours. Other observations included a generally reduced haematocrit and increased haemolysis, in agreement with the literature [176, 188].

In comparison, Eμ-*TCL1* mice demonstrated a peak of mortality at around 1 year of age, with the majority of terminal tumours presenting between 10 and 14 months of age. The median age of terminal tumour presentation was 352 days, with an SD of 54 days.

As demonstrated by Bichi et al [144], Eμ-*TCL1* mice present with an accumulation of $CD5^+$ B cells in the peripheral blood. These were detected by flow cytometry of blood staining for CD5 and B220. To demonstrate the typical profiles of the $CD5^+$ $B220^+$ tumour cell population, blood from Eμ-*TCL1* and WT mice at 3 months and Eμ-*TCL1* mice at 10 months was analysed, shown in **Figure 5.2**. At 3 months, the $CD5^+$ and $B220^+$ stained lymphocyte population of Eμ-*TCL1* mice demonstrated no difference to WT lymphocytes, in contrast with a clear tumour population at 10 months.

Figure 5.2A illustrates the presentation of a raised leukaemic population of $CD5^+$ $B220^+$ cells, intermediate in expression of both CD5 and B220 relative to T and B cells, respectively. **Figure 5.2B** plots the concentration of lymphocytes in the blood against the $CD5^+$ $B220^+$ for the 14 terminal Eμ-*TCL1* mice described in **Figure 5.1** relative to 3-month-old WT and Eμ-*TCL1* mice. The blood concentration of leukaemia cells did not consistently differentiate healthy from terminal mice, whereas $CD5^+$ $B220^+$ cell population percentage gave the clearest separation of the two terminal and non-terminal states. An approximately linear correlation was observed above that of a $CD5^+$ $B220^+$ cell population 60%. **Figure 5.2B** also demonstrates that at 3

months of age, Eμ-*TCL1* mice present with little difference to WT, for either the blood lymphocyte concentration or the percentage of $CD5^+$ $B220^+$ cells.

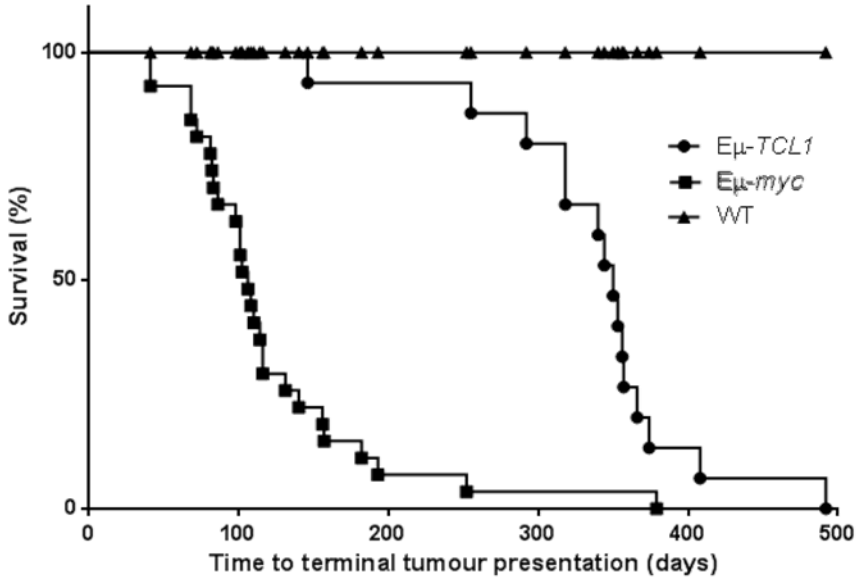


Figure 5.1. Eμ-*TCL1* and Eμ-*myc* terminal tumour presentation. Kaplan-Meier survival curve demonstrating terminal tumour presentation of 27 Eμ-*myc* mice and 14 Eμ-*TCL1* mice compared to WT littermates. All three genotypes demonstrate significantly different survival times from one another, $P<0.001$.

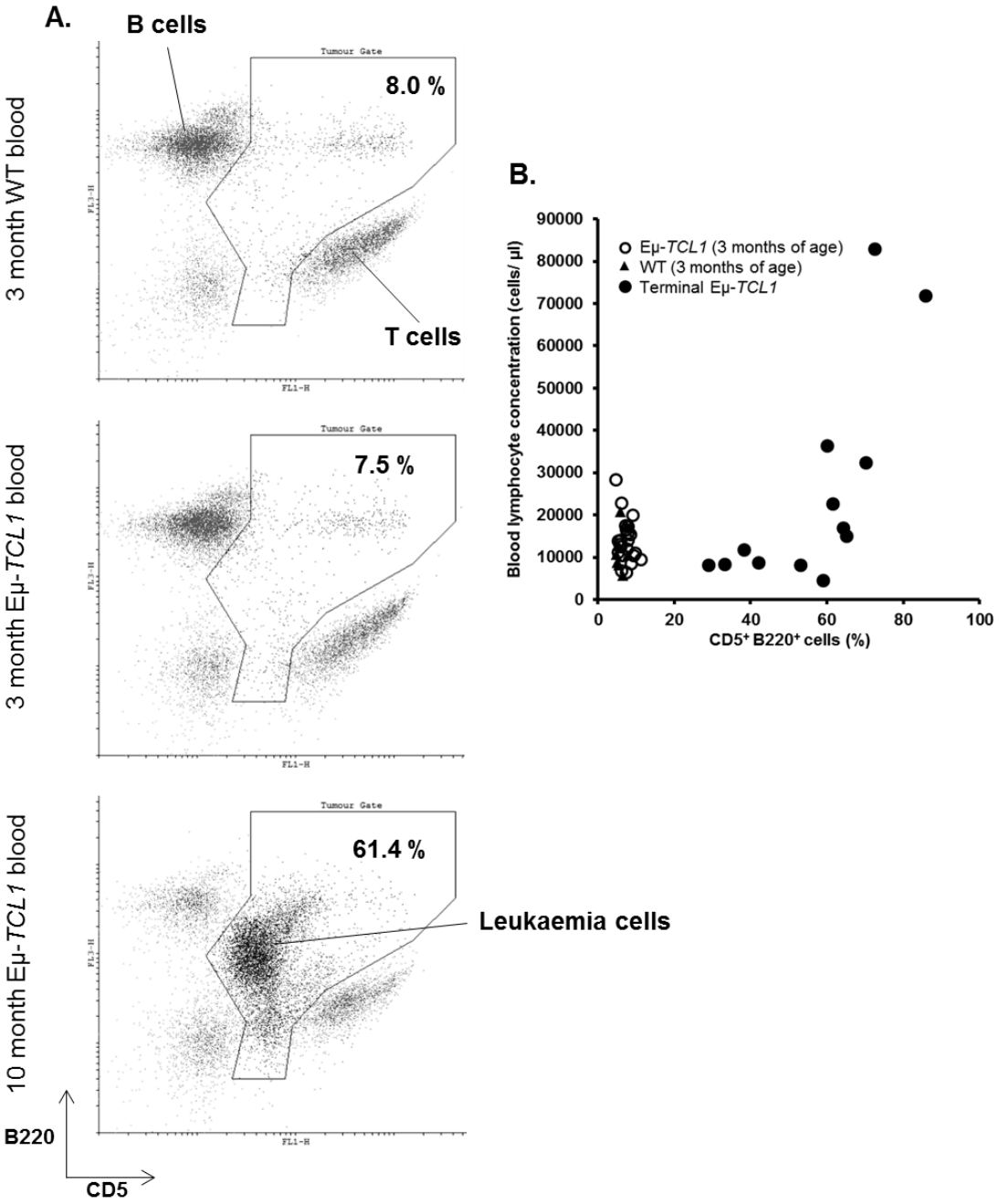


Figure 5.2. CD5$^+$ B220$^+$ leukaemic populations appearing in terminal Eμ-*TCL1* mice. A. PBMCs stained for CD5 and B220, where the gated population describes the common bounds of the leukaemic population of CD5$^+$B220$^+$ cells. For each condition, PBMCs from 4 mice were pooled. **B.** Blood from 22 Eμ-*TCL1* mice and 8 WT mice of 3 months of age and 14 terminal-stage Eμ-*TCL1* mice was analysed to determine the lymphocyte concentrations and CD5$^+$B220$^+$ cell percentage. Data from monthly screenings of the lymphocyte concentrations and CD5$^+$B220$^+$ cell percentage was plotted for terminal mice, with measurements taken no more than a month prior to terminal presentation.

5.3 Eμ-*MYC* AND Eμ-*TCL1* B-CELL CANCER QUANTITATIVE PROTEOMICS DESIGN AND WORKFLOW

Having characterised the time course of tumour development in the Eμ-*myc* and Eμ-*TCL1* models, an appropriate proteomics experiment was designed that could be accommodated within an iTRAQ 8-plex experiment. Consideration was given to understanding the protein expression of B-cell tumours relative to WT B cells as well as pre-tumour B cells from both models.

B cells were derived from the spleen as performed for the characterisation described in **Chapter 3**, and consistently provided enough material for proteomics analysis. MACS was used as detailed in **Section 2.8.1**, which again gave a suitable purity of B cells for proteomics. All samples were derived from female mice to avoid inconsistencies that may have arisen from gender dimorphism, as well as due to a higher availability because males were typically kept for breeding purposes. To reduce variability that might exist between individual samples while accommodating the experiment within a single iTRAQ 8-plex, multiple samples were pooled.

For the non-tumour control samples, due to low cell numbers and to minimise handling time, splenocytes were pooled in batches of three upon collection and processed as such. Given the previous observations of an increased number of differentially expressed proteins in tumours (**Chapter 3**), the Eμ-*myc* and Eμ-*TCL1* tumour samples were handled individually and characterised as pools of 2. Reproducibility of the tumour pools was assessed by the incorporation of a biological replicate for each tumour. The characteristics of these samples, including splenocyte cellularity and B-cell yield is detailed in **Figure 5.3** with additional details including the purity and viability of B-cell samples in **Appendix A7**.

To define differential protein expression from normal precursors, splenic B cells were isolated from WT littermates at 6 weeks of age to act as a control against which iTRAQ ratios would be defined. An aged WT B-cell control was from WT mice at 200 days of age also assessed; the approximate average age of tumour presentation across the two models. B cells were isolated from Eμ-*myc* and Eμ-*TCL1* mice at 6 weeks of age as a means of characterising the pre-tumour proteomes in these models (**Figure 5.3**).

Tumour and non-tumour B cells were isolated, washed and snap frozen as cell pellets using the workflows summarised in **Figure 5.3** and described in **Section 2.10** and **Section 2.11.2**. Once all samples had been procured, cell pellets were defrosted and lysed using a combination of trituration and sonication to solubilise proteins as thoroughly as possible.

For each pair of control sample replicates, each consisting of material derived from 3 mice, 50 ug of cell lysate was pooled. For the tumour samples, 50 ug of cell lysate from two

tumours was pooled. Accordingly, the 8 subsequent lysate pools were then subjected to reduction, alkylation and digestion by trypsin, as described in **Section 2.19.2**. The resulting peptides were allocated to iTRAQ 8-plex labels as described in **Figure 5.3**. The labelled peptides were then pooled, lyophilised and subjected to high-pH, C8, reverse phase chromatography to pre-fractionate the peptides as described in **Section 2.19.4**. Each of the subsequent 69 peak-dependent fractions were then further resolved by low-pH, C18, reverse phase chromatography and analysed by in-line electrospray ionisation and MS characterisation and MS/MS characterisation by DDA, as described in **Section 2.19.5**.

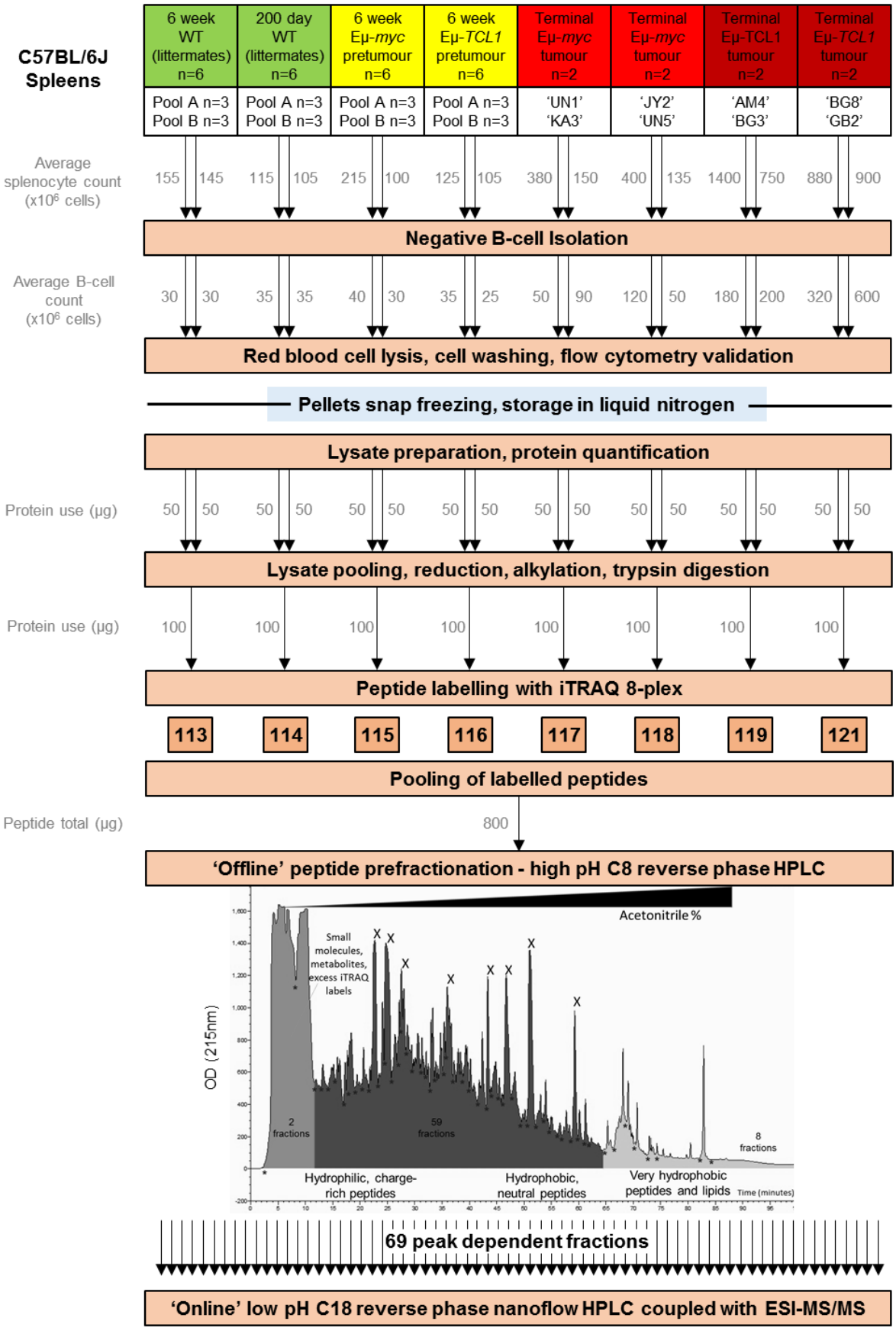

C57BL/6J Spleens
6 week WT (littermates) n=6
200 day WT (littermates) n=6
6 week Eμ-myc pretumour n=6
6 week Eμ-TCL1 pretumour n=6
Terminal Eμ-myc tumour n=2
Terminal Eμ-myc tumour n=2
Terminal Eμ-TCL1 tumour n=2
Terminal Eμ-TCL1 tumour n=2
Pool A n=3 Pool B n=3
Pool A n=3 Pool B n=3
Pool A n=3 Pool B n=3
Pool A n=3 Pool B n=3
'UN1' 'KA3'
'JY2' 'UN5'
'AM4' 'BG3'
'BG8' 'GB2'
Average splenocyte count (x10⁶ cells)
155 145 115 105 215 100 125 105 380 150 400 135 1400 750 880 900
Negative B-cell Isolation
Average B-cell count (x10⁶ cells)
30 30 35 35 40 30 35 25 50 90 120 50 180 200 320 600
Red blood cell lysis, cell washing, flow cytometry validation
Pellets snap freezing, storage in liquid nitrogen
Lysate preparation, protein quantification
Protein use (μg)
50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50
Lysate pooling, reduction, alkylation, trypsin digestion
Protein use (μg)
100 100 100 100 100 100 100 100
Peptide labelling with iTRAQ 8-plex
113 114 115 116 117 118 119 121
Pooling of labelled peptides
Peptide total (μg)
800
'Offline' peptide prefractionation - high pH C8 reverse phase HPLC
Acetonitrile %
Small molecules, metabolites, excess iTRAQ labels
OD (215nm)
2 fractions
59 fractions
8 fractions
Hydrophilic, charge-rich peptides
Hydrophobic, neutral peptides
Very hydrophobic peptides and lipids
Time (minutes)
69 peak dependent fractions
'Online' low pH C18 reverse phase nanoflow HPLC coupled with ESI-MS/MS

Figure 5.3. The experimental design and workflow for the 8-plex iTRAQ characterisation of the Eμ-*myc* and Eμ-*TCL1* B-cell proteome. This describes, in descending order: The 8 biological conditions or replicates for which samples were collected for MS characterisation including age, genotype and biological state – the 'traffic light' colours represent the severity of tumour risk/progression. Splenocytes were subjected to negative B-cell isolation using an antibody cocktail to label and deplete non-B cells. The isolated B cells were washed and validated for purity before snap freezing. After storage in liquid nitrogen, B-cell lysates were prepared by trituration and sonication of the B-cell pellets. Lysates were pooled to their respective conditions to give 8 100 μg pools. B-cell lysate pools were reduced, alkylated and proteolysed with trypsin. Peptides were labelled by iTRAQ 8-plex, pooled and subjected to prefractionation chromatography to reduce sample complexity generating 69 fractions. Each peptide fraction was subject to a second round of chromatography, in line with electrospray ionisation, MS detection and data-dependent MS/MS characterisation.

5.4 Eμ-*MYC* and Eμ-*TCL1* Protein Identification and Relative Quantification

Spectra produced by the MS analysis were analysed using target decoy searching as described in **Section 2.19.6** and **Figure 5.4**. CID and HCD spectra were subjected to separate searches due to the differences in detection methods. Spectra were subjected to iterative searches firstly identifying peptides based on highly stringent criteria followed by a subsequent relaxed criteria search for all unmatched spectra. Peptide identifications were accepted with an FDR of <1% and matched to 9260 protein groups. Data was then submitted to the SPIQuE tool, described in **Chapter 4**, quantifying a total of 7391 proteins.

5.5 Cluster Analysis Confirms Reproducible Protein Quantitations and Dysregulation of Eμ-*MYC* and Eμ-*TCL1* Tumours

The final outputs from the SPIQuE tool were clustered using Cluster 3.0 and Euclidian distance, showing a high degree of dysregulation as well as reproducibility between the tumour replicates for both B-cell cancer models (**Figure 5.4**). The Eμ-*myc* tumours demonstrated the greatest degree of protein dysregulation relative to the WT B-cell controls. The Eμ-*myc* 6 week, pre-malignant sample pool presented a highly similar pattern of dysregulation to that of the malignant stage of the disease and clustered strongly with the Eμ-*myc* tumours. The Eμ-*TCL1* tumours displayed a lesser, but still substantial and reproducible pattern of protein dysregulation. Clustering revealed similarities and differences to the Eμ-*myc* model, indicative of a common B-cell tumour signature and oncogene-specific trends, respectively. Both the 6-week Eμ-*TCL1* B cells and the 200 day WT B cells presented with minimal dysregulation relative to 6-week WT B cells with 250 (3%) and 90 (1%) of the proteins dysregulated more than 2-fold, respectively.

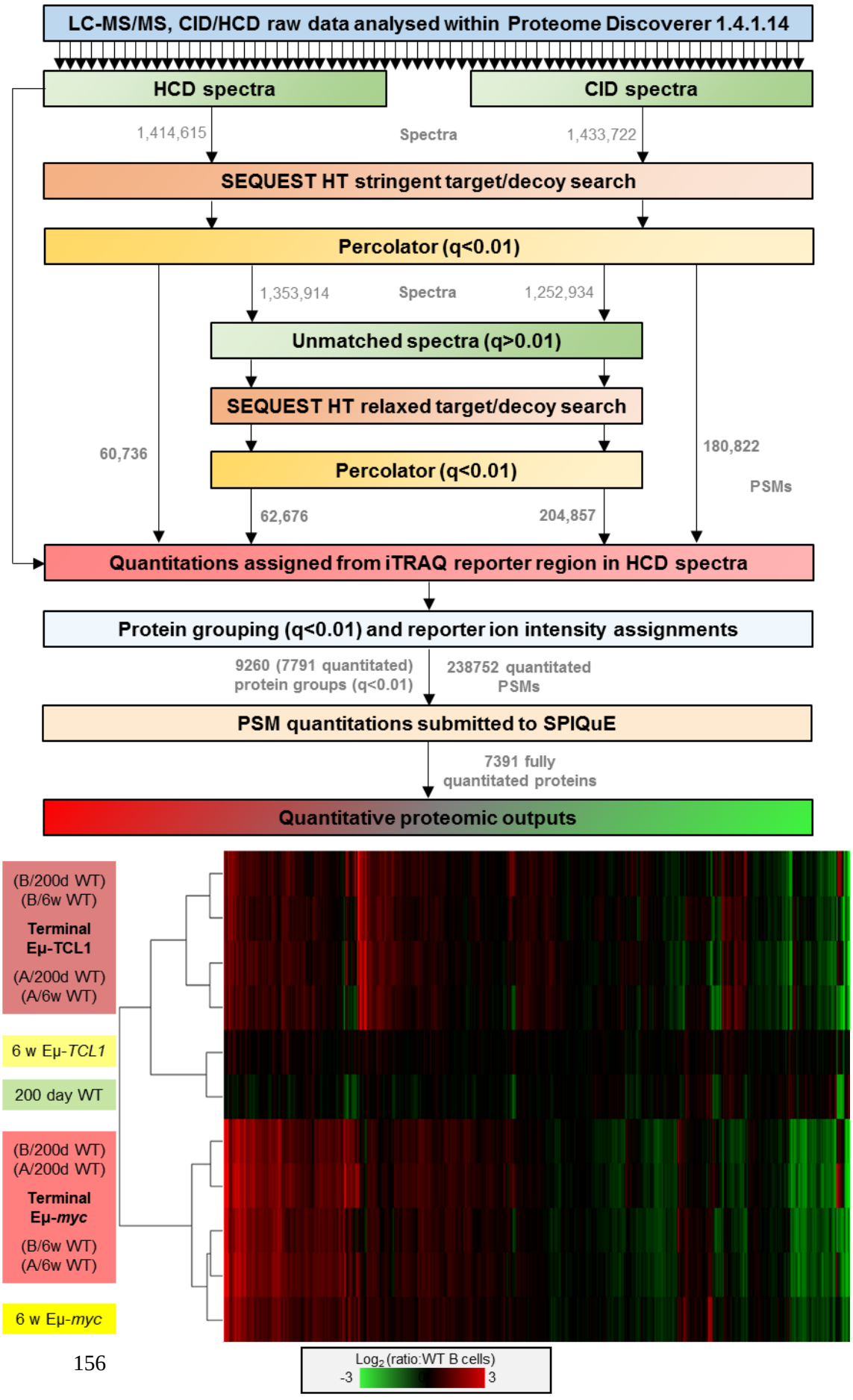

LC-MS/MS, CID/HCD raw data analysed within Proteome Discoverer 1.4.1.14
HCD spectra
CID spectra
1,414,615
Spectra
1,433,722
SEQUEST HT stringent target/decoy search
Percolator (q<0.01)
1,353,914
Spectra
1,252,934
Unmatched spectra (q>0.01)
SEQUEST HT relaxed target/decoy search
60,736
Percolator (q<0.01)
180,822
PSMs
62,676
204,857
Quantitations assigned from iTRAQ reporter region in HCD spectra
Protein grouping (q<0.01) and reporter ion intensity assignments
9260 (7791 quantitated) protein groups (q<0.01)
238752 quantitated PSMs
PSM quantitations submitted to SPIQuE
7391 fully quantitated proteins
Quantitative proteomic outputs
(B/200d WT)
(B/6w WT)
Terminal Eμ-TCL1
(A/200d WT)
(A/6w WT)
6 w Eμ-TCL1
200 day WT
(B/200d WT)
(A/200d WT)
Terminal Eμ-myc
(B/6w WT)
(A/6w WT)
6 w Eμ-myc
Log2 (ratio:WT B cells)
-3
0
3

Figure 5.4. Identification and quantification of differential protein expression in the Eμ-*myc* and Eμ-*TCL1* B-cell cancer models. Fractions were processed using two stage target decoy searches within Proteome Discoverer. Spectra were searched separately for CID and HCD fragmentation, and initially searched for peptides spectrum matches using stringent settings; allowing for no missed cleavages and no dynamic modifications. Spectra failing to match to a peptide with a percolator-determined false discovery rate of $q<0.01$ were subjected to a second, relaxed target decoy search allowing for missed cleavages and dynamic modifications. The PSM quantitations and features with the protein group assignments were exported for analysis with SPIQuE. Quantitative proteomic results were clustered using Cluster 3.0 and to represent the topological patterns in the differentially expressed data for each ratio.

5.6 IDENTIFICATIONS OF ANTICIPATED PROTEIN REGULATIONS

To initially confirm the quantitative accuracy of the 2D-LC MS/MS quantitative workflow and the subsequent data analysis with SPIQuE, a list of anticipated protein identifications and their respective quantitations was generated (**Figure 5.5A**). This described a selection of proteins with links to B cells and B-cell cancers and proteins with previous validation summarised in **Appendices A4** and **A8**. The expression of myc was confirmed to be greatly increased in all samples derived from the Eμ-*myc* model. An example of such a match is shown in **Figure 5.5B**. Myc was also observed to be upregulated in the Eμ-*TCL1* model, only explainable from murine myc protein expression. Only peptides specific to the human transgenic TCL1A protein were identified (**Figure 5.5A, D**) which consistently exhibited an Eμ-*TCL1* model-specific expression in both the pre-tumour and tumour samples (**Figure 5.5E)**.

Further to the expression of the model-specific transgene proteins, other anticipated proteins were identified and quantified. For example, CD5 was overexpressed specifically in Eμ-*TCL1* tumours, while B220 was observed downregulated, confirming the characteristic observations of the increased and decreased expression levels, respectively, previously observed by flow cytometry (**Figure 5.2**). IgM was overexpressed, specifically in Eμ-*TCL1* tumours, while IgD and CD200 were downregulated in agreement with **Appendix A8**. CD79a and CD79b were observed marginally downregulated in Eμ-*TCL1* tumours, approximately in agreement with the variable expression outlined in **Appendix A8.** In contrast, in Eμ-*myc* tumours the majority of these components were very downregulated. Other B-cell proteins were identified exhibiting tumour-specific downregulation, such as CD38, CD23 and CD22. CD19 expression appeared mostly unchanged in the tumours, while CD20 and CD5 downregulation was specific

to the Eμ-*myc* tumours. Of the proteins validated by WB in Chapter 3, all but MDH2 agreed with the directionality of regulation.

A single peptide matching p53 was identified, indicative of an overexpression in all 4 tumour pools, most substantially in Eμ-*myc* tumours, correlating with the observed expression of myc. GAPDH, a common loading control, was observed without regulation in any of the samples.

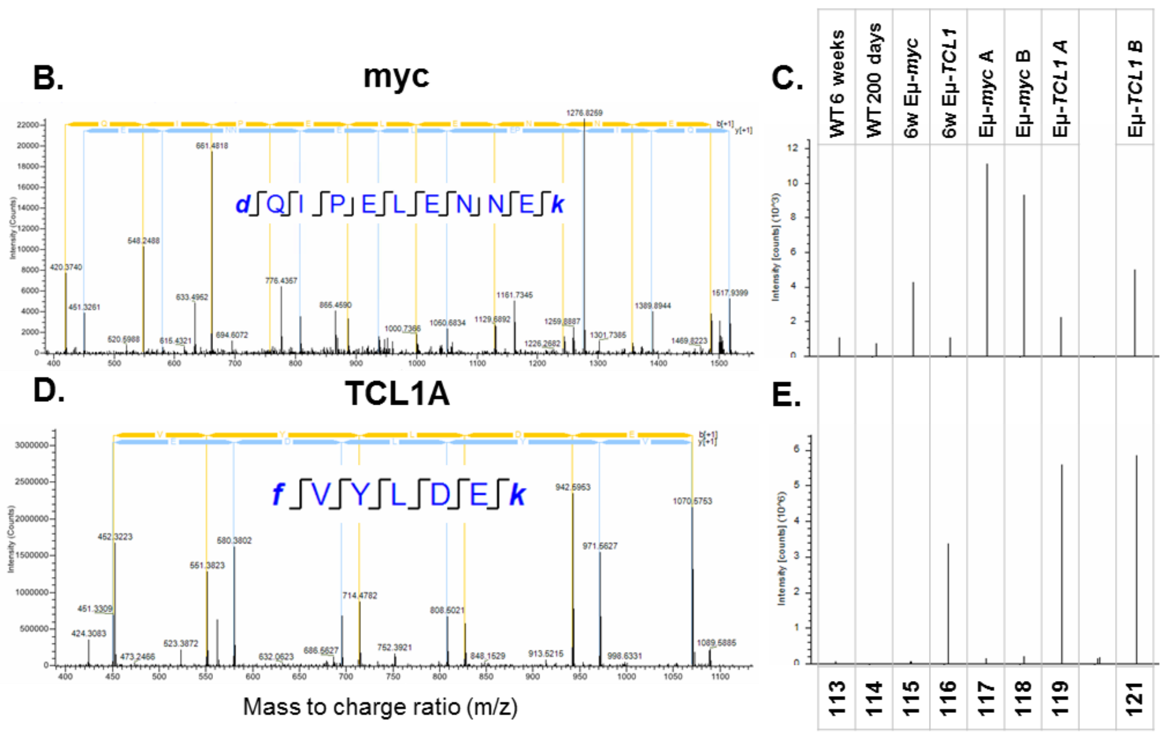
Together, these findings validated the ability of the 2D-LC MS/MS approach and subsequent data analysis methods to detect and reliably quantify oncogenic and tissue-specific proteins in B-cell tumours; the first mass spectrometry based observations of such characteristics in the Eμ-*myc* and Eμ-*TCL1* models.

Figure 5.5. Expected protein and peptide identifications and quantifications relating to B-cell tumours. A. Identifications and quantifications describing protein expression across the Eμ-*myc* and Eμ-*TCL1* B cells relative to healthy WT B cells collected at 6 weeks and 200 days of age. Proteins selected included; the driving oncogenes in each model, characteristic proteins immunophenotyped in **Figure 5.2**, previously evaluated proteins detailed in **Appendix A8**, B-cell proteins and the proteins investigated for expression in Eμ-*myc* tumours by WB in **Chapter 3**/**Appendix A4** (HSP90 was excluded due to isoforms falling into multiple protein groups). GAPDH illustrates an example of an unregulated protein. The final two right hand columns detail either anticipated upregulation (↑/red) or downregulation (↓/green) on the basis of the WB/flow cytometry results relative to WT B cells. **B.** A CID peptide spectrum match to the myc-specific peptide (surrogate to both the human transgene and murine myc proteins) DQIPELENNEK (N-terminally and K11 modified with iTRAQ 8-plex) was observed with an m/z of 969.01690 Da (theoretical m/z; 969.02545 Da). **C.** The iTRAQ reporter region of the HCD spectrum corresponding to the precursor fragmented in B. indicating the relative expression across 8 conditions for myc. **D.** A CID peptide spectrum match to the human TCL1A-specific peptide (surrogate to the human transgene TCL1 protein) FVYLDEK (N-terminally and K6 modified with iTRAQ 8-plex) was observed with an m/z of 761.43079 (theoretical m/z 761.44229). **E.** The iTRAQ reporter region of the HCD spectrum corresponding to the precursor fragmented in D. indicating the relative expression across 8 conditions for TCL1.

A.

Log$_2$ (ratio to WT B cells): -1 (green) — 0 (black) — 1 (red)

Description	Symbol	Unique Peptides	PSMs	WT 200d : 6w WT	6w Eμ-*myc* : 6w WT	6w Eμ-*TCL1* : 6w WT	Eμ-*myc* A : 6w WT	Eμ-*myc* A : 200d WT	Eμ-*myc* B : 6w WT	Eμ-*myc* B : 200d WT	Eμ-*TCL1* A : 6w WT	Eμ-*TCL1* A : 200d WT	Eμ-*TCL1* B : 6w WT	Eμ-*TCL1* B : 200d WT	Expected (Eμ-myc)	Expected (Eμ-*TCL1*)
				Log$_2$ (ratios)												
(Human/mouse) Myc proto-oncogene protein	**MYC**	2	3	-0.43	1.80	-0.41	2.98	3.41	2.89	3.32	0.60	1.03	1.78	2.21	↑	↑
(Human) T-cell leukemia/lymphoma protein 1A	**TCL1A**	5	150	0.13	-0.07	2.37	0.09	-0.02	-0.22	-0.40	2.58	2.36	2.58	2.32		↑
T-cell surface glycoprotein CD5	**CD5**	6	18	-0.63	-0.51	0.48	-1.10	-0.46	-0.91	-0.40	0.84	1.61	1.60	2.44		↑
Receptor-type tyr phosphatase C /CD45	**B220**	49	666	0.11	-0.95	0.01	-1.32	-1.43	-1.30	-1.42	-0.70	-0.81	-0.60	-0.71		↓
Ig mu chain C region	**IgM**	23	982	0.60	0.42	0.47	0.51	-0.09	-0.18	-0.80	2.80	2.26	2.25	1.72		↑
BCR complex-associated protein alpha chain	**CD79A**	5	83	0.20	-0.88	-0.12	-1.44	-1.64	-0.77	-1.00	-0.19	-0.39	-0.33	-0.55		
BCR complex-associated protein beta chain	**CD79B**	3	34	0.39	-0.99	-0.12	-1.70	-2.01	-0.65	-1.03	-0.07	-0.51	-0.62	-0.97		-
Ig delta chain C region secreted form	**IgD**	6	32	0.64	-2.04	-0.19	-2.78	-3.42	-2.59	-3.23	-4.09	-4.74	-3.23	-3.88		↓
OX-2 membrane glycoprotein	**CD200**	2	5	0.25	-2.22	-0.58	-2.73	-2.97	-2.40	-2.65	-1.48	-1.73	-1.88	-2.13		↓
Cellular tumor antigen p53	**TP53**	1	3	0.17	1.20	0.41	1.90	1.74	2.29	2.12	1.17	1.01	1.02	0.86		
B-lymphocyte antigen CD19	**CD19**	8	59	0.13	-0.25	0.04	-0.27	-0.40	-0.04	-0.17	0.17	0.04	-0.13	-0.25		
B-lymphocyte antigen CD20/MS4A1	**CD20**	9	99	0.20	-0.83	0.09	-1.48	-1.68	-1.28	-1.47	0.10	-0.09	-0.16	-0.41		
ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1	**CD38**	9	50	0.54	-2.09	0.13	-2.29	-2.81	-2.64	-3.16	-1.04	-1.56	-0.97	-1.50		
Low affinity Ig epsilon Fc receptor FCER2	**CD23**	7	23	0.26	-2.38	-0.50	-2.54	-2.85	-2.78	-3.06	-3.22	-3.52	-3.02	-3.30		
B-cell receptor CD22	**CD22**	10	122	0.40	-1.47	-0.03	-1.46	-1.79	-1.27	-1.58	-0.96	-1.36	-1.05	-1.45		
Alpha-enolase	**ENO1**	26	2667	0.12	0.18	[illegible]	0.81	0.72	0.80	0.69	-0.06	-0.16	0.30	0.22	↑	
Nucleophosmin	**NPM1**	24	813	0.13	0.68	-0.02	1.05	0.95	0.99	0.89	0.47	0.36	0.46	0.35	↑	
Malate dehydrogenase, mitochondrial	**MDH2**	20	862	0.08	0.07	0.06	0.14	0.04	[illegible]	-0.07	0.37	0.27	0.25	0.15	↑	
Tyr-protein phosphatase non-receptor 6	**SHP1**	35	1028	0.06	-0.75	-0.09	-1.25	-1.31	-0.91	-0.97	0.19	0.12	0.22	0.16	↓	
Coronin-1A	**COR1A**	25	760	0.13	-0.48	[illegible]	-0.72	-0.86	-0.55	-0.69	-0.03	-0.16	-0.11	-0.25	↓	
Capping protein (Actin filament), gelsolin-like	**CAPG**	12	288	0.42	-0.80	0.07	-1.10	-1.52	-0.78	-1.21	0.79	0.37	0.86	0.44	↓	
Plastin-2	**LCP1**	41	4374	0.18	-0.53	[illegible]	-0.82	-1.00	-0.68	-0.86	0.07	-0.10	0.02	-0.20	↓	
Myosin-9	**MYO9**	127	3800	-0.07	-0.22	[illegible]	-0.35	-0.27	-0.46	-0.39	0.26	0.33	0.06	0.14	↓	
Moesin	**MOE**	39	1751	[illegible]	-0.37	-0.14	-0.70	-0.68	-0.37	-0.36	-0.24	-0.24			↓	
Glyceraldehyde-3-phosphate dehydrogenase	***GAPDH***	13	1961	-0.09	-0.27		-0.15	-0.07	-0.18	-0.11		0.09	-0.05	[illegible]	-	-



5.7 CONFIDENT, CONSISTENT PROTEIN UPREGULATION IN TUMOURS OF THE Eμ-*MYC* AND Eμ-*TCL1* MODELS

Given that findings were made describing both the expected proteins and quantitations, the 7391 quantitated protein were evaluated to identify further biologically relevant protein dysregulation. To identify the most significant, consistently upregulated proteins in the two tumour models, as well as in both models, protein expression was evaluated for the four ratios; tumour A: WT 6 weeks, tumour B: WT 6 weeks, tumour A: WT 200 days, tumour B: WT 200 days. While such ratios contained technical replicates, they still provided the opportunity to identify the most consistently regulated tumour proteins relative to both WT samples. For these ratios, two scores were calculated; a $-\log_{10}$ (FDR-corrected p-value) (**Section 2.19.8**) and a measurement indicative of the magnitude and consistency of regulation termed the 'regulation score' (RS) (**Section 2.19.7**), determined by the formula: average / (SD + 1). Proteins were determined as significantly regulated in both tumours by the separate calculation of p-values and RS for all 8 $\log_2$ (ratios). In all cases, a p-value of <0.05 ($\log_{10}$ (0.05) → >1.301) was considered significant and an RS of >0.5 or <-0.5 was considered regulated.

The $-\log_{10}$ (p-value) and RS were plotted against one another in a manner described as volcano plots due to the characteristic shape. The complementarity of these two values allowed for consideration of the significance and magnitude of the differential regulation being described by iTRAQ to be simultaneously taken into consideration. The upper right and upper left regions of these volcano plots are indicative of those proteins with the most significant observations of differential expression.

Figure 5.6 represents the volcano plots of the 7391 quantitated proteins for the Eμ-*myc* tumours, the Eμ-*TCL1* tumours and for both tumours. Overall, Eμ-*myc* tumours presented the greatest degree of dysregulation with 1840 proteins significantly upregulated (p<0.05, RS>0.5) and 824 proteins significantly downregulated (p<0.05, RS<-0.5), totalling 36% of all quantitated proteins demonstrating dysregulation. The Eμ-*TCL1* tumours presented less dysregulation, with 324 down- and 1142 up-regulated proteins (20% of the quantified proteome). To derive p-values and regulation scores for each protein for both tumour types, the same equations were applied to all 8 ratios. 21% of the proteome was significantly dysregulated in both tumours with 1219 up- and 369 down-regulated proteins.

Those proteins with both the smallest p-values and the largest regulation scores in the upper right hand corner of each volcano plot (**Figure 5.6**) were plotted and annotated to give a detailed representation of the most upregulated proteins. For the Eμ-*myc* tumours, myc itself was amongst these top proteins including multiple KIFs (kinesin motor proteins) and CDCs (cell

division cycle proteins). The most upregulated proteins were ubiquitin-conjugating enzyme E2 C (UBE2C), kinesin-like protein KIF23, cyclin-dependent kinase 1 (CDK1), targeting protein for Xklp2 (TPX2) and transcription factor E2F8. Eμ-*TCL1* tumours were also determined to have their driving oncoprotein TCL1A substantially overexpressed. The mitochondrial protein enoyl-CoA hydratase domain-containing protein 3 (ECHDC3) was observed to be the most significantly and substantially overexpressed in Eμ-*TCL1* tumours alongside; neuroblast differentiation-associated protein 2 (AHNAK2), a titin-like protein, nicotinamide nucleotide transhydrogenase (NNT), ornithine carbamoyltransferase (OTC), pleckstrin homology domain-containing family A member 7 (PLEKHA7) and distinct subgroup of the Ras family member 2 (DIRAS2). Finally, for both tumours, several consistently upregulated proteins, including many of those highlighted for Eμ-*myc* and Eμ-*TCL1* tumours alone, were plotted, identifying the urea cycle enzymes argininosuccinate synthase 1 (ASS1) and OTC as the most prominent. **Figure 5.7** contains details relating to a selection of these proteins.

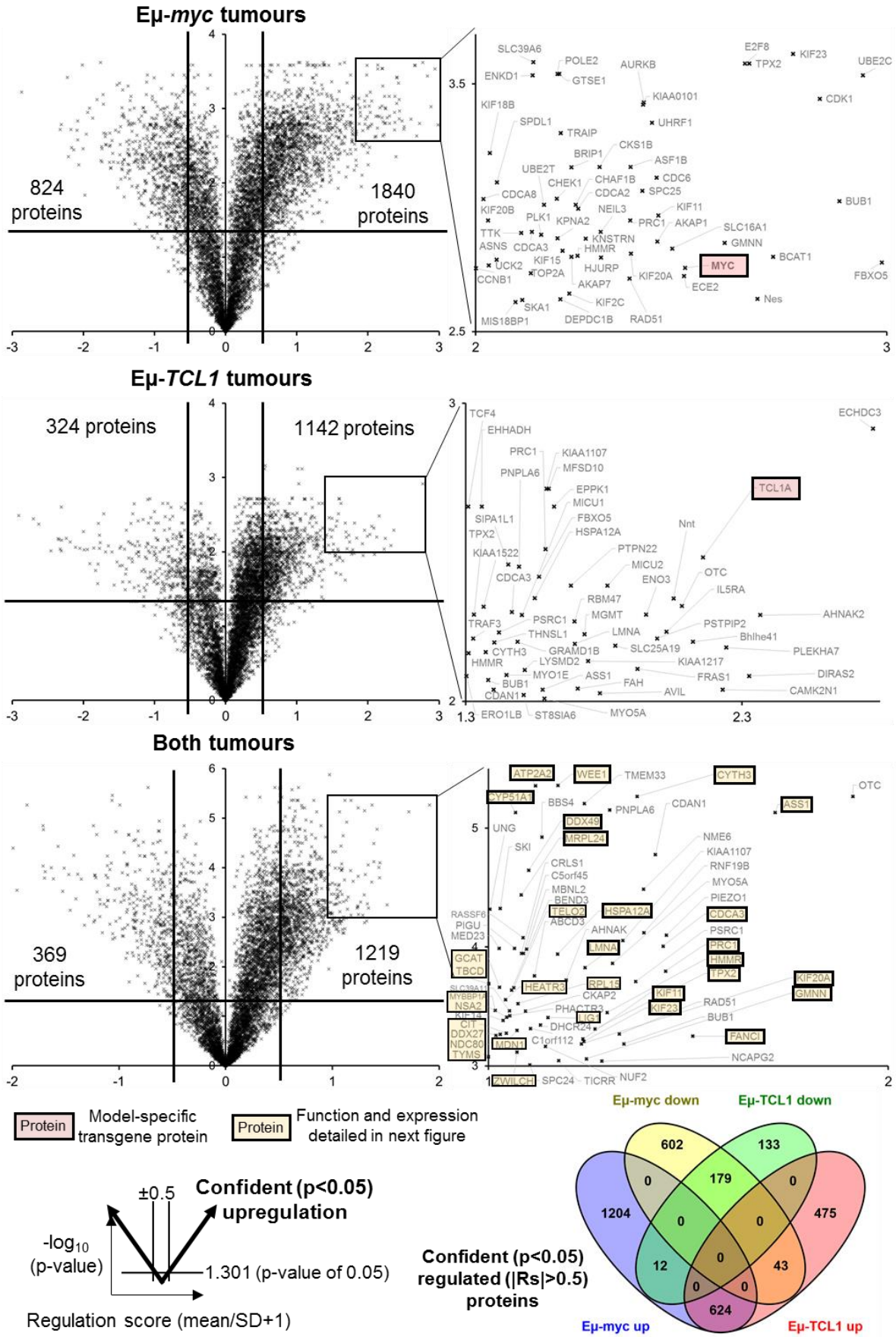
The top 50 most confidently upregulated proteins in both tumours, highlighted in yellow in **Figure 5.6**, were filtered on the basis of the number of unique peptides, PSMs and ratio consistency and were systematically evaluated for their function and relevance to cancers (**Figure 5.7**). Due to the majority of these tumour protein ratios meeting or exceeding a $\log_2$ (ratio) of 1, the threshold of colour saturation used in **Figure 5.5**, the colouring was set to be saturated at a $\log_2$ (ratio to WT B cells) of 2, the equivalent of a four-fold change.

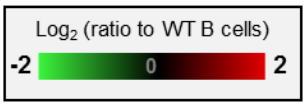
These 50 proteins predominantly demonstrated upregulation in the Eμ-*myc* tumours with, to a lesser extent, upregulation in Eμ-*TCL1* tumours. The exceptions to this included; cytohesis-3 (CYTH3), prelaminin-A/C (LMNA), wee1-like protein kinase (WEE1), heat shock 70 kDa protein 12A (HSPA12A), lanosterol 14-alpha demethylase (CYP51A1) and mitochondrial protein 2-amino-3-ketobutyrate A ligase (GCAT); for which Eμ-*TCL1* tumours protein upregulation was either equivalent to, or greater than, that of Eμ-*myc* tumours. Multiple proteins involved in processes such as cell cycle progression, DNA damage response and repair, amino acid metabolism and cytokinesis were observed.

The majority of the 50 proteins upregulated in both tumours were previously reported with other observations of upregulations or functional roles in cancers. Brief reviews were carried out for each protein searching for any of the protein names given on Uniprot in conjunction with the word ‘cancer’. While not fully comprehensive, this search gave an indication that approximately a fifth of these proteins were novel observations in cancer, whilst the remaining four fifths were proteins that could be linked to malignancies. For at least half of these, the proteins were linked to multiple cancer types. Of those without a clear published link to malignancy, zwilch homologue (ZWILCH), telomere regulation TEL2 homolog (TELO2) and

protein spindly (SPDL1) all had clear roles in cell cycle progression, 39S mitochondrial ribosomal protein L24 (MRPL24), Midasin (MDN1), GCAT and 60S ribosomal protein L13a (RPL13A) had roles in protein synthesis, leaving just TBCD and the uncharacterised protein HEATR3 without a published or functional link to cancer.

Figure 5.6. Significant and consistent protein overexpression in Eμ-*myc* and Eμ-*TCL1* B-cell cancers. Volcano plots demonstrating proteins with both a strong statistical significance and magnitude of regulation, determined by the regulation score (mean/SD+1) of the $\log_2$ (ratios) generated for each tumour type. The most highly significant and upregulated proteins are demonstrated for each comparison on the right. A Venn diagram comparing the overlap of the significantly up and downregulated proteins is shown. Proteins expected within each model due to the Eμ-driven transgenes are highlighted red. Those proteins identified as significantly upregulated in both tumours, with the highest quality of data, are highlighted yellow and detailed in the next figure.


Eμ-myc tumours
824 proteins
1840 proteins
Eμ-TCL1 tumours
324 proteins
1142 proteins
Both tumours
369 proteins
1219 proteins
Protein
Model-specific transgene protein
Protein
Function and expression detailed in next figure
±0.5
Confident (p<0.05) upregulation
-log10 (p-value)
1.301 (p-value of 0.05)
Regulation score (mean/SD+1)
Confident (p<0.05) regulated (|Rs|>0.5) proteins
Eμ-myc down
Eμ-TCL1 down
Eμ-myc up
Eμ-TCL1 up
602
133
179
1204
475
12
43
624



Description	Symbol	Unique Peptides	PSMs	WT 200d : 6w WT	6w Eμ-*myc* : 6w WT	6w Eμ-*TCL1* : 6w WT	Eμ-*myc* A : 6w WT	Eμ-*myc* A : 200d WT	Eμ-*myc* B : 6w WT	Eμ-*myc* B : 200d WT	Eμ-*TCL1* A : 6w WT	Eμ-*TCL1* A : 200d WT	Eμ-*TCL1* B : 6w WT	Eμ-*TCL1* B : 200d WT	Regulation score of both tumours	-log10(FDR p-value)	Function	Observations in cancers
Argininosuccinate synthase	ASS1	7	374												1.72	5.13	Urea cycle enzyme and arginine synthesis	Regulated in many tumours (up and down), possible correlation with cisplatin resistance/poor prognosis
Protein regulator of cytokinesis 1	PRC1	14	52												1.44	3.80	Microtubule crosslinker facilitating cytokinesis	Upregulated in breast and bladder cancers
Cell division cycle-associated protein 3	CDCA3	3	9												1.44	4.01	Mediates degradation of CDK inhibitors	Upregulated in oral and prostate cancer
Cytohesin-3	CYTH3	3	25												1.37	5.27	Integrin recycling and cell mobility, promotes reduced adhesion	Upregulated and increases proliferation in liver cancer
Hyaluronan mediated motility receptor	HMMR	17	57												1.30	3.45	(CD168) Cell motility, transformation and ERK regulation.	Prognostic in various cancers, previoulsly considered target in CLL and AML
Targeting protein for Xklp2	TPX2	28	88												1.27	3.31	Role in DNA damage response, chromosome stability and cytokinesis	Potential theraputic target in multiple cancers
Geminin	GMNN	6	15												1.25	3.06	Component of DNA replication licencing	Possible predictive biomarker in breast cancer, stem-like phenotype association
Prelamin-A/C	LMNA	38	511												1.24	3.83	Nuclear envelope component, role in chromatin organization	Previously observed upregulated in mesothelioma and breast cancer
Kinesin-like protein KIF23	KIF23	15	51												1.24	3.21	Component of the cytokinesis contractile ring	Increases glioma proliferation.
Kinesin-like protein KIF11	KIF11	33	167												1.24	3.23	Bipolar spindle formation, essential for centrosome migration	Clinical target by inhibitors in numerous cancers
Kinesin-like protein KIF20A	KIF20A	15	42												1.23	3.18	Required for cytokinesis	Target in pancreatic cancer
Wee1-like protein kinase	WEE1	7	14												1.18	5.36	Tyr kinase of CDK1, prevents premature entry into cytokinesis	Evaluated theraputic target in numerous cancers
Heat shock 70 kDa protein 12A	HSPA12A	8	29												1.17	3.94	Minimal characterisation, contains an atypical Hsp70 ATPase domain	Upregulated in liver cancer
DNA ligase	LIG1	8	199												1.15	3.41	Repairs nicks in dsDNA	Mutated in lung and other cancer
Protein CIP2A	CIP2A	8	36												1.13	2.99	Inhibits tumour supressor PP2A	Involvement in numerous cancers, expression linked with myc
Endoplasmic reticulum Ca2+ ATPase 2	ATP2A2	27	180												1.12	5.36	Transports Ca2+ into the ER	Involvement in numerous cancers
Ubiquitin-conjugating enzyme E2 C	UBE2C	7	48												1.12	2.77	Controls mitosis progression via the anaphase promoting complex	Potential biomarker, causes chromosome missegregation
Kinetochore protein Spc25	SPC25	5	12												1.10	2.95	Required for spindle checkpoints and chromosome segregation	Observed overexpressed in GI cancers
RNA helicase DDX49	DDX49	7	21												1.10	4.64	No published characterisations	Observed downregulated in irradiated healthy lymphocytes
Cyclin-dependent kinase 1	CDK1	13	184												1.10	2.76	Key cell cycle regulator	dysregulated in numerous tumours
39S ribosomal protein L24, mt	MRPL24	4	45												1.08	4.44	Mitochondrial ribosome protein, no published characterisations	**None reported**
60S ribosomal protein L15	RPL15	7	76												1.07	3.53	Ribosomal protein	Dysregulated in numerous cancers
Protein zwilch homolog	ZWILCH	9	28												1.07	3.11	Key for cell division (mitotic checkpoint), kinetochore component	**None reported**
Midasin	MDN1	66	237												1.07	3.30	Role in ribosome biogenesis	**None reported**
Lanosterol 14-alpha demethylase	CYP51A1	10	29												1.07	5.13	Steroid demethylation	Inhibition induced differentiation in neutrophil-derived leukaemic cells
E3 ubiquitin-protein ligase UHRF1	UHRF1	27	162												1.07	2.91	Involvement in DNA methylation and repair, chromatin modification	Upregulated in numerous cancers.
Telomere regulation TEL2 homolog	TELO2	4	16												1.06	3.66	Promotes cell cycle rate and lengthens telomeres	**None reported**
G2/mitotic-specific cyclin-B2	CCNB2	6	15												1.06	2.92	Regulates mitotic progression	Correlates with poor prognosis in lung and breast cancer
Ribosome biogenesis NSA2 homolog	NSA2	2	12												1.05	3.41	Ribosome biogenesis, regulates G1/S transition, promotes proliferation	Promotes proliferation and upregulated; in numerous cell lines
Citron Rho-interacting kinase	CIT	21	84												1.05	3.27	Role in cytokinesis	Overexpressed in ovarian carcinomas
HEAT repeat-containing protein 3	HEATR3	12	38												1.04	3.56	No annotated funtion	**None reported**
Chromatin assembly factor 1 subunit B	CHAF1B	12	45												1.04	2.89	Roles in double strand break DNA repair and chromasome assembly	associated with poor prostate cancer prognosis
Kinesin-like protein KIF15	KIF15	34	134												1.04	2.83	Kinesin-like motor enzyme. Interacts with TPX2. Intergrin trafficking.	Confers drug resistance to KIF11 inhibitors
RNA helicase DDX27	DDX27	23	96												1.03	3.26	A KO lethal (99%) RNA helicase, implicated in rRNA processing	Appears to reduce chemotherapy responses of DNA damage and apoptosis
Fanconi anemia group J homolog	BRIP1	4	18												1.03	2.92	DNA helicase and double strand break repair (with BRCA1)	Suggested role in breast cancer
2-amino-3-ketobutyrate A ligase, mt	GCAT	7	25												1.03	3.66	Metabolises gly, ser and thr in the mitochondria	**None reported**
Cell division cycle-associated protein 2	CDCA2	9	23												1.02	2.83	Essential for cell cycle progression, inhibits DNA damage response	Overexpressed in squamous cell carcinoma and other cancers
Kinetochore protein NDC80 homolog	NDC80	19	55												1.02	3.26	Spindle checkpoint signalling	Targeted in liver and breast cancer
Myb-binding protein 1A	MYBBP1A	44	515												1.02	3.47	Transcriptional regualator, circadian regulation	Tumour supressor in breast cancers
Tubulin-specific chaperone D	TBCD	14	61												1.00	3.69	Regulates tubulin (de)stablisation	**None reported**
Thymidylate synthase	TYMS	6	14												1.00	3.08	Thymine synthesis (dUMP->dTMP) maintains U:T ratio preventing DNA damage	Theraputic target in multiple cancers
Disks large-associated protein 5	DLGAP5	8	31												1.00	2.94	Regulates cell cycle progression via the kinetochore	Identifed as HURP (Hepatoma up-regulated protein), with other diagnostic and prognostic implications
60S ribosomal protein L13a	RPL13A	6	83												0.99	3.47	Translational inhibition in IFNγ-induced inflammation processes	**None reported**
Kinesin-like protein KIF2C	KIF2C	16	74												0.99	2.54	Microtubule-depolymerizing kinesin facilitating chromasome segregation	Involved in various tumours
BrCa type 1 susceptibility homolog	BRCA1	9	27												0.99	3.06	dsDNA break repair enzyme	Mutated in some breast and ovarian cancers among others
Exportin-T	XPOT	7	28												0.98	3.10	Transports tRNA-aa out of the nucleus	Dysregulated in mesothelioma
Importin subunit alpha-1	KPNA2	19	153												0.98	2.68	Nuclear pore complex protein import	Observed over-expressed in variouscancers, correlating with proliferation
Protein Spindly	SPDL1	10	28												0.97	2.88	Mitotic kinetochore regulation	**None reported**
Cyclin-dependent kinase 4	CDK4	8	90												0.96	3.10	Ser/Thr-kinase regulating G1-S phase progression via Rb/E2F	Inhibition target
MTHF dehydrogenase, mt	MTHFD2	10	73												0.95	3.18	Mitichondrial folate metabolism enzyme	Linked with poor prognosis

Figure 5.7. Consistently overexpressed proteins in both Eμ-*myc* and Eμ-*TCL1* B-cell tumours. The top 50 Proteins determined with significant upregulation in both tumours with at least 12 PSMs or 3 unique peptides, less than ±0.5 $\log_2$ (ratio) change between 6-week and 200 day WT controls, in descending order of regulation score. To emphasise the differences in differential expression, a colour saturation threshold of a $\log_2$ (ratio) of 2 (fold change of 4) was used.

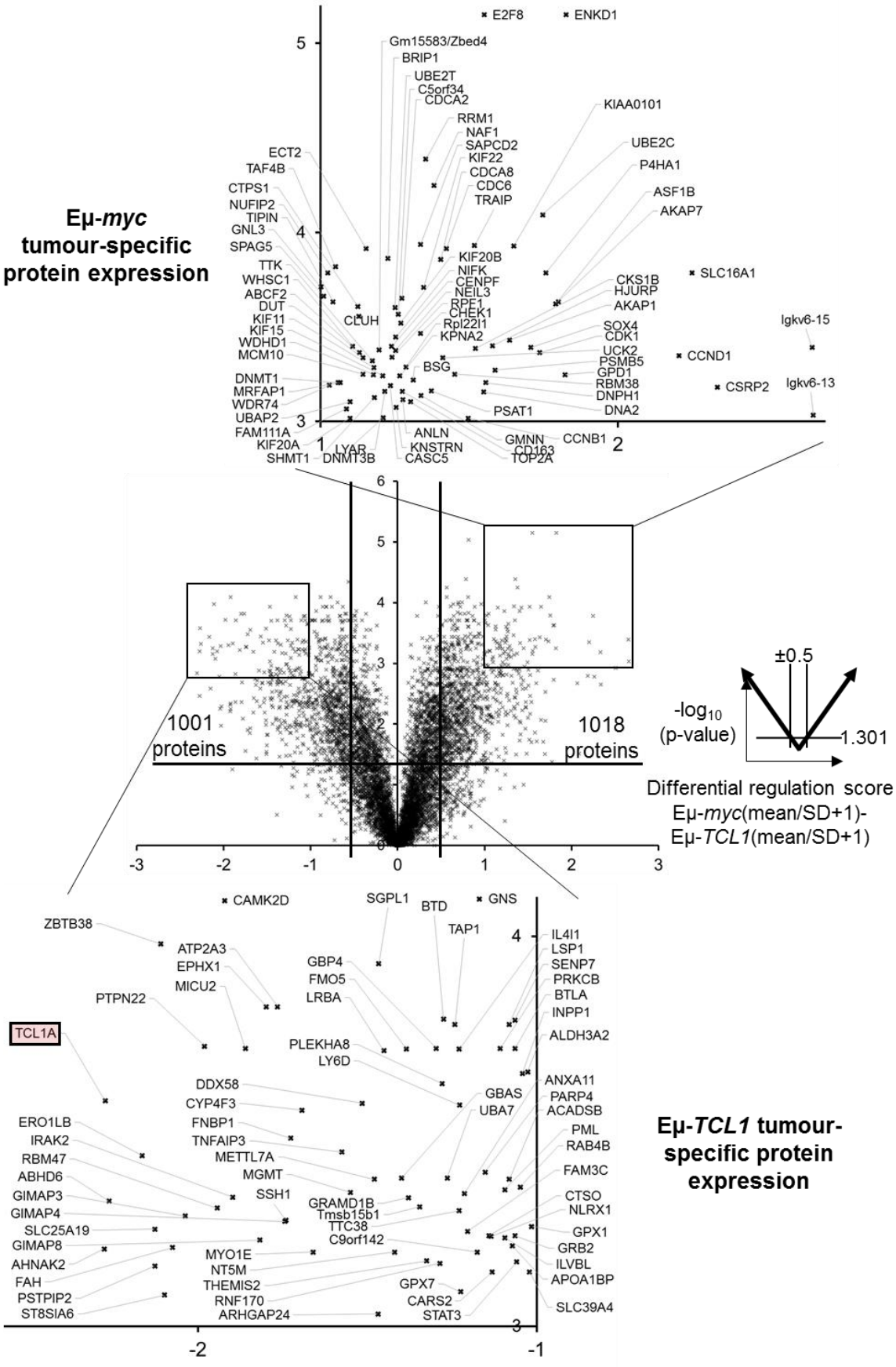
5.8 Model-Specific Protein Expression in the Eμ-*MYC* and Eμ-*TCL1* Tumours

Whilst **Figure 5.6** and **5.7** highlighted proteins with increased expression in both tumour types, to better understand proteins with a high degree of specificity to each tumour type, protein ratios were statistically compared and differential regulation scores calculated. These were plotted (**Figure 5.8**) as described in **Figure 5.6** and those proteins most specific to each model (with a low p-value and a high differential regulation score), plotted and annotated. The number of proteins specifically, statically significant to each models' tumours were 1018 and 1001 for Eμ-*myc* and Eμ-*TCL1*, respectively. For the Eμ-*myc* tumours, a number of those proteins observed upregulated in **Figure 5.6** were reproduced, where there was relatively little or no regulation present in the Eμ-*TCL1* tumours; most notably, E2F8, UBE2C, proliferating cell nuclear antigen-associated factor of 15 kDa (KIAA0101/PAF15) and monocarboxylate transporter 1 (SLC16A1). Other proteins appeared in this analysis, however, such as; enkurin domain-containing protein 1 (ENKD1), prolyl 4-hydroxylase subunit alpha-1 (P4HA1), ribonucleoside-diphosphate reductase large subunit (RRM1) and H/ACA ribonucleoprotein complex non-core subunit (NAF1), alongside multiple instances, as with **Figure 5.6**, of cell division cycle and kinesin proteins. A further observation was that of the two DNA methyltransferases; DNMT1 and DNMT3B, both with specificity to Eμ-*myc* tumours.

As described earlier, the expression of the TCL1 transgene protein was observed with a high degree of specificity to the Eμ-*TCL1* tumours (**Figure 5.8**). Only C-terminal binding protein-interacting BTB-domain containing zinc finger protein (ZBTB38) and calcium/calmodulin-dependent protein kinase type II subunit delta (CAMK2D) exhibited a greater degree of differential expression. Further proteins with an Eμ-*TCL1* tumour-specific expression included; monoacylglycerol lipase ABHD6, endoplasmic reticulum oxidoreductin-1-like protein B (ERO1LB), tyrosine-protein phosphatase non-receptor type 22 (PTPN22), mitochondrial calcium uptake protein (MICU2), epoxide hydrolase 1 (EPHX1) and sarcoplasmic/endoplasmic reticulum calcium ATPase 3 (ATP2A3). All 7 quantitated members of the GTP-binding and immunity-associated protein family (GIMAP), GIMAP1, 3, 4, 6, 7, 8 and 9 were observed significantly downregulated relative to WT B cells in Eμ-*myc* tumours,

while either unregulated or upregulated with a significant specificity to Eμ-*TCL1* tumours. Of these, just GIMAP4 was significantly upregulated in Eμ-*TCL1* tumours relative to WT B cells.

Figure 5.8. Eμ-*myc* and Eμ-*TCL1* B-cell tumour-specific protein expression. Volcano plot demonstrating the significance and magnitude of differential protein expression between the two B-cell cancer models, highlighting those proteins with the greatest degree of specificity to each tumour.


Eμ-*myc* tumour-specific protein expression
E2F8
ENKD1
Gm15583/Zbed4
BRIP1
UBE2T
C5orf34
CDCA2
KIAA0101
RRM1
NAF1
SAPCD2
KIF22
UBE2C
ECT2
TAF4B
CDCA8
P4HA1
CTPS1
CDC6
ASF1B
NUFIP2
TRAIP
AKAP7
TIPIN
GNL3
SPAG5
KIF20B
TTK
NIFK
SLC16A1
WHSC1
CENPF
CKS1B
ABCF2
NEIL3
HJURP
DUT
RPE1
AKAP1
KIF11
CHEK1
CLUH
Rpl22l1
SOX4
Igkv6-15
KIF15
KPNA2
CDK1
WDHD1
UCK2
CCND1
MCM10
BSG
PSMB5
GPD1
DNMT1
RBM38
CSRP2
Igkv6-13
MRFAP1
DNPH1
WDR74
PSAT1
DNA2
UBAP2
FAM111A
ANLN
GMNN
CCNB1
KIF20A
LYAR
KNSTRN
CD163
SHMT1
DNM1P35
CASC5
TOP2A
1001 proteins
1018 proteins
±0.5
-log10 (p-value)
1.301
Differential regulation score
Eμ-*myc*(mean/SD+1)-
Eμ-*TCL1*(mean/SD+1)
CAMK2D
SGPL1
BTD
GNS
ZBTB38
TAP1
IL4I1
ATP2A3
LSP1
EPHX1
GBP4
SENP7
MICU2
FMO5
PRKCB
PTPN22
LRBA
BTLA
INPP1
TCL1A
ALDH3A2
PLEKHA8
LY6D
ANXA11
DDX58
GBAS
PARP4
CYP4F3
UBA7
ACADSB
ERO1LB
FNBP1
PML
IRAK2
TNFAIP3
RAB4B
RBM47
METTL7A
FAM3C
ABHD6
MGMT
CTSO
GIMAP3
SSH1
GRAMD1B
NLRX1
GIMAP4
Tmsb15b1
TTC38
GPX1
SLC25A19
C9orf142
GRB2
GIMAP8
MYO1E
ILVBL
AHNAK2
NT5M
APOA1BP
FAH
THEMIS2
GPX7
SLC39A4
PSTPIP2
RNF170
CARS2
ST8SIA6
ARHGAP24
STAT3
Eμ-*TCL1* tumour-specific protein expression

5.9 TUMOUR-SPECIFIC PROTEIN EXPRESSION RELATIVE TO PRE-TUMOUROUS STATES IN Eμ-*MYC* AND Eμ-*TCL1* MODELS

To understand molecular processes contributing to the tumourigenic transition in the two B-cell cancer models, tumour protein expression was compared with that of the pre-cancer controls (**Figure 5.9**). Protein expression in the Eμ-*myc* pre-tumours was observed to be very similar to that of the tumours **(Figure 5.4**) therefore far fewer proteins appeared as significantly, differentially regulated, between tumours and pre-tumour controls compared to other comparisons. Myc was significantly overexpressed in tumours versus that of pre-tumours, observed previously by Western blotting in **Figure 3.6**. In total 100 proteins were overexpressed in the tumour condition and 254 underexpressed, relative to the pre-tumour state, just 5% of the quantitated proteome, compared to 36% of the quantitated proteome dysregulated in Eμ-*myc* relative to WT B cells. The proteins most upregulated in the isolated Eμ-*myc* tumour pools were zinc transporter Zrt- and Irt-like protein 6 (SLC39A6/ZIP-6), chromodomain-helicase-DNA-binding protein 7 (CHD7), nestin (NES) and three immunogloblulin proteins; Igkv 6-15, 13-84 and 6-13. Downregulated proteins included cofilin-2 (CFL2), dedicator of cytokinesis protein 5 (DOCK5), dystroglycan (DAG1) and two Igs; Igkv17-127 and Ighv11-1.

For the Eμ-*TCL1* pre-tumour B cells, protein expression was very minimally altered from that of the 6-week WT B cells. The most noticeable difference was the absence of TCL1 from this analysis, which was equivalently expressed in both pre-tumour and tumour samples (**Figure 5.5**), so was not highlighted by **Figure 5.9**.

Overall, 28 proteins were observed as significantly overexpressed in both tumours types, relative to their respective pre-tumour counterparts. Of these, 11 had minimal regulation ($<\pm 0.5$ $\log_2$ (ratio)) in the pre-tumour state relative to the WT controls, and were therefore highly specific to all 4 tumour pools. These included; hydroxymethylglutaryl-CoA synthase (HMCS1), heat shock factor-binding protein 1 (HSBP1), charged multivesicular body protein 7 (CHMP7), alpha-internexin (INA), zinc finger CCHC domain-containing protein 9 (ZCCHC9), COP9 signalosome complex subunit 7a (COPS7A), serum amyloid A-like 1 (SAAL1), intraflagellar transport protein 27 homolog (IFT27), proteasome assembly chaperone 4 (PSMG4), neuroblast differentiation-associated protein (AHNAK) and RNA polymerase B transcription factor 3 (BTF3).

Eμ-*myc*
Tumour vs 6 week Eμ-*myc*
(pre-tumour) protein expression

Eμ-*TCL1*
Tumour vs 6 week Eμ-*TCL1*
(pre-leukeamia) protein expression

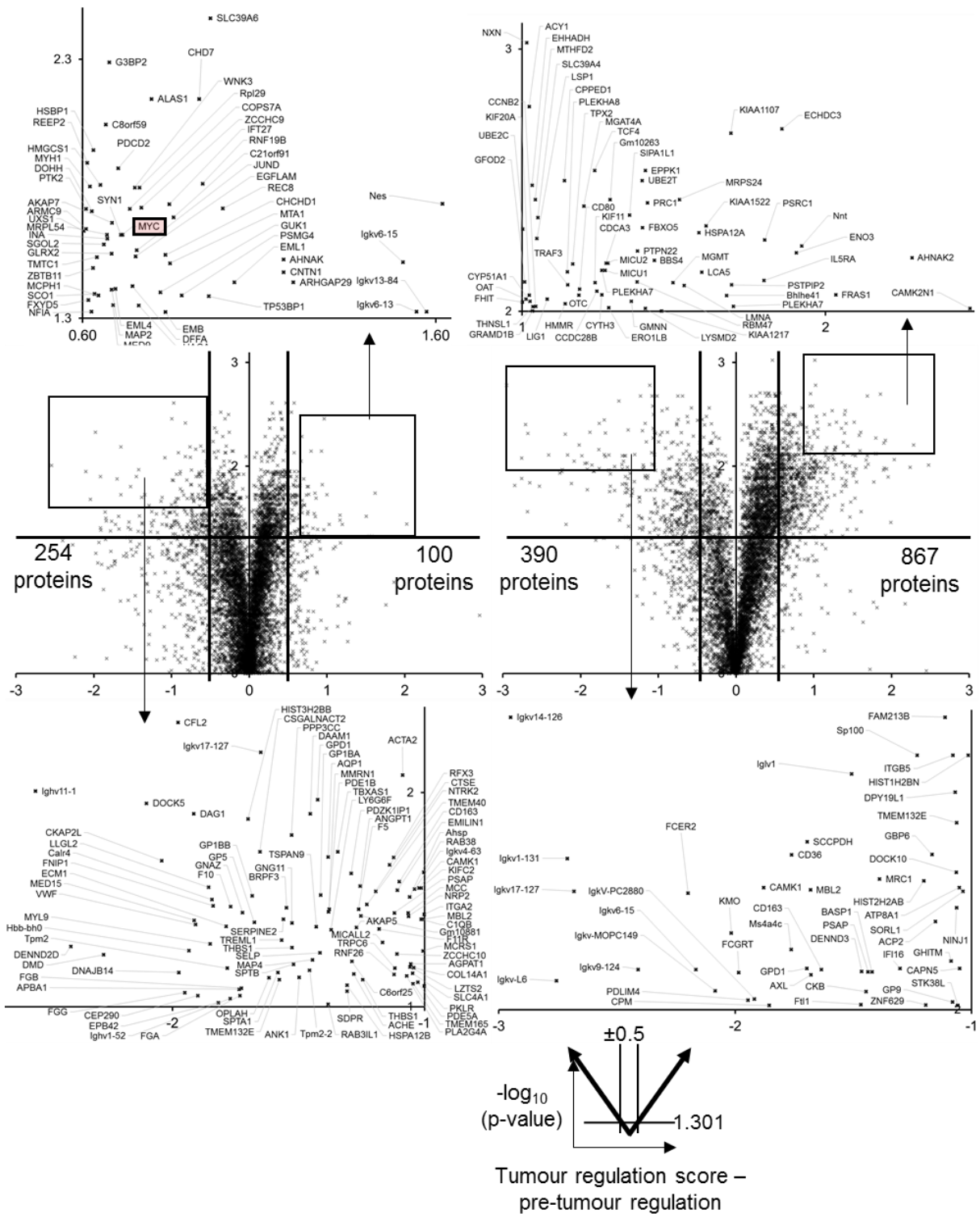


Figure 5.9. Comparing tumour with pre-tumour protein expression in the Eμ-*myc* and Eμ-*TCL1* models. For each plot, the respective tumour regulation scores were adjusted by subtracting the pre-tumour $\log_2$ (ratio) and p-values determined by one-sample T-test to the pre-tumour values.

5.10 Localisation and Functions of Differentially Expressed Proteins in Eμ-*MYC* and Eμ-*TCL1* Tumours

Proteins were annotated for their canonical localisations and functional categories using Ingenuity Pathway Analysis software. The number of proteins fitting into each of these categories for those proteins differentially regulated ($p<0.05$) in each tumour or both tumours, was determined and plotted against the numbers observed from all proteins to identify any patterns of enrichment. **Figure 5.10** represents the proteins up- or down-regulated in the Eμ-*myc*, Eμ-*TCL1* and both tumours annotated as canonically existing within the extracellular space, the plasma membrane, the cytoplasm or the nucleus – alongside the background percentage of all 7391 quantitated proteins.

A trend that was observed for those proteins differentially expressed ($p<0.05$) in both tumours, whether considered individually or combined, was that of a strong downregulation of membrane proteins, with on average, twice as many proteins as would be expected to be downregulated based on the total number of observed membrane proteins, and the inverse true for upregulated membrane proteins. For the Eμ-*myc* tumours, nuclear proteins were overrepresented amongst the upregulated proteins and underrepresented in the downregulated proteins. For Eμ-*TCL1* tumours, both up- and downregulated nuclear proteins were underrepresented. An overrepresentation of upregulated proteins and underrepresentation of downregulated proteins was observed for cytoplasmic proteins.

The same analysis was performed for the annotated protein classifications (**Figure 5.11**). This demonstrated a trend in the overall increased expression of proteins classified as enzymes, with almost 20% of identified enzymes upregulated while just 3% were downregulated in both tumours. Translational regulators were also overrepresented in the upregulated proteins, with 52 upregulated in Eμ-*myc* tumours out of the 78 identified. Transmembrane receptors made up a large proportion of downregulated proteins in both tumours; in accordance with the observation of downregulated plasma membrane proteins in **Figure 5.10**. Of the 100 identified, 30 of these were downregulated while just 7 were upregulated in both tumours.

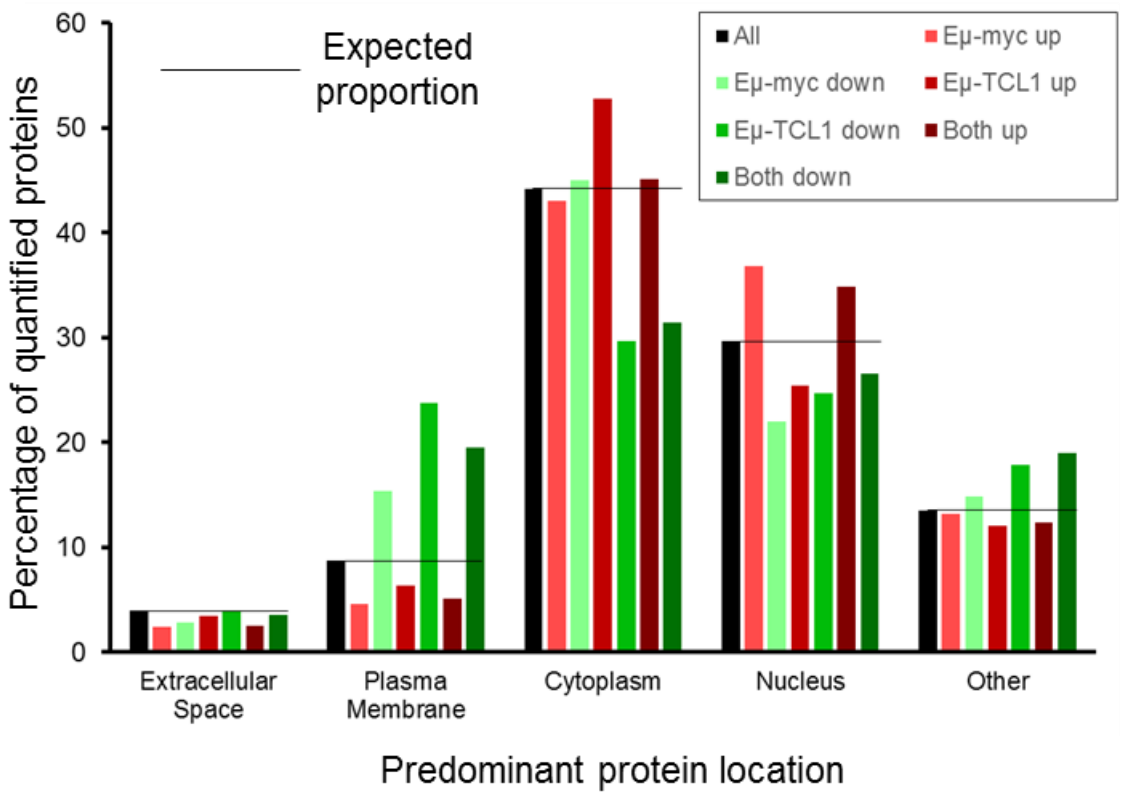


Figure 5.10. Annotation of protein localisations for the up and downregulated proteins in the Eμ-*myc* and Eμ-*TCL1* tumours. A. Proteins defined as up (RS>0.5, p<0.05) and down (RS< -0.5, p<0.05) regulated in either tumour, or when considering the RS and p-values derived for both tumours, were categorised for their predominant protein localisation and compared with all protein identifications.

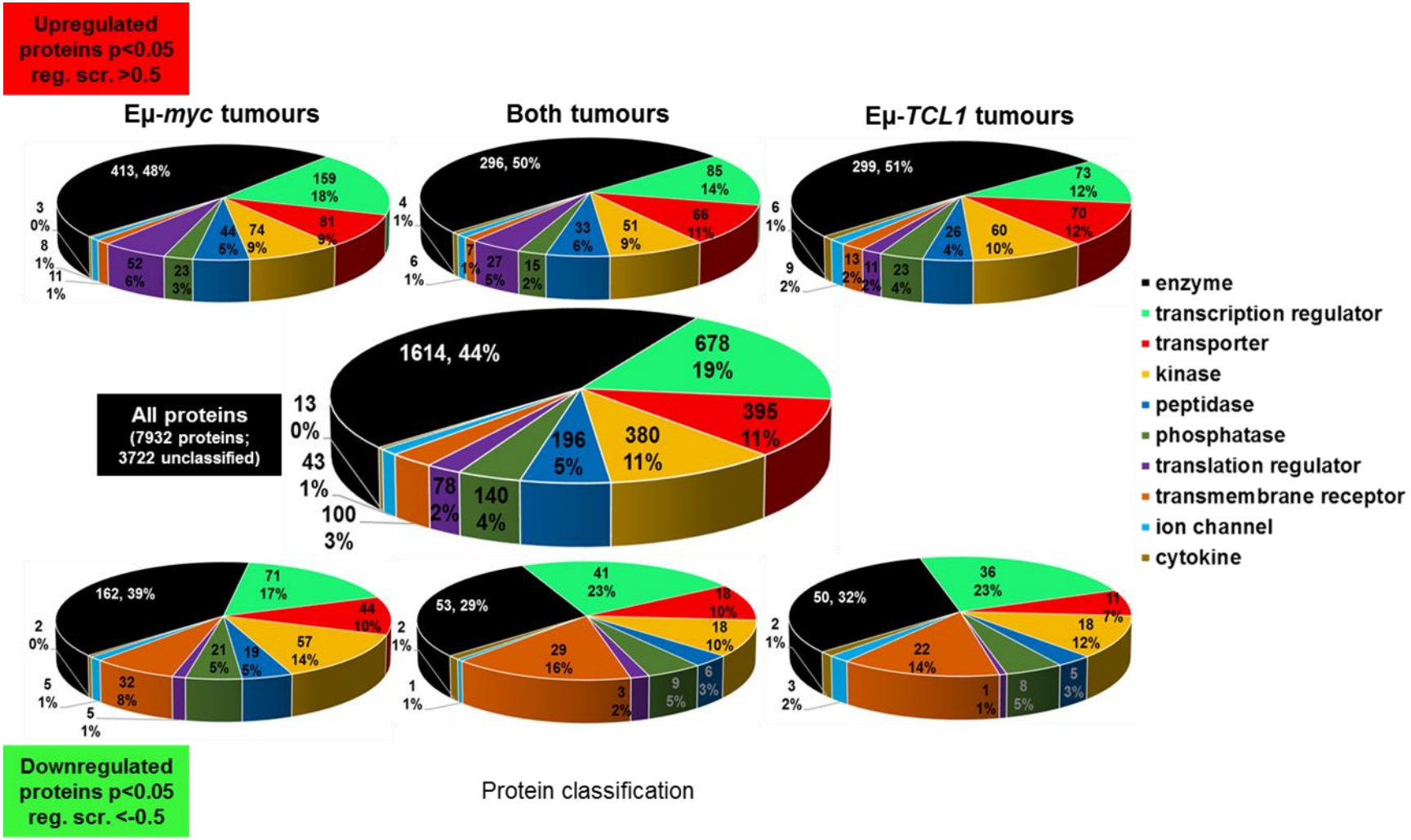

Upregulated proteins p<0.05 reg. scr. >0.5
Eμ-myc tumours
Both tumours
Eμ-TCL1 tumours
413, 48%
159 18%
81 9%
74 9%
44 5%
52 6%
23 3%
11 1%
8 1%
3 0%
296, 50%
85 14%
66 11%
51 9%
33 6%
27 5%
15 2%
7 1%
6 1%
4 1%
299, 51%
73 12%
70 12%
60 10%
26 4%
23 4%
13 2%
11 2%
9 2%
6 1%
All proteins (7932 proteins; 3722 unclassified)
1614, 44%
678 19%
395 11%
380 11%
196 5%
140 4%
78 2%
100 3%
43 1%
13 0%
enzyme
transcription regulator
transporter
kinase
peptidase
phosphatase
translation regulator
transmembrane receptor
ion channel
cytokine
162, 39%
71 17%
44 10%
57 14%
21 5%
19 5%
32 8%
5 1%
5 1%
2 0%
53, 29%
41 23%
18 10%
18 10%
29 16%
9 5%
6 3%
3 2%
2 1%
1 1%
50, 32%
36 23%
11 7%
18 12%
22 14%
8 5%
5 3%
1 1%
2 1%
3 2%
Downregulated proteins p<0.05 reg. scr. <-0.5
Protein classification

Figure 5.11. Annotation of protein subtypes for the up and downregulated proteins in the Eμ-myc and Eμ-TCL1 tumours. Proteins defined as up (RS>0.5, p<0.05) and down (RS< -0.5, p<0.05) regulated in either tumour, or when considering the RS and p-values derived for both tumours, were categorised for their annotated functional proteins categories. Unclassified proteins are not plotted and the percentages of these proteins do not greatly differ.

5.11 GENE ONTOLOGY ENRICHMENT AMONGST DIFFERENTIALLY EXPRESSED PROTEINS IN Eμ-*MYC* AND Eμ-*TCL1* TUMOURS

To gain a greater insight into any commonalities in the nature of the proteins identified as upregulated and downregulated a process called gene ontology term (GO term) enrichment was employed. GO terms offer descriptions of the proteins, similar to that outlined in **Figure 5.10** and **Figure 5.11**, however with far greater specificity and annotation depth. Using the online software tool 'DAVID' (the database for annotation, visualisation and integrated discovery) the significantly upregulated and downregulated proteins were assessed against all 7391 fully quantitated proteins (background) to identify which GO terms were significantly overrepresented in the up and down regulated groups. **Figure 5.12** and **Figure 5.13** demonstrate visual representations (using the tool REVIGO) of those GO terms determined as significantly enriched amongst the proteins up and down regulated commonly across both tumours, respectively. The plots describe the GO terms in three ways; the size is proportional to the $-\log_{10}$ (term enrichment p-value) for the input data, the colour is indicative of the overall frequency of each GO term amongst all mouse genes - red being frequent and blue being rare, and the axis of the plot are based on semantic space, clustering terms that are more frequently co-annotated together.

Due to sizeable overlap of the GO term enrichment for each tumour type with that of the analysis simultaneously for both tumours, the separate REVIGO outputs for Eμ-*myc* Eμ-*TCL1* tumours are presented in **Appendix A9**. In total 123 and 67 GO terms were significantly enriched amongst the 1219 upregulated and 369 downregulated proteins, respectively, amongst which 52, 47 and 24, and 51, 15 and 1 belonged to the three classifications; biological process, molecular function and cellular component, respectively.

The proteins upregulated in both tumours demonstrated several significantly enriched GO terms suggesting that B-cell tumours overexpress proteins driving cell proliferation **(Figure 5.12)**. Terms were indicative of a vastly increased rate of protein synthesis with the most significantly enriched term being 'translation (GO:0006412)'. In addition to this, terms including 'cellular amino acid metabolism (GO:0006520)', 'protein metabolism (GO:0019538)' and

‘cellular component biogenesis (GO:0044085)’ described an increase in translation at the levels of both metabolic precursors and macromolecular products. Terms relating to non-coding RNAs are also indicative of a translational regulation via tRNAs and rRNAs; ‘ncRNA processing (GO:0034470)’ and ‘ncRNA metabolism (GO:0034660)’ The second most significantly enriched ‘biological process’ term amongst the upregulated proteins was ‘cell cycle (GO:0007049)’. Several other terms were also related to cell cycle progression, suggesting coordinated division in the cell and its components, including; ‘organelle fission (GO:0048285)’, ‘nuclear division (GO:0000280)’, and ‘cell division (GO:0051301)’. Processes underlying the initiation and regulation of mitosis were also enriched, such as; ‘cell cycle phase (GO:0022403)’, ‘cell cycle (GO:0007049)’, ‘cell cycle process (GO:0022402)’ and ‘mitotic cell cycle (GO:0000278)’. Proteins promoting the replication and division of DNA were identified with a number of enriched terms, including; ‘DNA replication (GO:0006260)’, ‘DNA metabolism (GO:0006259)’, ‘DNA replication initiation (GO:0006270)’, ‘chromosome condensation (GO:0030261)’ and ‘chromosome segregation (GO:0007059)’.

For the ‘cellular component’ terms, enrichment amongst the upregulated tumour proteins highlighted the involvement of organelles without membrane association, such as ribosomes, indicated by the top three terms; ‘intracellular non-membrane-bounded organelle (GO:0043232)’, ‘non-membrane-bounded organelle (GO:0043228)’ and ‘ribonucleoprotein complex (GO:0030529)’. Further terms detailed the specificity of this enrichment for ribosomes with ‘preribosome (GO:0030684)’, ‘large ribosomal subunit (GO:0015934)’, ‘organellar ribosome (GO:0000313)’ and ‘mitochondrial ribosome (GO:0005761)’ enriched amongst the upregulated tumour proteins. Nuclear (GO:0044428, GO:0005634), and specifically nucleoli (GO:0005730) proteins were significantly enriched for, including terms relating to chromosomes (GO:0005694) and spindles (GO:0005819); ‘condensed chromosome (GO:0000793)’, ‘chromosome, centromeric region (GO:0000775)’ and ‘spindle microtubule (GO:0005876)’.

‘Molecular function’ GO terms again highlighted the predominance of ribosomal enrichment amongst the upregulated tumour proteins; with ‘structural constituent of ribosome (GO:0003735)’ the with the lowest enrichment p-value alongside the term ‘rRNA binding (GO:0019843)’. A number of transcription- and RNA-related functions were also observed as significantly enriched, including; ‘RNA polymerase activity (GO:0034062)’, ‘DNA-directed RNA polymerase activity (GO:0003899)’, ‘RNA binding (GO:0003723)’, ‘nucleoside binding (GO:0001882)’, ‘purine nucleotide binding (GO:0017076)’, ‘helicase activity (GO:0004386)’ and ‘ATP-dependent helicase activity (GO:0008026)’. Some specific metabolic functions were also overrepresented; ‘ligase activity (GO:0016874)’, ‘transferase activity, transferring one-carbon groups (GO:0016741)’, ‘methyltransferase activity (GO:0008168)’, ‘hydrolase activity,

acting on acid anhydrides (GO:0016817)' and 'nucleoside-triphosphatase activity (GO:0017111)'.

Those proteins significantly downregulated in across both pools of both Eμ-*myc* and Eμ-*TCL1* tumours were also submitted for GO term enrichment analysis (**Figure 5.13**). The two most significantly enriched, downregulated 'biological processes' were 'immune system process (GO:0002376)' and 'immune response (GO:0006955)'. As a subcategory of this loss of immune response, three GO terms described, more specifically, a downregulation of major histocompatibility complex (MHC) antigen processing and presentation (GO:0002495, GO:0019884, GO:0019886) as well as the related processes of 'endocytosis (GO:0006897)' and 'membrane invagination (GO:0010324)'. Additionally, a suppression of proteins annotated as having roles in 'signal transduction (GO:0007165)', 'response to stimulus (GO:0050896)', 'regulation of response to stimulus (GO:0048583)' and 'small GTPase mediated signal transduction (GO:0007264)' were observed, as well as terms relating to biological process regulation (GO:0065007, GO:0050789 and GO:0048518). Processes relating to differentiation (GO:0045597) and development (GO:0050793) were lost in tumours, as well as proteins relating to nucleosome assembly (GO:0006334) and DNA packaging (GO:0006323).

Downregulated 'cellular components' were dominated by terms describing proteins localised to the cell surface; 'cell surface (GO:0009986)', 'membrane (GO:0016020)', 'plasma membrane part (GO:0044459)', 'intrinsic component of membrane (GO:0031224)', 'membrane part (GO:0044425)', 'external side of plasma membrane (GO:0009897)', 'plasma membrane (GO:0005886)' and 'integral component of membrane (GO:0016021)'. In agreement with the biological processes, MHC (GO:0042611, GO:0042613) and nucleosomes (GO:0000786, GO:0032993) were underrepresented when considering their components. Just the single term 'antigen binding (GO:0003823)', describing the selective interaction between an antigen and any molecule capable of initiating a specific immune response, was significantly enriched for within the downregulated proteins for the category of 'molecular function'.

Figure 5.12. Gene ontology term enrichment analysis of the upregulated proteins across both Eμ-*myc* and Eμ-*TCL1* tumours. GO term enrichment p-values (Benjamini-corrected) were determined for those proteins with significant differential upregulation (RS>0.5, $p<0.05$) in all 4 tumour pools using DAVID with all fully quantitated proteins as background. Significantly enriched GO terms ($p<0.05$) were visualised with REVIGO. These are separated into biological processes, cell components and molecular function. The circle size is proportional to the $-\log_{10}$ (term enrichment p-value), the colour indicates the GO term frequency amongst all mouse genes, and the axis of the plot are based on semantic space, clustering terms that are more frequently co-annotated together.

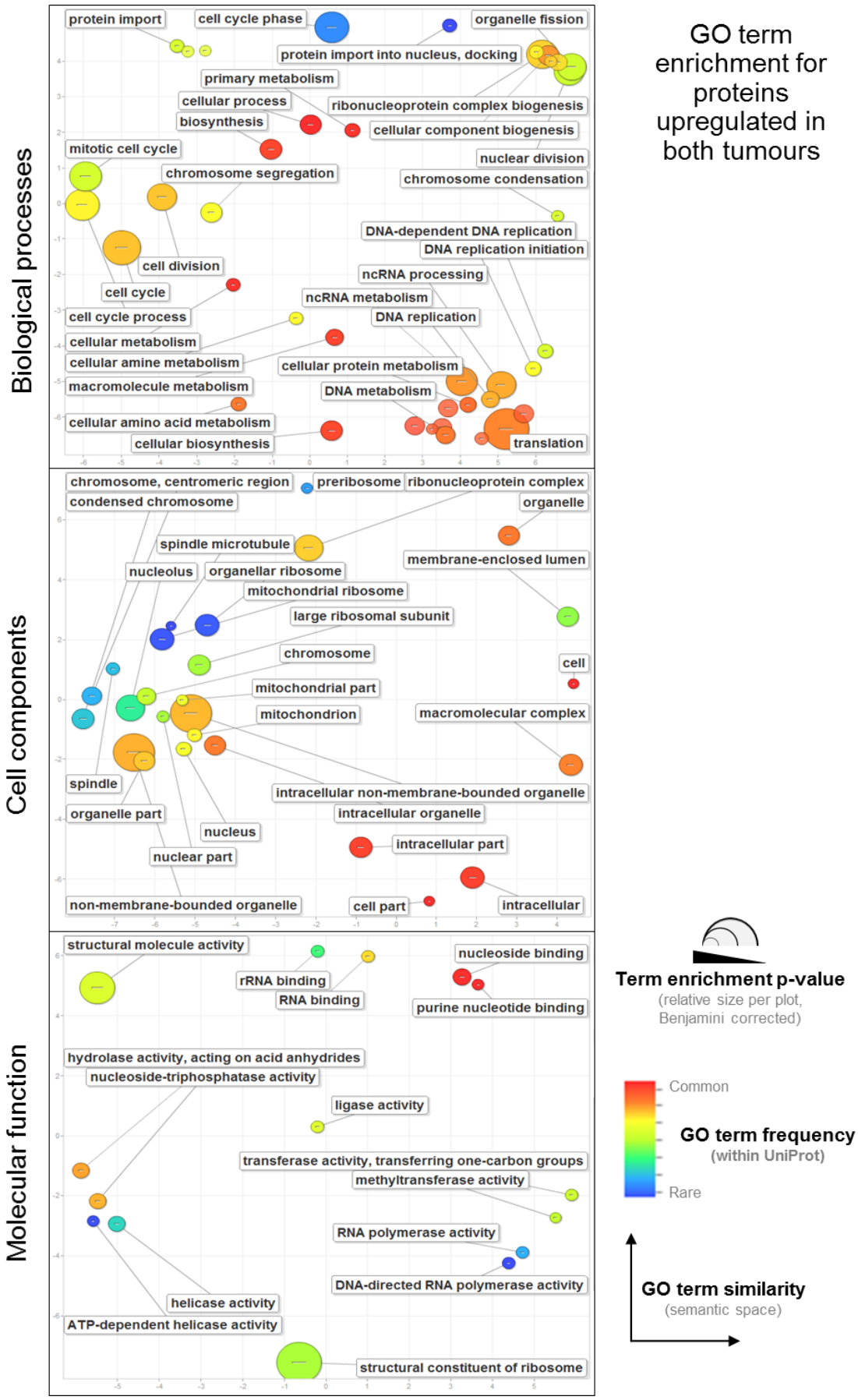
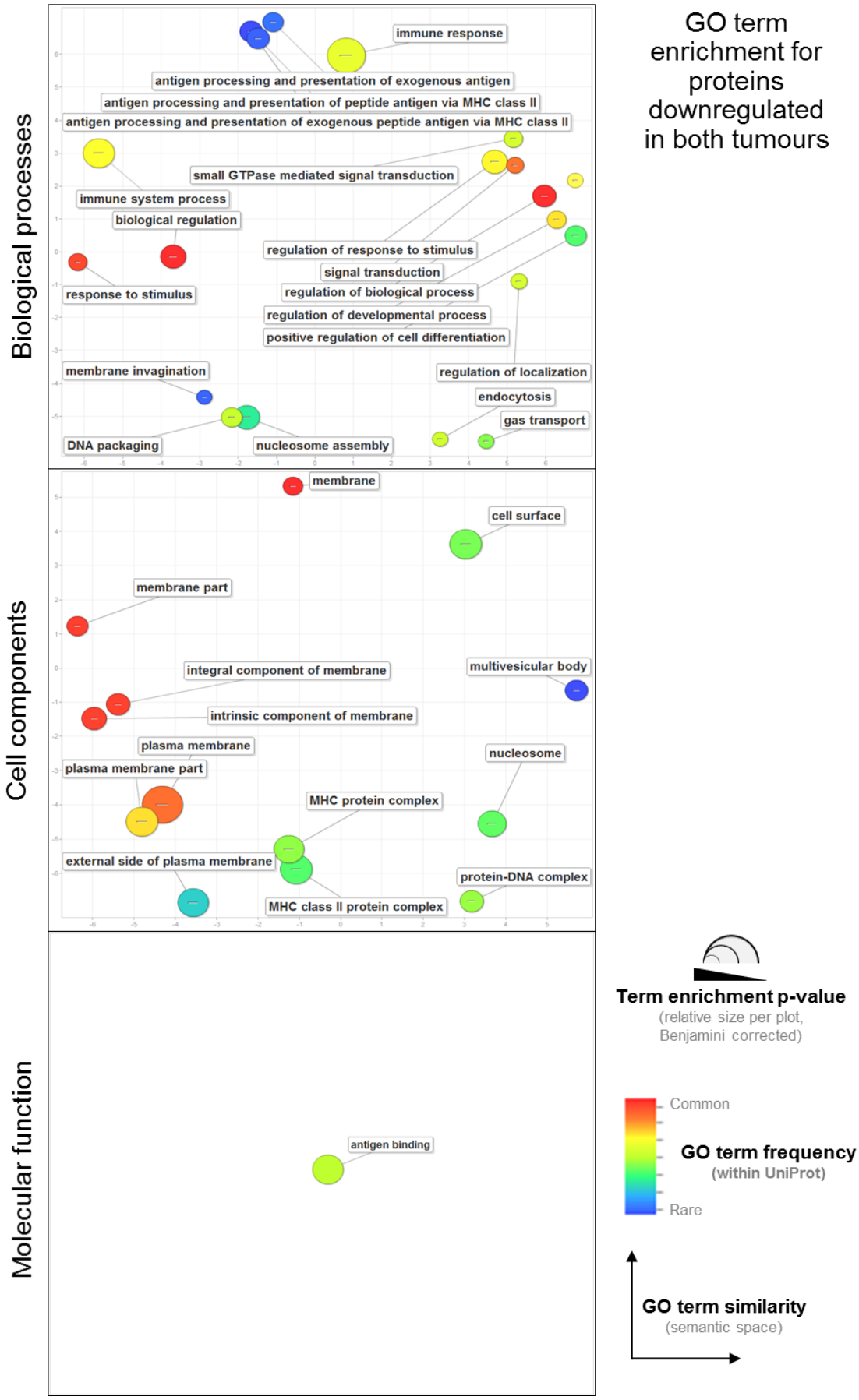

Biological processes
protein import
cell cycle phase
organelle fission
protein import into nucleus, docking
primary metabolism
cellular process
ribonucleoprotein complex biogenesis
biosynthesis
cellular component biogenesis
mitotic cell cycle
nuclear division
chromosome segregation
chromosome condensation
DNA-dependent DNA replication
DNA replication initiation
cell division
ncRNA processing
cell cycle
ncRNA metabolism
cell cycle process
DNA replication
cellular metabolism
cellular amine metabolism
cellular protein metabolism
macromolecule metabolism
DNA metabolism
cellular amino acid metabolism
cellular biosynthesis
translation
Cell components
chromosome, centromeric region
preribosome
ribonucleoprotein complex
condensed chromosome
organelle
spindle microtubule
membrane-enclosed lumen
nucleolus
organellar ribosome
mitochondrial ribosome
large ribosomal subunit
chromosome
cell
mitochondrial part
mitochondrion
macromolecular complex
spindle
intracellular non-membrane-bounded organelle
organelle part
intracellular organelle
nucleus
intracellular part
nuclear part
non-membrane-bounded organelle
cell part
intracellular
Molecular function
structural molecule activity
nucleoside binding
rRNA binding
RNA binding
purine nucleotide binding
hydrolase activity, acting on acid anhydrides
nucleoside-triphosphatase activity
ligase activity
transferase activity, transferring one-carbon groups
methyltransferase activity
RNA polymerase activity
DNA-directed RNA polymerase activity
helicase activity
ATP-dependent helicase activity
structural constituent of ribosome
GO term enrichment for proteins upregulated in both tumours
Term enrichment p-value
(relative size per plot, Benjamini corrected)
Common
GO term frequency
(within UniProt)
Rare
GO term similarity
(semantic space)

Figure 5.13. Gene ontology term enrichment analysis of the downregulated proteins across both Eμ-*myc* and Eμ-*TCL1* tumours. GO term enrichment p-values (Benjamini-corrected) were determined for those proteins with significant downregulation expression (RS< -0.5, $p<0.05$) in all 4 tumour pools using DAVID with all fully quantitated proteins as background. Significantly enriched GO terms ($p<0.05$) were visualised with REVIGO. These are separated into biological processes, cell components and molecular function. The circle size is proportional to the $-\log_{10}$ (term enrichment p-value), the colour indicates the GO term frequency amongst all mouse genes, and the axis of the plot are based on semantic space, clustering terms that are more frequently co-annotated together.


Biological processes
immune response
antigen processing and presentation of exogenous antigen
antigen processing and presentation of peptide antigen via MHC class II
antigen processing and presentation of exogenous peptide antigen via MHC class II
small GTPase mediated signal transduction
immune system process
biological regulation
regulation of response to stimulus
signal transduction
response to stimulus
regulation of biological process
regulation of developmental process
positive regulation of cell differentiation
membrane invagination
regulation of localization
endocytosis
gas transport
DNA packaging
nucleosome assembly
Cell components
membrane
cell surface
membrane part
integral component of membrane
multivesicular body
intrinsic component of membrane
plasma membrane
plasma membrane part
nucleosome
MHC protein complex
external side of plasma membrane
protein-DNA complex
MHC class II protein complex
Molecular function
antigen binding
GO term enrichment for proteins downregulated in both tumours
Term enrichment p-value
(relative size per plot, Benjamini corrected)
Common
GO term frequency
(within UniProt)
Rare
GO term similarity
(semantic space)

5.12 CANONICAL PATHWAY ENRICHMENT AMONGST DIFFERENTIALLY EXPRESSED PROTEINS IN Eμ-*MYC* AND Eμ-*TCL1* TUMOURS

Much like with GO terms, proteins classified as up or downregulated can be mapped to pathways to identify overrepresentation of discrete cellular functions. Pathways were analysed with 'DAVID' which evaluated KEGG, BioCarta and PANTHER pathways for overrepresentation in each given list of proteins. Pathways were considered for each tumour type separately to reflect the discrete biology of the two models. The full summary of significant ($p<0.05$) pathways determined for the significantly ($p<0.05$) up ($RS>0.5$) and down ($RS<-0.5$) regulated proteins in both tumours and the Eμ-*myc* and Eμ-*TCL1* tumours separately, is given in **Appendix A10**. IPA pathway analysis was evaluated, but due to the large number of differentially regulated proteins produced results which were difficult to interpret. DAVID analysis provided a more concise summary of pathways.

For the proteins upregulated in Eμ-*myc* tumours, 5 BioCarta pathways were significantly enriched **(Figure 5.14)**. **Figure 5.12A** summarises these alongside 10 other pathways determined as significant prior to multiple test correction, but still with a magnitude of enrichment. Of the 5 pathways, 4 (**Figure 5.14B,C,D,F)** were related to the cell cycle and the 5th (**Figure 5.14E)**, to protein synthesis. Of those pathways not reaching corrected significance, all 10 were directly relatable to tumourigenic mechanisms, such as DNA damage response, tumour suppressor pathways or further cell cycle promoting pathways.

13 KEGG pathways were identified as significantly enriched, with a further 11 when disregarding correction (**Figure 5.15A**). These pathways suggested a trend in increased rates of protein synthesis with 77 upregulated ribosome components (**Figure 5.15B**) demonstrating an enrichment p-value of 2.6×10^{-68}. Amino acid synthesis was also enriched for, with 5 pathways indicative of increased synthesis of 10 amino acids (Gly, Ser, Thr, Val, Leu, Iso, Ala, Asp, Glu and Selenocysteine (Sec)), and aa-tRNA regeneration ($p=4.5 \times 10^{-10}$). Protein degradation pathways were also enriched, with 'ubiquitin mediated proteolysis' ($p=0.0002$) and the proteasome appearing in this list, with 28 and 10 upregulated proteins mapping to these pathways respectively. 'Cell cycle' (**Figure 5.15C**) was the second most significantly enriched pathway (3.5×10^{-16}) with 46 upregulated members. This also identified that all 7 members of the mini-chromosome maintenance (MCM) complex were upregulated (including MCM10, not shown). DNA replication ($p=7.4 \times 10^{-12}$) was identified as an enriched pathway; emphasised by a significant enrichment of both pyrimidine ($p=1.0 \times 10^{-15}$) (**Figure 5.15D**) and purine ($p=2.2 \times 10^{-11}$) nucleotide synthesis pathways. DNA damage pathways were also significantly enriched

($p<0.01$), including nucleotide excision repair, mismatch repair and p53 signalling. Base excision repair, cytosolic DNA sensing and homologous recombination also demonstrated some enrichment.

With fewer significantly upregulated proteins, fewer pathways were significantly enriched for Eμ-*TCL1* tumours **(Figure 5.16A)**. Just one BioCarta pathway was significantly enriched after correction; 'retinoblastoma tumour suppressor/checkpoint signalling in response to DNA damage' ($p=0.0045$) (**Figure 5.16B**). However, several other BioCarta pathways were significantly enriched before multiple test correction. These were related to the cell cycle and its regulation, such as p53 and ataxia telangiectasia mutated (ATM) signalling, G2/M and G1/S checkpoints, p27 phosphorylation, E2F1 and cyclin E destruction, regulation of cyclins, CDKs and DNA replication, and regulation of mitosis by Ran and AKAP95. The pathway 'telomeres, telomerase, cellular aging, and immortality' was also enriched for. Overall the majority of enriched KEGG pathways were similar to that of the Eμ-*myc* tumours, but due to a lower number of significantly upregulated proteins fewer pathways had corrected significance; including the ribosome ($p=5.8 \times 10^{-6}$) (**Figure 5.16C**) and the cell cycle ($p=1.5 \times 10^{-5}$) (**Figure 5.16D**). Only two KEGG pathways reached significance which were not observed for the Eμ-*myc* tumours. Firstly valine, leucine and isoleucine biosynthesis reached corrected significance ($p=8.2 \times 10^{-4}$) and secondly N-glycan biosynthesis was observed to be significantly enriched uniquely to the Eμ-*TCL1* model. Three Panther pathways reached significance, uniquely in the Eμ-*TCL1* tumours, two of which described core p53 biology. The third, and a fourth - significant before correction, described the *de novo* biosynthesis of pyrimidines and purines, respectively.

Figure 5.14

A.

Biocarta Pathway	Fold Enrichment	p-value	Benjamini
Cyclins and Cell Cycle Regulation	**4.55**	**0.00004**	**0.00568**
Role of Ran in mitotic spindle regulation	**7.40**	**0.00007**	**0.00561**
Cell Cycle: G1/S Check Point	**4.19**	**0.00009**	**0.00450**
Eukaryotic protein translation	**8.16**	**0.00020**	**0.00766**
Cell Cycle: G2/M Checkpoint	**3.73**	**0.00141**	**0.04233**
RB Tumor Suppressor/Checkpoint Signaling in response to DNA damage	5.19	0.00313	0.07673
mTOR Signaling Pathway	3.46	0.00505	0.10468
AKAP95 role in mitosis and chromosome dynamics	5.29	0.00934	0.16425
Estrogen-responsive protein Efp controls cell cycle and breast tumors growth	6.35	0.01745	0.25862
p53 Signaling Pathway	3.57	0.01924	0.25718
CDK Regulation of DNA Replication	5.44	0.02824	0.32860
Regulation of p27 Phosphorylation during Cell Cycle Progression	3.97	0.02863	0.30948
Influence of Ras and Rho proteins on G1 to S Transition	2.78	0.03153	0.31411
ATM Signaling Pathway	3.17	0.03184	0.29790
Role of BRCA1, BRCA2 and ATR in Cancer Susceptibility	3.01	0.03971	0.33853

B.

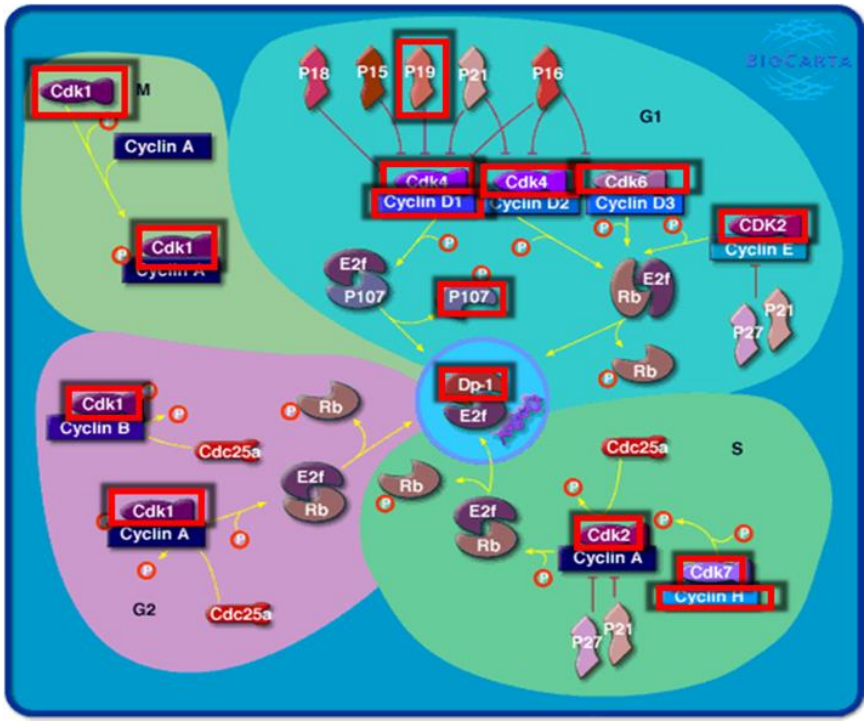
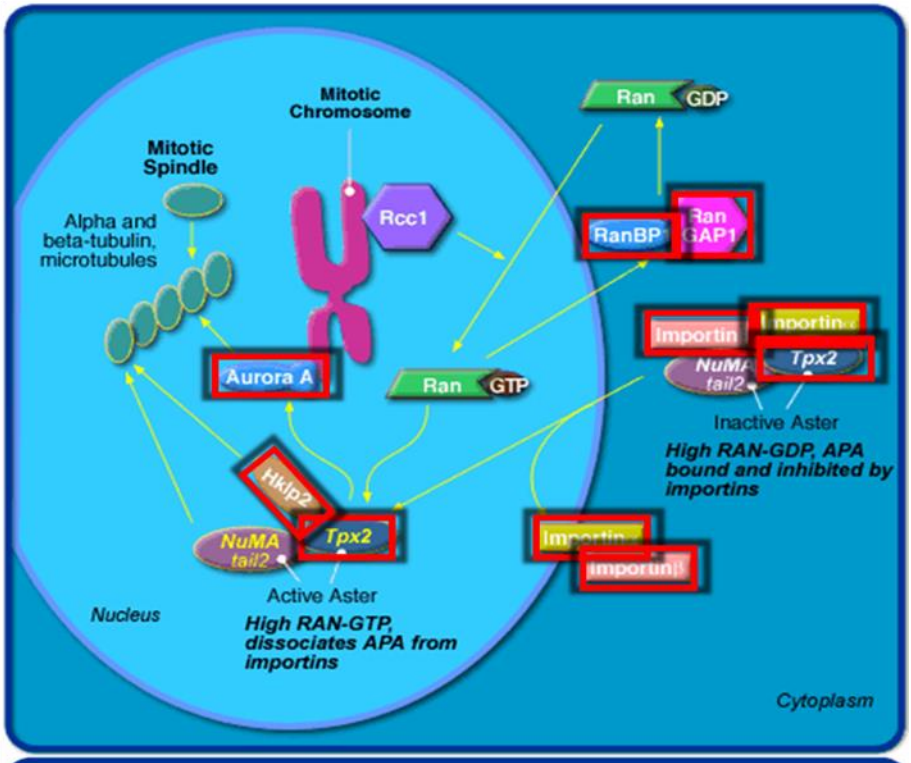


Figure 5.14 Biocarta pathways identified as significantly enriched from the proteins upregulated in Eμ-myc tumours. Significantly ($p<0.05$), differentially expressed ($RS>0.5$) proteins were analysed using DAVID to identify significantly enriched pathways. Red boxes highlight those overexpressed proteins mapping to these pathways. **A.** A summary of the fold enrichments, p-values and corrected p-values for the top 15 BioCarta pathways. **B.** Cyclins and cell cycle regulation. **C.** Role of Ran in mitotic spindle regulation. **D.** Cell cycle: G1/S check point, **E.** Eukaryotic protein translation. **F.** Cell cycle: G2/M checkpoint.

Figure 5.14

C.



D.

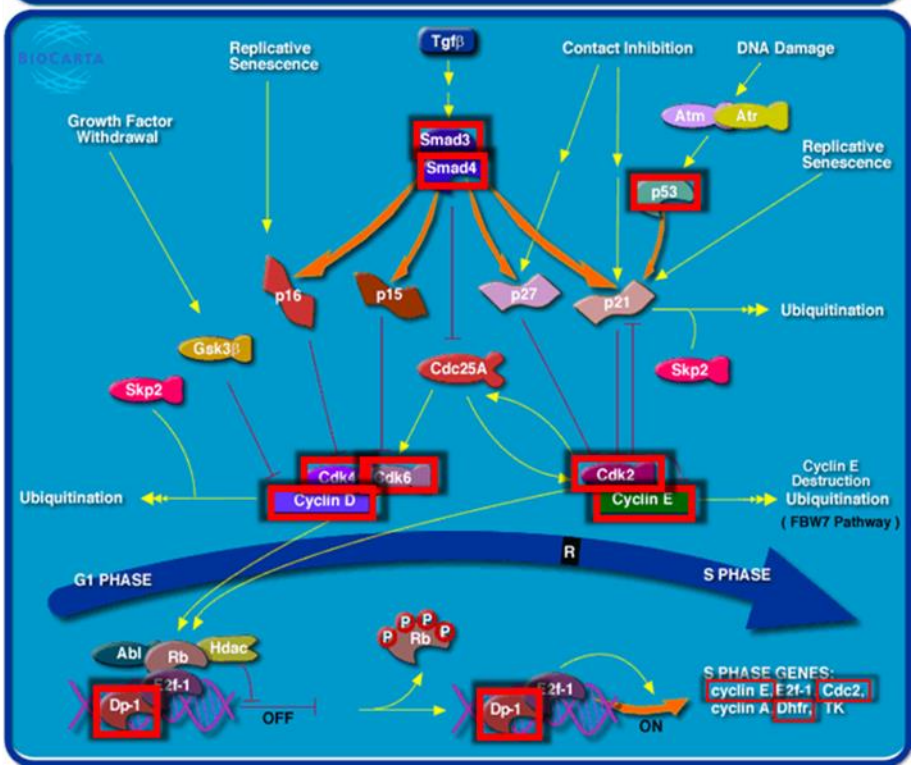
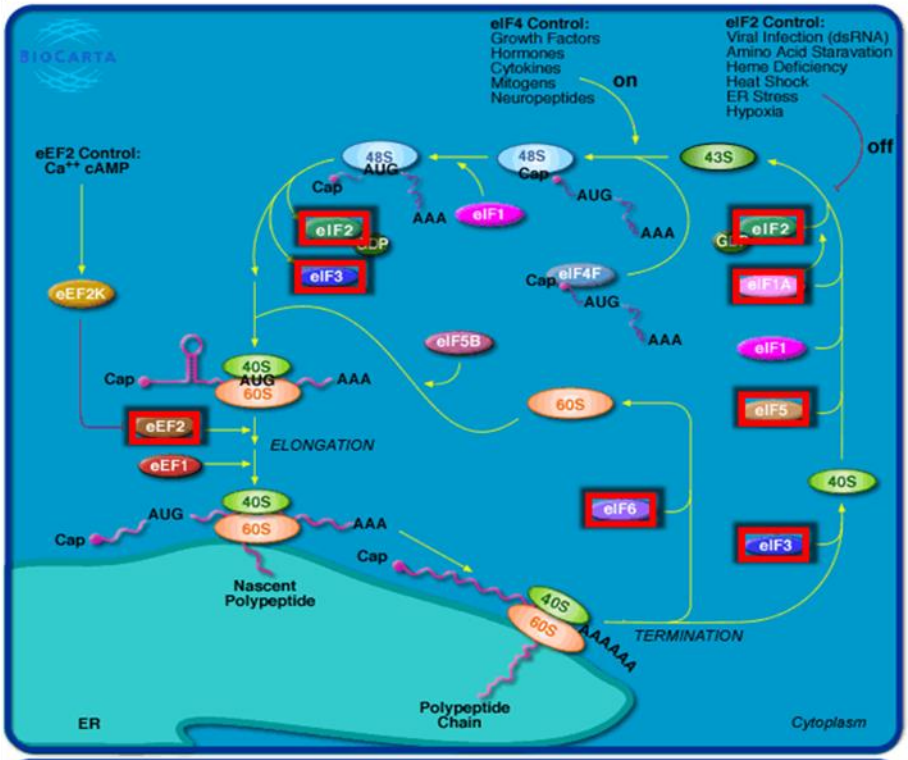


Figure 5.14

E.



F.

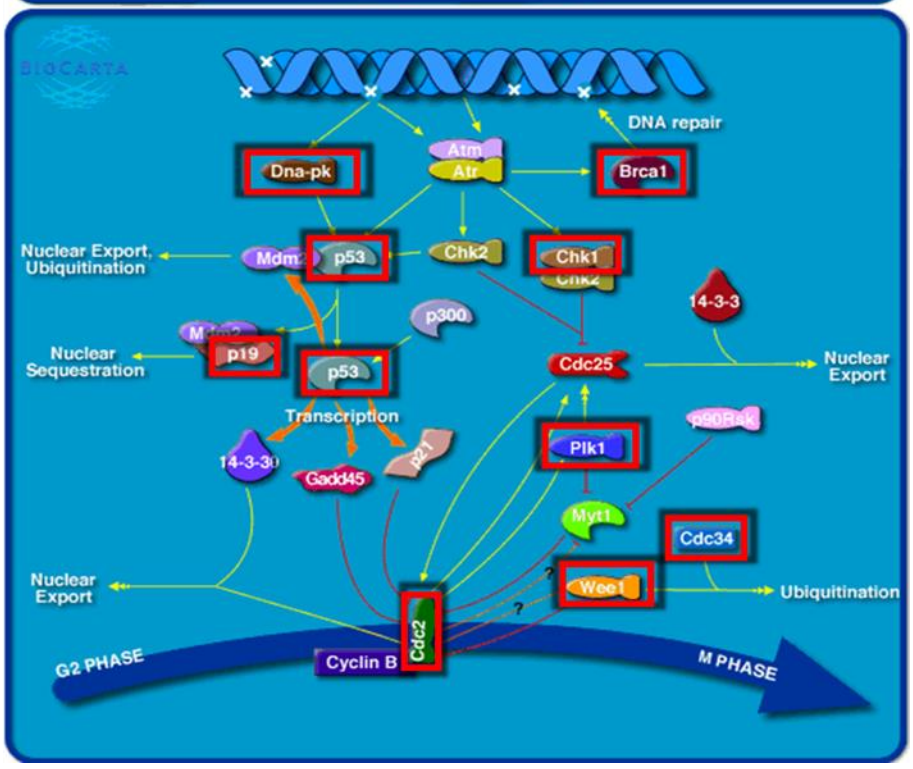


Figure 5.15

A.

KEGG Pathway	Fold Enrichment	p-value	Benjamini
Ribosome	**10.36**	**1.5E-70**	**2.6E-68**
Cell cycle	**4.30**	**4.1E-18**	**3.5E-16**
Pyrimidine metabolism	**4.87**	**1.8E-17**	**1.0E-15**
DNA replication	**7.19**	**1.8E-13**	**7.4E-12**
Purine metabolism	**3.36**	**6.7E-13**	**2.2E-11**
Aminoacyl-tRNA biosynthesis	**5.99**	**1.6E-11**	**4.5E-10**
RNA polymerase	**7.54**	**2.3E-11**	**5.6E-10**
Nucleotide excision repair	**4.46**	**8.8E-07**	**1.9E-05**
One carbon pool by folate	**7.49**	**1.2E-06**	**2.3E-05**
Ubiquitin mediated proteolysis	**2.47**	**1.5E-05**	**0.0002**
Mismatch repair	**5.45**	**3.3E-05**	**0.0005**
p53 signaling pathway	**2.78**	**0.0004**	**0.0060**
Progesterone-mediated oocyte maturation	**2.40**	**0.0015**	**0.0191**
Oocyte meiosis	1.98	0.0064	0.0739
Small cell lung cancer	2.11	0.0099	0.1050
Proteasome	2.55	0.0140	0.1376
Base excision repair	2.70	0.0156	0.1436
Porphyrin and chlorophyll metabolism	2.80	0.0344	0.2789
Pancreatic cancer	2.00	0.0346	0.2676
Cytosolic DNA-sensing pathway	2.18	0.0363	0.2667
Selenoamino acid metabolism	3.12	0.0376	0.2643
Drug metabolism	2.25	0.0427	0.2834
Gly, ser and thr metabolism	2.62	0.0457	0.2893
Glycerophospholipid metabolism	1.97	0.0491	0.2969
Val, leu and iso biosynthesis	4.36	0.0572	0.3271
Homologous recombination	2.66	0.0688	0.3691
Chronic myeloid leukemia	1.73	0.0978	0.4729
Ala, asp and glu metabolism	2.40	0.0993	0.4659

Figure 5.15 KEGG pathways identified as significantly enriched from the proteins upregulated in Eμ-myc tumours. Significantly ($p<0.05$), differentially expressed ($RS>0.5$) proteins in Eμ-myc tumours were analysed using DAVID to identify significantly enriched pathways. Red stars highlight those overexpressed proteins mapping to these pathways. **A.** A summary of the fold enrichments, p-values and corrected p-values for the top 28 KEGG pathways. **B.** Proteins mapping to the ribosome. **C.** Cell cycle pathway proteins. **D.** Pyrimidine nucleotide and nucleoside synthesis and metabolism.

Figure 5.15

B.

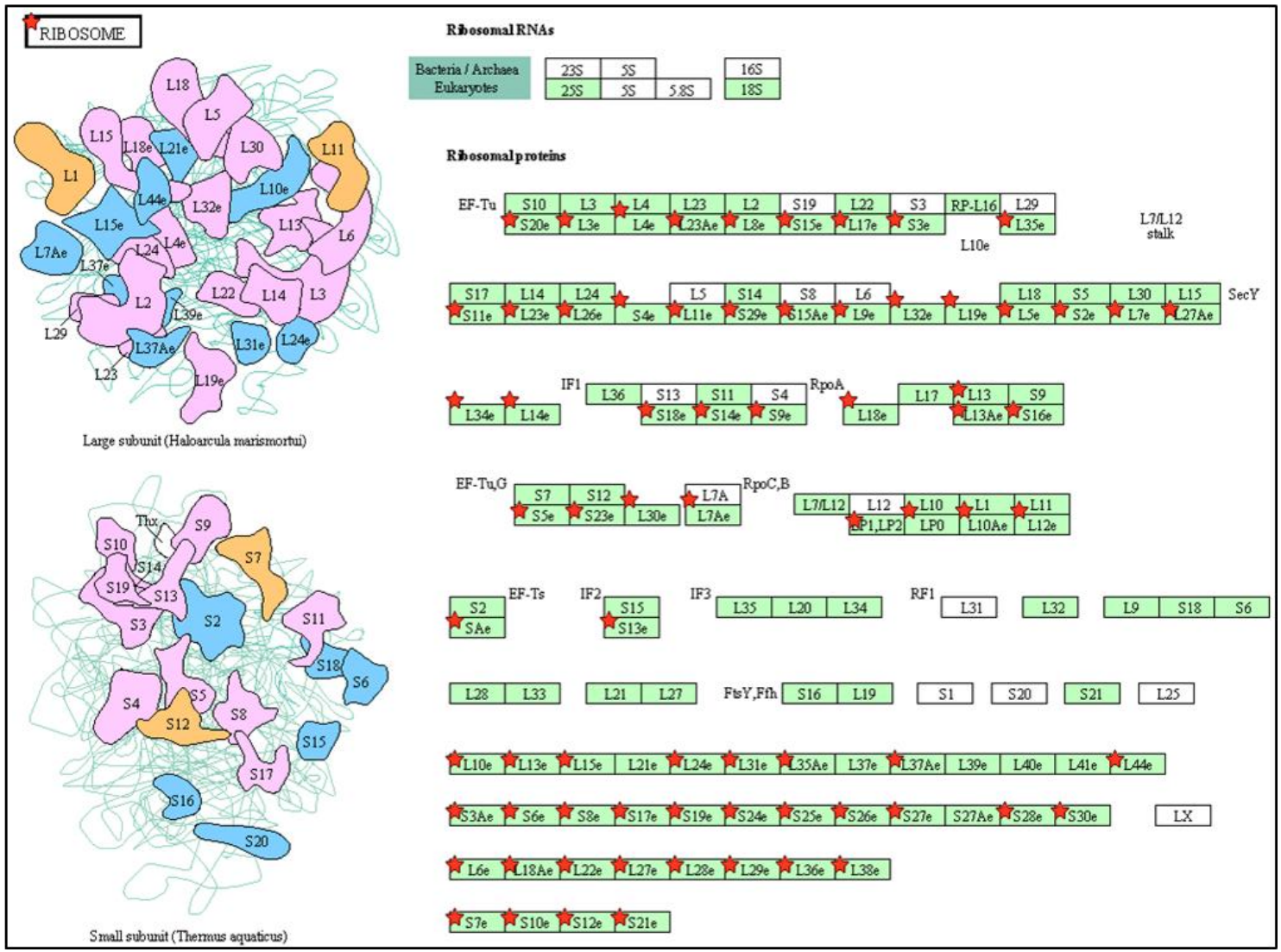

RIBOSOME
Large subunit (Haloarcula marismortui)
Small subunit (Thermus aquaticus)
Ribosomal RNAs
Bacteria / Archaea
Eukaryotes
23S 5S 16S
28S 5S 5.8S 18S
Ribosomal proteins
EF-Tu
L7/L12 stalk
L10e
SecY
IF1
RpoA
EF-Tu,G
RpoC,B
EF-Ts
IF2
IF3
RF1
FtsY,Ffh
LX

Figure 5.15

C.

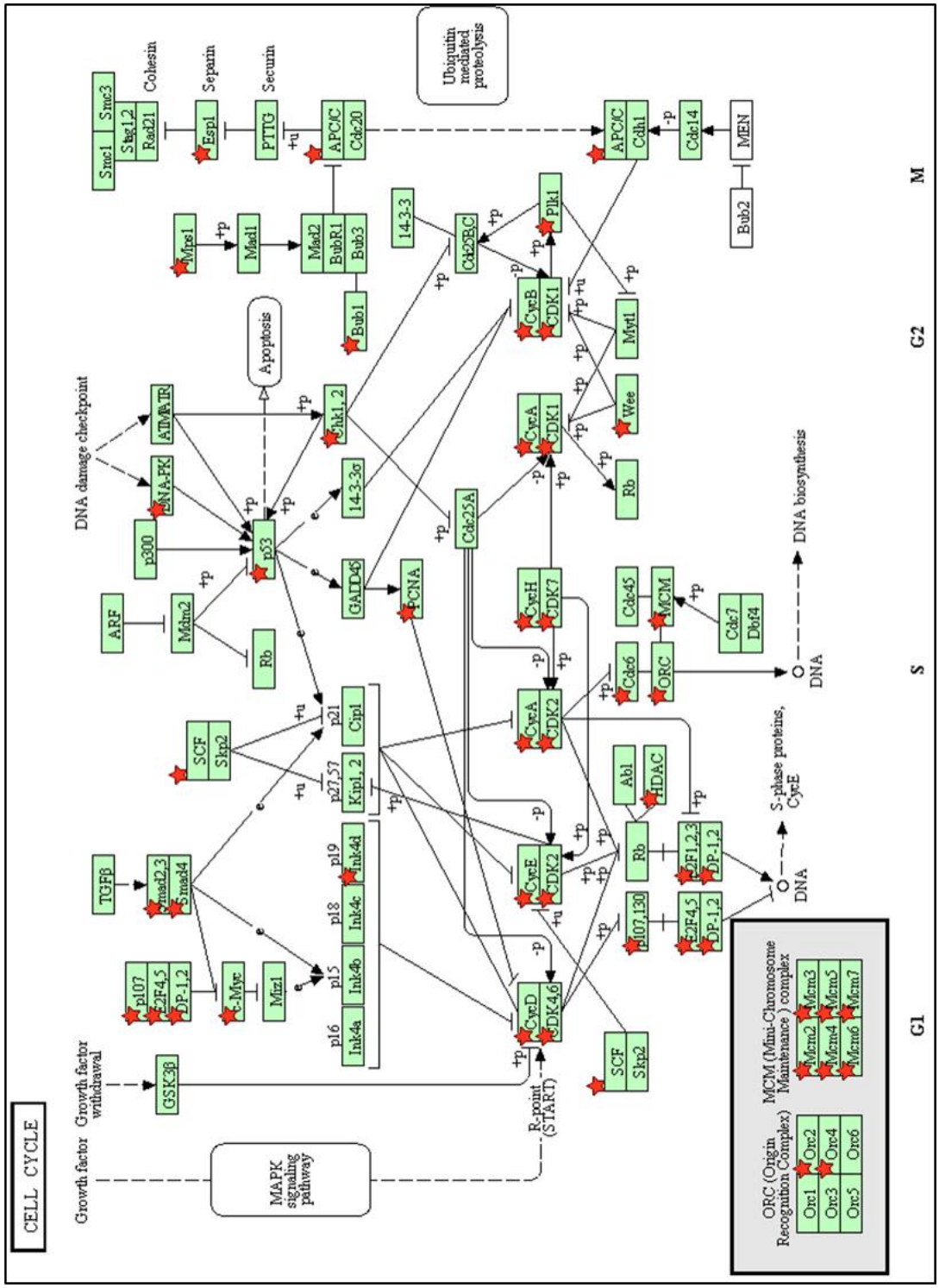

CELL CYCLE
G1
S
G2
M
Growth factor
Growth factor withdrawal
MAPK signaling pathway
GSK3β
TGFβ
Smad2,3
Smad4
c-Myc
Mxi1
p107
E2F4,5
DP-1,2
p16
Ink4a
p15
Ink4b
p18
Ink4c
p19
Ink4d
CycD
CDK4,6
R-point (START)
SCF
Skp2
p21
Cip1
p27,57
Kip1, 2
CycE
CDK2
CycA
CDK2
CycH
CDK7
CycA
CDK1
CycB
CDK1
Rb
E2F1,2,3
DP-1,2
E2F4,5
DP-1,2
p107,130
Abl
HDAC
S-phase proteins, CycE
DNA
Cdc45
MCM
Cdc6
ORC
Cdc7
Dbf4
DNA biosynthesis
DNA damage checkpoint
ATM/ATR
cHK-FK
p300
p53
ARF
Mdm2
Rb
Chk1, 2
14-3-3σ
GADD45
PCNA
Cdc25A
Apoptosis
Bub1
Mps1
Mad1
Mad2
BubR1
Bub3
14-3-3
Cdc25B/C
Plk1
Myt1
Wee
Rb
Ubiquitin mediated proteolysis
Cohesin
Smc1
Smc3
Stag1,2
Rad21
Separin
Esp1
Securin
PTTG
APC/C
Cdc20
APC/C
Cdh1
Cdc14
MEN
Bub2
ORC (Origin Recognition Complex)
Orc1
Orc2
Orc3
Orc4
Orc5
Orc6
MCM (Mini-Chromosome Maintenance) complex
Mcm2
Mcm3
Mcm4
Mcm5
Mcm6
Mcm7

Figure 5.15

D.

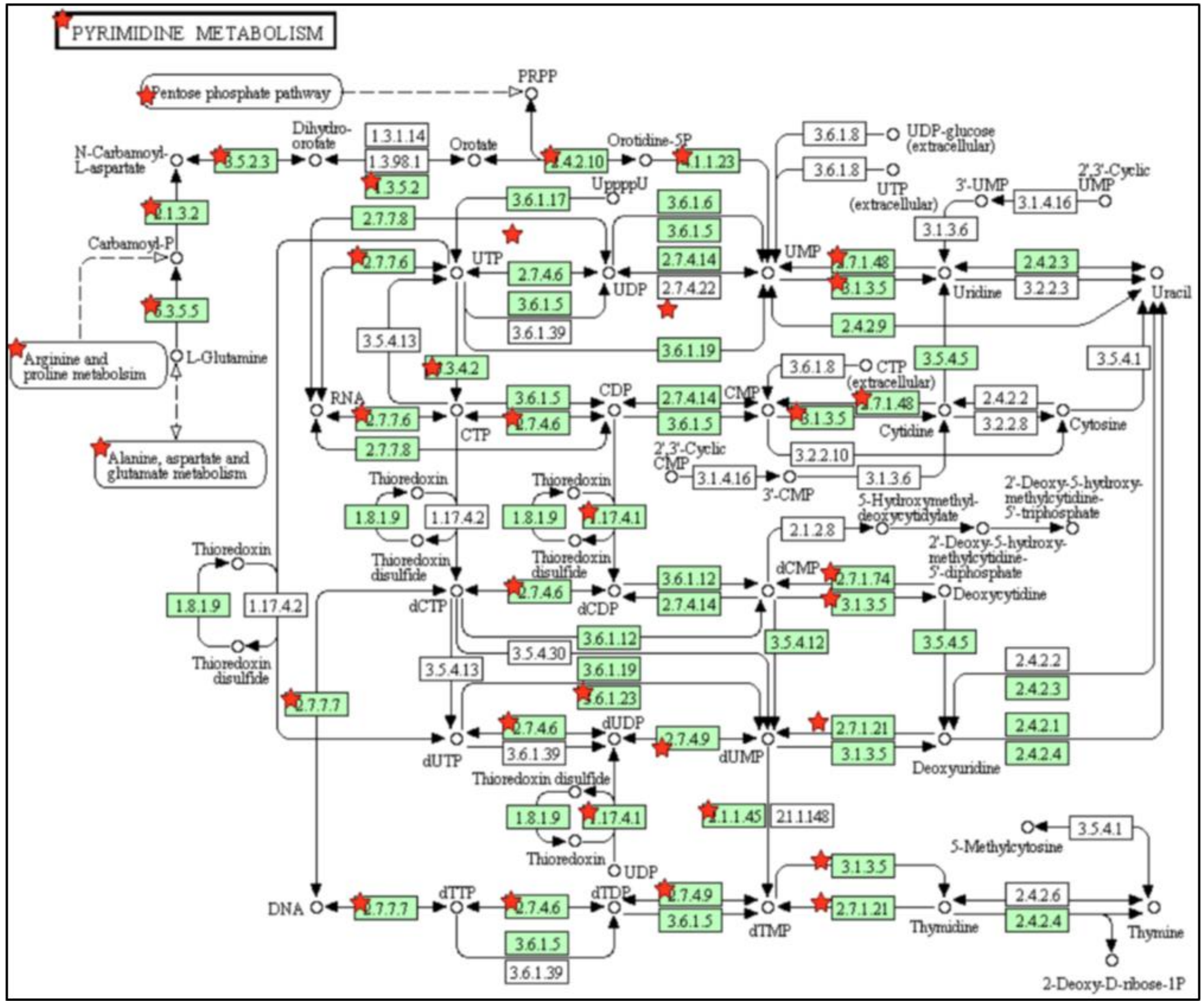

PYRIMIDINE METABOLISM
Pentose phosphate pathway
PRPP
N-Carbamoyl-L-aspartate
Dihydro-orotate
Orotate
Orotidine-5P
Carbamoyl-P
L-Glutamine
Arginine and proline metabolsim
Alanine, aspartate and glutamate metabolism
UDP-glucose (extracellular)
UTP (extracellular)
3'-UMP
2',3'-Cyclic UMP
UpppU
UTP
UDP
UMP
Uridine
Uracil
CTP (extracellular)
RNA
CTP
CDP
CMP
Cytidine
Cytosine
2',3'-Cyclic CMP
3'-CMP
5-Hydroxymethyl-deoxycytidylate
2'-Deoxy-5-hydroxy-methylcytidine-5'-triphosphate
2'-Deoxy-5-hydroxy-methylcytidine-5'-diphosphate
Thioredoxin
Thioredoxin disulfide
dCTP
dCDP
dCMP
Deoxycytidine
dUTP
dUDP
dUMP
Deoxyuridine
5-Methylcytosine
DNA
dTTP
dTDP
dTMP
Thymidine
Thymine
2-Deoxy-D-ribose-1P

Figure 5.16

A.

BIOCARTA	Fold Enrichment	p-value	Benjamini
RB Tumor Suppressor/Checkpoint Signaling in response to DNA damage	9.09	3.2E-05	0.0045
p53 Signaling Pathway	5.36	0.0033	0.2114
Cyclins and Cell Cycle Regulation	4.35	0.0035	0.1548
Estrogen-responsive protein Efp controls cell cycle and breast tumors growth	9.52	0.0055	0.1773
Cell Cycle: G1/S Check Point	4.00	0.0056	0.1465
Telomeres, Telomerase, Cellular Aging, and Immortality	4.76	0.0059	0.1300
Regulation of p27 Phosphorylation during Cell Cycle Progression	5.95	0.0069	0.1304
CDK Regulation of DNA Replication	8.16	0.0091	0.1500
Cell Cycle: G2/M Checkpoint	3.73	0.0175	0.2432
AKAP95 role in mitosis and chromosome dynamics	6.35	0.0198	0.2469
Cyclin E Destruction Pathway	6.35	0.0198	0.2469
Role of Ran in mitotic spindle regulation	6.35	0.0198	0.2469
E2F1 Destruction Pathway	5.71	0.0269	0.2963
ATM Signaling Pathway	3.97	0.0308	0.3098

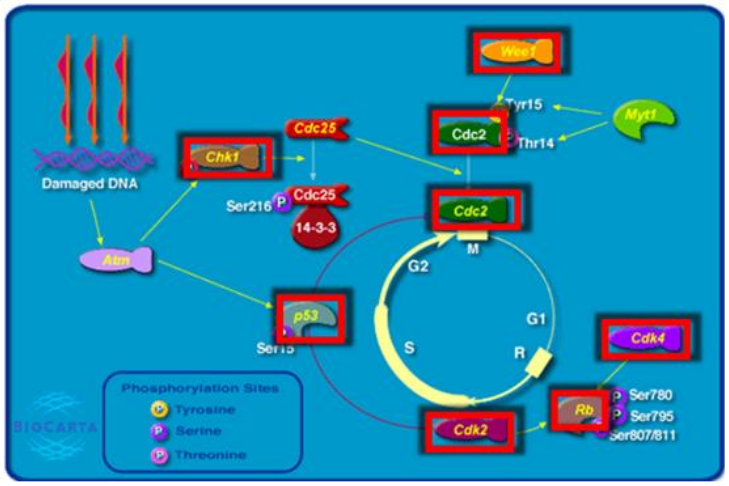
KEGG_PATHWAY			
Ribosome	4.34	3.4E-08	5.8E-06
Aminoacyl-tRNA biosynthesis	6.13	1.7E-07	1.4E-05
Cell cycle	3.45	2.6E-07	1.5E-05
DNA replication	5.78	1.0E-05	4.4E-04
Valine, leucine and isoleucine degradation	4.80	2.4E-05	8.2E-04
Pyrimidine metabolism	3.26	4.8E-05	0.0014
p53 signaling pathway	3.73	6.7E-05	0.0016
RNA polymerase	5.45	4.4E-04	0.0095
N-Glycan biosynthesis	4.00	6.7E-04	0.0126
Purine metabolism	2.11	0.0047	0.0768
Small cell lung cancer	2.60	0.0058	0.0865
Arginine and proline metabolism	3.12	0.0071	0.0969
Non-small cell lung cancer	3.07	0.0080	0.1001
Oocyte meiosis	2.24	0.0089	0.1038
One carbon pool by folate	5.75	0.0092	0.0997
Glycerophospholipid metabolism	2.47	0.0273	0.2563
Insulin signaling pathway	1.87	0.0356	0.3055
Apoptosis	2.11	0.0449	0.3535
Butanoate metabolism	2.98	0.0475	0.3549
Phosphatidylinositol signaling system	2.21	0.0489	0.3484

PANTHER_PATHWAY			
p53 pathway	2.89	2.8E-04	0.0305
p53 pathway feedback loops 2	3.78	4.6E-04	0.0250
De novo pyrimidine deoxyribonucleotide biosynthesis	6.73	0.0014	0.0494
De novo purine biosynthesis	3.05	0.0771	0.8876

Figure 5.16 Pathways identified as significantly enriched from the proteins upregulated in Eμ-TCL1 tumours. Significantly ($p<0.05$), differentially expressed ($RS>0.5$) proteins in Eμ-TCL1 tumours were analysed using DAVID to identify significantly enriched BioCarta, KEGG and Panther pathways. Red stars highlight those overexpressed proteins mapping to these pathways. **A.** A summary of the fold enrichments, p-values and corrected p-values for the top pathways identified by DAVID. **B.** Retinoblastoma tumour suppressor/checkpoint signalling in response to DNA damage. **C.** Proteins mapping to the ribosome. **D.** Cell cycle pathway proteins.

Figure 5.16

B.



C.

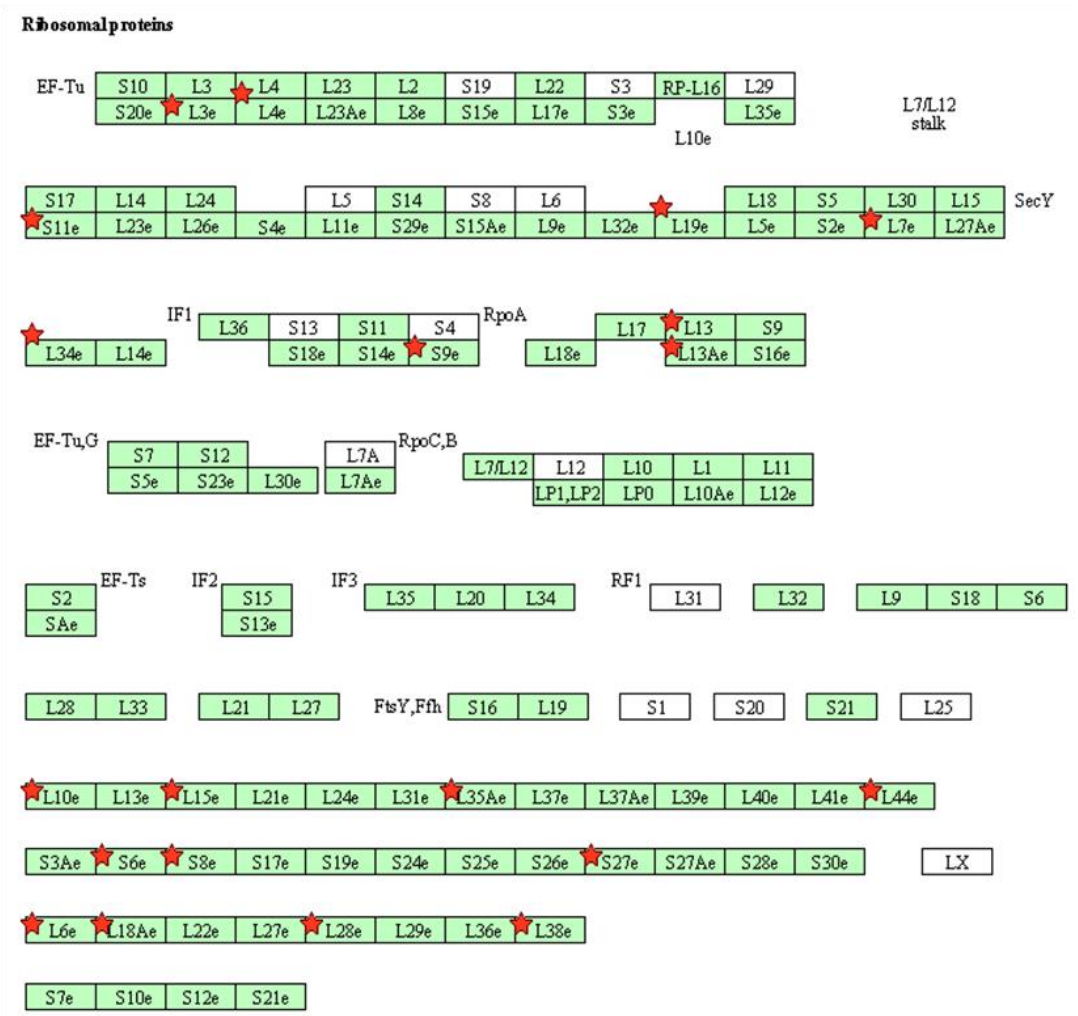
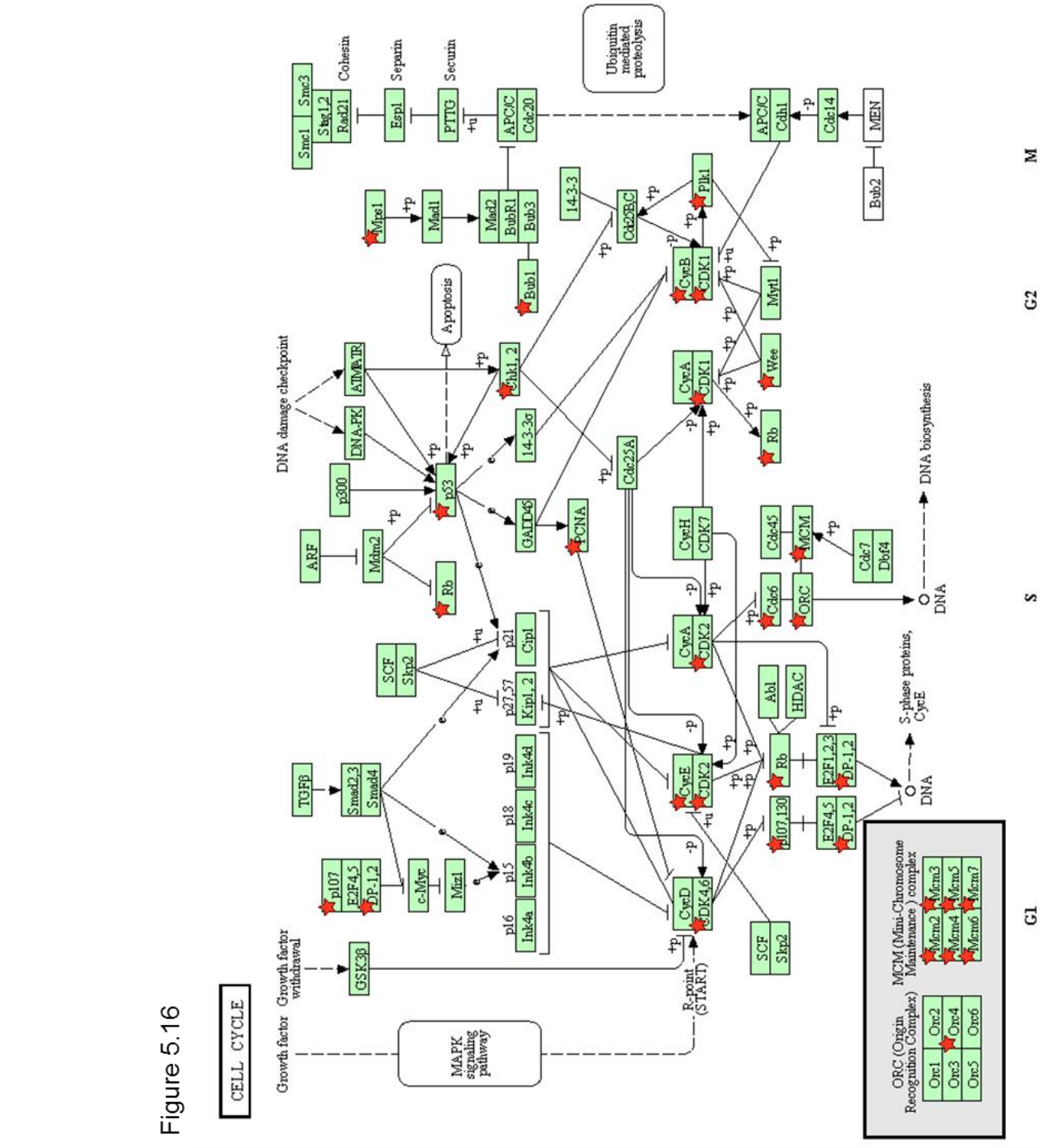


Figure 5.16

D.


CELL CYCLE
Growth factor
Growth factor withdrawal
MAPK signaling pathway
GSK3β
R-point (START)
TGFβ
Smad2,3
Smad4
p107
E2F4,5
DP-1,2
c-Myc
Miz1
p15
Ink4b
p16
Ink4a
p18
Ink4c
p19
Ink4d
CycD
CDK4,6
SCF
Skp2
p21
Cip1
p27,57
Kip1,2
CycE
CDK2
p107,130
E2F4,5
DP-1,2
E2F1,2,3
DP-1,2
Rb
Abl
HDAC
CycA
CDK2
CycH
CDK7
Cdc25A
PCNA
GADD45
p53
14-3-3σ
Chk1,2
ATM&TR
DNA-PK
DNA damage checkpoint
p300
ARF
Mdm2
Rb
Apoptosis
Mps1
Mad1
Mad2
BubR1
Bub3
Bub1
14-3-3
Cdc25BC
CycB
CDK1
Plk1
Myt1
Wee
Rb
CycA
CDK1
Cdc45
MCM
Cdc7
Dbf4
Cdc6
ORC
DNA
DNA biosynthesis
S-phase proteins, CycE
DNA
Ubiquitin mediated proteolysis
Smc1
Smc3
Scc1,2
Rad21
Cohesin
Esp1
Separin
PTTG
Securin
+u
APC/C
Cdc20
APC/C
Cdh1
-p
Cdc14
MEN
Bub2
ORC (Origin Recognition Complex)
Orc1
Orc2
Orc3
Orc4
Orc5
Orc6
MCM (Mini-Chromosome Maintenance) complex
Mcm2
Mcm3
Mcm4
Mcm5
Mcm6
Mcm7
G1
S
G2
M

5.13 ANNOTATION OF UPREGULATED PROTEINS IN Eμ-*MYC* AND Eμ-*TCL1* TUMOURS WITH PRE-EXISTING THERAPEUTIC REAGENTS

Overall, these observations demonstrate widespread systemic protein dysregulation of mapping to several functions and cellular pathways in the two tumours models. To understand the potential clinical relevance of the proteomics data produced here, proteins upregulated in either or both tumours were annotated using IPA software to describe the potential therapeutic reagents available which may be able to interfere with or target these proteins.

In total, 72 and 46 significantly upregulated proteins represented potential drug targets in the Eμ-*myc* and Eμ-*TCL1* tumours, respectively. 51 targetable proteins were significantly upregulated when considering both tumours. **Figure 5.17** presents those 28, 10 and 31 targetable proteins identified with at least 3 PSMs for Eμ-*myc*, Eμ-*TCL1*, and both tumours, respectively. This analysis highlighted the potential of the flavonoid alkaloid kinase inhibitor alvocidib, capable of inhibiting cyclin dependent kinases 1, 2 and 4 (CDK1, CDK2 and CDK4) in both tumours, as well as CDK6 specifically in Eμ-*myc*. Gemcitabine, a nucleoside analogue, was also observed with three potential targets in both tumours; ribonucleotide reductase M1 and M2 (RRM1 and RRM2), and DNA polymerase ε2 (POLE2). The additional target phosphoribosylglycinamide formyltransferase (GART) was also targetable with gemcitabine, primarily to tumours of Eμ-*myc* mice, but potentially also in those of Eμ-*TCL1* (RS=0.45). Thioguanine, another nucleic acid (nucleobase) analogue, was observed with two upregulated targets; phosphoribosyl pyrophosphate amidotransferase (PPAT) and inosine 5'-monophosphate dehydrogenase 2 (IMPDH2), with a third, ephrin receptor B4 (EPHB4), again with specificity to Eμ-*myc*. The two canonical aurora kinases AURKA and AURKB were identified as upregulated in both tumour types both targetable by inhibitors such as tozasertib and danusertib. Three histone deacetylases (HDACs) were identified as potential, upregulated drug targets in the two tumour types and annotated with inhibitors such as tributyrin, belinostat and vorinostat. HDAC8 was upregulated in both tumours, while HDAC2 and HDAC4 had a high degree of specificity to Eμ-*myc* and Eμ-*TCL1* tumours, respectively. Specifically to Eμ-*myc* tumours, three potential targets of a further nucleoside analogue, cytarabine, were upregulated; deoxycytidine kinase (DCK), DNMT1 and primase, DNA, polypeptide 2 (58kDa) (PRIM2).

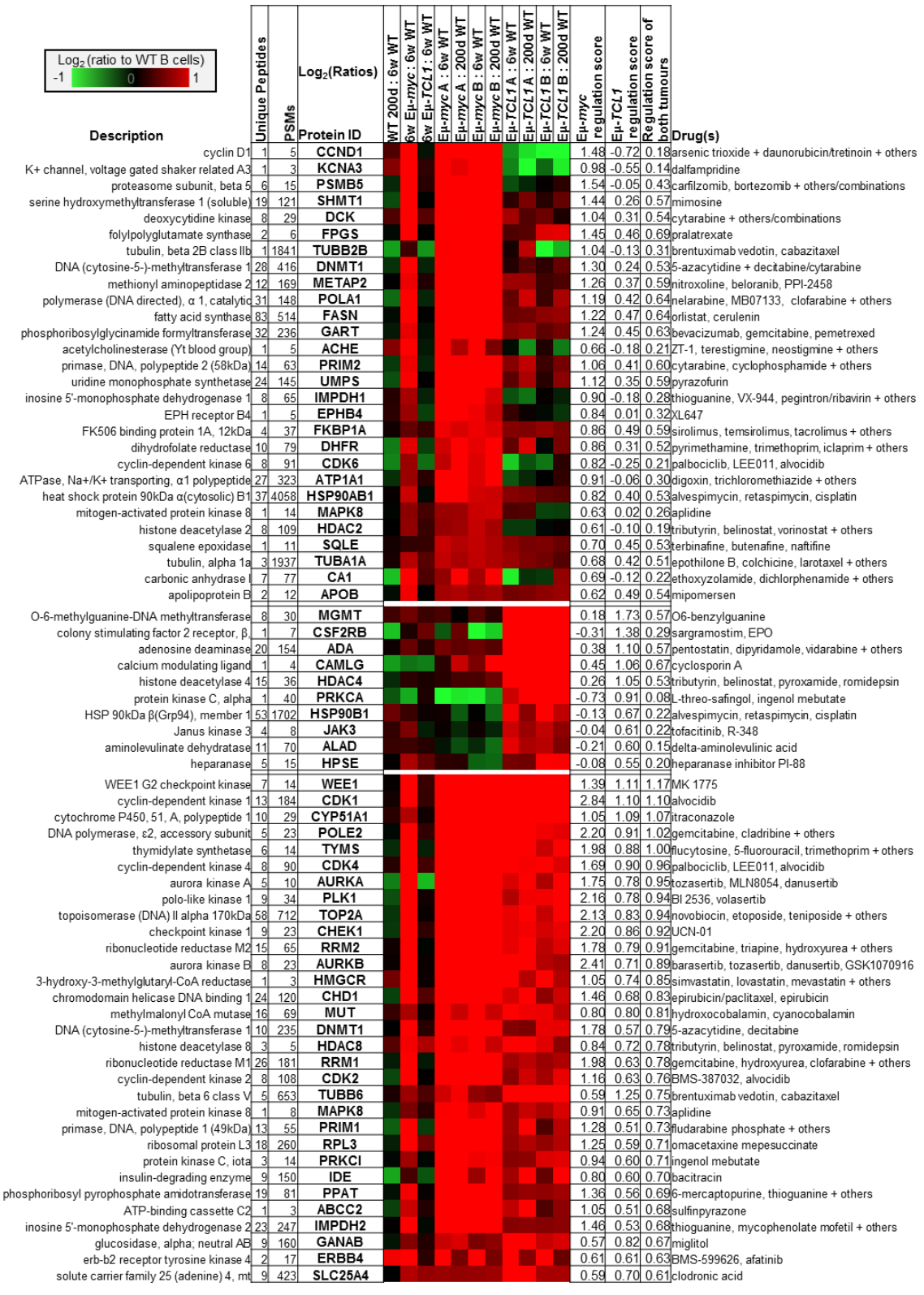


Figure 5.17. Proteins upregulated in either or both tumours with potential for therapeutic targeting. Proteins with IPA-annotated drug targets demonstrating a significant differential expression in either or both of the B-cell tumours. These were filtered for proteins with at least 3 PSMs.

5.14 Evaluation of Upregulated Proteins Annotated with Cell Surface Expression and Immunotherapy Potential

The expression of proteins at the surface of a cancer cell offers the potential for diagnosis and treatment and as such, surface protein expression was investigated by filtering all proteins on the basis of GO term annotation for cell surface expression. While the majority of these proteins were surface-expressed, no GO term exists to specifically describe transmembrane proteins as constitutively localised to the cell surface, therefore all membrane proteins were identified and those with cellular component GO terms for ER, mitochondrial or nuclear expression were excluded (detailed in full in **Section 2.21.6**). In total, 515 quantitated proteins were determined to have an annotation of surface expression. The regulation scores of these proteins were plotted against their $-\log_{10}$(p-values) calculated for each tumour type and for both tumours to represent the surface profiles of these cells (**Figure 5.18A**). A selection of the most confidently upregulated proteins with at least three PSMs are detailed for each tumour and for both tumours in **Figure 5.18B**.

Eμ-*myc* tumours presented 64 proteins with significant upregulation at the cell surface. Amongst these were 16 proteins annotated with receptor activity (CD163, EphB4, CD95/FAS, SLAMF7, CRCP, HMMR, INSR, ITGB1, ITGA6, IL1RAP, IL17RA, PTPRS, SLC20A1, SLC7A1, CD71/TFRC and TRPC6), 17 with transmembrane transporter activity (ABCC2, ATP1A1, ATP1B3, STEAP3, KCNK18, KCNA3, SLC3A2, SLC7A1, SLC39A11, SLC16A1, SLC20A1, SLC38A2, SLC39A6, SLC4A11, SLC35B2, MFSD10 and TRPC6) and 6 cell adhesion molecules (ITGB1, ITGA6, HMMR, PTPRS, VCAM1 and VEXT). The most upregulated surface protein was the zinc transporter SLC39A6/ZIP6, which was observed alongside another significantly upregulated zinc transporter SLC39A11/ZIP11. Secondary to this were the proteins; hyaluronan mediated motility receptor (HMMR), monocarboxylate transporter 1 (SLC16A1) and endothelin-converting enzyme 2 (ECE2). These observations also included the upregulation of two Na^+-exporting, K^+-importing transporters; alpha 1 and beta 3, and a combination of integrins (alpha 6 and beta 1) reported to bind laminin on platelets as well as neuropilin-2 (NPR2).

Of the 109 cell surface proteins downregulated on Eμ-*myc* tumours, 10 proteins had roles in cell adhesion; CD2, CD11a (ITGAL), CD22, CD31 (PECAM1), CD36, CD40, CD45 (PTRPC/B220), CD47, CD84 (SLAM5) and CD166 (ALCAM). Also with roles in cell-cell interactions were several MHC proteins; isoforms 1 and 2 of CD74, beta-2-microglobulin and 16 HLA isoforms (6 class I and 10 class II). 15 proteins with a function in the determination of haematopoietic cell lineages were downregulated, including the CD antigens; CD1d, CD2, CD5, CD13 (ANPEP), CD20, CD21 (CR2), CD22, CD23 (FCER), CD36, CD37, CD38, CD55 and

CD124 (IL4RA) and two HLA class II proteins. Lymphocyte antigen 6D (LY6D), a protein with a role at the earliest stage of B-cell specification, was the most downregulated, alongside CD272 (B-/T-lymphocyte attenuator (BTLA)), an inhibitor of immune cell activation.

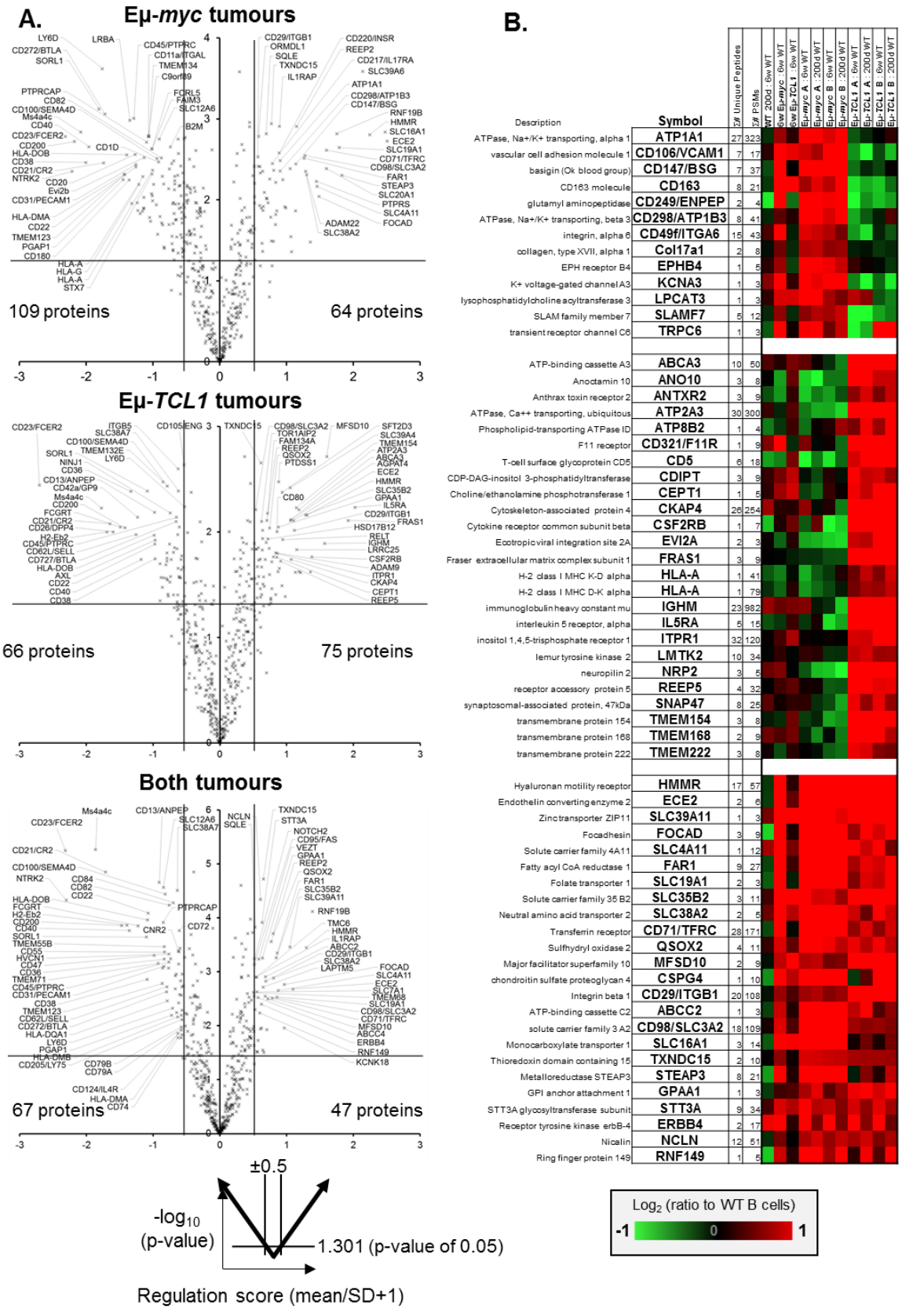
75 proteins exhibited upregulation at the cell surface of Eμ-*TCL1* tumours, with 21 common to Eμ-*myc* tumours. Again, transporter proteins were a common class, with 15 demonstrating upregulation; ZIP11, SLC19A1, SLC35B2, SLC38A2, ABCC2, SLC3A2 (CD98) and MFSD10 were also observed on the Eμ-*myc* tumours, with ABCA3, ATP2A3, ATP2C1, SLC2A1, ZIP4, MAGT1, ITPR1 and TFRC, being unique in their upregulation on Eμ-*TCL1* tumours. 15 upregulated surface proteins had receptor activity; HMMR, ITGB1 and CD71 (TFRC); also seen with the Eμ-*myc* tumours and; ATP9B, CD5, CD80, CD86, NOTCH2, RELT, ANTXR2, CSFR2RB, ITPR1, IL17RE, IL5RA and NPR2 specific to Eμ-*TCL1*. Seven proteins with annotated surface expression and a role in lipid biosynthesis were upregulated; 1-acyl-sn-glycerol-3-phosphate acyltransferase delta (AGPAT4), CD81, CDP-diacylglycerol inositol 3-phosphatidyltransferase (CDIPT), choline/ethanolaminephosphotransferase 1 (CEPT1), fatty acyl-CoA reductase 1 (FAR1), estradiol 17-beta-dehydrogenase 12 (HSD17B12) and phosphatidylserine synthase 1 (PTDSS1) with all but FAR1, specific to Eμ-*TCL1* tumours. Of these, the greatest specificity was seen in those proteins also relating to phospholipid synthesis. This may constitute erroneous annotation give that several of these proteins localise to the ER and golgi. Upregulated cell adhesion molecules included; CD80, CD86, CD321/F11R, ITGB1, two MHC proteins and CD43 (sialophorin/leukocialin). The most upregulated surface protein was major facilitator superfamily domain-containing protein 10 (MFSD10), a tetracycline transporter analogue with putative drug efflux capabilities. Secondary to this, was the interleukin 5 receptor alpha subunit, observed alongside its common beta cytokine co-receptor (CSF2RB/CD131), both with specific upregulation in the Eμ-*TCL1* tumours. The developmental extracellular matrix-interacting protein FRAS1 was also substantially, significantly upregulated with a high degree of specificity in Eμ-*TCL1* tumours.

Eμ-*TCL1* tumours had 66 downregulated surface proteins, of which 39 were also observed underexpressed in the Eμ-*myc* tumours. Cell adhesion molecules were also well-represented amongst these proteins, with 21 exhibiting a downregulation; CD2, CD22, CD36, CD40, CD45, CD47 and CD84 were downregulated as with Eμ-*myc* tumours; while additionally, CD11d (ITGAD), CD26 (DPP4), CD42a,c,d (GP9,1BB,5), CD48, CD49b (ITGA2), CD61 (ITGB3), CD62L (L-selectin), CD100 (Semaphorin-4D) and CD105 (endoglin) were downregulated with specifically to Eμ-*TCL1* tumours, alongside the non-CD proteins, synaptobrevin-3 (VAMP3), ITGB5 and the tyrosine kinase receptor AXL. 22 proteins were annotated with involvement in immune system process regulation including; ANO6, BTLA,

complement receptor 2 (CR2), cannabinoid receptor 2 (CNR2), CD23 (FCER), CD38, TYROBP, CD300a, FCGRT, CD55, CLEC2D, SLAMF7 and three class II HLA proteins, in addition to those also with a role in adhesion above; CD26, CD36, CD45, CD47, CD49b, CD61, CD62L and AXL.

9 surface proteins demonstrated inverse expression with significant upregulation in one tumour and significant downregulation in the other. Six proteins were observed upregulated in the Eμ-*TCL1* tumours which were downregulated in the Eμ-*myc* tumours; the predominantly intracellular ion channel anoctamin 10 (ANO10), sarcoplasmic/ER calcium ATPase 3 (ATP2A3), CD5, a putative receptor component; ecotropic viral integration site 2A (EVI2A), neuropilin-2 (NPR2) and the predominantly ER-localised regulator of inositol 1,4,5-trisphosphate receptor type 1 (ITPR1); E3 ubiquitin-protein ligase (RNF170). Three proteins were significantly up- and downregulated at the cell surface of Eμ-*myc* and Eμ-*TCL1*, respectively; CD163 (scavenger receptor cysteine-rich type 1 protein M130), K^+ voltage-gated channel A3 (KCNA3) and signalling lymphocytic activation molecule family member 7 (SLAM7).

Figure 5.18. Cell surface expression of proteins in Eμ-*myc* and Eμ-*TCL1* tumours. Fully quantitated proteins were filtered on the basis of GO annotated localisation to membranes (GO:0005886, GO:0005887, GO:0009986, GO:0031226, GO:0009897, GO:0016020 and GO:0016021) and subsequently filtered for any proteins also annotated with localisation to the endoplasmic reticulum (GO:0005783), mitochondria (GO:0005739), or nucleus (GO:0005634). This list was then combined with those proteins described by the MS surface atlas [368]. This resulting list was used to filter for those proteins characterised here and the remaining proteins filtered against the surfaceome database [369], a tool combining surface expression observations from several resources. In total, this process enriched for a list of 516 predominantly surface-expressed, transmembrane proteins. **A.** Volcano plots highlighting proteins with a strong statistical significance and magnitude of regulation. **B.** The most upregulated proteins in either and both tumours, filtered for those with at least 3 PSMs.

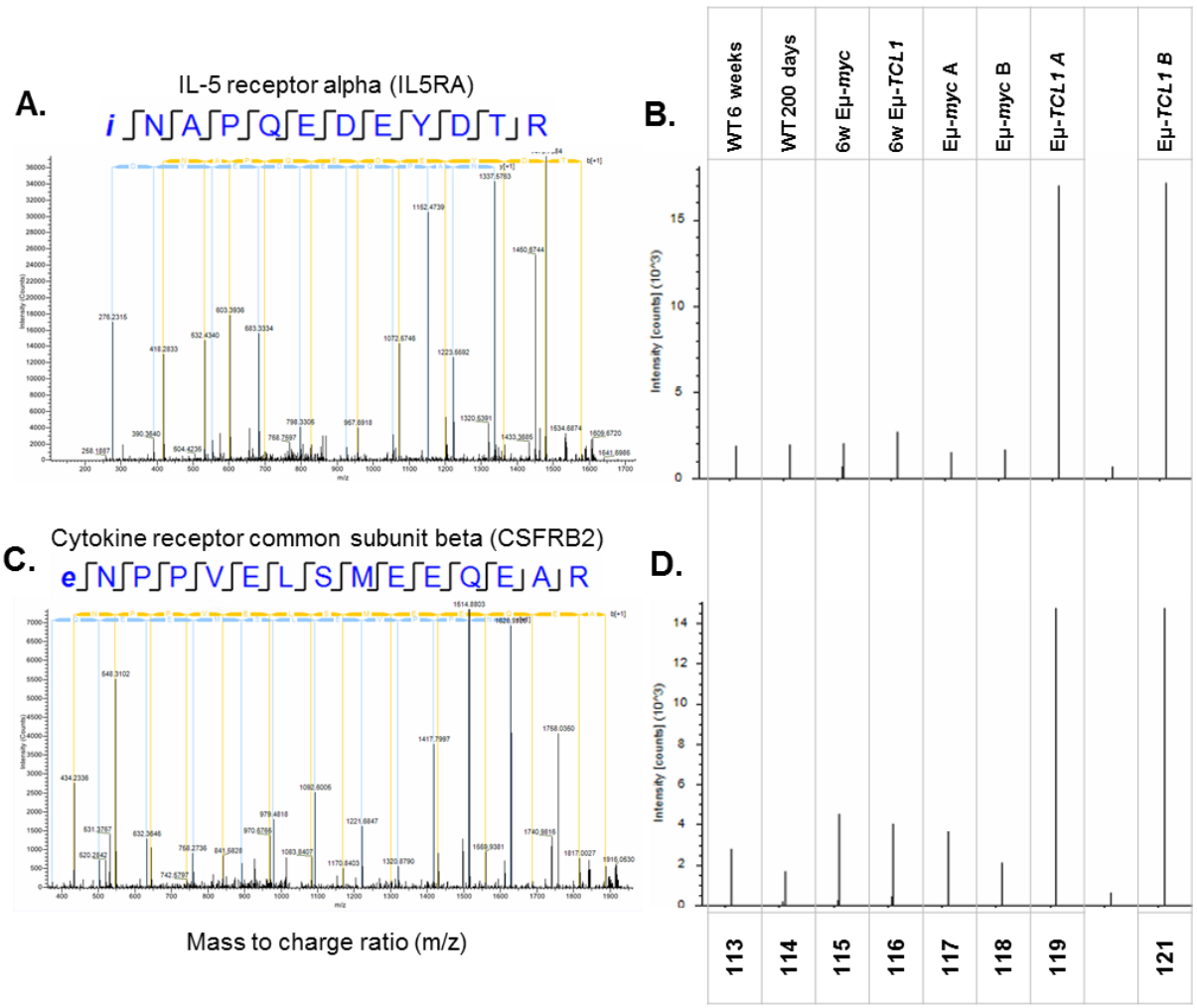

A.
Eμ-myc tumours
109 proteins
64 proteins
Eμ-TCL1 tumours
66 proteins
75 proteins
Both tumours
67 proteins
47 proteins
B.
Description
Symbol
Unique Peptides
PSMs
WT : 200d : 6w WT
6w Eμ-myc : 6w WT
6w Eμ-TCL1 : 6w WT
Eμ-myc A : 6w WT
Eμ-myc A : 200d WT
Eμ-myc B : 6w WT
Eμ-myc B : 200d WT
Eμ-TCL1 A : 6w WT
Eμ-TCL1 A : 200d WT
Eμ-TCL1 B : 6w WT
Eμ-TCL1 B : 200d WT
ATPase, Na+/K+ transporting, alpha 1 ATP1A1 27 323
vascular cell adhesion molecule 1 CD106/VCAM1 7 17
basigin (Ok blood group) CD147/BSG 7 37
CD163 molecule CD163 8 21
glutamyl aminopeptidase CD249/ENPEP 2 4
ATPase, Na+/K+ transporting, beta 3 CD298/ATP1B3 8 41
integrin, alpha 6 CD49f/ITGA6 15 43
collagen, type XVII, alpha 1 Col17a1 2 8
EPH receptor B4 EPHB4 1 5
K+ voltage-gated channel A3 KCNA3 1 3
lysophosphatidylcholine acyltransferase 3 LPCAT3 1 3
SLAM family member 7 SLAMF7 5 12
transient receptor channel C6 TRPC6 1 3
ATP-binding cassette A3 ABCA3 10 50
Anoctamin 10 ANO10 3 8
Anthrax toxin receptor 2 ANTXR2 3 9
ATPase, Ca++ transporting, ubiquitous ATP2A3 30 300
Phospholipid-transporting ATPase ID ATP8B2 1 4
F11 receptor CD321/F11R 1 9
T-cell surface glycoprotein CD5 CD5 6 18
CDP-DAG-inositol 3-phosphatidyltransferase CDIPT 3 9
Choline/ethanolamine phosphotransferase 1 CEPT1 1 5
Cytoskeleton-associated protein 4 CKAP4 26 254
Cytokine receptor common subunit beta CSF2RB 1 7
Ecotropic viral integration site 2A EVI2A 2 3
Fraser extracellular matrix complex subunit 1 FRAS1 3 9
H-2 class I MHC K-D alpha HLA-A 1 41
H-2 class I MHC D-K alpha HLA-A 1 79
immunoglobulin heavy constant mu IGHM 23 982
interleukin 5 receptor, alpha IL5RA 5 15
inositol 1,4,5-trisphosphate receptor 1 ITPR1 32 120
lemur tyrosine kinase 2 LMTK2 10 34
neuropilin 2 NRP2 3 5
receptor accessory protein 5 REEP5 4 32
synaptosomal-associated protein, 47kDa SNAP47 8 25
transmembrane protein 154 TMEM154 3 8
transmembrane protein 168 TMEM168 2 9
transmembrane protein 222 TMEM222 3 8
Hyaluronan motility receptor HMMR 17 57
Endothelin converting enzyme 2 ECE2 2 6
Zinc transporter ZIP11 SLC39A11 1 3
Focadhesin FOCAD 3 9
Solute carrier family 4A11 SLC4A11 1 12
Fatty acyl CoA reductase 1 FAR1 9 27
Folate transporter 1 SLC19A1 2 3
Solute carrier family 35 B2 SLC35B2 3 11
Neutral amino acid transporter 2 SLC38A2 2 5
Transferrin receptor CD71/TFRC 28 171
Sulfhydryl oxidase 2 QSOX2 4 11
Major facilitator superfamily 10 MFSD10 2 9
chondroitin sulfate proteoglycan 4 CSPG4 1 10
Integrin beta 1 CD29/ITGB1 20 108
ATP-binding cassette C2 ABCC2 1 3
solute carrier family 3 A2 CD98/SLC3A2 18 109
Monocarboxylate transporter 1 SLC16A1 3 14
Thioredoxin domain containing 15 TXNDC15 2 10
Metalloreductase STEAP3 STEAP3 8 21
GPI anchor attachment 1 GPAA1 1 3
STT3A glycosyltransferase subunit STT3A 9 34
Receptor tyrosine kinase erbB-4 ERBB4 2 17
Nicalin NCLN 12 51
Ring finger protein 149 RNF149 1 5
±0.5
-log10 (p-value)
1.301 (p-value of 0.05)
Regulation score (mean/SD+1)
Log2 (ratio to WT B cells)
-1
0
1

5.15 Eμ-*TCL1* Tumour-specific Upregulation of the Interleukin 5 Receptor

Of the observations made of upregulated cell surface proteins, the most biologically interesting finding related to the IL5 receptor. Both components of the interleukin 5 receptor, IL5RA and the common receptor subunit, CSF2RB, which also has receptor activity to IL3 and granulocyte-macrophage colony-stimulating factor, were strongly and consistently upregulated on both Eμ-*TCL1* tumour pools, as well as marginally upregulated in the pre-tumour B cells isolated at 6 weeks of age. Prior to the biological validation of this target, due to a relatively low number of quantitations for these proteins, the PSMs and quantitations were manually inspected to increase the confidence in these findings. **Figure 5.19A-D** summarises two PSMs and their respective iTRAQ quantitations for these proteins, demonstrating a clear, specific overabundance of iTRAQ reporter masses 119 and 121, corresponding to the two Eμ-*TCL1* tumour pools. **Figure 5.19E** shows all PSM-derived quantitation ratios; in all but a single case, a strong and reproducible trend of upregulation for the peptides of these proteins was observed.

Given the strength of this evidence, the expression of IL5RA was validated using flow cytometry (**Figure 5.20**). Peripheral blood from an advanced Eμ-*TCL1* mouse with an advanced tumour and a WT littermate was evaluated for the expression of the standard phenotyping markers B220 and CD5 to define the tumour population, as well as for IL5RA. Cells of the $CD5^+$ $B220^+$ tumour population corresponded to those exhibiting IL5RA expression. Three terminal, tumourous Eμ-*TCL1* splenocyte suspensions were evaluated by the same means, evaluated for peripheral leukaemic IL5RA expression. All three tumours demonstrated a high degree of staining for IL5RA.

Figure 5.19. iTRAQ 2DLC-MS/MS identification and quantification of peptides uniquely matching to IL5RA and it receptor partner CSF2RB. A. The CID PSM aligning to the IL5RA-specific peptide INAPQEDEYDTR (N-terminally modified with iTRAQ 8-plex), observed with a precursor m/z of 877.92230 (theoretical m/z, 877.92858). **B.** The respective, HCD-derived iTRAQ precursor region for this IL5RA peptide. **C.** The CID PSM aligning to the CSF2RB-specific peptide ENPPVELSMEEQEAR (N-terminally modified with iTRAQ 8-plex), observed with a precursor m/z of 1031.49597 (theoretical m/z, 1031.50689). **D.** The respective, HCD-derived iTRAQ precursor region for this CSF2RB peptide. **E.** The full list of individual PSM quantitations matching to IL5RA and CSF2RB.



E.

Log_2 (ratio to WT B cells)

-1 0 1

Protein/peptides	WT 200d : 6w WT	6w Eμ-*myc* : 6w WT	6w Eμ-*TCL1* : 6w WT	Eμ-*myc* A : 6w WT	Eμ-*myc* A : 200d WT	Eμ-*myc* B : 6w WT	Eμ-*myc* B : 200d WT	Eμ-*TCL1* A : 6w WT	Eμ-*TCL1* A : 200d WT	Eμ-*TCL1* B : 6w WT	Eμ-*TCL1* B : 200d WT
IL5RA	-0.32	-0.21	0.23	-0.21	0.10	-0.88	-0.64	2.28	2.70	2.35	2.77
iNAPQEDEYDTR	-0.40	-0.93	0.06	-0.49	-0.12	-1.55	-1.18	2.23	2.61	2.49	2.87
iNAPQEDEYDTR	0.16	-0.25	0.45	-0.89	-1.09	-0.95	-1.14	2.59	2.40	2.57	2.39
aTGLAQVLLHWDPNPDQEQR		0.29	0.75	0.67	0.64	-0.34	-0.37	2.24	2.21	2.05	2.03
iNAPQEDEYDTR	-0.54	-0.85	0.28	-0.69	-0.18	-1.45	-0.94	2.51	3.03	2.63	3.15
iNAPQEDEYDTR	-0.98	-0.62	-0.33	-0.63	0.32	-1.06	-0.10	1.62	2.57	1.69	2.66
dLPVVTEYEkPSNETk	0.39	0.11	0.19	0.40	0.03	-0.11	-0.53	2.69	2.27	2.51	2.10
hVDLEyHVk	-0.28	-1.24	0.37	-0.54	-0.30	-1.82	-1.57	2.68	2.93	2.59	2.85
dLPVVTEYEkPSNETk	-0.35	-0.17	-0.16	-0.10	0.23	-0.64	-0.31	-0.24	0.09	-0.35	
CSF2RB	-0.76	0.11	0.42	-0.34	0.41	-1.57	-0.81	1.82	2.58	1.49	2.25
eNPPVELSMEEQEAR	-0.82	-0.69	0.24	-1.14	-0.35	-3.06	-2.27	1.78	2.58	1.18	1.98
eNPPVELSMEEQEAR	-0.69	0.31	0.41	-0.21	0.45	-1.08	-0.42	1.78	2.44	1.75	2.41
eNPPVELSMEEQEAR	-0.75	0.42	0.56	-0.06	0.66	-0.64	0.09	1.40	2.13	0.94	1.67

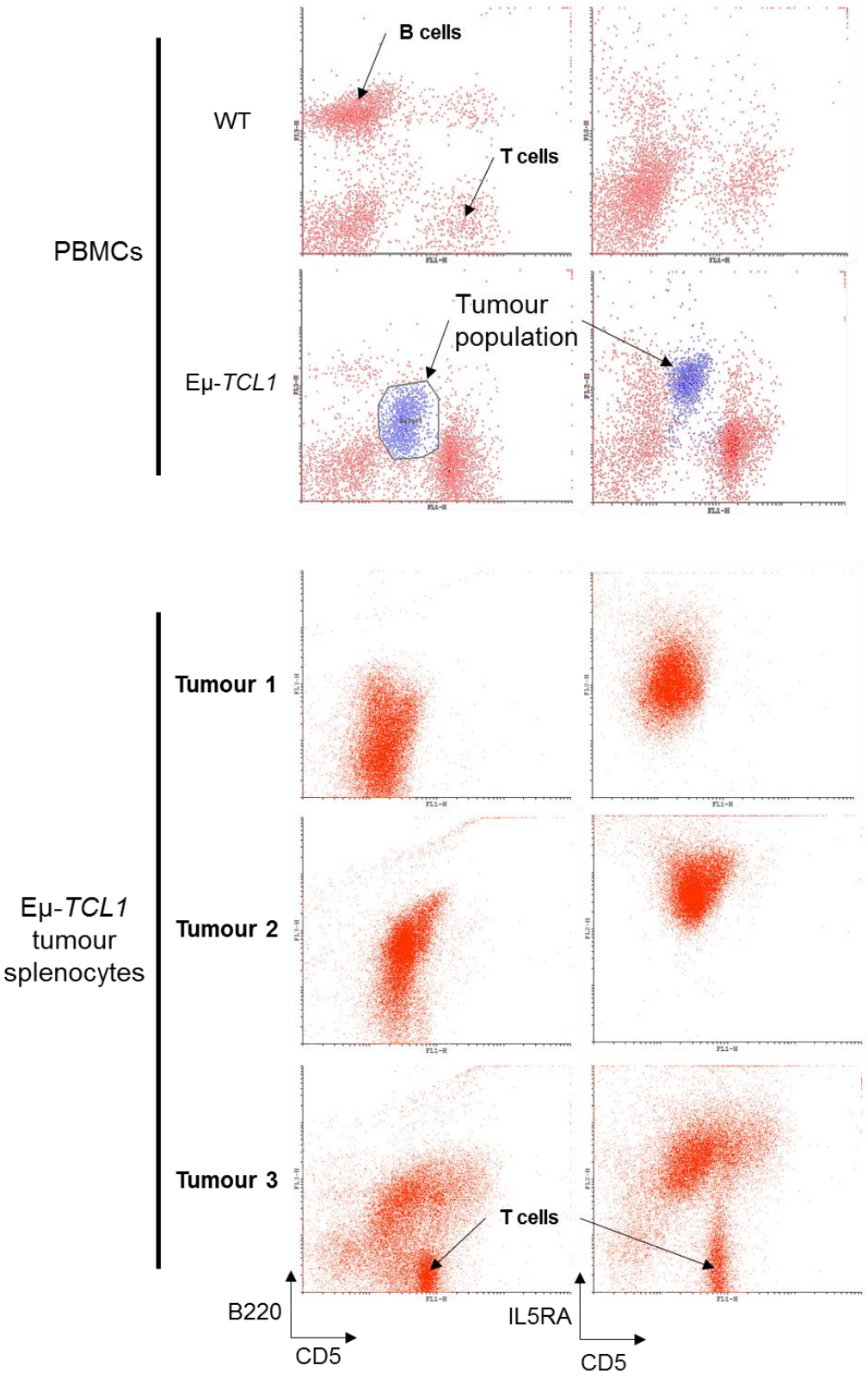


Figure 5.20. Flow cytometry validation of IL5RA expression on Eμ-*TCL1* tumours. The expression of CD5, B220 and IL5RA on peripheral blood mononuclear cells (PBMCs) and tumourous splenocytes from Eμ-*TCL1* mice.

5.16 INTERLEUKIN 5 INDUCES DOSE-DEPENDENT CELL SURVIVAL, CELL CYCLE PROGRESSION AND MITOSIS IN Eμ-*TCL1* TUMOURS *IN VITRO*

The observation of the substantial upregulation of the common beta subunit of the IL5 receptor, alongside IL5RA, suggested that the IL5 receptor may have a functional role in Eμ-*TCL1* tumours. To assess the inferred functionality of the IL5 receptor, tumourous splenocytes from two terminal Eμ-*TCL1* mice were first co-cultured with three concentrations (10, 50 and 100 ng/ml) of interleukin 5 and the cell density observed after 48 hours (**Figure 5.21**). A significant increase in cell number was observed for both tumours at all three concentrations of IL5 ($p<0.05$), with significance and average cell numbers increasing with higher doses. These data indicated a clear, dose-dependent proliferative effect of IL5 on the tumours of the Eμ-*TCL1* model.

To evaluate these observations in more detail, splenocytes were subjected to the same three doses of IL5 for 48 hours and were analysed by flow cytometry-based methods (**Figure 5.22**). Hypotonic propidium iodide staining was used to evaluate the cell cycle stage based on cell DNA content. Untreated Eμ-*TCL1* tumours had 4.9% of cells in S/G_2/M phases of the cell cycle. This increased to 6.9% upon treatment with 10ng/ml of IL5, and further to 8.8% at 50 ng/ml. Treatment with 100 ng/ml tripled that of the untreated rate of cell cycle progression to 14.5%. Also investigated was the percentage of viable cells, indicative of the rate of apoptosis under each treatment condition, determined using propidium iodide exclusion, where the intact membrane of viable cells prevents DNA staining. 17.9% of the untreated cells cultured *in vitro* for 48 hours, were undergoing apoptosis, a figure which fell to 9.9% upon treatment with 10ng/ml of IL5, and again to 8.8% at 50 ng/ml. Just 3.3% of cells were PI-permeable when 100 ng/ml of IL5 were included in the culture. These data indicate that IL5 promotes cell cycle proliferation in the Eμ-*TCL1* tumour B cells.

Finally, CFSE membrane staining was employed, which allowed tracking of the percentage of cells from this population which had undergone cell division. A similar, dose-dependent response was observed for this assay with approximately 10.9% of the untreated cells due to division; which increase to 23.7% upon treatment with 10ng/ml IL5 treatment. For the 50 and 100 ng/ml treatments, 42.5% and 43.8% were divided, respectively.

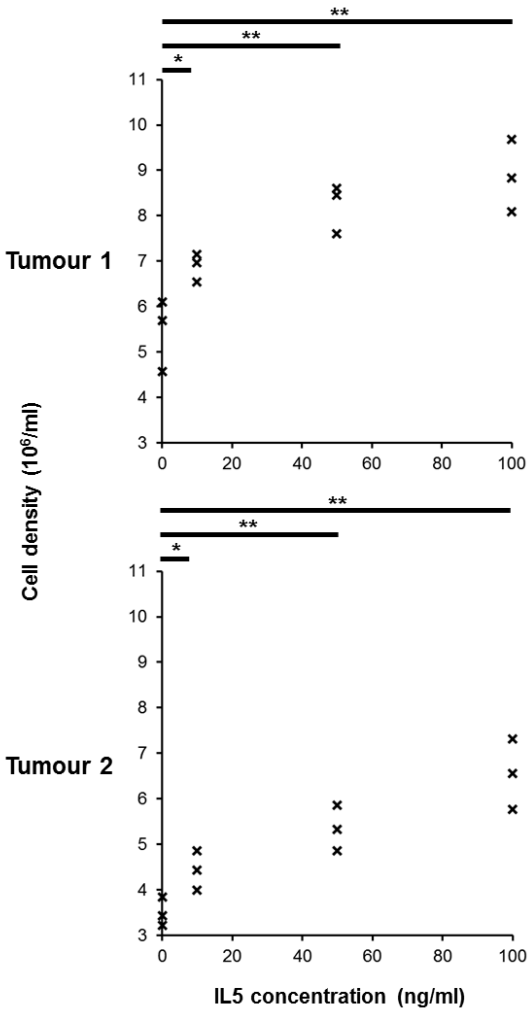
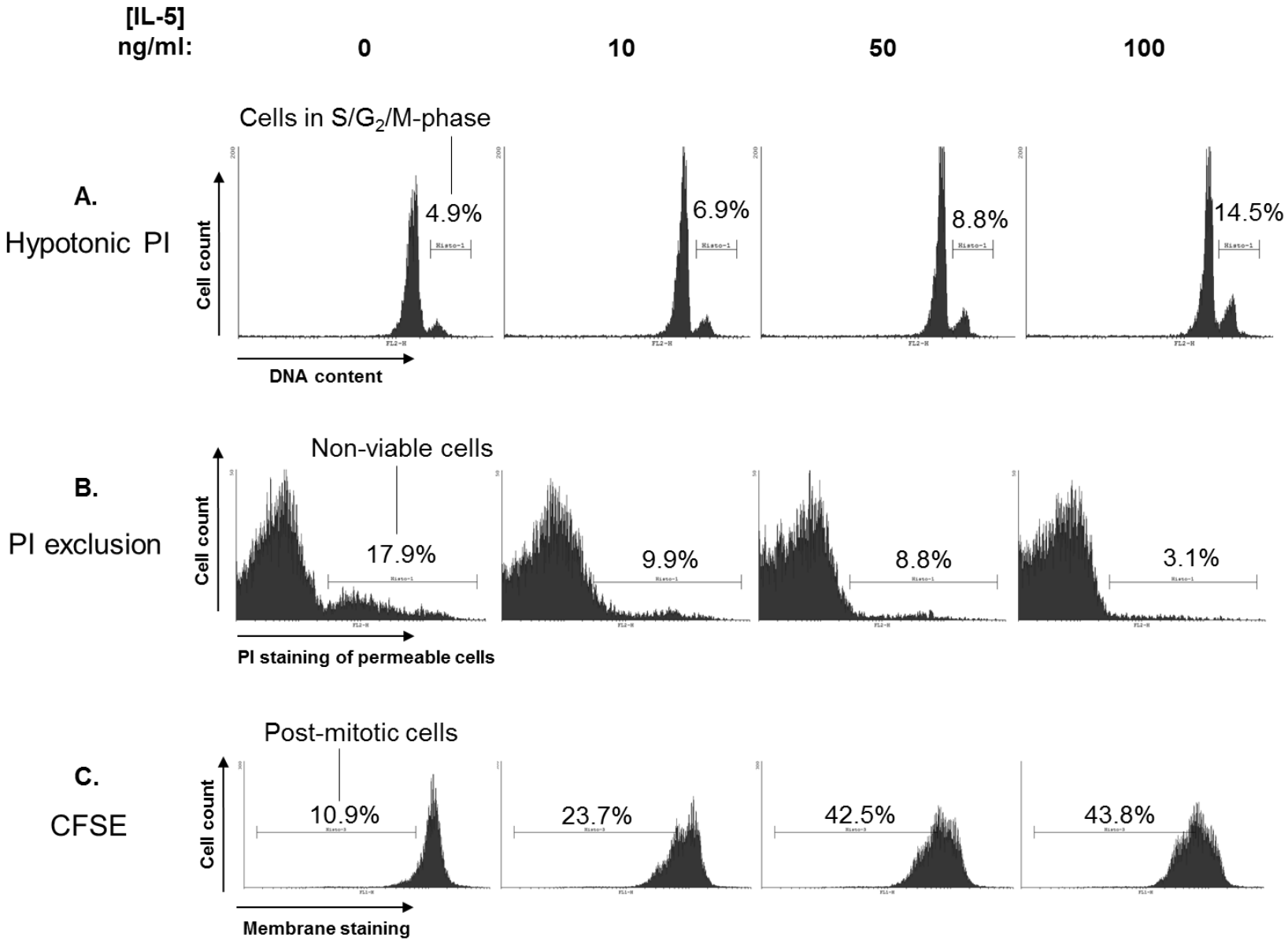


Figure 5.21. The effect of interleukin 5 on Eμ-*TCL1* tumour cell density *in vitro*. Eμ-*TCL1* tumour cells isolated by lymphoprep from the spleens of two terminal mice (corresponding to tumour 1 and tumour 2 evaluated in **Figure 5.20** with >95% tumour cell purity) were cultured at an initial density of $5x10^6$ and treated in triplicate with 0, 10, 50 or 100 ng/ml of IL5. Cell density was measured by a Coulter Counter after 48h. Significance was determined by t-test compared with the untreated (0 ng/ml) control. * - $p<0.05$, ** - $p<0.01$.

Figure 5.22. The effect of IL5 on proliferation and survival in Eμ-*TCL1* tumours *in vitro*. Tumourous Eμ-*TCL1* splenocytes were cultured for 48 h with 0, 10, 50 or 100 ng/ml of IL5. **A.** Cells were stained with hypotonic propidium iodide and DNA content measured with flow cytometry. **B.** Cells were stained with propidium iodide (isotonic) to measure exclusion by intact membranes indicative of cell viability. **C.** Prior to the culturing described above, tumourous Eμ-*TCL1* cells were stained with the fluorescent dye CFSE. The reduction in fluorescence induced by the physical division of the cell membranes was determined by flow cytometry.


[IL-5]
ng/ml:
0
10
50
100
A.
Hypotonic PI
Cells in S/G2/M-phase
4.9%
6.9%
8.8%
14.5%
Cell count
DNA content
B.
PI exclusion
Non-viable cells
17.9%
9.9%
8.8%
3.1%
Cell count
PI staining of permeable cells
C.
CFSE
Post-mitotic cells
10.9%
23.7%
42.5%
43.8%
Cell count
Membrane staining

5.17 CHAPTER DISCUSSION

This chapter aimed to combine and implement several of the latest advances in quantitative MS proteomics to comprehensively characterise the contrasting mouse B-cell cancer models Eμ-*myc* and Eμ-*TCL1*, observed with characteristics of BL and CLL, respectively. The resulting proteomic characterisation exceeded expectations raising the number of quantified proteins from 215 (peptide FDR<1%) in **Chapter 3** to 7391; a 34-fold increase. The combination of these approaches has provided one of the most comprehensive characterisation of primary B-cell cancers to date.

The samples characterised by this experiment were derived from Eμ-*myc* and Eμ-*TCL1* mice demonstrating phenotypes consistent with the initial tumour model characterisations [144, 176, 188]. The median time to presentation of terminal Eμ-*myc* tumours of 106 days, illustrated by the Kaplan-Meier survival curve in **Figure 5.1**, agreed closely with the generally accepted figure of 100 days [176, 188]. The median time to presentation for the Eμ-*TCL1* model of approximately a year, was slightly less than the figure of 13-18 months proposed by Bichi et al. [144]. It is however very closely in agreement with a later, more comprehensive characterisation performed by Johnson et al [145], suggesting a median age of death of 11.93 months. Survival times in the Eμ-*TCL1* model, for example, may be influenced by environmental factors, such as exposure to antigens potentially playing a role in tumour progression [392]. Stress to these animals is anecdotally observed to increase the progression of the disease. However, with each model following the expected phenotypes, such institute-specific factors were unlikely to have had a substantial influence on the MS-characterised tumours.

Biological replicates demonstrated reproducibility and clustering amongst the 7391 proteins for the Eμ-*myc* and Eμ-*TCL1* models (**Figure 5.4**). This topologically confirms the success of this quantitative experiment. The trend also implies that both models presented with reproducible differential protein expression in their respective tumours. The averaging effects of pooling cannot be ruled out as a factor in this, however.

The pattern of protein expression in Eμ-*myc* 6-week samples was surprisingly similar to that of the Eμ-*myc* tumours. While no tumours or signs of sickness were observed in any of the 6-week Eμ-*myc* mice, it was possible that in one or more of the mice had progressed to the early stages of lymphoma, potentially contaminating the analysis with tumour material. The use of pooling, however should have substantially reduced this observed trend. Given that the 6-week Eμ-*myc* 'pool A' had double the cellularity of pool B (**Figure 5.3**), even if all 100 million of the additional cells were tumourous, this would only have contributed to 25% of the protein

analysed after pooling. The SD of differential expression for 6-week Eμ-*myc* B cells was however 80% that of the terminal Eμ-*myc* tumours. This extent of differential protein abundance could not have been attributable to tumour contamination unless this contamination was present in at least 5 of the 6 samples. Therefore it remains possible that this protein expression was partly attributable to all the 6-week Eμ-*myc* B cells. The observation of myc protein upregulation in this sample however (**Figure 5.5**), is inconsistent with the literature suggesting myc protein expression is not elevated prior to tumour development [377], confirmed by WB in **Figure 3.6**. However, **Figure 5.9** does present some specific changes in protein expression, including further myc upregulation, between the 6-week Eμ-*myc* B cells and terminal, malignant Eμ-*myc* B cells which could therefore be specifically attributable to a malignant phenotype. Overall, this suggests that the pre-tumour Eμ-*myc* B-cell protein expression may very well be very similar to that which emerges during malignancy. Given the uncertainty, detailed conclusions relating to the 6-week Eμ-*myc* B cells were not drawn.

Contrasting the 6-week Eμ-*myc* protein expression profile; the 6-week Eμ-*TCL1* B cells demonstrated a minimal extent of dysregulation, even less than that of the 200 day WT B cells. This reflects the reduced aggression of the Eμ-*TCL1*-derived cancer and suggested that TCL1, even when substantially overexpressed (**Figure 5.5**), has little effect on the immediate phenotype of B cells.

Expected myc and TCL1 protein overexpression was detected in both the Eμ-*myc* and Eμ-*TCL1* terminal tumours, relative to both the WT B cells and respective non-model controls (**Figure 5.5**). In addition to these expected proteins, several other B-cell and B-cell cancer related proteins were observed with anticipated regulations. The WB conducted on Eμ-*myc* tumours in **Chapter 3** also served as validation for this proteome. These observations acted as positive controls, indicating that the 2D LC MS/MS workflow, the data analysis workflow and the implementation of SPIQuE had successfully derived an accurate representation of the characterised cell lysates. This represented the first characterisation of either myc or TCL1 by global proteomics in these respective models.

Myc overexpression in the Eμ-*TCL1* tumours, due to the lack of a myc human transgene, could only be traceable to murine myc upregulation, suggesting a role for myc and myc dysregulation as a downstream component of TCL1-induced signalling or as a bystander effect of lymphomagenesis. This is potentially expected given the activation effect of TCL1 on the mitogenic AKT protein [211, 212] which can facilitate the inhibition of myc degradation via glycogen synthase kinase 3 beta (GSK3β) phosphorylation [393]. CD19 signalling has been shown to have a BCR-independent role in myc upregulation via PI3K-directed AKT activation in lymphoma [183, 184]. CD19 expression is maintained in Eμ-*TCL1* tumours, compared to

several other downregulated B-cell surface components (**Figure 5.5A**), and multiple PI3Ks are upregulated. These observations makes it difficult to conclude a predominant mechanism of AKT-directed myc stabilisation, but suggest a convergence of both pathways. It is also possible that several other mitogenic pathways such as BCR signalling-induced RAS/RAF/MEK signalling influence the upregulation of myc. The genetic deletion of CD19, which induced longer survival in Eμ-*myc* mice [183], in the context of the Eμ-*TCL1* model may inform the importance of TCL1 activation of AKT in myc stabilisation in Eμ-*TCL1* tumours by eliminating this potential source of PI3K signalling. The lack of myc upregulation in 6-week Eμ-*TCL1* B cells suggests that other oncogenic events are required for the effective amplification of myc expression.

Upregulation of p53 correlating with myc expression (**Figure 5.5A**) was an expected observation given the p21 ARF myc-p53 feedback pathway. It also agreed with previous observations of p53 upregulation [394]. This suggests that myc expression in both models inhibits MDM2 allowing p53 to accumulate. Unfortunately, neither p19ARF nor MDM2 were identified by the proteomics, however, MDM2 binding protein (MTBP), a stabilising factor of MDM2 exhibited upregulation correlating with myc and p53 upregulation. While in theory MTBP overexpression would promote MDM2-directed degradation of p53, p19ARF expression likely counteracted this. In addition, a secondary p53 ubiquitin ligase (HUWE1), also inhibited by p19ARF [395], was upregulated correlating with myc expression. The upregulation and stabilisation of two E3 ubiquitin ligases of p53 degradation, correlating with p53 expression, suggest a possible feedback mechanism. Given that loss of functionality of the downstream effects of p53 typically results from mutations, it was difficult to derive any conclusions from the proteomics data as to how the tumours avoided p53-induced apoptosis.

Global analysis revealed a far greater extent of protein dysregulation in the Eμ-*myc* tumours, compared with Eμ-*TCL1* tumours (**Figure 5.6**). This was an anticipated observation given the contrast in phenotypes between the models and aggression of the oncogenes. The observation of common signatures between the two models was less expected, however, for these same reasons.

For those proteins exhibiting upregulation in both myc- and TCL1-driven B-cell tumours, the vast majority were far more upregulated in Eμ-*myc* tumours, highlighted by **Figure 5.7**. This suggests that while a common B-cell cancer protein signature may exist, the signature may be dependent on other factors. One such factor which correlates with these patterns of greater protein expression in Eμ-*myc* tumours is, indeed, myc itself, a likely candidate given the vastly pleiotropic nature of this transcription factor.

The most confidently upregulated protein across both tumours was ASS1 (**Figure 5.7**). This was an interesting observation on the basis that arginine dependence is observed in several cancer types where ASS1 is downregulated [396-399]. ASS1 expression has also been shown to correlate with the aggression of tumours, where upregulation confers treatment resistance and poor prognosis [400, 401]. ASS1 expression has been shown to be regulated by myc in a melanoma cell line, demonstrating that the ASS1 promoter region contains a binding site for myc [402]. This suggests that ASS1 upregulation in these B-cell tumours may be a side effect of myc overexpression rather than a necessary feature of tumourigenesis. The correlation between ASS1 and myc reinforces the suggestion that myc may be playing a proportional role in both tumour phenotypes. Unfortunately a large scale evaluation of myc-specific genes was not possible due to the current unavailability of the myc-target gene database [162].

Amongst the other proteins highlighted by **Figure 5.7**, there was a strong trend of proteins with roles relating to the hallmarks of cancer (**Figure 1.1**) [1, 2] and with published links to cancers. Even the proteins amongst this list with no clear links to cancer were implicated in oncogenic processes. This strongly reinforces the suggestion that this proteomics experiment provides an unprecedented characterisation of these B-cell cancers, offering potentially, hundreds of novel hypotheses relating to Eμ-*myc* tumours, Eμ-*TCL1* tumours and other cancer types.

To elucidate trends from the tumour data besides the evaluation of single protein regulations, enrichment analyses were employed which highlighted, for example, the enrichment of proteins involved in processes relating to an increased cell metabolism (**Figure 5.12**). Given the nature of these terminal tumours, this was an expected observation with protein synthesis, DNA replication, organelle fission required for increased cell division. Processes specifically related to the promotion of cell cycle progression were also enriched for. This trend of cell proliferation and protein metabolism was reinforced by several pathways relating to these processes, highlighted by **Figures 5.14**, **5.15** and **5.16**.

The enrichment of upregulated mitochondrial proteins suggested the mechanism by which energy was being derived by the tumour cells to maintain cell growth and metabolism. Greater numbers of mitochondria or mitochondrial proteins might suggest that these tumours have a lesser dependence on the ‘Warburg effect’, a description of cancer cells adopting mitochondrial-independent aerobic glycolysis [403]. However, lactate dehydrogenase A, the enzyme responsible for the mitochondria-independent ATP synthesis from pyruvate, and transmembrane monocarboxylate (lactate) exporters were also substantially upregulated, suggesting an interplay between these two means of energy production. Overall this agrees with the hallmark describing metabolic dysregulation in cancer [1, 2] and suggests some specific

means by which it may be occurring. The increase number of mitochondrial proteins may also be due to myc induction of metabolic proteins in response to rapidly dividing cells.

Downregulated protein trends predominantly indicated immune evasion with the enrichment of underexpressed MHC complex proteins and terms such as ‘signal transduction’ and ‘plasma membrane’ also suggesting a reduced interaction with the exterior of cells. This downregulation of interaction with the immune system and extracellular environment is consistent with the hallmark of cancer describing immune system evasion [1, 2] and previous observations of MHC downregulation in cancers [404, 405]. The downregulations also suggests a trend towards pathways and proteins that may be responsible for the loss of contact inhibition by intercellular signalling [1, 2]. These GO terms were also consistent with the loss of B-cell-specific attributes such as the downregulations described in **Figure 5.5**. This agrees with the proposed concept of de-differentiation especially in the context of myc expression, a pluripotency factor and inhibitor of terminal differentiation [406, 407]. As B cells progress to cancers, it would make sense that metabolism and protein synthesis of B-cell-specific attributes would be evolutionarily lost to favour cells with proliferative potential. This was not so much the case with Eμ-*TCL1* tumours, especially for BCR components supporting suggestions of the involvement of BCR signalling in proliferation and development [408].

Evaluation of the drug targets confirmed the efficacy of the proteomics approach in the identification of potential targets of small molecular inhibitors of cancer pathways, with several known targets observed for both tumours and each tumour type (**Figure 5.17**). The predominance of drug targets relating to DNA synthesis and cell cycle progression confirms the general trend of cells maintaining a proliferative state, but may also have reflected the predominance of drug development in areas such as nucleoside analogues [409, 410]. Given that the standard therapies for both BL and CLL frequently incorporate nucleoside analogues or DNA synthesis inhibitors, such observations are expected [411]. Another well-represented class of drug were kinase inhibitors, again, related to the cell cycle in the context of cyclin-dependent kinase and aurora kinase inhibitors. Such drugs offer the potential to interfere with the core mechanisms of cell proliferation, including those described in the pathways in **Figures 5.14**.

Histone deacetylase inhibitors (HDACIs) offer an alternative means of cancer treatment. While currently pan-HDACi are used (**Figure 5.17**), the observation of HDAC8 upregulated in both B-cell cancer types suggests a potentially highly specific means of therapeutic intervention. HDAC8 is a ubiquitously expressed HDAC with a role in cancer survival suggested by the reduced *in vitro* proliferation resulting from HDAC8 siRNA knockdown in hepatocellular carcinoma [412]. Several oncogenic mechanisms appear to be enhanced by HDAC8, such as telomerase activity [413], tumour suppressor downregulation [414, 415] and cytokine signalling

[416], either by the deacetylation of histone lysine residues as a transcriptional regulator or by the direct deacetylation of lysine residues on non-histone proteins [417, 418]. The exact effects of HDAC8 inhibition by several of the specific inhibitors [419-421] is unclear as they do not appear to have been examined specifically in B-cell cancers. The observation of upregulation in both the 6-week model samples, additionally implies that myc and TCL1 may have a regulator role in the upregulation of HDAC8, however, no published links are apparent. HDAC2 and DNA methyltransferase 1, specific to Eμ-*myc* tumours and HDAC4, specific to Eμ-*TCL1* tumours, suggests differential roles for epigenetic regulation between these two tumours.

While only currently annotated drugable targets are described in **Figure 5.17**, several other proteins such as kinases identified by proteomics may suggest targets of small molecular inhibitors, such as those described in **Figure 5.7**.

Membrane protein analysis identified several dysregulated mechanisms at the cell surface in both tumours (**Figure 5.18**). Despite the GO term enrichment detailed in **Figure 5.11**, suggesting an overall downregulation of surface proteins, a selection of receptors and transmembrane transporters were still upregulated. A strong signature common to both tumours was that of solute carrier channels involved in the transport of zinc, folate, monocarboxylates (lactate/pyruvate), neutral amino acids, organic anions and large neutral amino acids. This trend describing an increased transmembrane transport of compounds was consistent with the observations of increased cell metabolism (**Figure 5.10**). In order for translation and DNA synthesis to take place, prior to cell division, an uptake of amino acids and nucleotide precursors is required. These data suggests that a mechanism by which this is occurring is through the upregulation of transporters proteins. This mechanism is well characterised in the context of the glucose transporter [422], monocarboxylate transporters [423, 424] and amino acid transporters [425-427], offering inhibition targets capable of impeding tumour proliferation in several tumour types.

The observation of two zinc transporters ZIP6 and ZIP11 upregulated in Eμ-*myc* tumours and ZIP4 and ZIP11 in Eμ-*TCL1* tumours suggested a greater requirement for zinc in these tumour cells. Zinc transporter upregulation is less reported and less consistent than those transporters above, with some reports of downregulation [428-431]. Instances have been observed however with upregulation of the zinc transporter ZIP5 which appeared to increase cell proliferation and migration, reduced upon knockdown [432]. ZIP11 has been reported with a possible role in bladder cancer based upon variant-related correlations with prognosis [433]. ZIP4 was shown to dramatically increase the aggression of pancreatic cancer [434], which has been suggested to function via a zinc-dependent zinc finger transcription factor [435]. This finding suggests that zinc may play a role in the maintenance of B-cell tumour growth and could

provide a therapeutic target. Supplementation or withdrawal of zinc from *in vitro* Eμ-*myc* tumour cultures might provide further insight into the dependency of B-cell cancers upon Zn^{2+} ions for transcription factors as well as other zinc-dependent proteins, an interesting example of which is HDAC8.

The most upregulated cell surface protein across both cancers was HMMR, also known as CD168, which is a receptor for the extracellular matrix component hyaluronic acid. CD168 is overexpressed in several malignancies, including B-cell cancers, such as CLL where it has been suggested to have prognostic utility, correlating with aggressive disease [436]. While CD168 has been rejected as an immunotherapy target in acute myeloid leukaemia [437], observations in other cancers suggests the potential of anti-CD168 therapies and its value in prognostic use [438-442]. This analysis presented several other surface proteins potentially worthy of investigation for biological and therapeutic relevance to B-cell cancers.

The combined, Eμ-*TCL1*-specific upregulations of both interleukin 5 receptor subunits, IL5RA and CSF2RB (**Figure 5.19** and **Figure 5.20**), strongly implied that IL5 has a role in signal transduction in Eμ-*TCL1*-derived B-cell cancers. This hypothesis was supported by previous observations of murine B-cell cancers responding to IL5: A cell line, BCL1, derived from a spontaneous murine B-cell leukaemia with CLL-like characteristics [443], was observed with a dependency on IL5 for proliferation [444, 445]. An IL5-dependent follicular B-cell lymphoma line, cRCS-X, was also described [446]. Additionally, a mouse strain with constitutive expression of IL5 demonstrated the emergence of a $CD5^+$ CLL-like B-cell leukaemia [447].

Murine splenic B cells were previously investigated for IL5RA expression, revealing an $IL5RA^+$ subpopulation of 2-4% [448]. This work also determined a strong link between IL5RA and IgM surface expression, a co-expression also observed on Eμ-*TCL1*-derived cancers. IL5 has also been described with a role in inducing proliferation of murine B cells, previously known as B-cell growth factor II [449, 450]. A deficiency of a B-cell subset described as $CD5^+$ B-1 cells was also observed in a mouse knock out model of IL5 [451]. Taken together these observations strongly supported the potential for a role for IL5 in the development, proliferation and survival of Eμ-*TCL1* tumours, from $CD5^+$ B-1-cell precursor cells.

Figure 5.21 identified a dose-dependent increase in cell number in Eμ-*TCL1*-derived tumour splenocytes upon treatment with IL5. While this suggested that proliferation was occurring, further assays were conducted to evaluate the specific effects of IL5 upon cell division, cell cycle phase and cell viability (**Figure 5.22**). Overall, these assays demonstrated a dose-dependent response to IL5, promoting cell survival, cell cycle entry and cell division. This

confirms that, like with the other murine B-cell tumours described above, IL5 signalling plays a role in the proliferation of Eμ-*TCL1* tumours. IL5 signalling may also be critical to the development of Eμ-*TCL1* tumours.

The above observations have several implications for the Eμ-*TCL1* tumour model. The response to IL5 appears to function by downstream signalling predominantly via the intracellular domain of CSF2RB which can induce activation of Lyn, JAK2, Syk, BTK, NF-κB and PI3K [452-454] which demonstrates considerable crossover with BCR signalling. Amongst these pathways, the only proteins demonstrating upregulation similar to that of the IL5R were members of PI3K family. While this doesn't rule out signalling via other pathways, the selective upregulation of these PI3Ks could suggest a selective advantage of signalling induced by these molecules.

Phosphoinositide-3-kinase, regulatory subunit 6 (PIK3R6) and phosphatidylinositol-4-phosphate 3-kinase, catalytic subunit type 2 beta (PIK3C2B) both presented as candidates for this signalling, with highly specific upregulation in Eμ-*TCL1* tumours (RS>1) and no regulation in any other samples. Phosphoinositide-3-kinase, regulatory subunit 5 (PIK3R5) also followed this specific expression but was less upregulated (RS>0.5). As discussed previously, PI3K signalling induces AKT which is amplified by TCL1 overexpression. It follows that the action of TCL1 might therefore occur partly though the amplification of the IL5RA signalling pathway, potentially inducing the expression of myc (described in **Figure 5.5**) promoting aberrant cell growth and tumourigenesis. This hypothesis is partly supported by the recent observation of PI3K inhibitors exhibiting antitumour effects in Eμ-*TCL1* tumours [230]. These observations also suggest that the IgM^+ $IL5RA^+$ B cells [448] is a very likely B-cell subtype from which Eμ-*TCL1* tumours originate.

The data suggest Eμ-*TCL1* tumours may offer more potential for *in vitro* study when supplemented with IL5, such as those previously characterised tumours [443, 446] to enable proliferation and survival of these cells, which may simply be recapitulating the *in vivo* conditions. Further work would need to be conducted to identify the concentration required to best reproduce the *in vivo* growth rate *in vitro*.

It was considered, given the CLL-like nature of Eμ-*TCL1* tumours, that IL5 may be involved in a similar manner in CLL. However, flow cytometry did not identify any upregulation of IL5RA relative to healthy B cells in CLL and IL5 treatment did not appear to induce proliferation or survival in CLL cells and, in fact, has previously been observed to induce spontaneous apoptosis in CLL [455].

The evidence regarding the effects of IL5 is conflicted between humans and mice, with reports that suggest IL5 has no effect on human B cells [450, 456]. The role of IL5 and potentially the IL5RA-expressing B-cell subset does not appear to be conserved across species, however does appear to be conserved with regards to eosinophil proliferation and differentiation [450, 456].

The observation of an alternative pathway and cell of origin driving Eμ-*TCL1* tumour development suggests that the model may have limitations in the recapitulation of CLL. However, given that the pathways activated appear to have several parallels to BCR signalling, the model may still hold potential for furthering understanding of many aspects of the human disease.

This chapter aimed to generate high-quality quantitative proteomics data describing the tumours and pre-tumours of the Eμ-*myc* and Eμ-*TCL1* B-cell cancer models. Emphasis was placed upon using the use of the latest advances in proteomics, such as the use of DMSO in the mobile phase, to maximise the potential of identifying differential protein expression relevant to the understanding of these models, and potentially other B-cell cancers. Overall, these results represent the most comprehensive phenotypic characterisations of these models to date. While limitations such as sample pooling existed, this was overcome by the use of replicate pools and the rejection of variable observations. This has resulted in the novel characterisation of a cytokine signalling pathway driving the proliferation of Eμ-*TCL1* tumours. Additionally, given the vastly different presentation and aetiology of the Eμ-*myc* and Eμ-*TCL1* B-cell tumours, there is a strong possibility that many of the molecular characteristics common to both tumours, may have commonality with other B-cell cancers.

6.0 QUANTITATIVE PROTEOMIC CHARACTERISATION OF Eμ-*MYC* AND Eμ-*TCL1* BLOOD PLASMA

6.1 CHAPTER INTRODUCTION

In addition to those samples collected for the characterisation of the B-cell proteome of the Eμ-*myc* and Eμ-*TCL1* models, detailed in **Chapter 5**, blood plasma was simultaneously isolated for proteomics.

Plasma proteomics offers the potential to identify biomarkers indicative of the presence, as well as progression, of diseases, due to the minimally-invasive nature of sample collection. In cancer, mechanisms such as cell lysis, exosome production, secretion and shedding can allow tumour-derived proteins to be detectable in the blood plasma and provide signatures of the presence and/or stage of a cancer [316]. Immune response to cancer can also be detected in the plasma providing early, non-invasive indications of the state of a cancer. Other systemic responses and the microenvironment also likely contribute to protein presence in the blood [316].

Two major obstacles have hindered the identification of robust biomarkers for cancer diagnosis and prognosis from serum and plasma. Firstly, the methods of detection of proteins by approaches such as proteomics is hampered by the disproportionate abundances of certain

proteins in plasma and secondly the collection of well controlled samples is challenging in pre-symptomatic stages of a disease.

In an attempt to overcome some of these issues with plasma protein identifications, an adapted version of SuPrE-SEC, described in **Section 2.18**, was employed. The use of the Eμ-*myc* and Eμ-*TCL1* mouse models offered an opportunity to study disease progression with a model in which cancer development was both predictable but also spontaneous. This offered a degree of control over the characteristics and sample quality which would be very challenging to obtain from human samples.

This chapter describes the characterisation of plasma collected from the same mice for which the tumour and B-cell proteomes are described in **Chapter 5**. An additional cohort was also used to generate intermediate stage plasma samples from Eμ-*TCL1* mice, where a leukaemia was established but not yet terminal. Characterising both the cells and plasma of these tumours offered an opportunity to gain an insight into the complex interactions between the impacts of these tumours on the whole organism, as well as the systemic effect of the organism on the tumours. Furthermore, correlations between the B-cell cancers and blood plasma proteomes had the potential to reveal a mechanism and explanation behind how these markers were appearing in the blood.

6.2 DEFINING AND COLLECTING PRE-TERMINAL, INTERMEDIATE-STAGE Eμ-*TCL1* PLASMA

Plasma samples were collected in tandem with those derived for B-cell characterisation in **Chapter 5** using an adapted method (detailed in **Section 2.12**), drawing heparinised blood directly from the inferior vena cava. Plasma was then isolated by centrifugation, to ensure the removal of any cellular material. The effects of an advanced terminal tumour on plasma were substantial enough to be observed by eye, with a reduced haematocrit and clear discolouration evident (**Section 5.2**). For the Eμ-*myc* model, given that substantial protein dysregulation was observed at the 6 week, pre-tumour period, these samples were capable of providing insight into the Eμ-*myc* model prior to terminal tumour development. For the Eμ-*TCL1* model, however, due to the far slower progression, combined with the minimal dysregulation of proteins at 6 weeks, it was concluded that the characterisation of an intermediate state would be more informative. An intermediate state also more accurately reflects the clinical stage at which CLL is frequently diagnosed. An additional condition was therefore considered for the Eμ-*TCL1* plasma samples, with material derived when a $CD5^+$ $B220^+$ blood-borne leukaemic population was initially established. Based on the characterisation of $CD5^+$ $B220^+$ populations in WT and Eμ-*TCL1* mice (**Section 5.2**), it was concluded that a threshold of at least 15% above that of the WT $CD5^+$

$B220^+$ percentage at any monthly screening, to account for variability in phenotyping, would provide such samples. In all cases, this equated to a $CD5^+$ $B220^+$ population representing 30% of the lymphocyte gate, and the condition was named as such (Eμ-*TCL1* '30%') for simplicity. As with the B-cell-based proteomics study, only female mice were used to generate samples for quantitative proteomics.

A cohort of 22 Eμ-*TCL1* mice (8 male, 14 female) and 8 WT female littermates were screened for a $CD5^+$ $B220^+$ leukaemia cell population using the flow cytometry gating described in **Figure 5.2** on a monthly basis. On each screening the $CD5^+$ $B220^+$ population was determined for each animal and Eμ-*TCL1* percentages compared to WT percentages. To avoid bias that might be introduced by selecting the 6 earliest tumours emerging within the cohort, 6 terminal plasma samples were collected from every other animal reaching a 30% $CD5^+$ $B220^+$ population.

Figure 6.1 describes the percentage of $CD5^+$ $B220^+$ cells in the circulation of the 8 WT littermates and the 16 tracked Eμ-*TCL1* mice over the course of 12 months, as well as the 6 mice for which samples were taken for quantitative proteomics. For the WT mice, as determined by genotyping described in **Section 2.4**, there was consistently less than a 15% $CD5^+$ $B220^+$ population with the exception of a single aberrant phenotyping at 6 months. For the monitored Eμ-*TCL1* mice, all but 'BC4' at some stage exceeded all $CD5^+$ $B220^+$ cell percentages characterised in the WT samples, which became terminal shortly after the 12 month time-point. The first lethal tumour of this cohort (BF7) appeared at 5 months, which presented atypically with peripheral tumours and no splenomegaly, and was not used for proteomics. From this point on, every other animal reaching the 30% threshold was used for the proteomic analysis. For the 6 samples collected due to a 30% threshold being reached, the average age of sample collection was 207 days (SD=38), ranging from 182 to 283 days. The range of $CD5^+$ $B220^+$ cell percentages for these samples was 33.2% to 48.7% with an average and SD of 36.8% and 6.5%, respectively. All other Eμ-*TCL1* mice succumbed to a terminal tumour within 2 months of the end of this 12 month observation period.

The $CD5^+$ $B220^+$ cell percentages determined just prior to the collection of the '30%' samples were significantly greater than all WT screenings ($p<0.00018$). The '30%' percentages were also significantly less than those determined just prior to terminal tumour development in the animals allowed to progress past 30% ($p=0.0069$). Finally, a significant difference was also observed between the '30%' $CD5^+$ $B220^+$ cell percentages and the Eμ-*TCL1* percentages initially observed at 3 months ($p=3.9x10^{-5}$).

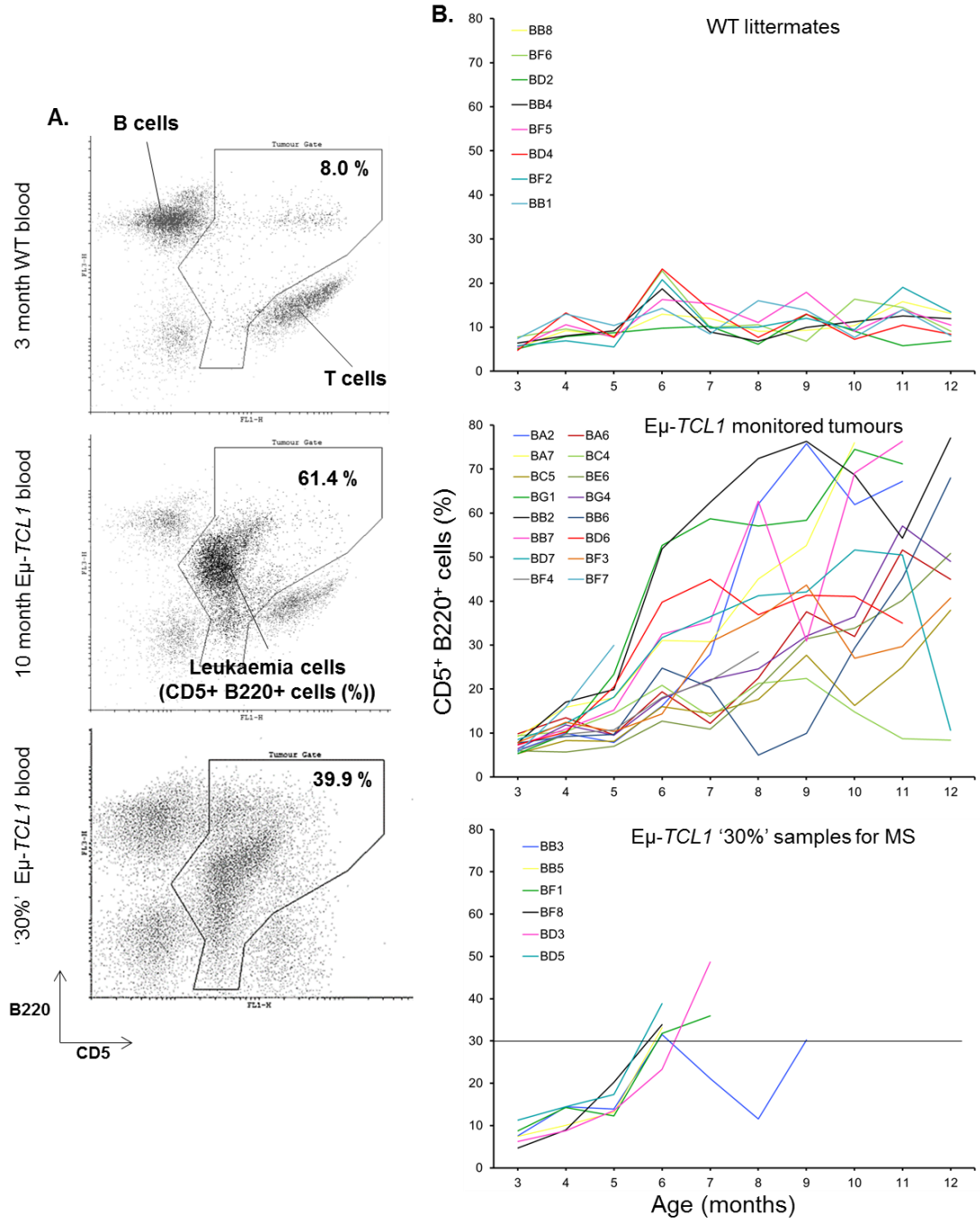


Figure 6.1. Longitudinal observations of CD5+ B-cell percentages in Eμ-TCL1 and WT mice. A. Representative lymphocyte populations describing the $CD5^+$ $B220^+$ cell percentage in the blood of WT, '30%' Eμ-*TCL1* and terminal Eμ-*TCL1* mice. **B.** 22 Eμ-*TCL1* mice (8 male, 14 female) and 8 WT mice (all female) were immunophenotyped for $CD5^+$ $B220^+$ cell percentage on a monthly basis from 3 months of age. The '30%' samples describe every other Eμ-*TCL1* mouse presenting with >30% $CD5^+$ $B220^+$ cell percentage, which were euthanised for sample collection. The remaining animals were monitored by monthly screening, alongside their WT littermates. BA, BC and BG were male animals, BB, BD and BF, female.

6.3 DESIGN AND WORKFLOW FOR QUANTITATIVE PLASMA PROTEOMICS OF THE Eμ-*MYC* AND Eμ-*TCL1* MODELS

A proteomics experiment was designed that incorporated the same conditions described in **Chapter 5**, while additionally incorporating the intermediate, '30%' Eμ-*TCL1* samples, described in **Section 6.2**. The optimisation and confirmation of the capabilities of the newly-available TMT 10-plex reagents, described in **Chapter 4**, accommodated for these additional samples, while maintaining the experiment within a single isobaric tag experiment. Six samples were collected for each of the 7 biological conditions, which in the case of the WT and 6-week pre-tumour controls were from the same 6 mice from which the B cells were collected (**Chapter 5**). For the terminal tumours, due to challenges in the collection of specimens without haemolysis, some plasma samples from the tumours used in the B-cell proteomics were not of sufficient quality for plasma proteomics. Additional samples were therefore collected to reach a total of 6 samples for each condition. For the non-tumour controls, 20 µl of each of the six plasma samples were combined into a single 120 µl pool. For the samples derived from terminal and intermediate Eμ-*TCL1* '30%' mice, two 120 µl pools of 3 x 40 µl were generated to provide biological replicates. This allocation of samples and the specific tumours used in each pool is described in **Figure 6.2**. Pooled plasma samples were immediately diluted with 6 M guanidine hydrochloride, the mobile phase used for SEC; a strong chaotropic solution capable of denaturing almost all non-covalent bonds within the plasma, which subsequently prevents any proteolysis occurring in the samples.

Plasma pools were subjected to SEC with the aim of isolating the low molecular weight sub-proteome for each of the 10 pools. The low M_w sub-proteome contains the lowest percentage of albumin, based on spectral counting from previous experiments, thereby giving the lowest levels of potential interference from albumin peptides when undergoing LC-MS analysis. While excluding many large proteins, this sub-proteome provided the greatest number of protein identifications and gave a representative signature of large proteins, partially due to the enrichment of proteolysis products. Using identical retention time cut points, the eluting protein was isolated by dialysis and lyophilisation. This protein was solubilised and treated as a cell lysate (**Section 2.19**) with 30 µg being the lowest yielded mass of protein and therefore the quantity used from all samples. As with **Chapter 5**, the protein solution was reduced, alkylated and trypsinised and the resulting peptides were, in this case, labelled with TMT 10-plex according to **Figure 6.2**. The peptides were then pooled and subjected to analysis by 2D-LC MS/MS; first prefractionating the peptides with high-pH reverse phase HPLC, then resolving each fraction with low-pH reverse phase HPLC in-line with MS analysis, described in **Section 2.19.5**.

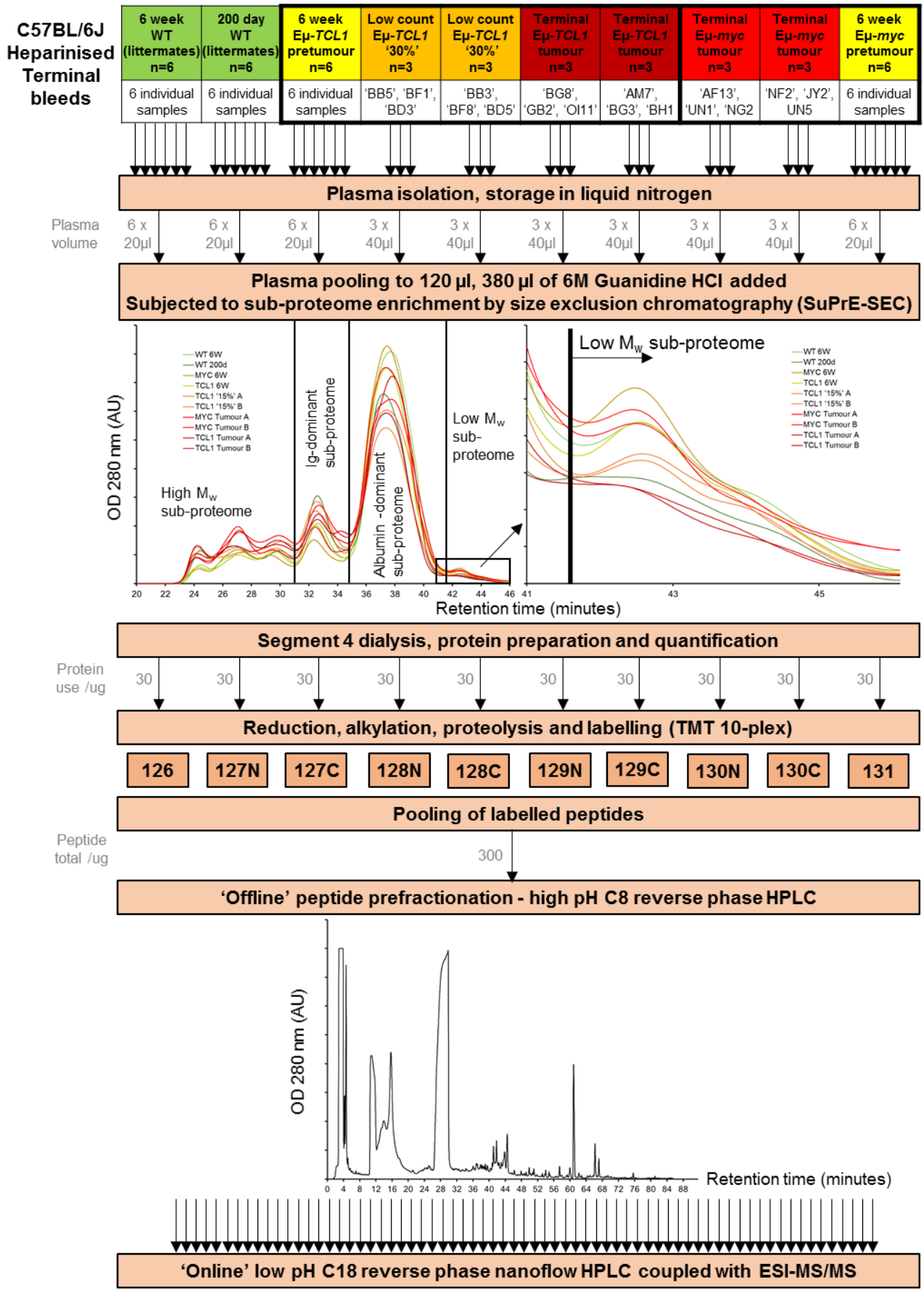

C57BL/6J Heparinised Terminal bleeds
6 week WT (littermates) n=6
200 day WT (littermates) n=6
6 week Eμ-TCL1 pretumour n=6
Low count Eμ-TCL1 '30%' n=3
Low count Eμ-TCL1 '30%' n=3
Terminal Eμ-TCL1 tumour n=3
Terminal Eμ-TCL1 tumour n=3
Terminal Eμ-myc tumour n=3
Terminal Eμ-myc tumour n=3
6 week Eμ-myc pretumour n=6
6 individual samples
6 individual samples
6 individual samples
'BB5', 'BF1', 'BD3'
'BB3', 'BF8', 'BD5'
'BG8', 'GB2', 'OI11'
'AM7', 'BG3', 'BH1
'AF13', 'UN1', 'NG2
'NF2', 'JY2', UN5
6 individual samples
Plasma isolation, storage in liquid nitrogen
Plasma volume
6 x 20μl
3 x 40μl
Plasma pooling to 120 μl, 380 μl of 6M Guanidine HCl added
Subjected to sub-proteome enrichment by size exclusion chromatography (SuPrE-SEC)
OD 280 nm (AU)
High Mw sub-proteome
Ig-dominant sub-proteome
Albumin -dominant sub-proteome
Low Mw sub-proteome
Low Mw sub-proteome
WT 6W
WT 200d
MYC 6W
TCL1 6W
TCL1 '15%' A
TCL1 '15%' B
MYC Tumour A
MYC Tumour B
TCL1 Tumour A
TCL1 Tumour B
Retention time (minutes)
Segment 4 dialysis, protein preparation and quantification
Protein use /ug
30
Reduction, alkylation, proteolysis and labelling (TMT 10-plex)
126
127N
127C
128N
128C
129N
129C
130N
130C
131
Pooling of labelled peptides
Peptide total /ug
300
'Offline' peptide prefractionation - high pH C8 reverse phase HPLC
OD 280 nm (AU)
Retention time (minutes)
'Online' low pH C18 reverse phase nanoflow HPLC coupled with ESI-MS/MS

Figure 6.2. The experimental design and workflow for the TMT 10-plex characterisation of the Eμ-*myc* and Eμ-*TCL1* plasma proteomes. The workflow describes, in descending order: The 10 biological conditions or replicates for which plasma samples were collected for MS characterisation including age, genotype and biological state – the 'traffic light' colours represent the severity of tumour risk/progression. For each biological condition, 6 samples were collected for pooling. For the '30%' and terminal tumours 6 samples were pooled into two biological replicates of 3 samples. Blood was drawn under terminal anaesthesia from the inferior vena cava and plasma isolated by centrifugation and stored in liquid nitrogen. For each of the 10 conditions/replicates, a pool of 120 µl was derived from either 6 x 20 µl or 3 x 40 µl, which was then diluted to 500 µl with 6 M guanidine. Each plasma pool was subjected to size exclusion chromatography to isolate the low molecular weight sub-proteome for each sample. The proteins enriched within this sub-proteome were extracted by dialysis, concentration determined and 30 µg for each pool subjected to peptide preparation and labelling with TMT 10-plex. The labelled peptides were pooled and prefractionated and each fraction subjected to a second round of chromatography, in line with electrospray ionisation, MS detection and data-dependent MS/MS characterisation.

6.4 Eμ-*MYC* and Eμ-*TCL1* Protein Identification and Relative Quantification

Approximately 3 million spectra were subjected to two target decoy searches, relating to CID and HCD spectra as described in **Section 2.19.6** and **Figure 6.3**. As with **Chapter 5**, spectra were subjected to iterative searches firstly identifying peptides based on highly stringent criteria followed by a subsequent, relaxed criteria search for all unmatched spectra, identifying a total of 92,368 PSMs (FDR <1%) matching 2568 protein groups. PSM quantitation data was subjected to analysis by the SPIQuE tool, fully quantifying a total of 2095 proteins.

The terms differential expression or up or downregulated, were considered inaccurate for describing plasma protein quantities. Processes such as cell lysis can influence the protein profile, which is neither technically an expression nor a regulation. The term 'abundance', 'overabundant' and 'underabundant' were considered more appropriate for these proteins.

6.5 Cluster Analysis of Differential Protein Abundance in the Eμ-*MYC* and Eμ-*TCL1* Models

Cluster 3.0 using Euclidian distance was employed to analyse and visualise the global relative protein abundance across the 15 $\log_2$ (ratios) determined by SPIQuE (**Figure 6.3**). As with **Chapter 5**, the biological replicates relating to the terminal tumours demonstrated both strong clustering and a high degree of reproducibility, with regions of similarities and differences appearing in the plasma between the two models. As with the protein expression of the B-cell tumour material (relative to WT controls) terminal Eμ-*myc* plasma had far greater differential abundance in the proteome. Many of these proteins appeared to also be differentially abundant in Eμ-*TCL1* terminal plasma but to a lesser extent. The Eμ-*myc* terminal plasma also had a large region of proteins specifically, and reproducibly overabundant. Only a small region of proteins appeared specifically overabundant only in plasma from Eμ-*TCL1* terminal tumours, however of these proteins, many also appeared overabundant in this region for the '30%' Eμ-*TCL1* samples. Variability was observed between underabundant proteins when comparing to the 6-week or 200 day WT control plasma. Relative to the 6-week WT plasma, the 200 day plasma demonstrated a region of 138 proteins twice as abundant, 77 of which were Ig components. This enrichment is likely indicative of a background artefact caused by Ig reduction allowing appearance in the low M_w sub-proteome. The use of the regulation score combined with statistics identified such inconsistencies and removed these proteins from the subsequent analysis.

The 6-week Eμ-*TCL1* plasma presented with minimal differential abundance relative to 6-week WT plasma with only 69 (3.2%) and 59 (2.8%) of the proteins more than twice as abundant, respectively. Of the overabundant proteins, several were in a region which was also

differentially abundant for the ‘30%’ Eμ-*TCL1* plasma. The Eμ-*myc* 6 week, pre-malignant plasma pool presented a pattern of dysregulation which clustered with the Eμ-*TCL1* samples at all stages but most strongly with the 6-week Eμ-*TCL1* plasma.

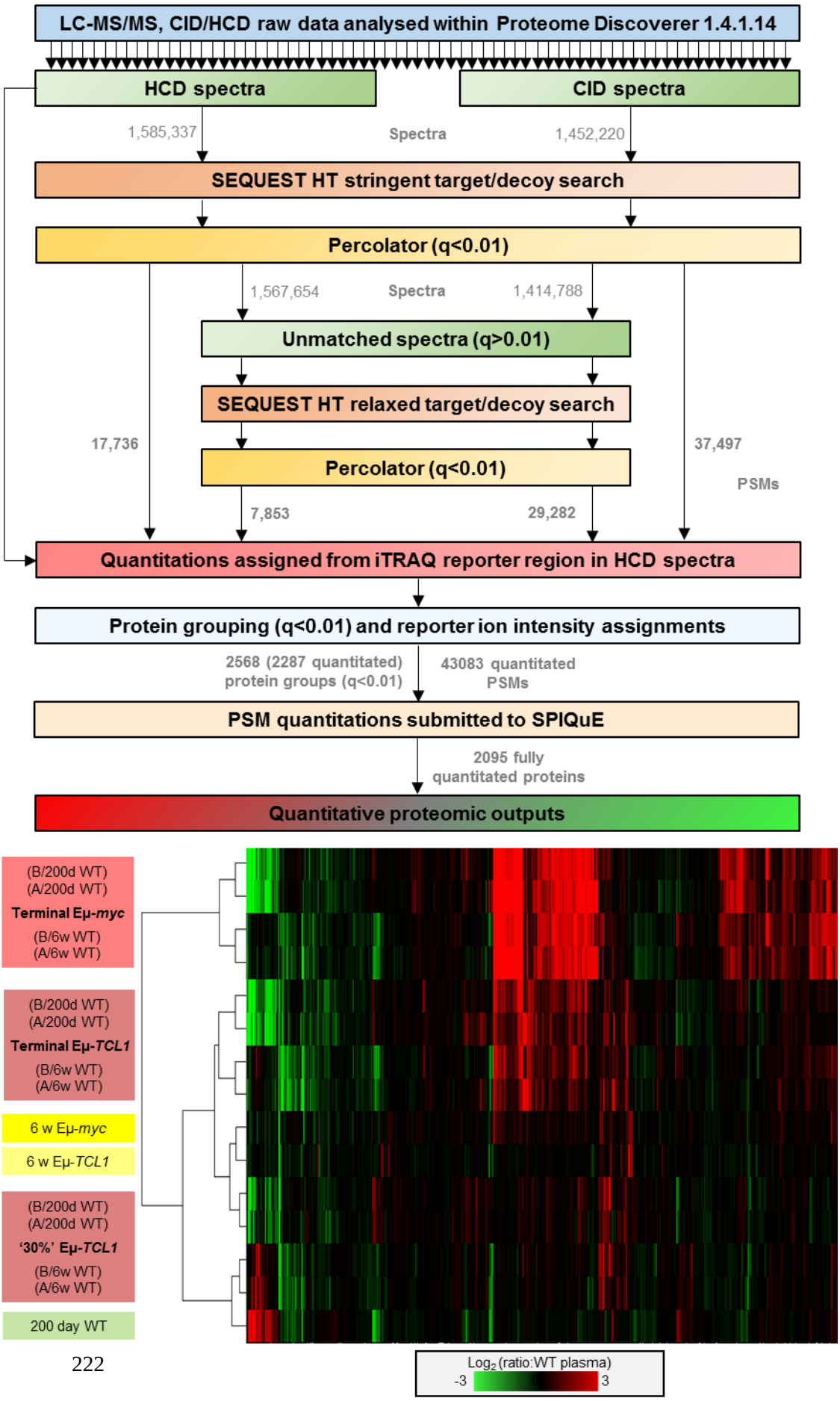

LC-MS/MS, CID/HCD raw data analysed within Proteome Discoverer 1.4.1.14
HCD spectra
CID spectra
1,585,337
Spectra
1,452,220
SEQUEST HT stringent target/decoy search
Percolator (q<0.01)
1,567,654
Spectra
1,414,788
Unmatched spectra (q>0.01)
SEQUEST HT relaxed target/decoy search
17,736
Percolator (q<0.01)
37,497
PSMs
7,853
29,282
Quantitations assigned from iTRAQ reporter region in HCD spectra
Protein grouping (q<0.01) and reporter ion intensity assignments
2568 (2287 quantitated) protein groups (q<0.01)
43083 quantitated PSMs
PSM quantitations submitted to SPIQuE
2095 fully quantitated proteins
Quantitative proteomic outputs
(B/200d WT)
(A/200d WT)
Terminal Eμ-myc
(B/6w WT)
(A/6w WT)
(B/200d WT)
(A/200d WT)
Terminal Eμ-TCL1
(B/6w WT)
(A/6w WT)
6 w Eμ-myc
6 w Eμ-TCL1
(B/200d WT)
(A/200d WT)
'30%' Eμ-TCL1
(B/6w WT)
(A/6w WT)
200 day WT
Log2 (ratio:WT plasma)
-3
0
3

Figure 6.3. Identification and quantification of differential protein abundance in the plasma of Eμ-*myc* and Eμ-*TCL1* models. Fractions were processed using two stage target decoy searches within Proteome Discoverer 1.4. Spectra were searched separately for CID and HCD fragmentation, and initially searched for peptide spectrum matches using stringent settings; allowing for no missed cleavages and no dynamic modifications. Spectra failing to match to a peptide with a percolator-determined false discovery rate of $q<0.01$ were subjected to a second, relaxed target decoy search allowing for missed cleavages and dynamic modifications. The PSM quantitations and features with the protein group assignments were exported for analysis with SPIQuE. Quantitative proteomic results were clustered using Cluster 3.0 and Euclidian distance to represent the topological patterns of relative protein abundance for each biological condition.

6.6 PLASMA PROTEIN OVERABUNDANCE IN TUMOURS OF THE Eμ-*MYC* AND Eμ-*TCL1* MODELS

Given the potential technical limitations of the low M_w sub-proteome analysis, such as the variability of cut points [321] or the sensitivity to biological or technical variation, the use of efficient data analysis of the biological replicates was all the more important to identify only those robustly reproducible candidates demonstrating differential protein abundance. With regards to the anomalies seen with the 200 day control, this presented an opportunity to identify proteins, such as several Igs, susceptible to variation, which could therefore be excluded from further analysis. As described in **Chapter 5**, **Section 2.19.7** and **Section 2.19.8**, all four ratios for each terminal tumour, as well as the combination of all 8 ratios, were analysed by quantitative and statistical analysis. The resulting regulation scores (average / (SD + 1)) and -$\log_{10}$ (FDR-corrected p-values) were plotted to give volcano plots to identify the most significantly upregulated proteins in the plasma of each, and both, of the terminal tumours (**Figure 6.4**). This analysis focused on the upregulated proteins and those with a larger number of PSMs, therefore indicative of higher protein concentration and proteins with traceability to the tumours.

Quantitative analysis of the Eμ-*myc* plasma identified 727 overabundant proteins ($p<0.05$, $RS>0.5$), 34.7% of all quantitated proteins, with 113 underabundant (5.4%). Amongst the overabundant proteins, 4 were clear outliers; translation machinery-associated protein 7 (TMA7), THO complex subunit 4 (Alyref), DNA replication licensing factor, minichromosome maintenance 4 (MCM4) and tyrosine-protein kinase bromodomain adjacent to zinc finger domain protein 1B (BAZ1B), resulting from very consistently, high fold changes relative to both WT controls. When examining the upper right hand region of the volcano plot, other proteins which appeared with substantial overabundance included; nucleolin (NCL), thyroid hormone

receptor-associated protein 3 (THRAP3), heterogeneous nuclear ribonucleoprotein A0 (HNRNA0), actin-related protein 2/3 complex subunit 5-like protein (ARPC5L), leydig cell tumor 10 kDa protein homolog (C19orf53) and complexin-1 (CPLX1). Also amongst these annotated proteins were 9 ribosomal subunits (RP); RPL24, RPL13, RPL28, RPL23, RPS19, RPL5, RPS5, RPL23a and RPL11, three translational regulators; eukaryotic elongation factors (EEF) EEF1B2 and EEF1A1 and eukaryotic initiation factor 4G1 (EIF4G1) and, in addition to HNRNA0, were four further ribonucleoproteins; heterogeneous nuclear ribonucleoprotein U (HNRNPU), heterogeneous nuclear ribonucleoproteins C (HRNRPC), SAP domain-containing ribonucleoprotein (SARNP) and U5 small nuclear ribonucleoprotein 200 kDa helicase (SNRNP200). The partner protein which enables the transcriptional activity of myc - myc-associated factor X (MAX) - was detected as a significantly overabundant protein in Eμ-*myc* plasma. However the myc protein was not itself identified in the plasma.

The Eμ-*TCL1* plasma contained 344 (16.5%) significantly overabundant and 119 (5.7%) underabundant proteins with two overabundant outliers; D-3-phosphoglycerate dehydrogenase (PHGDH) and serum amyloid A3 (SAA3). Other proteins exhibiting the most substantial significant overabundance in Eμ-*TCL1* plasma included; ubiquitin-like protein 5 (UBL5), ubiquitin-fold modifier 1 (UFM1), UV excision repair protein RAD23 homolog B (RAD23B), Kunitz-type protease inhibitor 2 (SPINT2), heterogeneous nuclear ribonucleoproteins A2/B1 (HNRNPA2B1), RNA-binding protein neuro-oncological ventral antigen 2 (NOVA2) and the histone proteins HIST1H1C and HIST2H2AC. Two further histone proteins appeared in this region; HIST1H2AH and HIST1H1A, alongside non-histone chromosomal proteins; high mobility group nucleosome-binding domain-containing proteins (HMGN), HMGN1 and HMGN2.

When considering the $\log_2$ (ratios) for the plasma of terminal tumours from both the Eμ-*myc* and Eμ-*TCL1* models, 625 (29.8%) proteins reached significant overabundance, with 193 (9.2%) underabundant. Coagulation factor V (F5), galectin-3-binding protein (LGALS3BP), Tyrosine-protein phosphatase non-receptor type 18 (PTPN18), Coactosin-like protein (COTL1), thioredoxin-related transmembrane protein 1 (TMX1) and ribosome-binding protein 1 (RRBP1) all presented as outliers with higher $-\log_{10}$ (p-values) indicative of significant overabundance. Other overabundant proteins in terminal plasma included; component of Sp100-rs (CSPRS), nuclear ubiquitous casein and cyclin-dependent kinase substrate 1 (NUCKS1), coiled-coil domain-containing protein 23 (CCDC23), cysteine-rich protein 1 (CRIP1), napsin-A (NAPSA), UBL5, SAA3 and HNRNPA2B1. One observable trend amongst the most overabundant proteins was that of three S100 proteins; S100A6, S100A8 and S100A9.

To better represent the individual ratios for those proteins overabundant in terminal plasma, as well as with a high protein concentration, a filtered list of the most confident proteins was generated based upon PSMs, unique peptides and regulation scores (**Figure 6.5**). PSMs were used as an imprecise, but approximate indication of total protein concentration in all 10 samples. As with **Figure 5.7**, the colouration threshold was raised to a $\log_2$ (ratio) of 2 to emphasise differences between the higher ratios. This colour threshold was exceeded for the majority of the ratios for proteins derived from terminal Eμ-*myc* plasma. Histone H2A type 2-C, for example, had the highest overall average ratios for Eμ-*myc* plasma proteins, with average $\log_2$ (ratios) of 5.76 and 4.35 for plasma pools A and B, equivalent of fold changes of 54 and 20, respectively. Additionally histone H2A type 2-C was observed with, and quantitated from, 786 PSMs. Five other histone proteins were in this list, H1.3, H2A.Z, H3.1, H1.2 and H2A type 1-H with between 219 and 961 PSMs, all exhibiting at least a 15-fold increase in terminal Eμ-*myc* plasma, relative to WT plasma. Four members of the U6 snRNA-associated Sm-like proteins; LSM2, LSM6, LSM7 and LSM8; three ribosomal proteins; RPL29, RPL32 and RPL7; and two proteasome subunits; PSMB10 and PSMA6 were all significantly overabundant in the plasma of Eμ-*myc* and Eμ-*TCL1* mice bearing terminal tumours.

Proteins which were equivalently overabundant in terminal Eμ-*TCL1*, compared to Eμ-*myc* plasma included; TMX1, PTPN18, RRBP1 and F5, which were highlighted previously due to an increased significance in **Figure 6.4**. Amongst this list, serum amyloid A-3 was the only protein which exhibited greater overabundance in terminal Eμ-*TCL1* than Eμ-*myc* plasma, relative to WT plasma.

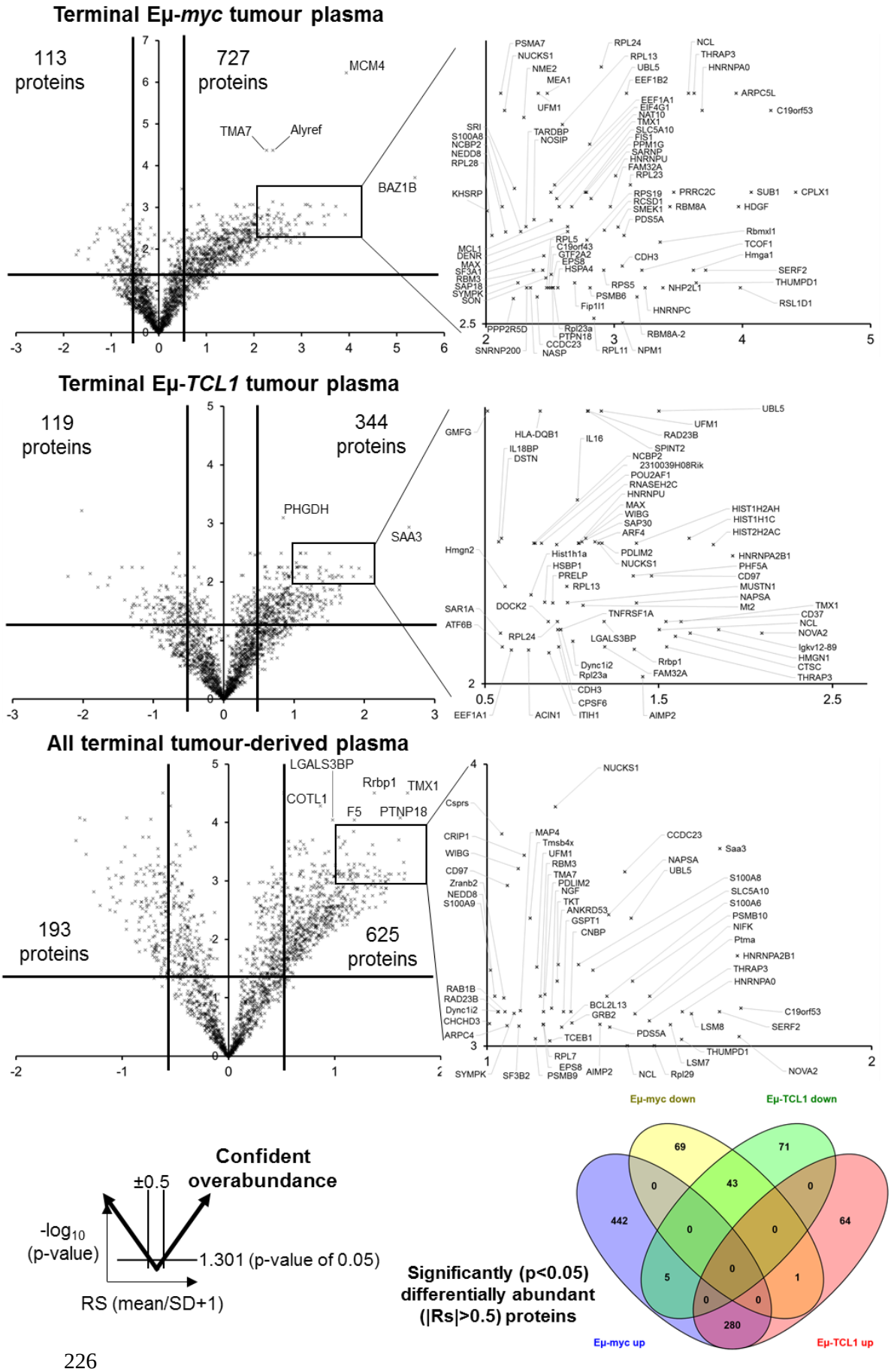
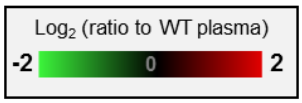

Terminal Eμ-*myc* tumour plasma
113 proteins
727 proteins
MCM4
TMA7
Alyref
BAZ1B
Terminal Eμ-*TCL1* tumour plasma
119 proteins
344 proteins
PHGDH
SAA3
All terminal tumour-derived plasma
193 proteins
625 proteins
LGALS3BP
Rrbp1
TMX1
COTL1
F5
PTNP18
Confident overabundance
±0.5
-log10 (p-value)
1.301 (p-value of 0.05)
RS (mean/SD+1)
Significantly (p<0.05) differentially abundant (|Rs|>0.5) proteins
Eμ-myc down
Eμ-TCL1 down
Eμ-myc up
Eμ-TCL1 up
69
71
43
0
0
442
0
0
64
5
0
1
0
0
280

Figure 6.4. Significant and consistent protein overabundance in terminal Eμ-*myc* and Eμ-*TCL1* plasma. Volcano plots demonstrating proteins with both a significant and large differential abundance in the plasma of Eμ-*myc* and Eμ-*TCL1* mice bearing terminal tumours. The most significantly overabundant proteins are highlighted for each comparison on the right. A Venn diagram comparing the overlap of the significantly over and underabundant proteins for the two models is shown.



Description	Symbol	Unique Peptides	PSMs	WT 200d : 6w WT	6w Eμ-*TCL1* : 6w WT	Eμ-*TCL1* 30% A : 6w WT	Eμ-*TCL1* 30% A : 200d WT	Eμ-*TCL1* 30% B : 6w WT	Eμ-*TCL1* 30% B : 200d WT	Eμ-*TCL1* A : 6w WT	Eμ-*TCL1* A : 200d WT	Eμ-*TCL1* B : 6w WT	Eμ-*TCL1* B : 200d WT	Eμ-*myc* A : 6w WT	Eμ-*myc* A : 200d WT	Eμ-*myc* B : 6w WT	Eμ-*myc* B : 200d WT	6w Eμ-*myc* : 6w WT	Regulation score of both tumours	$-\log_{10}$(FDR p-value)
Thioredoxin-related transmembrane protein 1	**TMX1**	2	5																1.69	4.51
Leydig cell tumor 10 kDa protein homolog	**C19orf53**	2	8																1.66	3.13
Heterogeneous nuclear ribonucleoproteins A2/B1	**HNRNPA2B1**	5	37																1.65	3.32
Tyr phosphatase non-receptor type 18	**PTPN18**	4	29																1.62	4.08
Serum amyloid A-3 protein	**Saa3**	3	27																1.61	3.70
Small EDRK-rich factor 2	**SERF2**	2	20																1.60	3.12
U6 snRNA-associated Sm-like protein LSm8	**LSM8**	6	77																1.53	3.11
Heterogeneous nuclear ribonucleoprotein A0	**HNRNPA0**	2	8																1.51	3.12
Histone H1.3	**HIST1H1D**	3	255																1.49	2.88
Deoxyuridine triphosphatase	**DUT**	7	48																1.48	2.99
Cell division cycle-associated protein 3	**CDCA3**	2	5																1.48	2.88
U6 snRNA-associated Sm-like protein LSm7	**LSM7**	5	27																1.48	3.08
Ribosomal L1 domain-containing protein 1	**RSL1D1**	5	15																1.46	2.79
60S ribosomal protein L29	**Rpl29**	5	20																1.43	3.00
Polyadenylate-binding protein-interacting 2	**PAIP2**	2	29																1.43	2.78
MKI67 FHA-interacting nucleolar phosphoprotein	**NIFK**	3	6																1.42	3.18
Prothymosin alpha	**Ptma**	3	77																1.38	3.11
Proteasome subunit beta type-10	**PSMB10**	4	67																1.38	3.23
Ribosome-binding protein 1	**Rrbp1**	11	47																1.37	4.51
Nucleolin	**NCL**	23	99																1.36	3.00
Non-histone chromosomal protein HMG-14	**HMGN1**	4	85																1.36	2.75
Coiled-coil domain-containing protein 23	**CCDC23**	2	5																1.36	3.62
RNA binding motif protein, X-linked-like-1	**Rbmxl1**	8	25																1.34	2.93
Histone H2A type 2-C	**HIST2H2AC**	2	786																1.33	2.54
Napsin-A	**NAPSA**	4	15																1.32	3.47
Protein S100-A6	**S100A6**	3	44																1.31	3.18
Zinc finger protein 706	**ZNF706**	2	9																1.31	2.87
aa-tRNA synthase complex-interacting 2	**AIMP2**	3	11																1.29	3.08
U6 snRNA-associated Sm-like protein LSm6	**LSM6**	7	79																1.28	2.64
High mobility group protein HMG-I/HMG-Y	**Hmga1**	6	80																1.27	2.65
Histone H2A.Z	**H2AFZ**	3	534																1.26	2.52
Histone H3.1	**Hist1h3a**	2	327																1.22	2.43
Growth factor receptor-bound protein 2	**GRB2**	2	5																1.22	3.08
Proteasome subunit alpha type-6	**PSMA6**	4	30																1.21	2.89
Lamin-B1	**LMNB1**	11	97																1.21	2.81
Histone H1.2	**HIST1H1C**	4	219																1.20	2.45
Protein dpy-30 homolog	**DPY30**	4	34																1.19	2.69
60S ribosomal protein L32	**Rpl32**	3	7																1.19	2.92
Coagulation factor V	**F5**	5	10																1.18	4.04
Transketolase	**TKT**	3	7																1.18	3.23
NHP2-like protein 1	**NHP2L1**	2	22																1.16	2.54
U6 snRNA-associated Sm-like protein LSm2	**LSM2**	4	19																1.15	2.68
Protein PRRC2C	**PRRC2C**	6	13																1.15	2.58
60S ribosomal protein L7	**RPL7**	2	5																1.15	3.08
SNW domain-containing protein 1	**SNW1**	2	13																1.14	2.57
Putative RNA-binding protein 3	**RBM3**	4	27																1.14	3.18
Ubiquitin-fold modifier 1	**UFM1**	2	5																1.13	3.28
Enhancer of rudimentary homolog	**ERH**	3	20																1.12	2.77
Activated RNA pol II coactivator p15	**SUB1**	6	94																1.12	2.33
Histone H2A type 1-H	**HIST1H2AH**	4	961																1.12	2.25

Figure 6.5. Reproducibly overabundant proteins in plasma derived from terminal, tumour bearing Eμ-*myc* and Eμ-*TCL1* mice. The 50 proteins exhibiting a combination of high abundance and significant overabundance in the terminal plasma of both tumour models relative to WT plasma. Proteins were filtered to include the top 50 proteins on the basis of regulation score for both models, with 5 PSMs, 2 unique peptides and a less than 1.5 fold change observed between 6-week and 200 day WT controls. A threshold of saturated colouration of $\log_2$ (ratio) of 2 was used.

6.7 Relative Plasma Protein Abundance Across the Eμ-*TCL1* Model Time Course

As mentioned in **Section 6.2**, the terminal plasma of Eμ-*myc* and Eμ-*TCL1* mice is not typically representative of cancers observed in a clinical setting, as tumours are rarely allowed to progress this far without intervention. Consideration was therefore given to better understand the differential abundances occurring at the intermediate, '30%' stage of Eμ-*TCL1* cancer development, where leukaemia percentage were significantly different to both the 6-week and terminal Eμ-*TCL1* mice.

A volcano plot representative of the differential abundances of the '30%' Eμ-*TCL1* plasma proteins, identified 43 overabundant and 29 underabundant proteins ($p<0.05$), respectively (**Figure 6.6A**). Additionally, to represent proteins with differential abundance at the terminal stage of the Eμ-*TCL1* model, the '30%' plasma proteins were overlaid with three groups of data identifying the 344 significantly overabundant proteins, the 119 significantly underabundant proteins and the proteins without significant regulation amongst the terminal Eμ-*TCL1* plasma proteins described in **Figure 6.4**. 22 proteins significantly overabundant in terminal Eμ-*TCL1* plasma mapped to the significantly overabundant '30%' Eμ-*TCL1* plasma, including; prolargin (PRELP), neurotrophin receptor-alike death domain protein (NRADD), small EDRK-rich factor 2 (SERF2), family with sequence similarity 32, member A (FAM32A), gastric inhibitory polypeptide (GIP) and the previously described SPINT2 and NUCKS1. Of the underabundant proteins, 12 were common to both '30%' and terminal Eμ-*TCL1* mouse plasma, most notably; desmoglein-2 (DSG2), uncharacterized protein C17orf78 homolog (C17orf78), IgHV1-5, calcium-transporting ATPase type 2C member 2 (ATP2C2) and apolipoprotein C-III (APOC3).

Figure 6.6B demonstrates the same volcano plot as **Figure 6.6A,** overlaid with the proteins determined to be over or underabundant in the 6-week Eμ-*TCL1* plasma pool. As no regulation score or statistics could be calculated for a single ratio, a two-fold change (FC) cut-off

was used to define differential abundance, with 69 and 59 proteins considered over and underabundant by this criteria, respectively. Four of the proteins significantly overabundant in '30%' Eμ-*TCL1* plasma were overabundant at 6 weeks; cysteine and glycine-rich protein 3 (CSRP3), Ly6/PLAUR domain-containing protein 3 (LYPD3), gastric inhibitory polypeptide receptor (GIPR) and epididymal secretory protein E1 (NPC2); with CSPR3 the only protein not demonstrating significant overabundance at the terminal stage, due to a lower overabundance in terminal Eμ-*TCL1* plasma pool B. Proteins underabundant in both '30%' and 6-week Eμ-*TCL1* plasma included; DSG2, U1 small nuclear ribonucleoprotein 70 kDa (SNRNP70), misshapen-like kinase 1 (MINK1), APOC3, ATP2C2, coiled-coil domain-containing protein 113 (CCDC113), 'DEK', vezatin (VEZT) and insulin-like growth factor-binding protein 1 (IGFBP1).

Of the 43 proteins overabundant in the '30%' Eμ-*TCL1* plasma, 20 were uniquely overabundant in this state, however, of these just 7 proteins had no differential abundance in the terminal plasma (RS<0.2) or 6-week plasma (FC>2) an; coagulation factor XIII B (F13B), protein strawberry notch homolog 1 (SBNO1), cytochrome c oxidase subunit N69DUFA4 (NDUFA4), peroxiredoxin-6 (PRDX6), tetratricopeptide repeat protein 36 (TTC36), ATPase inhibitor, mitochondrial (ATPIF1) and corticosteroid-binding globulin (SERPINA6).

To better described those plasma proteins with abundance emerging over the Eμ-*TCL1* time course, proteins were filtered to include those with a stepwise increase from WT and 6-week Eμ-*TCL1* to '30%' Eμ-*TCL1* and further to terminal Eμ-*TCL1* (**Figure 6.7**). This analysis identified 43 proteins exhibiting an indication of a terminal tumour signature emerging at the intermediate '30%' stage of tumour development, that was minimally, or was not at all, traceable to the 6-week Eμ-*TCL1* plasma. Members of previously described protein groups, such as ribosome proteins (RPS5, RPL13, RPL29 and RPL35), eukaryotic initiation factors (eIF3A and eIF4G1), proteasome components (PSMA6 and PSMB10), chromosomal proteins (histone H2AFZ and HMGN1) and the LSM protein, LSM7, were present in this list. Other emerging proteins types included the chemokines CCL2 and CCL21 and the CD molecules H-2 class II histocompatibility antigen gamma chain (CD74) and the leukocyte migration factor CD97. While the majority (38) of these proteins were simultaneously observed significantly overabundant in Eμ-*myc* terminal plasma (RS>0.5, p<0.05), the remaining 5 demonstrated specificity or greater consistency of upregulation in '30%' and terminal plasma Eμ-*TCL1* than that of terminal or 6-week Eμ-*myc* plasma. These 5 proteins included haptoglobin (HP), CD74 and Igkv12-89, with some overabundance in Eμ-*myc*, and myosin regulatory light chain 2, skeletal (MYLPF) and programmed cell death 1 ligand 2 (PDCD1LG2) with no overabundance traceable to the plasma of the Eμ-*myc* model.

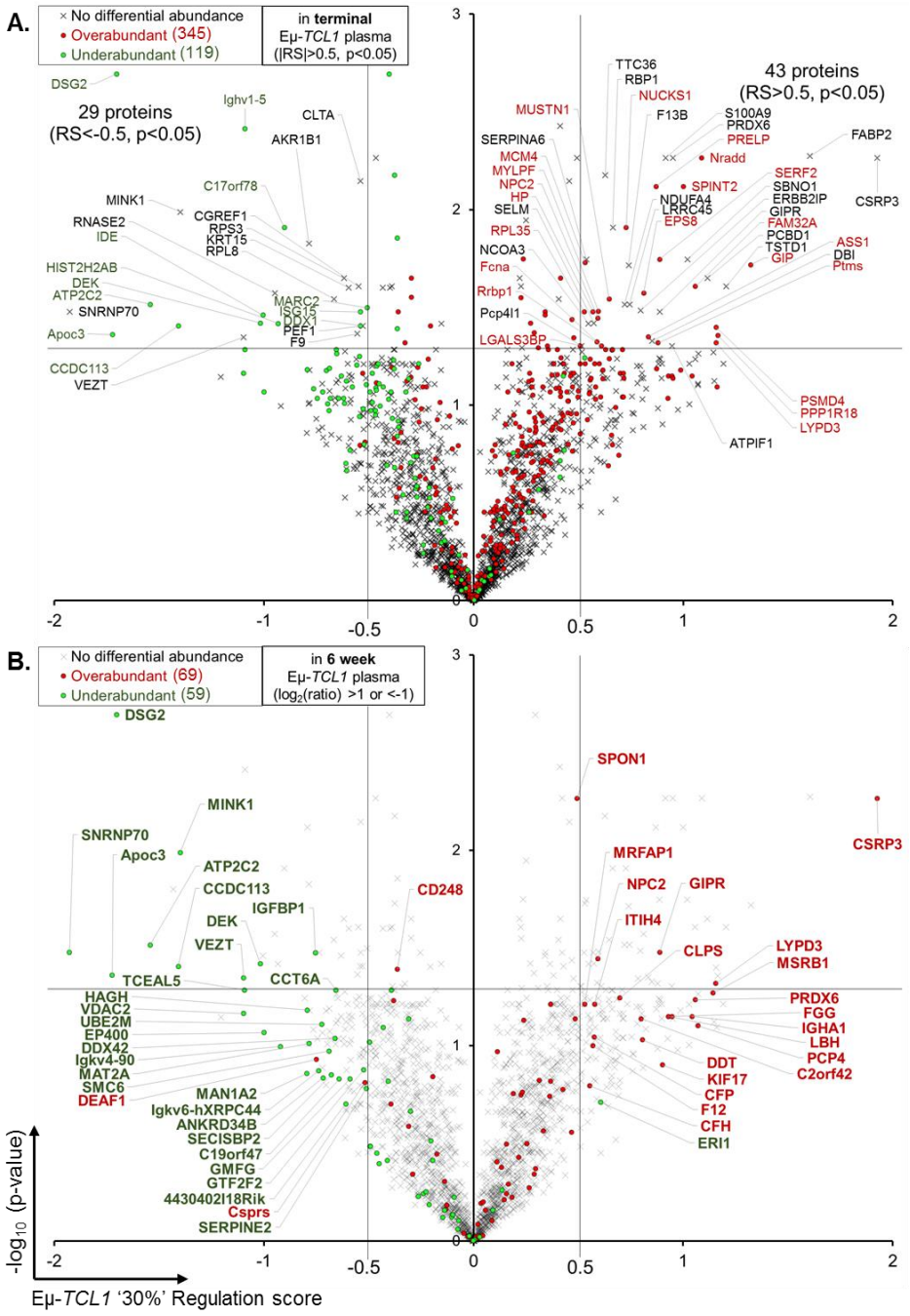
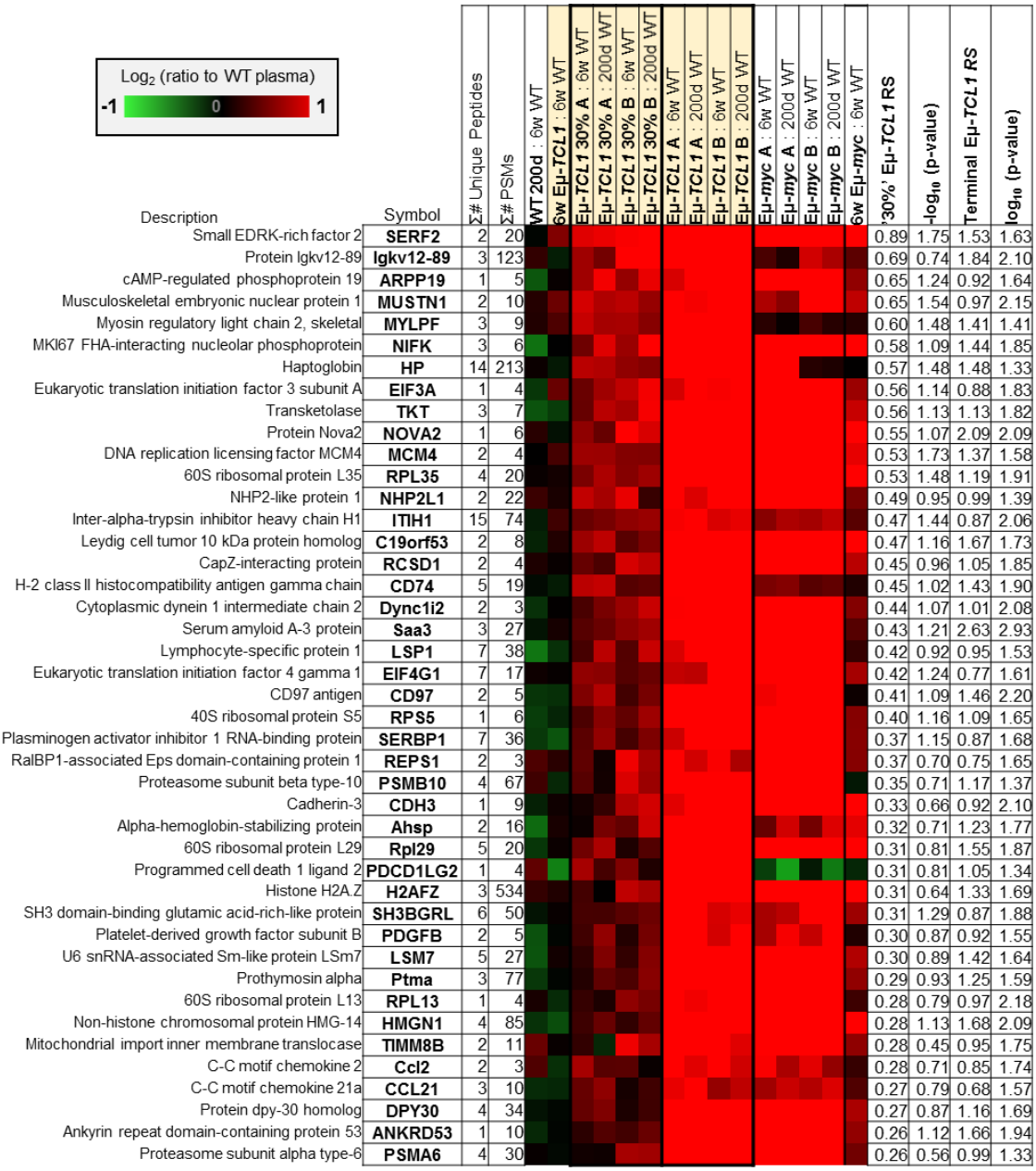


Figure 6.6. Plasma protein abundances at an intermediate stage of tumour development, compared to pre-tumour and terminal stages of Eμ-*TCL1* mice. Volcano plots demonstrating the regulation scores and -$\log_{10}$(p-values) determined for the intermediate '30%' Eμ-*TCL1* plasma proteins relative to WT. These plots are overlaid with; **A.** the terminal Eμ-*TCL1* plasma proteins with significant overabundance (red) (RS>0.5, p<0.05), underabundance (green) (RS<-0.5, p<0.05) or no significant abundance change (black) (0.5>RS>-0.5 or p>0.05), **B.** the 6-week Eμ-*TCL1* plasma proteins classed as overabundant (red) (FC>1), underabundant (green) (FC<-1) or with no abundance change (black) (1>FC>-1).



Description	Symbol	Σ# Unique Peptides	Σ# PSMs	WT 200d : 6w WT	6w Eμ-*TCL1* : 6w WT	Eμ-*TCL1* 30% A : 6w WT	Eμ-*TCL1* 30% A : 200d WT	Eμ-*TCL1* 30% B : 6w WT	Eμ-*TCL1* 30% B : 200d WT	Eμ-*TCL1* A : 6w WT	Eμ-*TCL1* A : 200d WT	Eμ-*TCL1* B : 6w WT	Eμ-*TCL1* B : 200d WT	Eμ-*myc* A : 6w WT	Eμ-*myc* A : 200d WT	Eμ-*myc* B : 6w WT	Eμ-*myc* B : 200d WT	6w Eμ-*myc* : 6w WT	'30%' Eμ-*TCL1* RS	$-\log_{10}$ (p-value)	Terminal Eμ-*TCL1* RS	$\log_{10}$ (p-value)
Small EDRK-rich factor 2	SERF2	2	20																0.89	1.75	1.53	1.63
Protein Igkv12-89	Igkv12-89	3	123																0.69	0.74	1.84	2.10
cAMP-regulated phosphoprotein 19	ARPP19	1	5																0.65	1.24	0.92	1.64
Musculoskeletal embryonic nuclear protein 1	MUSTN1	2	10																0.65	1.54	0.97	2.15
Myosin regulatory light chain 2, skeletal	MYLPF	3	9																0.60	1.48	1.41	1.41
MKI67 FHA-interacting nucleolar phosphoprotein	NIFK	3	6																0.58	1.09	1.44	1.85
Haptoglobin	HP	14	213																0.57	1.48	1.48	1.33
Eukaryotic translation initiation factor 3 subunit A	EIF3A	1	4																0.56	1.14	0.88	1.83
Transketolase	TKT	3	7																0.56	1.13	1.13	1.82
Protein Nova2	NOVA2	1	6																0.55	1.07	2.09	2.09
DNA replication licensing factor MCM4	MCM4	2	4																0.53	1.73	1.37	1.58
60S ribosomal protein L35	RPL35	4	20																0.53	1.48	1.19	1.91
NHP2-like protein 1	NHP2L1	2	22																0.49	0.95	0.99	1.39
Inter-alpha-trypsin inhibitor heavy chain H1	ITIH1	15	74																0.47	1.44	0.87	2.06
Leydig cell tumor 10 kDa protein homolog	C19orf53	2	8																0.47	1.16	1.67	1.73
CapZ-interacting protein	RCSD1	2	4																0.45	0.96	1.05	1.85
H-2 class II histocompatibility antigen gamma chain	CD74	5	19																0.45	1.02	1.43	1.90
Cytoplasmic dynein 1 intermediate chain 2	Dync1i2	2	3																0.44	1.07	1.01	2.08
Serum amyloid A-3 protein	Saa3	3	27																0.43	1.21	2.63	2.93
Lymphocyte-specific protein 1	LSP1	7	38																0.42	0.92	0.95	1.53
Eukaryotic translation initiation factor 4 gamma 1	EIF4G1	7	17																0.42	1.24	0.77	1.61
CD97 antigen	CD97	2	5																0.41	1.09	1.46	2.20
40S ribosomal protein S5	RPS5	1	6																0.40	1.16	1.09	1.65
Plasminogen activator inhibitor 1 RNA-binding protein	SERBP1	7	36																0.37	1.15	0.87	1.68
RalBP1-associated Eps domain-containing protein 1	REPS1	2	3																0.37	0.70	0.75	1.65
Proteasome subunit beta type-10	PSMB10	4	67																0.35	0.71	1.17	1.37
Cadherin-3	CDH3	1	9																0.33	0.66	0.92	2.10
Alpha-hemoglobin-stabilizing protein	Ahsp	2	16																0.32	0.71	1.23	1.77
60S ribosomal protein L29	Rpl29	5	20																0.31	0.81	1.55	1.87
Programmed cell death 1 ligand 2	PDCD1LG2	1	4																0.31	0.81	1.05	1.34
Histone H2A.Z	H2AFZ	3	534																0.31	0.64	1.33	1.69
SH3 domain-binding glutamic acid-rich-like protein	SH3BGRL	6	50																0.31	1.29	0.87	1.88
Platelet-derived growth factor subunit B	PDGFB	2	5																0.30	0.87	0.92	1.55
U6 snRNA-associated Sm-like protein LSm7	LSM7	5	27																0.30	0.89	1.42	1.64
Prothymosin alpha	Ptma	3	77																0.29	0.93	1.25	1.59
60S ribosomal protein L13	RPL13	1	4																0.28	0.79	0.97	2.18
Non-histone chromosomal protein HMG-14	HMGN1	4	85																0.28	1.13	1.68	2.09
Mitochondrial import inner membrane translocase	TIMM8B	2	11																0.28	0.45	0.95	1.75
C-C motif chemokine 2	Ccl2	2	3																0.28	0.71	0.85	1.74
C-C motif chemokine 21a	CCL21	3	10																0.27	0.79	0.68	1.57
Protein dpy-30 homolog	DPY30	4	34																0.27	0.87	1.16	1.69
Ankyrin repeat domain-containing protein 53	ANKRD53	1	10																0.26	1.12	1.66	1.94
Proteasome subunit alpha type-6	PSMA6	4	30																0.26	0.56	0.99	1.33

Figure 6.7. Emerging markers correlating with Eμ-*TCL1* tumour development. Proteins were filtered to generate a list containing those proteins for which the relative abundance incrementally increased between 6 weeks and '30%' leukaemia and again at a terminal stage of Eμ-*TCL1*. Proteins were first filtered for those with significant overabundance in the terminal pools (RS>0.5, $p<0.05$) and for no overabundance at 6 weeks, to remove proteins relating to the Eμ-*TCL1* model prior to any tumourigenesis. The remaining proteins were filtered to include those with a RS of at least 0.25 for the comparison of '30%' to WT and a differential RS of 0.25 between terminal and '30%', indicative of greater abundance in terminal Eμ-*TCL1* plasma. Proteins were also filtered to exclude those with >1.5-fold regulation in 200 day WT plasma and with less than 3 PSMs.

6.8 RELATIVE PROTEIN ABUNDANCE IN PRE-TERMINAL Eμ-*MYC* PLASMA

Given the more aggressive nature of the Eμ-*myc* model and the observation of substantial protein dysregulation in the pre-tumour samples at 6 weeks, a signature at this stage was anticipated to be indicative of the most robust plasma markers of Eμ-*myc* tumours. An analysis like that of 6-week Eμ-*TCL1* plasma, described in **Figure 6.6B**, was therefore conducted. **Figure 6.8** highlights plasma proteins with progressive abundance in the Eμ-*myc* model, overlaying the volcano plot describing terminal Eμ-*myc* plasma with proteins with either a 2-fold over or underabundance in Eμ-*myc* plasma at 6 weeks. Of the 128 proteins with >2-fold overabundance in Eμ-*myc* plasma at 6 weeks, 72 were also traceable to significant overabundance in terminal Eμ-*myc* plasma ($RS>0.5$, $p<0.05$) including; ribosomal proteins; RPS23, RSP24, RSP30, RPL23, RPL29, RPL30 and RPL35; and histone proteins; H1C, H1D, H2AC, H2AFX, H2AFZ, H3A and H3F3C. Additional groups of proteins included fatty acid binding proteins; FABP1 (liver), FABP2 (intestinal) and FABP5 (epidermal), all of which were observed with a minimum of 4 unique peptides and 29 PSMs; and growth factor and hormone-related proteins; epidermal growth factor receptor kinase substrate 8 (ESP8), growth hormone receptor (GHR) hepatoma-derived growth factor (HDGF) and thyroid hormone receptor-associated protein 3 (THRAP3). 12 proteins had greater abundance at 6 weeks, on average, than that of terminal Eμ-*myc* plasma, including, most notably, all three FABPs; 1, 2 and 5; syndecan-1 (SDC1), tRNA:m(4)X modification enzyme TRM13 homolog (TRMT13), cysteine-rich protein 2 (CRIP2), SUMO-activating enzyme subunit 1 (SAE1), antileukoproteinase (SLPI) and another FA-related protein; non-specific lipid-transfer protein (SPC2). Two proteins were observed with significant underabundance in terminal Eμ-*myc* plasma ($RS<-0.5$, $p<0.05$), while >2-fold overabundant in Eμ-*myc* plasma at 6 weeks; hyaluronan-binding protein 2 (HABP2) and the growth factor and possible cytokine; granulins (GRN).

58 proteins were >2-fold underabundant in 6-week Eμ-*myc* plasma, and of these, 18 mapped to the significantly underabundant proteins in terminal Eμ-*myc* plasma ($RS<-0.5$, $p<0.05$). Only one protein, adenomatous polyposis coli protein (APC), demonstrated reproducible underabundance and further underabundance, comparing WT to 6-week Eμ-*myc* and 6-week Eμ-*myc* to terminal Eμ-*myc* plasma, respectively. The three proteins demonstrating the greatest underabundance in 6-week Eμ-*myc* plasma relative to underabundance in terminal Eμ-*myc* plasma were CCDC113, VDAC2 and DSG2, all three of which were also observed as underabundant in all Eμ-*TCL1* plasma samples. One underabundant 6-week Eμ-*myc* plasma protein was observed which demonstrated significant overabundance when terminal; gasdermin-D (GSDMD)

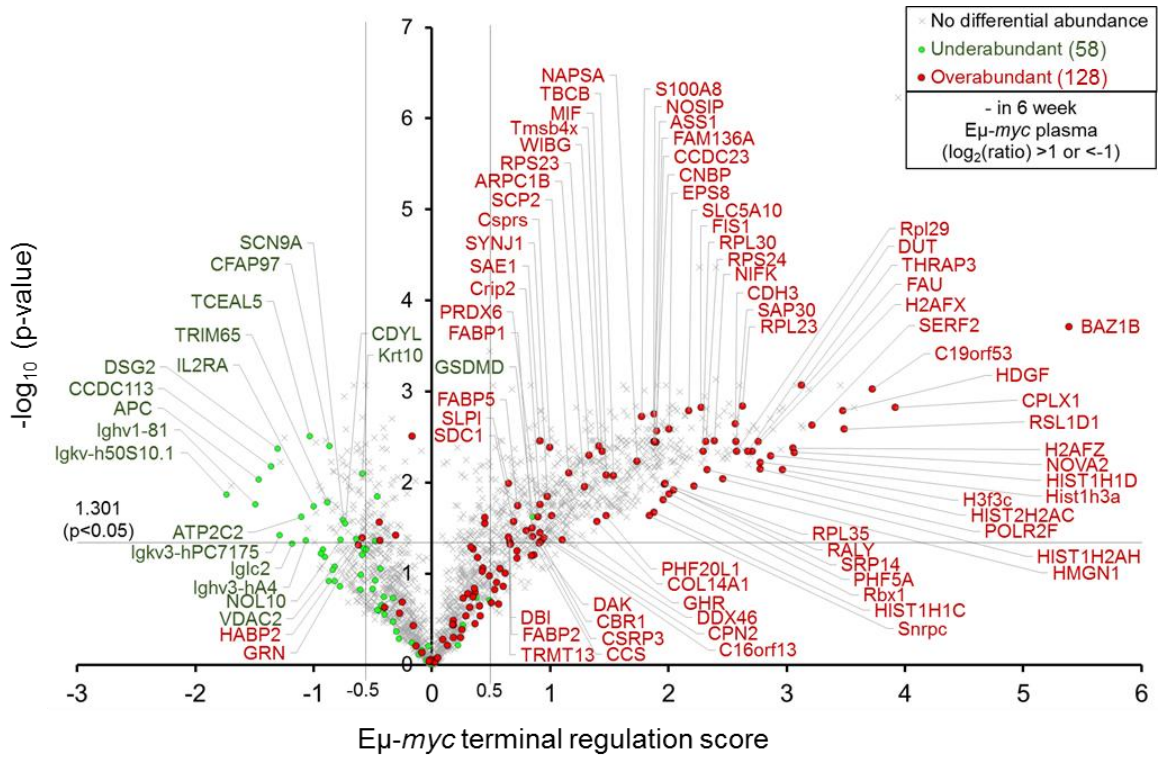


Figure 6.8. Protein abundances in pre-tumour, compared to terminal, Eμ-*myc* plasma. Volcano plot demonstrating the regulation scores and -$\log_{10}$(p-values) determined for the terminal Eμ-*myc* plasma proteins relative to WT. The plot is overlaid with the 6-week Eμ-*myc* plasma proteins classed as overabundant (red) (FC>1), underabundant (green) (FC<-1) or with no abundance change (grey) (1>FC>-1).

6.9 Differentially Abundant, High-Concentration Proteins in Tumour Plasma

When considering proteins as potential biomarkers of disease, an important consideration is the concentration of the proteins in plasma. While no concentration can be directly inferred from MS data, an approximate indication can be derived from the number of PSMs with which a protein was identified. To better understand the effects of the terminal tumours on the higher abundant proteins detected in plasma, the regulation scores for the '30%' Eμ-*TCL1*, terminal Eμ-*myc* and terminal Eμ-*TCL1* plasma were plotted against the number of PSMs for each protein (**Figure 6.9**). **Figure 6.9A** therefore highlighted those proteins with both differential abundance and high concentration in the terminal Eμ-*myc* plasma. The most prominent cluster contained 9 histone proteins each with over 300 PSMs, with an average RS of 2.7. A second cluster of proteins, all with more than 100 PSMs and an average RS of 2, was again dominated by histone proteins, but also contained two proteasome components; PSMB6 and PSMB1; two nucleoside diphosphate kinases (NME); NME1 and NME2; and the ribosome

protein; RPLP2. A third cluster was observed with several proteins all with less than, or around, 100 PSMs and a RSs ranging from 2 to 4. These proteins were all nuclear proteins capable of nucleic acid or histone binding; heterogeneous nuclear ribonucleoproteins C1/C2 (HNRNPC), laminin B1 (LMNB1), high mobility group nucleosome-binding domain-containing protein 1 (HMGN1), HMGA1, nucleophosmin, nucleolin, prothymosin alpha (PTMA), LSM6, LSM8 and activated RNA polymerase II transcriptional coactivator p15 (SUB1). The highest concentration proteins appeared predominantly underabundant, with three apolipoproteins; APOA1, APOA2 and APOC3; significantly underabundant in terminal Eμ-*myc* plasma relative to WT plasma. Three alpha-1-antitrypsin proteins were underabundant, but only SERPINA1D was significantly so. The high-concentration plasma proteins transthyretin (TTR) and albumin (ALB) also appeared significantly underabundant. These observations may have been an artefact of the normalisation caused by the high overabundance of the histone proteins.

For terminal Eμ-*TCL1* plasma, 10 histones were also amongst the high-concentration, overabundant plasma proteins (**Figure 6.9B**). Alongside these were proteins normally located in the plasma; ceruloplasmin (CP), haptoglobin (HP), ITIH4 and all three isoforms of fibrinogen; FGA, FGB and FGG; all with more than 100 PSMs. High-concentration plasma proteins without differential abundance included albumin, complement C3, alpha-2-macroglobulin (A2M) and beta-2-microglobulin (B2M). Transthyretin was again significantly underabundant, as well as the apolipoproteins; APOA2, APOC3, APOA1, APOC3B and APOC2. As with Eμ-*myc* plasma, the three alpha-1-antitrypsin proteins SERPINA1 A, B and D were underabundant, with only SERPINA1D reaching significance. Several haemoglobin proteins were underabundant in terminal Eμ-*TCL1*, compared to WT, plasma, but without significance due to variability in the WT controls.

Overall, the higher-concentration proteins were not significantly differentially abundant in the ‘30%’ Eμ-*TCL1* plasma (**Figure 6.9C**). The proteins formed a triangular distribution with the highest abundant proteins exhibiting less regulation, resulting from averaging a larger number of observations. The protein with the highest predicted concentration with significant differential regulation was APOC3, underabundant by almost 4-fold relative to WT plasma. Significantly overabundant, high-concentration proteins included superoxide dismutase [Cu-Zn] (SOD1), the chemokine, CCL8, the putative pro-hormone, secretogranin (CHGB), HP, ITIH1, SERPINA6 and acyl-CoA-binding protein (DBI).

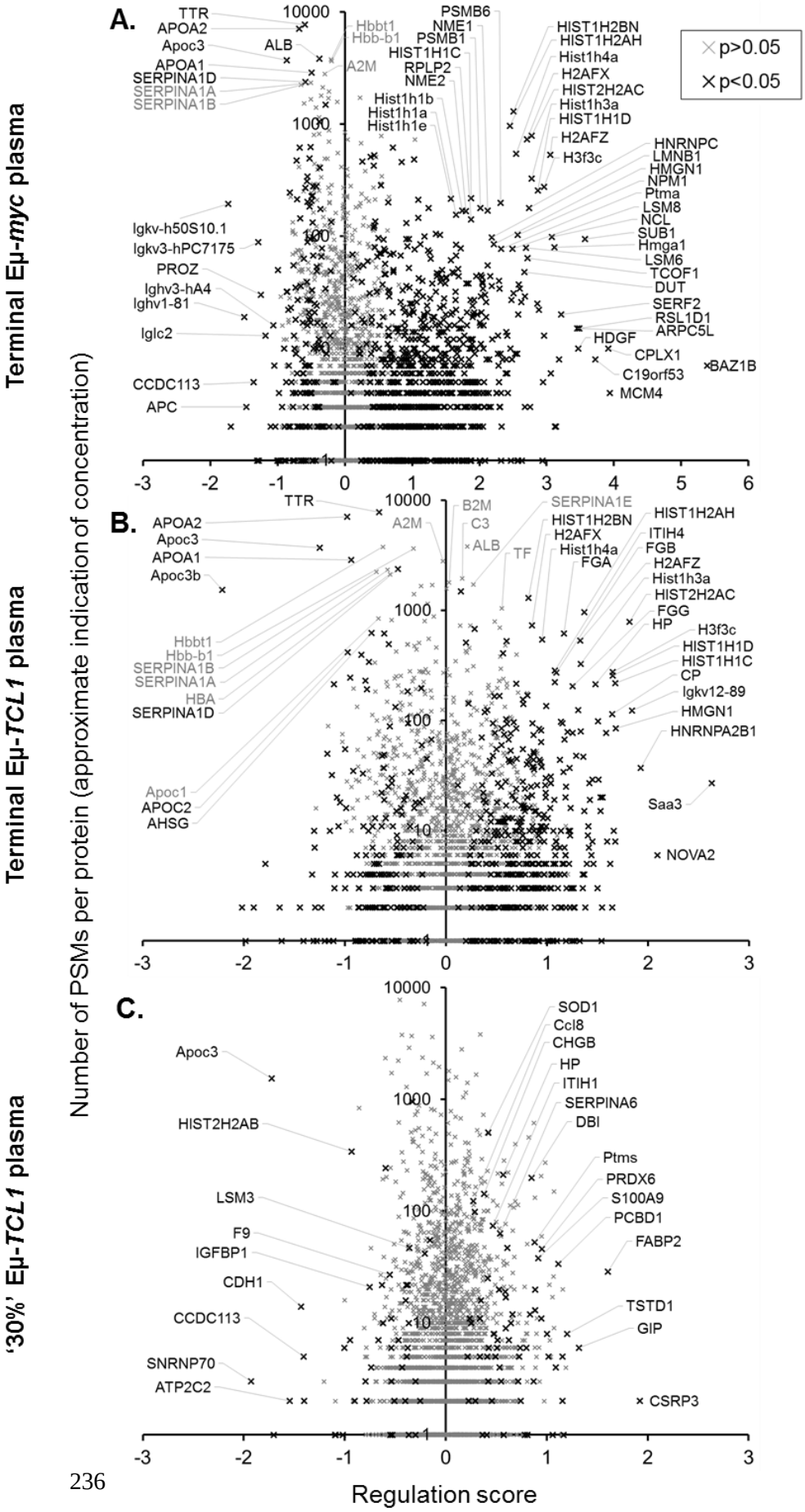

A.
B.
C.
p>0.05
p<0.05
Terminal Eμ-myc plasma
Terminal Eμ-TCL1 plasma
'30%' Eμ-TCL1 plasma
Number of PSMs per protein (approximate indication of concentration)
Regulation score
TTR
APOA2
Apoc3
APOA1
SERPINA1D
SERPINA1A
SERPINA1B
ALB
A2M
Hbbt1
Hbb-b1
PSMB6
NME1
PSMB1
HIST1H1C
RPLP2
NME2
Hist1h1b
Hist1h1a
Hist1h1e
HIST1H2BN
HIST1H2AH
Hist1h4a
H2AFX
HIST2H2AC
Hist1h3a
HIST1H1D
H2AFZ
H3f3c
HNRNPC
LMNB1
HMGN1
NPM1
Ptma
LSM8
NCL
SUB1
Hmga1
LSM6
TCOF1
DUT
SERF2
RSL1D1
ARPC5L
HDGF
CPLX1
C19orf53
MCM4
BAZ1B
Igkv-h50S10.1
Igkv3-hPC7175
PROZ
Ighv3-hA4
Ighv1-81
Iglc2
CCDC113
APC
B2M
C3
TF
SERPINA1E
ITIH4
FGB
FGA
H2AFZ
Hist1h3a
HIST2H2AC
FGG
HP
H3f3c
HIST1H1D
HIST1H1C
CP
Igkv12-89
HMGN1
HNRNPA2B1
Saa3
NOVA2
HBA
Apoc3b
Apoc1
APOC2
AHSG
SOD1
Ccl8
CHGB
HP
ITIH1
SERPINA6
DBI
Ptms
PRDX6
S100A9
PCBD1
FABP2
TSTD1
GIP
CSRP3
HIST2H2AB
LSM3
F9
IGFBP1
CDH1
SNRNP70
ATP2C2

Figure 6.9. Approximate protein concentrations of differentially abundant plasma proteins. The regulation scores plotted against the number of PSMs for; **A.** terminal Eμ-*myc*; **B.** '30%' Eμ-*TCL1* and; **C.** terminal Eμ-*TCL1*; plasma proteins. The plot is overlaid with those proteins reaching significance in each condition relative to WT ($p<0.05$) (black), and those without significance ($p>0.05$) (grey) relative to WT plasma.

6.10 Model-Specific Plasma Protein Signatures of Terminal Eμ-*TCL1* Tumours

To better understand the differences observed between the plasma derived from Eμ-*myc* and Eμ-*TCL1* mice bearing terminal tumours, the regulation scores for each model were plotted against one another (**Figure 6.10**). Tumour-specific signatures emerged, which demonstrated groups of proteins predominantly overabundant in each tumour, but with overabundance that was proportional, to a lesser extent, in the other tumour type. **Figure 6.10** highlights these two signatures specific to Eμ-*myc* and Eμ-*TCL1* plasma, using RS-defined cut-offs, to identify 747 and 127 proteins with a predominant specificity to each tumour, respectively. These groups of proteins were considered terminal Eμ-*TCL1*-'dominant' and terminal Eμ-*myc*-'dominant' signatures and were evaluated for GO term enrichment, to gain an insight into their biological meaning.

Figure 6.11 demonstrates the GO term enrichments derived from the DAVID analysis of the 127 terminal Eμ-*TCL1*-dominant plasma proteins analysed and presented with REVIGO. The biological processes were significantly enriched for parent terms relating to immune system processes; 'immune system process (GO:0002376)' and 'positive regulation of immune system process (GO:0002684)', which also encompassed, most significantly, the terms; 'immunoglobulin mediated immune response (GO:0016064)', 'B-cell mediated immunity (GO:0019724)', 'lymphocyte mediated immunity (GO:0002449)' and 'adaptive immune response (GO:0002250)'. Additionally, separate immune-related terms suggesting a loss of homeostasis and response to tissue damage were significantly enriched; 'response to stress (GO:0006950)', 'defense response (GO:0006952)', 'response to stimulus (GO:0050896)', 'response to external stimulus (GO:0009605)', 'regulation of response to stimulus (GO:0048583)', 'wound healing (GO:0042060)', 'response to wounding (GO:0009611)' and 'acute inflammatory response (GO:0002526)'. The terms 'hyaluronan metabolic process (GO:0030212)' and 'aminoglycan metabolic process (GO:0006022)' were both enriched due to the inter alpha-trypsin inhibitor, heavy chains; ITIH2, ITIH4 and ITIH5, with 'aminoglycan metabolic process (GO:0006022)' additionally enriched with fibrillin 1 (FBN1) and lymphatic vessel endothelial hyaluronan receptor 1 (LYVE1).

Cellular component enrichment identified 49 out of 127 proteins with an annotation of 'extracellular region (GO:0005576)' ($p=2.9x10^{-17}$), indicative of proteins with an association, but no covalent attachment to the plasma membrane. Alongside this were similar terms indicative of an enrichment of proteins with a canonical localisation at the surface of cells, including; 'cell surface (GO:0009986)', 'extracellular region part (GO:0044421)', 'external side of plasma membrane (GO:0009897)' and 'plasma membrane (GO:0005886)'. 'Extracellular space (GO:0005615)', a term suggesting no canonical association with the cell membrane, was also significantly enriched ($p=3.7x10^{-6}$). The terms 'vacuole (GO:0005773)', 'multivesicular body (GO:0005771)' and 'lysosome (GO:0005764)', all indicative of intracellular components, were enriched by proteins relating to MHC processing; H-2 class II histocompatibility antigen A beta chain (H2-AB1) and H-2 class II histocompatibility antigen gamma chain (CD74); and peptide cleavage; cathepsin B (CTSB), C (CTSC) and S (CTSS), gamma-glutamyl hydrolase (GGH) and legumain (LGMN).

A trend of protein binding (GO:0005515) emerged from the molecular function enrichment analysis of the 127 terminal Eμ-*TCL1*-dominant plasma proteins, with 59 annotated with this parent term ($p= 0.011$). This term could be attributed to several other, more minor terms relating to binding, including; 'pattern binding (GO:0001871)', 'carbohydrate binding (GO:0030246)', 'glycosaminoglycan binding (GO:0005539)' and 'polysaccharide binding (GO:0030247)'; terms common to proteins; C-C motif chemokine 8 (CCL8), connective tissue growth factor (CTGF), fibronectin (FN1), ITIH4, LYVE1 and vitronectin (VTN). 'Receptor binding (GO:0005102)' was enriched due to a number of hormones and cytokines, including CCL8 and CTGF, in addition to GIP, glycoprotein hormones alpha chain (CGA), inhibin beta C chain (INHBC), prolactin (PRL), resistin-like alpha (RETNLA) and tumour necrosis factor ligand superfamily member 13 (TNFSF13). Fibrinogen beta chain (FGB), fibrinogen gamma chain (FGG) and fibrinogen-like protein 2/fibroleukin (FGL2) were also annotated with this 'Receptor binding'. 'Antigen binding (GO:0003823)' was enriched due to the MHC and Ig components; H-2 class I histocompatibility antigen, K-B alpha chain (H2-K1), H2-AB1, Ig alpha chain C region (IGHA), Ig mu chain C region (IGHM) and Ig J chain (IGJ). The significance of the opposing terms 'enzyme inhibitor activity (GO:0004857)' and 'peptidase activity (GO:0008233)' was partly attributable to the three ITIH proteins and the three cathepsin proteins, respectively.

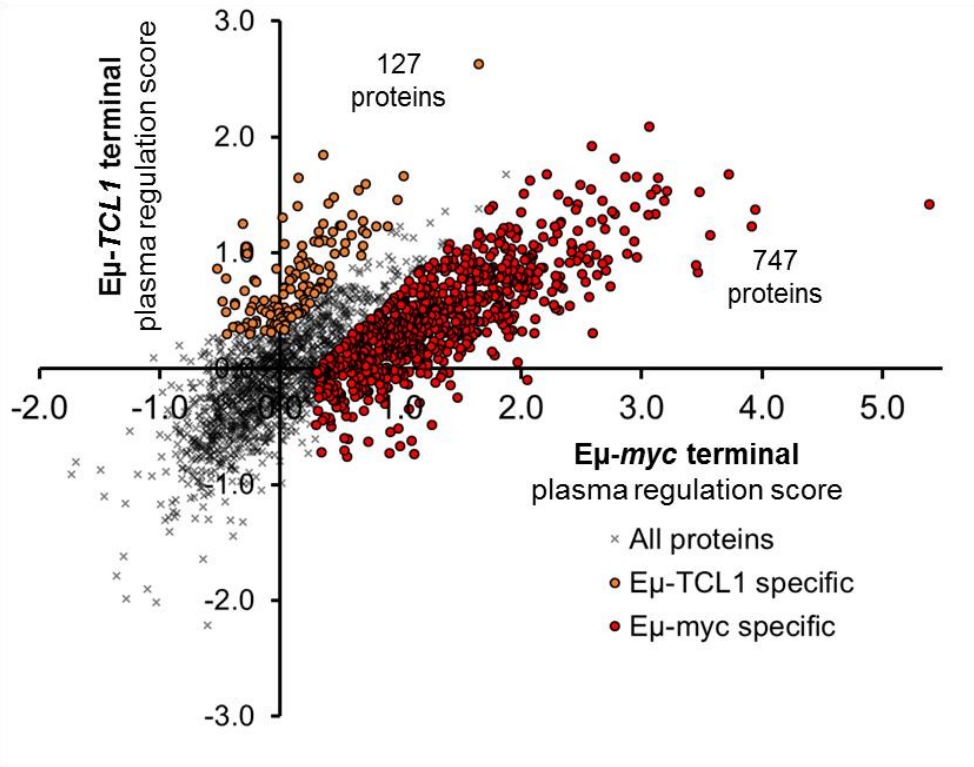


Figure 6.10. Comparison of Eμ-*myc* and Eμ-*TCL1* terminal plasma protein regulation scores. The regulation scores for the terminal tumour-derived Eμ-*myc* and Eμ-*TCL1* plasma plotted against one another demonstrating two distinct signatures. Proteins relating to each signature were selected on the basis of having at least a regulation score of 0.3 for each model as well as the protein regulation score exceeding that of the other model by at least 0.3. The plot was overlaid with those proteins matching these criteria for Eμ-*myc* (red) and Eμ-*TCL1* (orange).

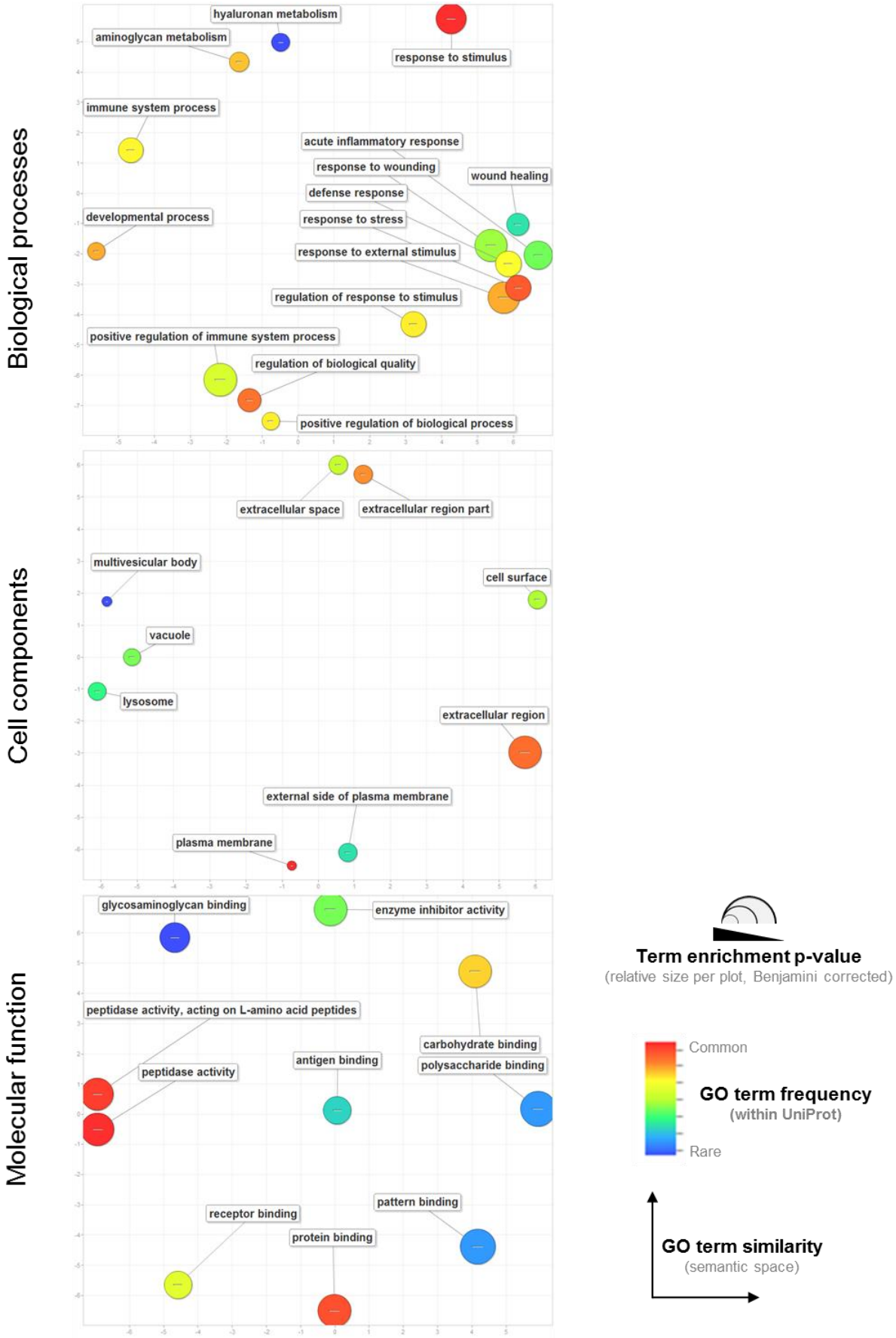

Biological processes
hyaluronan metabolism
aminoglycan metabolism
response to stimulus
immune system process
acute inflammatory response
response to wounding
wound healing
defense response
developmental process
response to stress
response to external stimulus
regulation of response to stimulus
positive regulation of immune system process
regulation of biological quality
positive regulation of biological process
Cell components
extracellular space
extracellular region part
multivesicular body
cell surface
vacuole
lysosome
extracellular region
external side of plasma membrane
plasma membrane
Molecular function
glycosaminoglycan binding
enzyme inhibitor activity
peptidase activity, acting on L-amino acid peptides
carbohydrate binding
antigen binding
polysaccharide binding
peptidase activity
pattern binding
receptor binding
protein binding
Term enrichment p-value
(relative size per plot, Benjamini corrected)
Common
GO term frequency
(within UniProt)
Rare
GO term similarity
(semantic space)

Figure 6.11. Gene ontology term enrichment analysis for the Eμ-*TCL1*-specific signature in terminal tumour plasma. GO term enrichment p-values (Benjamini-corrected) were determined for those proteins specifically overabundant in Eμ-*TCL1* plasma (RS>0.3, RS>Eμ-*myc* RS + 0.3) using DAVID. Significantly enriched GO terms ($p<0.05$) were visualised with REVIGO. These are separated into biological processes, cell components and molecular function.

6.11 MODEL-SPECIFIC PLASMA PROTEIN SIGNATURES OF TERMINAL Eμ-*MYC* TUMOURS

Based upon those 747 proteins highlighted with a predominant specificity to Eμ-*myc* terminal plasma in **Figure 6.10**, the DAVID and REVIGO analysis, like that of **Figure 6.11**, was conducted (**Figure 6.12)**.

For biological processes, the analysis identified a strong trend indicative of intracellular processes, with the most significantly enriched terms being; 'translation (GO:0006412)' and 'cellular process (GO:0009987)' ($p<10^{-30}$). Other significantly enriched terms ($p<10^{-20}$) included; 'RNA splicing (GO:0008380)', 'cellular macromolecule metabolic process (GO:0044260)', 'RNA processing (GO:0006396)', 'gene expression (GO:0010467)' and 'mRNA processing (GO:0006397)'; again processes canonically located within cells. The remaining enriched biological processes followed this trend, with all terms attributable to intracellular processes.

Cellular component GO term enrichment further identified the abundance of cell-derived proteins amongst the Eμ-*myc*-predominant terminal plasma, with the terms 'intracellular (GO:0005622)' and 'intracellular part (GO:0044424)' reaching a p-value of less than 10^{-55}. Three quarters of the 747 analysed (566) were annotated with these terms. Several terms were also enriched describing subcellular components; 'cytoplasm (GO:0005737)', 'nucleus (GO:0005634)', nuclear part (GO:0044428)', 'organelle (GO:0043226)', 'intracellular organelle (GO:0043229)', 'intracellular non-membrane-bounded organelle (GO:0043232)' and 'non-membrane-bounded organelle (GO:0043228); all significantly enriched ($p<10^{-25}$). Again, all the remaining terms were relatable to cell components, such as those describing chromatin, the ribosome and the spliceosome.

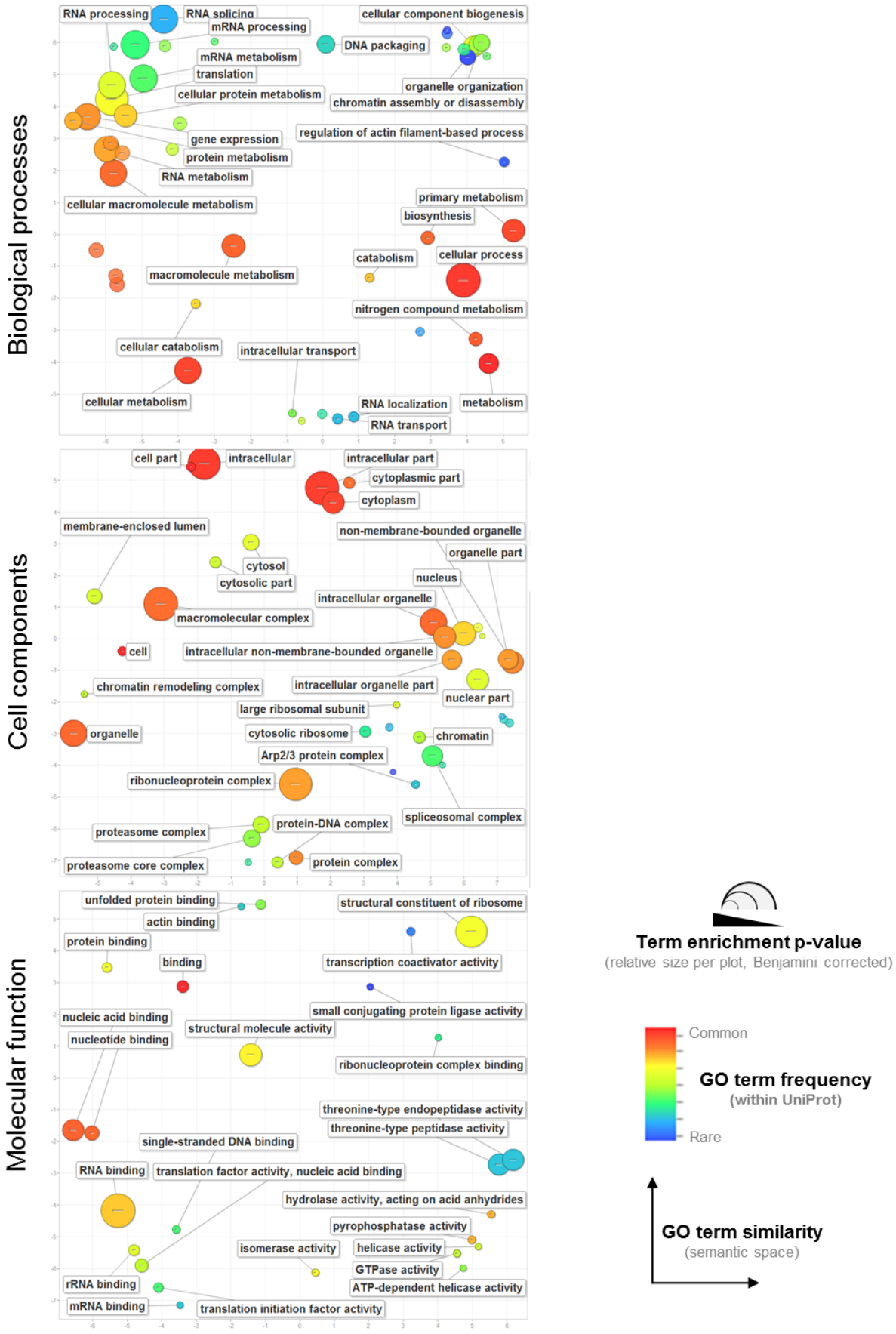
A trend of cellular proteins was consistent amongst the enriched GO terms related to molecular function, with 'structural constituent of ribosome (GO:0003735)' and 'RNA binding (GO:0003723)' both substantially enriched ($p<10^{-30}$). Additional terms identified intracellular functions including transcription ('transcription coactivator activity (GO:0003713)') translation ('translation initiation factor activity (GO:0003743)', 'mRNA binding (GO:0003729)' and 'translation factor activity, nucleic acid binding (GO:0008135)') and the proteasome ('threonine-

type endopeptidase activity (GO:0004298)' and 'threonine-type peptidase activity (GO:0070003)').

Amongst the significantly enriched GO terms derived from Eμ-*myc*-predominant terminal plasma proteins, several appeared in a similar manner to those derived from the analysis of the proteins significantly upregulated in Eμ-*myc* terminal tumour tissue (described in **Chapter 5, Section 5.11**). To investigate the overlap of these terms, Venn diagrams were generated to demonstrate the number of terms common and distinct between the two datasets (**Figure 6.13A**). This analysis identified a sizeable overlap of GO terms between tumour tissue and the plasma derived from terminal Eμ-*myc* mice. For the biological processes, approximately one third (27) of the 82 terms significantly enriched for plasma were common with the B-cell tumour material. Cell component terms exhibited the greatest overlap, with 32 out of the 56 (57%) plasma-derived GO terms common to those derived from tumour cells. Approximately half of the 28 plasma-derived terms (15) were common to both datasets for molecular function.

The $-\log_{10}$ (p-values) for the GO terms, observed for both the terminal Eμ-*myc* tumour and plasma proteins, were plotted to illustrate the similarities and differences in enrichment between the two differentially abundant proteomes (**Figure 6.13B**). Overall, low correlations were observed with R^2 values of 0.23, 0.15 and 0.37 for biological processes, cell component and molecular function, respectively. The predominant trend that emerged was that of the ribosome and translation being more significantly overrepresented in the plasma, resulting from a lower number of input proteins, and therefore higher proportion of enrichment; 'ribosome (GO:0005840)', 'ribosomal subunit (GO:0033279)', 'small ribosomal subunit (GO:0015935)', 'large ribosomal subunit (GO:0015934)', 'structural constituent of ribosome (GO:0003735)' and 'translation (GO:0006412)'. Terms relating to localisation such as 'intracellular (GO:0005622)', 'intracellular part (GO:0044424)', 'nucleus (GO:0005634)', 'organelle (GO:0043226)' and 'intracellular organelle (GO:0043229)' were less enriched in the terminal Eμ-*myc*-dominant plasma proteins.

Figure 6.12. Gene ontology term enrichment analysis for the Eμ-*myc*-specific signature in terminal tumour plasma. GO term enrichment p-values (Benjamini-corrected) were determined for those proteins specifically overabundant in Eμ-*myc* plasma (RS>0.3, RS>Eμ-*TCL1* RS + 0.3) using DAVID. Significantly enriched GO terms ($p<0.05$) were visualised with REVIGO.


Biological processes
RNA processing
RNA splicing
mRNA processing
mRNA metabolism
translation
cellular protein metabolism
gene expression
protein metabolism
RNA metabolism
cellular macromolecule metabolism
macromolecule metabolism
cellular catabolism
cellular metabolism
cellular component biogenesis
DNA packaging
organelle organization
chromatin assembly or disassembly
regulation of actin filament-based process
primary metabolism
biosynthesis
catabolism
cellular process
nitrogen compound metabolism
intracellular transport
RNA localization
RNA transport
metabolism
Cell components
cell part
intracellular
intracellular part
cytoplasmic part
cytoplasm
membrane-enclosed lumen
cytosol
cytosolic part
macromolecular complex
cell
non-membrane-bounded organelle
organelle part
nucleus
intracellular organelle
intracellular non-membrane-bounded organelle
chromatin remodeling complex
intracellular organelle part
nuclear part
large ribosomal subunit
organelle
cytosolic ribosome
chromatin
Arp2/3 protein complex
ribonucleoprotein complex
spliceosomal complex
proteasome complex
protein-DNA complex
proteasome core complex
protein complex
Molecular function
unfolded protein binding
actin binding
protein binding
binding
structural constituent of ribosome
transcription coactivator activity
small conjugating protein ligase activity
nucleic acid binding
structural molecule activity
nucleotide binding
ribonucleoprotein complex binding
threonine-type endopeptidase activity
threonine-type peptidase activity
single-stranded DNA binding
RNA binding
translation factor activity, nucleic acid binding
hydrolase activity, acting on acid anhydrides
pyrophosphatase activity
isomerase activity
helicase activity
GTPase activity
rRNA binding
ATP-dependent helicase activity
mRNA binding
translation initiation factor activity
Term enrichment p-value
(relative size per plot, Benjamini corrected)
Common
GO term frequency
(within UniProt)
Rare
GO term similarity
(semantic space)

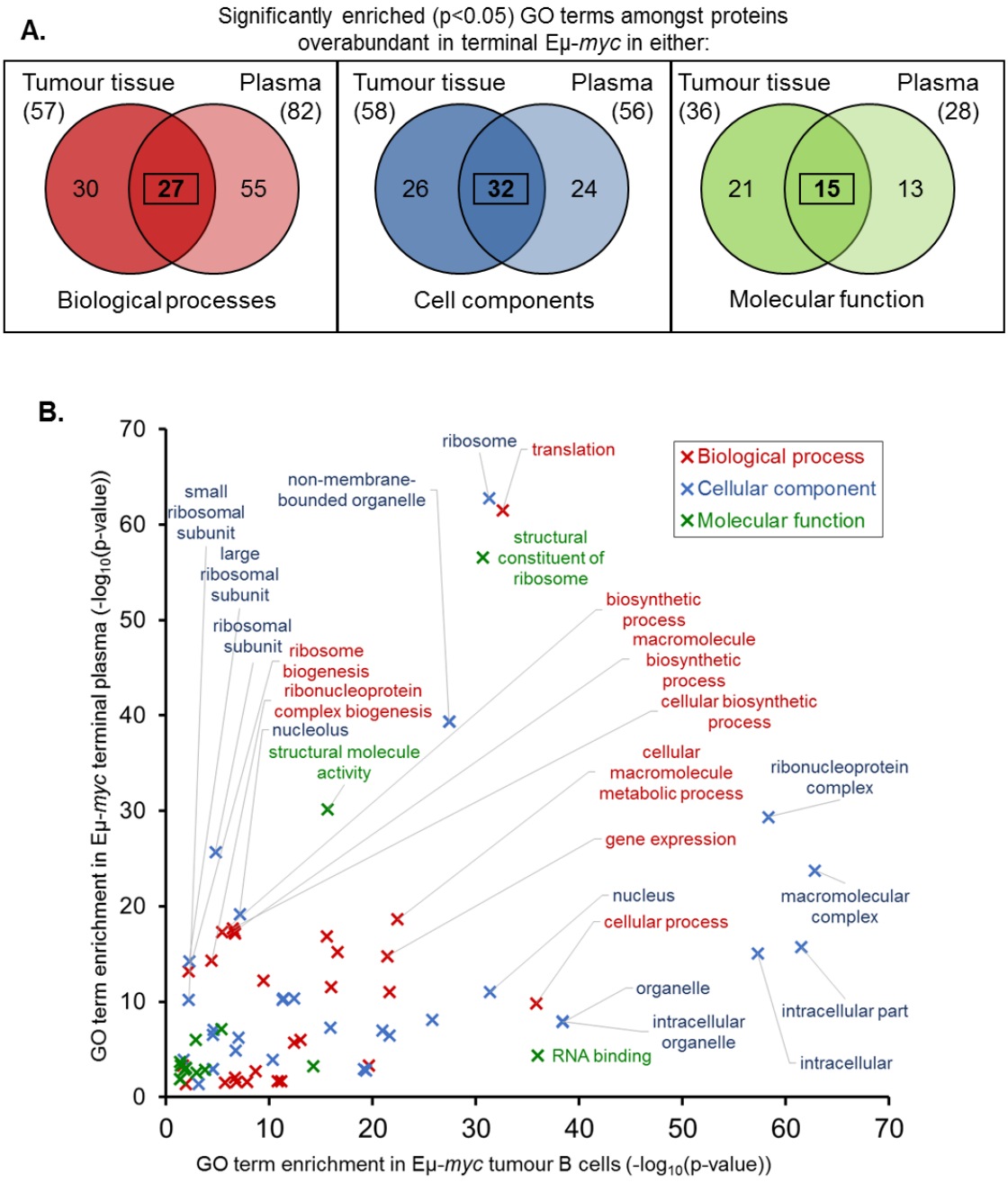


Figure 6.13. Comparison of GO term enrichment between the Eμ-*myc*-specific signature in terminal tumour plasma and Eμ-*myc* B-cell tumours. Significantly enriched GO terms ($p<0.05$) for both those proteins specifically overabundant in Eμ-*myc* plasma (RS>0.3, RS>Eμ-*TCL1* RS + 0.3) and proteins with significant overexpression in Eμ-*myc* tumour B cells (**Chapter 5**) were compared by Venn diagram (**A.**) and the $-\log_{10}$ (GO term enrichment p-values) of those common to both were plotted against one another (**B.**).

6.12 CANONICAL LOCALISATIONS OF PLASMA-IDENTIFIED PROTEIN SIGNATURES

Given that GO term enrichment suggested the overabundant proteins in the plasma of terminal Eμ-*myc* mice predominantly originated from cells, the 2095 proteins quantified in plasma were annotated for canonical localisation with IPA software (**Section 2.21.5**). The canonical localisations of proteins defined as significantly overabundant in the terminal plasma of each model (defined in **Figure 6.4**), the two model-specific signatures (defined in **Figure 6.10**) and the significantly overabundant proteins in the '30%' Eμ-*TCL1* plasma (defined in **Figure 6.6**) were plotted alongside the annotation of all 2095 quantified proteins (**Figure 6.14**). This firstly demonstrated the distribution of canonical protein localisations across all quantitated plasma proteins; identifying the cytoplasm as the most frequent, normal location of over a third (749) of the detected proteins. A fifth (448) of the proteins were annotated as normally located in the extracellular space, more than the 361 (17%) and the 270 (13%) of proteins typically found in the nucleus or localised to the plasma membrane, respectively. For the 43 proteins determined to be overabundant in the plasma of the Eμ-*TCL1* mice with '30%' leukaemia these proteins were distributed in almost exactly equal percentages compared with all protein quantified. Both the protein lists determined as significantly overabundant (727) and those 'dominantly' overabundant (747) in terminal Eμ-*myc* plasma demonstrated identical distributions with 80% of the proteins annotated as intracellular. 35% and 45% of the proteins were annotated as nuclear and cytosolic, respectively, with 7% traceable to the cell membrane and 5% normally found outside cells. A much larger difference was observed between the significantly overabundant Eμ-*TCL1* terminal plasma proteins (344) and those with the Eμ-*TCL1*-'dominant' pattern of expression (127). For the significantly overabundant, terminal Eμ-*TCL1* proteins, a similar trend to that of Eμ-*myc*, was observed, with 70% of proteins annotated as intracellular; 37% and 33% as canonically nuclear and cytosolic, respectively. 11 and 13% of the identified proteins were typically located at the cell surface or in the extracellular space. The Eμ-*TCL1*-'dominant' signature had 6% and 22% of proteins derived from a canonically nuclear or cytosolic origin, respectively. 26% of the proteins were typically plasma membrane proteins, a two-fold enrichment over the background percentage. Proteins annotated as extracellular accounted for 51 of the 127 proteins (40%) in the Eμ-*TCL1*-'dominant' signature, more than any other signature, in both number and percentage.

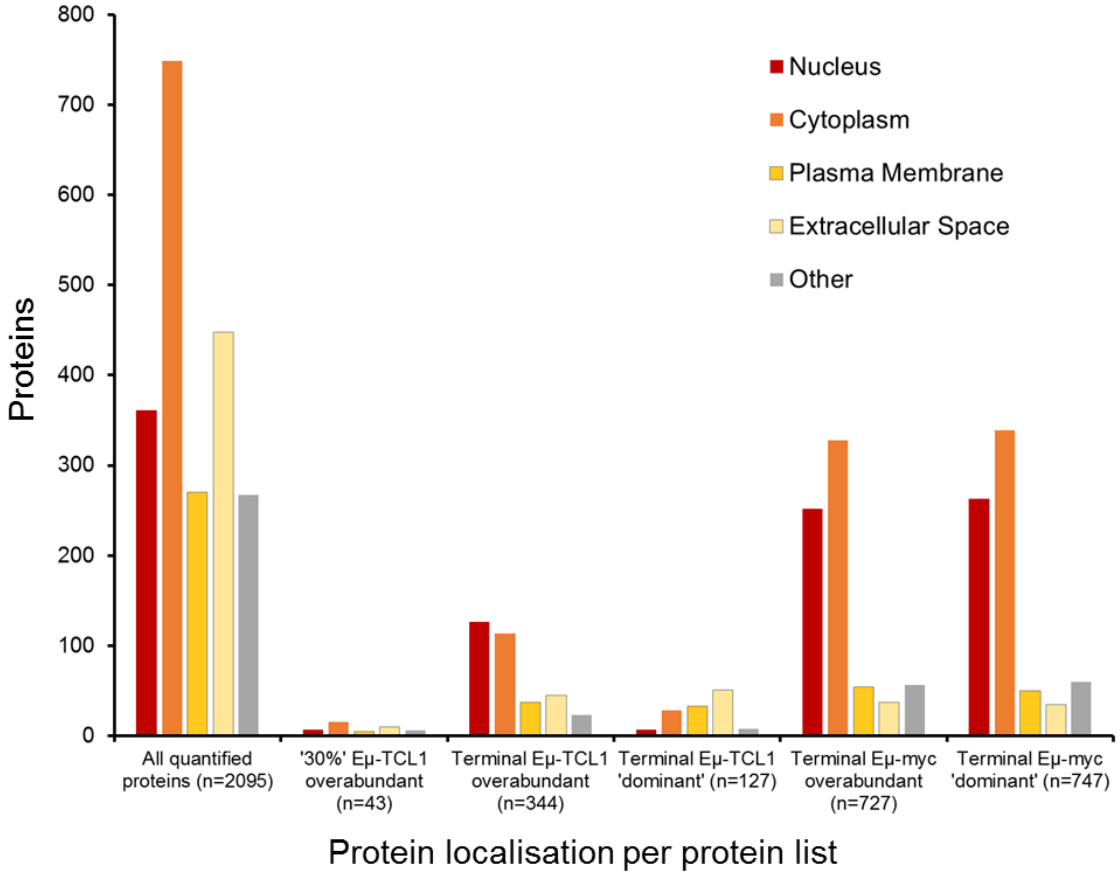


Figure 6.14. Annotation of protein localisations for the signatures of Eμ-*myc* and Eμ-*TCL1* tumours. Proteins defined as significantly overabundant in either tumour plasma (RS>0.5, p<0.05), specific 'dominant' to each tumour plasma (RS>0.3, RS> RS (other tumour) + 0.3) or in the '30%' Eμ-*TCL1* plasma (RS>0.5, p<0.05) were categorised for their predominant protein localisation and compared with all protein identifications.

6.13 Inter-proteome Comparison Between B cells and Plasma

With 70% and 80% of the proteins overabundant in terminal Eμ-*TCL1* and Eμ-*myc* plasma, respectively, annotated as canonically derived from the interior of cells, the question arose as to how many of these proteins were directly traceable to origins within tumour tissue. The overlap between proteins quantified in the plasma and in B cells was evaluated using a Venn diagram (**Figure 6.15A**). This identified that 1156 of the 2095 proteins (55%) were quantified in both proteomes. **Figure 6.15B** presents the distribution of these subsets of proteins amongst the canonical localisations for those proteins identified in; B cells only, plasma only and both B cells and plasma. Those proteins unique to B cells had the greatest degree of cellularity, as expected, with three quarters (4651 out of 6235 proteins) localised to the inside of cells. A further 8%

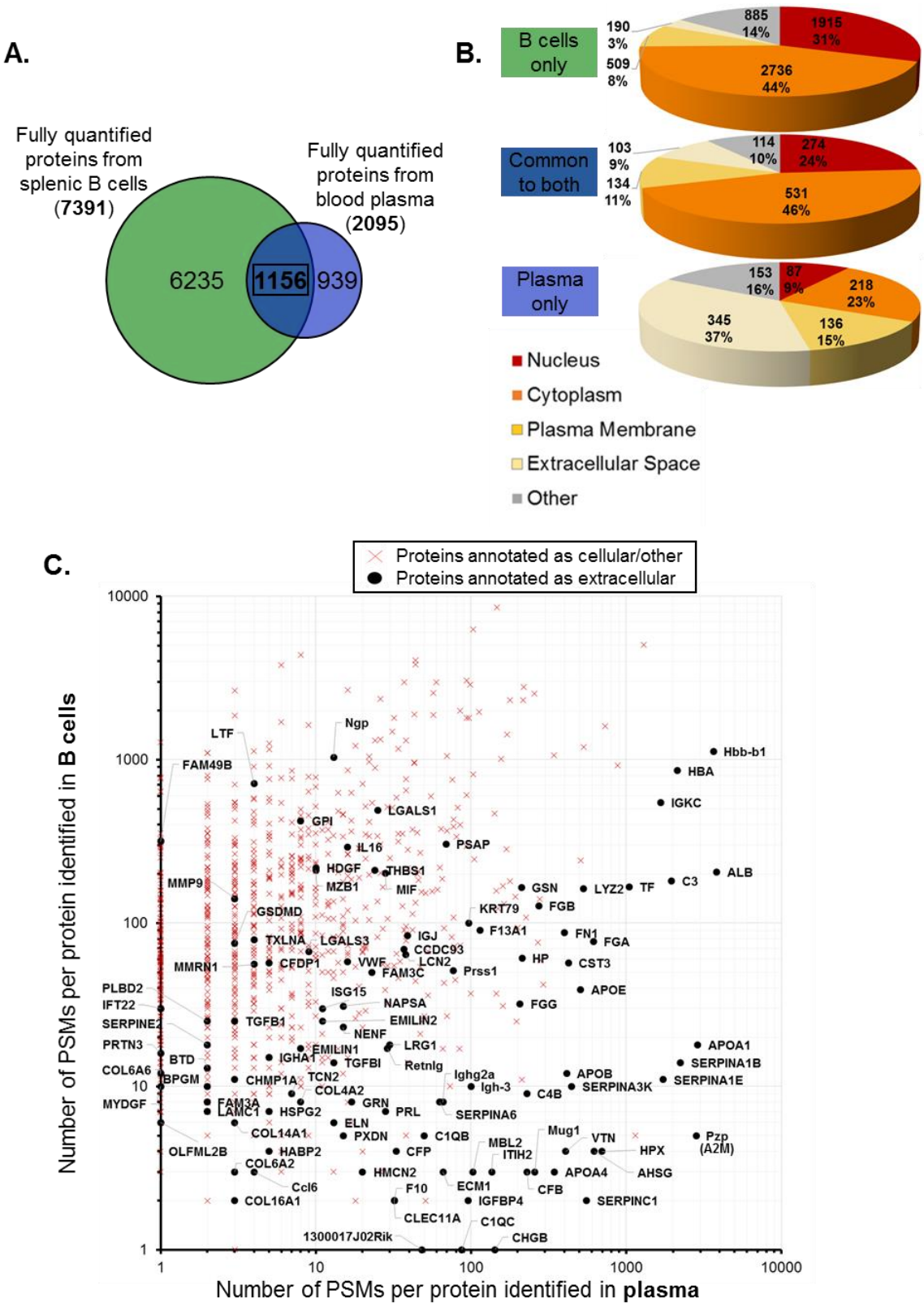
(509) were membrane proteins and 190 proteins (3%) were annotated as extracellular. For the 1156 proteins commonly observed in both proteomes, a similar distribution was observed, with 531 proteins (46%) and 274 proteins (24%) cytoplasm- and nuclear-localised, respectively. A greater percentage of proteins were extracellular; 9% (103) and; membrane-derived; 11% (134). The plasma-specific proteins demonstrated a distinct pattern of localisation, with more than a third of the 939 proteins identified annotated as primarily extracellular. Membrane localised proteins also had the highest proportion of these three groups with a total of 136 proteins (15%). 305 intracellular proteins (32%) were unique to plasma, with 87 and 218 annotated as nuclear and cytoplasmic, respectively.

To better understand how the two proteomes compared, the number of PSMs for those proteins with identifications in both plasma and B cells were plotted (**Figure 6.15C**). Plotting PSMs gave an approximate indication of the overall protein abundance in each proteome, i.e. irrespective of any iTRAQ or TMT quantitations. To identify which of these proteins were expected to have predominant abundance in plasma, those 103 proteins annotated as extracellular (**Figure 6.15B**) were highlighted. Overall the proteins with extracellular annotation were skewed towards plasma-specific expression, whereas, the cellular proteins were dominantly abundant in B cells, with less than 10% with 50+ PSMs. Three clusters of proteins with high abundance in plasma, with high, medium and low abundance in B cells were present. Haemoglobin (HB) subunits HBA and HBB-B1 were observed as contaminating proteins in both proteomes, predominantly in plasma. Alongside these was the antibody component Ig kappa chain C region (IGKC) with around 4-fold greater overall abundance in plasma, compared to B cells. Albumin and complement C3 had approximately 10-fold greater expression in plasma. A group of proteins with, on average 3-10-fold higher abundance in B cells, included; fibronectin (FN1), coagulation factor XIII A chain (F13A1), all three fibrinogens; FGA, FGB and FGG, lysozyme C2 (LYZ2), gelsolin (GSN), serotransferrin (TF), haptoglobin, cystatin-C (CST3) and apolipoprotein E. Five SERPIN proteins A6, C1, A3K, A1B and A1E; and three apolipoproteins; APOA1, APOB and APOA4 were observed in a cluster with, on average approximately 100-fold greater abundance in plasma. Alpha-2-macroglobulin was the most plasma-specific protein observed in both proteomes; in plasma with 2819 PSMs and just 5 PSMs in B cells.

Several proteins, annotated as extracellular, were identified with a greater abundance in B-cell material than plasma, indicative of B cells as a potential secretory source of these protein. These included, most predominantly; protein FAM49B, lactotransferrin (LTF), neutrophilic granule protein (NGP), the tumour-secreted angiogenic factor, glucose-6-phosphate isomerase (GPI), the B220 inhibitor, galectin-1 (LGALS1), IL16, matrix metalloproteinase-9 (MMP9), marginal zone B- and B1-cell-specific protein (MZB1), gasdermin-D (GSDMDC1) and HDGF.

Of these proteins; HDGF and GPI were overabundant in the terminal plasma from both tumours, with no differential expression in any B-cell type; IL16 was significantly underexpressed in both terminal tumours, while significantly overabundant in the terminal plasma and gasdermin-D was upregulated in Eμ-*TCL1* tumours, while only overabundant in the plasma of terminal Eμ-*myc* tumours. MZB1 was specifically overabundant in both the tumours and plasma derived from terminal Eμ-*TCL1* mice. LGALS1 was significantly overabundant in both plasma and B cells of both terminal tumours relative to the respective WT controls.

Figure 6.15. Comparison between the B-cell and plasma proteomes. An inter-proteome comparison between proteins fully quantified in plasma or B-cell characterisations. **A.** The overlap of proteins from the B-cell proteome and plasma proteome. **B.** The canonical localisations of the proteins determined as B-cell specific, plasma specific or common to both proteomes. **C.** The number of PSMs, plotted on a logarithmic scale, identified in plasma and B cells for those proteins commonly quantified in both proteomes. The 103 proteins annotated as extracellular are highlighted (black), while all other proteins are red.

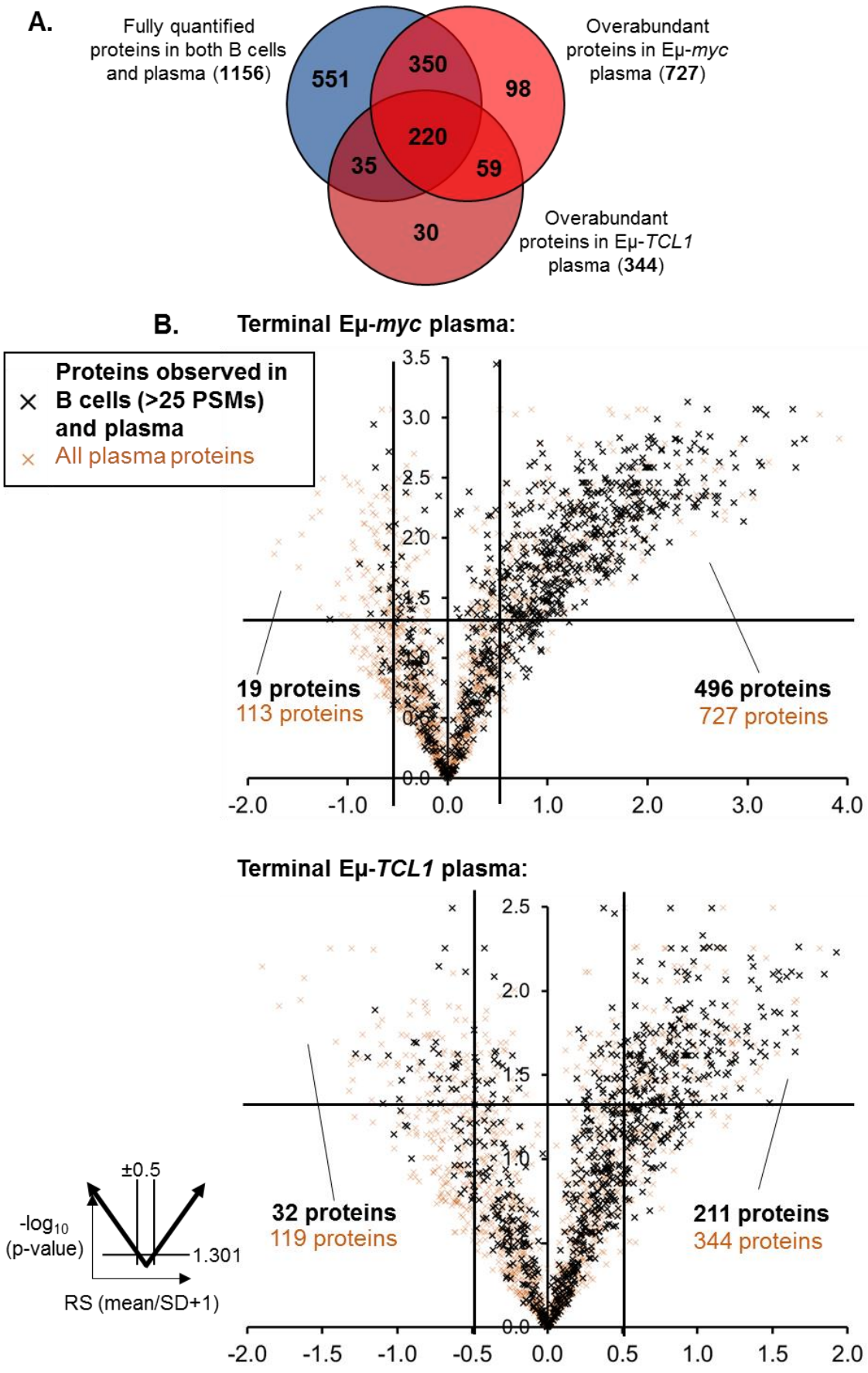

A.
Fully quantified proteins from splenic B cells (7391)
Fully quantified proteins from blood plasma (2095)
6235
1156
939
B.
B cells only
190 3%
885 14%
1915 31%
509 8%
2736 44%
Common to both
103 9%
114 10%
274 24%
134 11%
531 46%
Plasma only
153 16%
87 9%
218 23%
345 37%
136 15%
Nucleus
Cytoplasm
Plasma Membrane
Extracellular Space
Other
C.
Proteins annotated as cellular/other
Proteins annotated as extracellular
Number of PSMs per protein identified in B cells
Number of PSMs per protein identified in plasma
10000
1000
100
10
1
LTF
Ngp
FAM49B
Hbb-b1
HBA
IGKC
LGALS1
GPI
IL16
PSAP
HDGF
THBS1
MZB1
MIF
MMP9
GSN
LYZ2
TF
C3
ALB
KRT79
FGB
GSDMD
F13A1
FN1
FGA
TXLNA
LGALS3
IGJ
CCDC93
HP
CST3
CFDP1
VWF
LCN2
FAM3C
Prss1
MMRN1
APOE
PLBD2
ISG15
NAPSA
FGG
IFT22
SERPINE2
TGFB1
EMILIN2
NENF
PRTN3
EMILIN1
LRG1
APOA1
BTD
IGHA1
TGFBI
Retnlg
SERPINA1B
COL6A6
BPGM
CHMP1A
TCN2
Ighg2a
APOB
SERPINA3K
SERPINA1E
COL4A2
Igh-3
C4B
MYDGF
FAM3A
GRN
LAMC1
HSPG2
PRL
SERPINA6
ELN
COL14A1
PXDN
C1QB
Mug1
VTN
Pzp (A2M)
MBL2
OLFML2B
HABP2
CFP
ITIH2
HPX
COL6A2
HMCN2
APOA4
AHSG
Ccl6
F10
ECM1
CFB
COL16A1
IGFBP4
SERPINC1
CLEC11A
C1QC
1300017J02Rik
CHGB

6.14 TRACING THE OVERABUNDANT PLASMA PROTEOME TO THE B-CELL PROTEOME

To gain an insight into the extent of plasma proteins originating from tumour cells, the overlap of proteins categorised as overabundant in the plasma of each terminal tumour was compared to the 1156 proteins detected in both plasma and B-cell material (**Figure 6.16A**). Of the 727 proteins significantly overabundant in terminal Eμ-*myc* plasma, 570 (78%) were amongst those proteins observed in B cells, approximately half of all proteins detected in both proteomes. 255 (74%) of the 344 proteins overabundant in terminal Eμ-*TCL1* plasma were common with the proteins present in both plasma and B-cell proteomes. Of these 255 proteins, 86% (220) were also overabundant in terminal Eμ-*myc* plasma. Just 187 (24%) of the 792 proteins overabundant in terminal plasma of either tumour were not characterised in B cells.

To demonstrate the distribution of B-cell proteins detected in plasma, those B-cell proteins with >25 PSMs were overlaid onto the volcano plots of differential plasma protein abundance for each terminal tumour described in **Figure 6.4** (**Figure 6.16B**). A 25-PSM threshold was used to filter out proteins which the abundance was low in B cells, and were likely to be predominantly plasma proteins, such as those described in **Figure 6.15C**. The B-cell proteins mapped strongly to the significantly overabundant proteins in terminal Eμ-*myc* plasma, with 68% (496) of the 727 proteins with an abundant origin in B cells. Of the 113 proteins underabundant in terminal Eμ-*myc* plasma, just 19 (17%) had a possible B-cell origin. The same trend was true, but to a lesser extent for terminal Eμ-*TCL1* plasma, with 211 (61%) of the 344 overabundant and 32 (27%) of the 119 underabundant proteins traceable to B cells. In all the cases, the anticipated percentage based on random sampling would have been 40.7%. The trend was also visible outside the defined bounds of significance and regulation score, especially for proteins exhibiting a regulation score of <-0.5, but a p-value of >0.05.

Figure 6.16. Comparison between differentially abundant plasma proteins and B-cell tumour proteins. Proteins considered differentially abundant in either Eμ-*myc* or Eμ-*TCL1* terminal plasma were compared with those proteins with quantitations in both the plasma and B-cell proteomes; **A.** comparing the significantly overabundant proteins in each tumour by Venn diagram and; **B.** overlaying the volcano plots of the terminal plasma proteins with those proteins identified in B cells with at least 25 PSMs.


A.
Fully quantified proteins in both B cells and plasma (1156)
Overabundant proteins in Eμ-myc plasma (727)
Overabundant proteins in Eμ-TCL1 plasma (344)
551
350
98
220
35
59
30
B.
Terminal Eμ-myc plasma:
Proteins observed in B cells (>25 PSMs) and plasma
All plasma proteins
19 proteins
113 proteins
496 proteins
727 proteins
Terminal Eμ-TCL1 plasma:
32 proteins
119 proteins
211 proteins
344 proteins
±0.5
-log10 (p-value)
1.301
RS (mean/SD+1)

6.15 QUANTITATIVE COMPARISON BETWEEN DIFFERENTIALLY ABUNDANT PROTEINS IN THE B-CELL AND PLASMA PROTEOMES

Given that B-cell tumour proteins appeared strongly enriched amongst the overabundant terminal plasma proteins, it was of interest to see which of these proteins were also overexpressed, or underexpressed, in the tumours themselves. **Figure 6.17** describes those proteins determined as either significantly differentially abundant in terminal tumour plasma, which were also detected with significant differential expression in B-cell tumours, for each respective model. In total a signature of over 200 proteins was observed significantly overabundant for both terminal Eμ-*myc* plasma and tumours (**Figure 6.17A**), the most noteworthy of which were; cell division cycle-associated protein 3 (CDCA3), DNA replication licensing factor MCM4, several ribosomal proteins, deoxyuridine triphosphatase (DUT), targeting protein for Xklp2 (TPX2), probable ATP-dependent RNA helicase, DEAD box protein 27 (DDX27), MKI67 FHA domain-interacting nucleolar phosphoprotein (NIFK), nucleolin (NCL), DNA topoisomerase 2-alpha (TOP2A) and ASS1. The majority (120) of these proteins were annotated as cytoplasmic with a further 62 having a likely, nuclear origin. Receptor-type tyrosine-protein phosphatase S (PTPRS) was the only one of these 204 proteins with cell surface expression. Macrophage migration inhibitory factor (MIF) was the only secreted protein, with 202 PSMs in the B-cell proteome and 28 in the plasma proteome.

10 proteins were observed to be significantly underabundant in both terminal Eμ-*myc* tumour B cells and plasma, 5 of which were Ig isoforms and two were structural proteins alpha-synuclein (SNCA) and filamin-B (FLNB). Gamma-soluble NSF attachment protein (NAPG), biotinidase (BTD), H-2 class II histocompatibility antigen and A-U alpha chain (HLA-DQA1) were also observed with significant underabundance in both proteomes. 9 proteins exhibited upregulation in B cells, while still appearing significantly underabundant in plasma. This included three cell surface proteins; sodium-dependent phosphate transporter 1 (SLC20A1), 4F2 cell surface antigen heavy chain (SLC3A2/CD98) and interleukin-1 receptor accessory protein (IL1RAP). The inverse trend was true for 59 proteins with overabundance in plasma and downregulation in tumour cells, amongst which were 9 histone proteins, three actin-related protein 2/3 complex subunits (ARPC) (ARPC4, ARPC5 and ARPCL5) and two proteasome subunits. Other noteworthy proteins included HMGA1, translation machinery-associated protein 7 (TMA7), treacle protein (TCOF1) mitochondrial fission 1 protein (FIS1), LMNB1, sorcin (SRI), the cell surface protein sortilin-related receptor (SORL1), IL16, leukocyte antigen CD37 and the migratory and adhesion promoting proteins PDZ and LIM domain protein 2 (PDLIM2) and LIM domain-containing protein 2 (LIMD2).

Figure 6.17B describes those proteins with significant differential abundance in both the tumours and terminal plasma of Eμ-*TCL1* mice. 41 proteins were present as overabundant in both proteomes including, most substantially; IGKV12-89, CDCA3, ASS1, tyrosine-protein phosphatase non-receptor type 18 (PTPN18), S100A6 and cysteine-rich protein 1 (CRIP1). Also amongst these proteins were four ribosomal subunits (L3, L4, L7 and L34) and two mitochondrial import inner membrane translocase subunits (TIMM8B and TIMM13). As with Eμ-*myc*, the majority of proteins were traceable to cells, with 18 and 13 proteins annotated as cytoplasmic and nuclear. Two cell surface proteins were in this category; neurogenic locus notch homolog protein 2 (NOTCH2) and integral membrane protein 2B (ITM2B). NOTCH2 was identified in plasma with 3 unique peptides matching to aa 1566 to 1622. These peptides all matched to part of the cleavable extracellular domain (aa 26 - 1677). 7 unique peptides in plasma spanned aa 93 to 244 of ITM2B. These peptides were all traceable to the C-terminal extracellular domain of ITM2B (aa 76-266). Four proteins annotated as extracellular were overabundant in both the plasma and B cells of terminal Eμ-*TCL1* mice; LGALS1, MZB1, napsin-A (NAPSA) and prolactin (PRL).

Terminal Eμ-*TCL1* plasma and B-cell tumours had 10 proteins with consistent inter-proteome underabundance. Of these proteins, the most noteworthy were SNCA, the plasma protein alpha-2-HS-glycoprotein (AHSG) and the GTPase ras-related protein Rap-1b (RAP1B). Disintegrin and metalloproteinase domain-containing protein 19 (ADAM19), histone H2A-2B, tripartite motif-containing protein 65 (TRIM65), alpha globin 1 (HGA1), C6orf25 and three Igs were also amongst these proteins. 6 proteins were upregulated in tumours while significantly underabundant in plasma; SLC3A2, YTH domain-containing family protein 2 (YTHDF2), RPS6, hydroxyacylglutathione hydrolase, mitochondrial (HAGH), replication initiation factor MCM10 homolog (MCM10) and voltage-dependent anion-selective channel protein 2 (VDAC2). 16 proteins were oppositely differentially abundant; 9 of which were chromosomal proteins (HMGA1, HMGN2 and 7 histone isoforms), alongside 5 other intracellular proteins (cyclin-dependent kinase inhibitor 1B (CDKN1B), coiled-coil domain-containing protein 23 (CCDC23), 3-hydroxyanthranilate 3,4-dioxygenase (HAAO), NIFK and LMNB1. A single plasma membrane protein, lymphocyte antigen 6D (LY6D), was identified overabundant in Eμ-*TCL1* plasma with 2 peptides (aa 34-60) matching the mature protein region (aa 21-98). A single peptide (aa 47-60) was downregulated in Eμ-*TCL1*, as well as Eμ-*myc* B-cell tumours. Two secreted protein were observed overabundant in terminal Eμ-*TCL1* plasma while downregulated in the tumour tissue; coagulation factor V (F5) and IL16. For the '30%' Eμ-*TCL1* plasma proteins, just ASS1 and MCM4 were significantly overabundant, while also significantly overexpressed in terminal Eμ-*TCL1* tumours.

To summarise these finding, those 24 proteins reaching significant overabundance in both the terminal tumours and plasma of both the Eμ-*myc* and Eμ-*TCL1* models were tabulated (**Figure 6.18**). This again highlighted both the dominance of Eμ-*myc* tumours and the frequency of intracellular proteins, with 13 cytoplasmic proteins and 11 nuclear proteins. The signature of upregulated translation was apparent in this list with 6 proteins; eukaryotic translation initiation factor 3 subunit A (EIF3A), aminoacyl tRNA synthase complex-interacting protein 2 (AIMP2) and 4 ribosomal proteins (RPL3, 4, 7 and 34). 5 mitochondrial proteins were present; iron-sulphur cluster assembly enzyme, mitochondrial (ISCU), ASS1, DUT, TIMM8B and TIMM13. A reproducible, pre-terminal signature of marginal overabundance was observed in the '30%' Eμ-*TCL1* plasma for 6 proteins; EIF3A, TIMM13, ASS1, MCM4, Huntingtin-interacting protein K (HYPK) and ICSU.

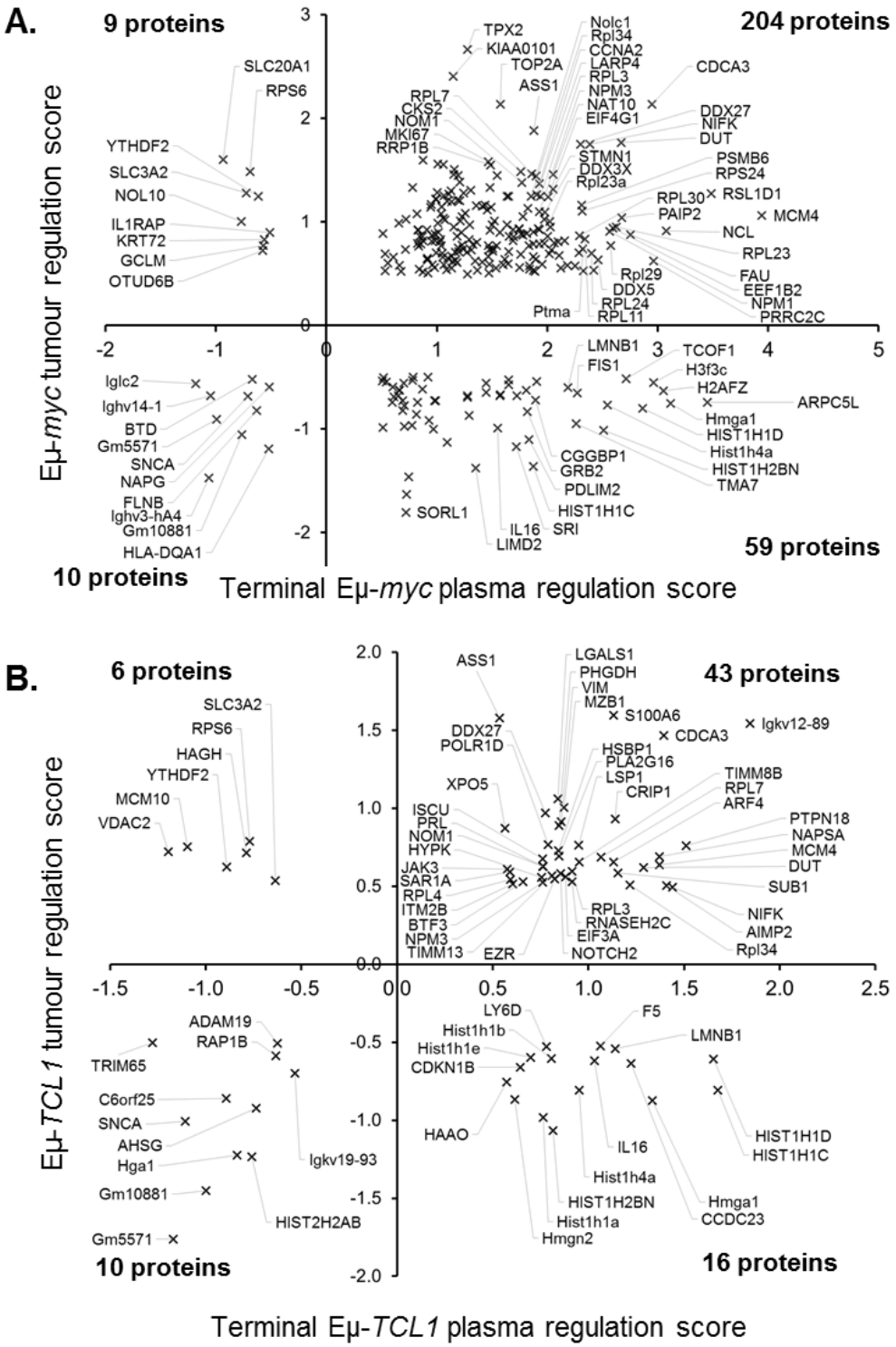


Figure 6.17. Significantly differentially abundant proteins in both plasma and B-cell tumours. Proteins with either significant overabundance (RS>0.5, p<0.05) or underabundance (RS<-0.5, p<0.05) in the terminal plasma and B-cell proteomes were filtered and plotted for each B-cell cancer model; **A.** Eμ-*myc* and **B.** Eμ-*TCL1*. The number of proteins in each quadrant is shown.

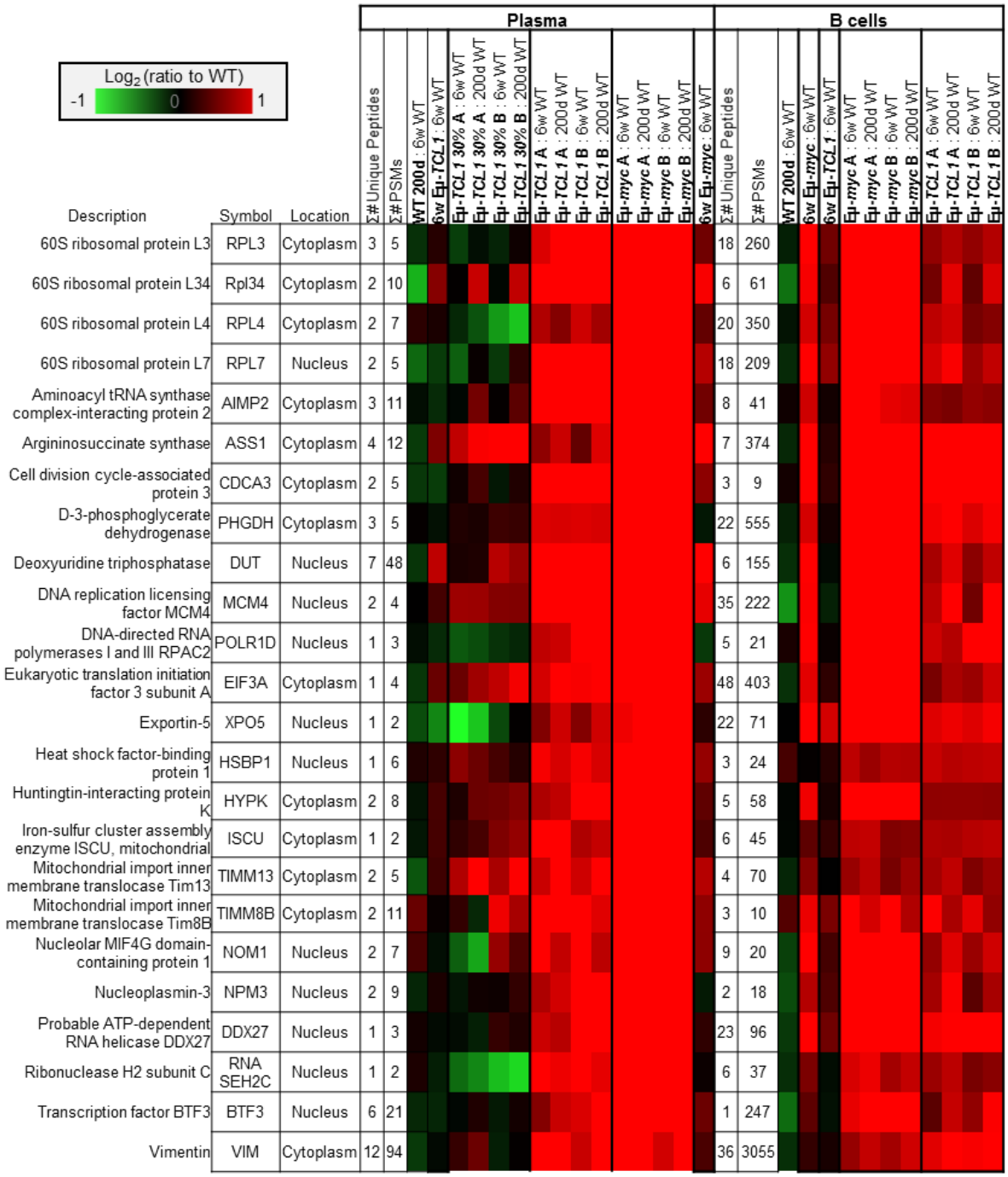


Figure 6.18. A plasma signature of B-cell tumours. Proteins reaching significant overabundance (RS>0.5, p<0.05) in both the terminal plasma and tumours of both the Eμ-*myc* and Eμ-*TCL1* B-cell cancer models.

6.16 DISCUSSION

The characterisation described in this chapter aimed to utilise methodological developments in plasma proteomics to capture protein signatures of B-cell cancers detectable in the blood plasma. Such observations held potential to provide further understanding of tumourigenesis in the context of the whole organism. This analysis presents one of the most comprehensive proteomic characterisations of plasma to date, with a greater depth of coverage than any other published plasma proteomes focusing on cancer biomarker discovery. While other proteomes have exceeded this coverage, they have either focused on non-cancer samples [457] or have used lower thresholds of false discovery ($q<0.05$) [321], compared with the more stringent FDR threshold ($q<0.01$) applied in this analysis.

The implementation of SuPrE-SEC for the isolation of the low molecular weight proteome offered an effective means of depleting albumin and other high abundant proteins. This was firstly demonstrated by the number of PSMs matching to albumin. While still high (3835), the number of PSMs was substantially lower than that of transthyretin (7769) and apolipoprotein A-II (7129). These were both proteins with molecular weights below 16 kDa; anticipated to be enriched in the low M_w sub-proteome. This enrichment and depletion enabled a more effective shotgun proteomics analysis, given that otherwise high abundant protein peptides would have caused signal suppression and DDA bias.

Limitations were present in this SuPrE-SEC analysis, however: Substantial abundance of Ig in the 200 day WT sample, observable in the clustering (**Figure 6.3**), was likely an artefact of background reduction caused by the age and storage of the 200 day WT samples, which for logistical reasons were kept at -80°C for a month prior to transfer to liquid nitrogen, where other samples were stored immediately. This artefact highlights the sensitivity of plasma to storage conditions and the need for rigorously controlled sample procurement and handling. This may have partly explained the previously observed variability when quantifying the low M_w proteome [321], in addition to the potential of minor cut point variation induced by the relatively crude nature of SEC compared with other chromatographic methods. The biological replicates allowed these issues to be circumvented by the stringent use of statistics and RS, conserving only the most robust observations.

Another limitation was the incomplete nature of this characterisation, focusing only on the low M_w sub-proteome. While this may have missed several interesting findings, it was considered that this isolated analysis of the low M_w sub-proteome provided the greatest depletion of high abundant proteins providing the greatest possible depth of proteome coverage. Interrogating the other sub-proteomes would have exponentially increased the analysis time and

was further hampered by poor protein solubility in several of these fractions. Another consideration was that the low M_w proteome contained the degradome and peptidome, a portion of plasma containing partially digested and degraded proteins which have been suggested to be a source of biomarkers [458-460]. While these proteins may not have functional relevance within the blood, they present a signature capable of acting as a biomarker for the prognosis and diagnosis of cancers and other diseases.

The use of pooling also substantially limited the statistical power of the analysis and potentially introduced artefacts from single samples, such as the variable pre-tumour controls. While the aim was to generate a robust analysis, for the tumours, biological replicates were used to minimise this effect and increase the confidence in findings. Overall, despite the limitations, the results provided a highly detailed quantitative analysis of plasma, likely far in excess than what might have been derived from the less focused analysis of all four sub-proteomes, or by immunodepletion strategies.

The quantitative proteomics identified a strong signature for the Eμ-*myc* terminal plasma with over a third of quantified proteins exhibiting overabundance (**Figure 6.3** and **Figure 6.4**). While a strong signature was expected, this signature was almost certainly attributable to proteins negligible or not present in the control plasma. The predominant source of these proteins was expected to be from tumour lysis products generated by the high rate of apoptosis observed in Eμ-*myc* tumours. This substantial overabundance will have biased the analysis for the detection of these proteins, explaining the magnitude of this signature. The observation of tumour lysis products in the plasma was confirmed and evaluated by several analyses.

Figure 6.4 and **Figure 6.5** presented general observations of the most overabundant proteins belonging to typically intracellular protein families, such as ribosome and proteasome components. **Figure 6.9** explored the differential abundance of proteins in Eμ-*myc* tumour plasma compared to WT plasma in relation to their experiment-wide abundance. This analysis confirmed that the most abundant, terminal Eμ-*myc*-overabundant proteins in the total analysed plasma were cellular-derived, strongly emphasised by several histones. The trend was explored further by demonstrating and evaluating the signature of proteins which were dominantly overabundant in Eμ-*myc* terminal plasma (**Figure 6.10**).

GO term enrichment of this Eμ-*myc* terminal plasma-dominant signature (**Figure 6.12**) revealed a trend which was similar in many ways to that of the GO term enrichment for upregulated tumour proteins in **Figure 5.12**. This trend demonstrated a sizeable overlap and some correlation between GO terms when comparing upregulated tumour proteins with overabundant terminal plasma proteins for Eμ-*myc* mice (**Figure 6.13**). The resulting GO term

signature, especially the cell component enrichment confirmed an overrepresentation of intracellular parts. Evaluation of the canonical localisations of the Eμ-*myc*-overabundant proteins (**Figure 6.14**) also demonstrated this trend of intracellular proteins.

Direct comparison between the proteomes describing plasma and B cells identified that over two thirds of the significantly overabundant terminal Eμ-*myc* plasma proteins were traceable to proteins confidently identified in the B-cell proteome (**Figure 6.16B**). Quantitative comparison highlighted the signature of more than a quarter of significantly overabundant proteins in terminal Eμ-*myc* plasma were also identified with significant upregulation in the Eμ-*myc* tumours (**Figure 6.17**).

Together, these observations confirmed that substantial tumour lysis was occurring in the Eμ-*myc* model. Tumour lysis is a well-documented occurrence especially in the context of aggressive therapies and lymphomas [461]. Tumour lysis syndrome describes a rapid cell death of tumour cells which can overwhelm the circulatory system with intracellular contents [462, 463]. The detection of this background level of tumour lysis confirms the principle that tumour lysis products have potential in the diagnosis of cancer through non-invasive testing of blood products. However, this may be exaggerated in the mouse model given that the Eμ-*myc* tumours can typically make up a large proportion, in excess of 10%, of a normal mouse body weight. Another potential mechanism which may explain the emergence of these proteins in the circulation is that of exosomes [464].

The Eμ-*myc* 6-week plasma signature was surprisingly minimal relative to the terminal signature, when considering the B-cell signature seen in **Chapter 5**. Given the tumour-like trends observed for the Eμ-*myc* 6-week B cells (**Figure 5.4**), a far greater clustering was expected than that seen in **Figure 6.3**. While these plasma and B-cell samples were procured form the same 6-week Eμ-*myc* mice, this trend was likely due to the minimal advancement of any tumour material in these animals. Given that the average splenocyte count per 6-week Eμ-*myc* spleen was ~50 x 10^6 cells compared with ~250 x 10^6 cells in the terminal spleens, it was unlikely that the extent of tumour lysis was anywhere near the extent observed in the terminal tumours. Some signatures of protein overabundance were observable in the Eμ-*myc* 6-week plasma, however, which corresponded with some of those proteins most consistently observed in terminal tumour plasma (**Figure 6.5**). This indicated that a low level of background lysis was occurring prior to any advanced tumourigenesis which could be detected predominantly as an overabundance of, for instance, histone proteins. This was further emphasised in **Figure 6.7**, an analysis focused on the identification of pre-terminal markers in the Eμ-*TCL1* model. **Figure 6.8** evaluated this trend more widely, considering the terminal differential abundance of proteins which had differential abundance at 6 weeks. This additionally suggested a pre-terminal

signature of tumour development. The proteins detailed in **Figure 6.18** confirmed that for several proteins there was traceability of prospective pre-terminal biomarkers between the tumours and plasma. The use of pooling and a lack of technical replicates, however, limits the conclusions that can be drawn from this analysis, as a single plasma sample in the pool could have substantially biased the pooled average.

The trend observed for Eμ-*TCL1* terminal plasma also demonstrated a signature of tumour lysis products. However, this trend was substantially less than that seen with Eμ-*myc* tumours. This is demonstrated by the observation of fewer differentially abundant proteins with a lesser extent of differential abundance in the Eμ-*TCL1* terminal plasma (**Figure 6.4**). Much like the trend observed in **Chapter 5** (**Figure 5.7**) for the tumour cell material, protein expression for commonly overabundant proteins was consistently lower for Eμ-*TCL1* terminal plasma than that of Eμ-*myc* terminal plasma (**Figure 6.5**). The nature of the Eμ-*TCL1* terminal plasma signature was again demonstrated to be related to tumour lysis by the observation of several histone proteins amongst the high abundance, overabundant proteins (**Figure 6.9B**). To understand how this signature differed from the Eμ-*myc* terminal plasma signature, the two RS were plotted against one another, revealing two very specific trends of protein expression. While these trends were not completely discrete to either tumour, they demonstrated a signature that emerged with a dominant expression in each tumour (**Figure 6.10**). While for Eμ-*myc* terminal plasma this was highly related to cell lysis, for Eμ-*TCL1* terminal plasma, the proteins were predominantly extracellular and suggested immune function was upregulated in the plasma (**Figure 6.11**). Closer evaluation (**Figure 6.14**) demonstrated that, while overall Eμ-*TCL1* terminal plasma contained many proteins with an intracellular localisation, the Eμ-*TCL1*-dominant terminal plasma signature consisted of predominantly extracellular proteins. This implied that for Eμ-*TCL1* tumours, an immune response is present which, while also present in Eμ-*myc* terminal plasma is far less apparent. This included the overabundance of proteins such as fibrinogens, ITIH proteins, cathepsins, Igs, cytokines and hormones indicative of immune system activation. A possible explanation for this may lie in the speed at which Eμ-*myc* tumours develop which can progress from asymptomatic to terminal in days [176, 188]. For Eμ-*TCL1* tumours this process is far slower with a time course of many months, which may provide a backdrop for the accumulation of several proteins involved in chronic inflammation. This presents the possibility that cancers with a slower development may have greater potential in being characterised by biomarkers derived from the immune response to cancers.

Another trend which emerged was that of cell surface proteins present in the plasma of terminal Eμ-*TCL1* mice (**Figure 6.11**). This observation of increased surface protein entry into the circulation, suggested that exosomes and shedding may have been mechanisms of plasma

protein appearance from Eμ-*TCL1* tumours. NOTCH2 and ITM2B were upregulated in the Eμ-*TCL1* tumours and peptides matching the extracellular domains of these proteins were identified as overabundant in the Eμ-*TCL1* terminal plasma (**Figure 6.17**). The observation of only the extracellular domains, coupled with the previous observations of the NOTCH2 extracellular domain cleavage [465] strongly supports shedding as a mechanism of tumour proteins appearing in the circulation and provides two confident examples of this process.

This was reinforced by the observation of similar GO term enrichments for the downregulated B-cell tumour proteins. The observation of the lymphocyte antigen, LY6D, downregulated in Eμ-*TCL1* tumours, while appearing overabundant in the plasma (**Figure 6.17B**) suggested that the extracellular domain of this protein was shed into the circulation. The suggestion that a lack of LY6D expression in lymphocytes correlates with increased pluripotency [466] suggests that this shedding may be a critical mechanism in tumourigenesis, which could offer potential as a mechanistic biomarker. The interference of this shedding mechanism could also be investigated as a means of reducing tumourigenesis, such as specific protease inhibitors.

Protein secretion was also an apparent mechanism, with several protein identification suggesting tumour cells as the most likely origin of plasma proteins, based on PSM numbers suggesting cell-dominant abundance. These proteins were identified as overabundant in both the tumours and tumour plasma and are canonically annotated as extracellular proteins. The example of MZB1, a known secreted protein with a high degree of specificity for B-cell expression, supported by the relative abundances based on PSM numbers demonstrated by **Figure 6.15**, indicates that proteins upregulated and secreted by the Eμ-*TCL1* tumour can be detected in the plasma. This highlighted a particularly interesting mechanism, given that MZB1 has been observed as an inhibitor of B-cell proliferation [467]. Secretion may therefore be acting as a mechanism by which the attenuation of MZB1 is avoided. Combined with the observation of prognostic value in the prediction of CLL aggression [468], the secretion of this protein may have prognostic and diagnostic value in human B-cell cancers.

Another interesting correlation observed between the B cells and plasma proteomes was that of hyaluronan metabolism enrichment observed in the Eμ-*TCL1*-dominant terminal plasma signature (**Figure 6.11**), in combination with the upregulation of the hyaluronan receptor, HMMR/CD168 as the most consistently upregulated cell surface protein in both tumours (**Figure 5.18**). The plasma enrichment was based upon the overabundance of several ITIH proteins, capable of stabilising the ECM by binding hyaluronan, which exhibited significant overabundance in the plasma of both terminal tumours, but to a greater extent in the Eμ-*TCL1* terminal plasma. The ITIH proteins have been proposed to be tumour suppressors which are

downregulated in many cancers [469, 470], also observed here, at least for ITIH2 in the B-cell tumours. While the mechanisms relating to the ITIH proteins are unclear, the observation of an overabundance of both extracellular hyaluronan binding proteins and a tumour-overexpressed hyaluronan receptor suggests a role for these proteins in supporting cancer hallmarks. The interference of hyaluronan binding may have some potential for further investigation. In addition to the biology, the ITIH proteins appeared with high concentration in the plasma suggesting potential for easy detection in the context of biomarkers. The ITIH proteins also appeared overabundant in some regard in the plasma of all pre-terminal states of the transgenic mice.

This investigation proved effective in the identification of a pre-terminal plasma signature of Eμ-*TCL1* tumours. As demonstrated by **Figure 6.6A,** this signature was minimal but had a sizable overlap with that of the terminal signature. This was highlighted in detail by **Figure 6.7** which suggested a list of proteins which were significantly overabundant in terminal Eμ-*TCL1* plasma, which also demonstrated overabundance in the '30%' Eμ-*TCL1* plasma. Interestingly, this highlighted that the majority of these proteins also demonstrated overabundance in the Eμ-*myc* terminal plasma. This implied that these proteins may have utility for identifying the early stages of other B-cell cancers.

Despite these similarities to the terminal signatures, the majority of these proteins were relatively low concentration in the plasma; inferred from their PSM number. **Figure 6.9** demonstrates this issue, highlighting that only a few significant overabundant proteins were observed with a confident number of PSMs. While there are a number of other proteins that exhibit a marginal overabundance, this result highlighted the issues faced by biomarker discovery. The signatures presented by these results are low abundance and in most cases only marginally detectable as differentially abundant. While proteins were detected in the plasma of the 6-week Eμ-*TCL1* mice with differential abundance (**Figure 6.6B**), the correlation of these proteins with both terminal and '30%' abundances were low. Given this low correlation and the lack of replicates for this condition, conclusions were not drawn for the 6-week Eμ-*TCL1* plasma.

Overall, this chapter has successfully demonstrated the characterisation of proteins in plasma explained by immune response to tumours, tumour lysis products, tumour-secreted proteins and cell surface proteins shedding from tumours. In varying combinations these mechanisms appear differentially in the two tumour types, in relation to tumour aggression and phenotype. This supports the potential of the implemented methods to provide a high depth of quantitative proteomic characterisation of plasma. Additionally, it strongly suggests that continued biomarker discovery, with continued methodological developments, which to date claims only minimal success, will yield discoveries that can help in the diagnosis and prognosis

of cancers. Given that the method presented here analysed a total of only 1.2 ml of plasma, enrichment strategies using greater plasma volumes, combined with the sensitivity of the latest MS instrumentations therefore promises huge potential in biomarker discovery. While limitations still appear in these strategies, this study presents some proof-of-principle findings that should be attempted to be reproduced in human cancers, on larger scales and using more rigorously procured samples – to avoid the issues seen with Ig in the 200 day WT samples. The combined biological characterisation of plasma and tumour samples also provides a unique insight into tumour mechanisms such as protein secretion and shedding, immune system responses and potential regulation of cellular interactions with the microenvironment.

7.0 QUANTITATIVE PROTEOMIC CHARACTERISATION OF CLL

7.1 CHAPTER INTRODUCTION

Proteomics of the plasma and B cells of the Eμ-*myc* and Eμ-*TCL1* mouse models (**Chapters 5 and 6**) proved to be an effective means of biological characterisation. Quantitative proteomics successfully profiled several previously described model-specific and cancer-related protein expression changes. The resulting data also identified novel mechanisms, such as the role of IL5 and IL5RA in the proliferation and survival of Eμ-*TCL1* tumour cells. Further investigation, however, demonstrated that this mechanism was not recapitulated in human CLL, with neither differential IL5RA expression, nor IL5 sensitivity. This suggested that critical differences exist between the biology of mouse and human B-cell cancers. Given the success of the proteomic methodologies, but the limitations of modelling CLL with the Eμ-*TCL1* mouse model, it was concluded that characterisation of human CLL samples would provide a greater deal of insight into CLL biology.

Several studies had previously characterised CLL using proteomics; with the latest, a characterisation of U-CLL versus M-CLL, fully quantitating 2024 proteins; the most comprehensive published study to date [352]. Proteomics characterisation of primary human CLL samples, however, presented several challenges, not present with the mouse study. The

acquisition and processing of human CLL samples required clinical expertise and an established infrastructure of ethical approval, standardised collection, characterisation and storage. Human CLL samples also exhibit a far greater degree of disease heterogeneity.

With the Eμ-*TCL1* mouse, a single transgene was used to elicit a phenotype modelling CLL. This incorporation of a clear transgene-induced cancer aetiology in an inbred black 6 mouse strain, kept in tightly controlled environmental conditions, meant pooling was likely to be sufficient to overcome the majority of inter-sample differences. Far less control could be exerted over the environmental factors, the underlying genetics, oncogenic factors and disease stages of human samples. Combined with the known heterogeneity of CLL, this posed an issue when faced with the limited capacity of single iTRAQ or TMT reagent sets. To overcome these capacity limits, common controls, termed 'bridging' controls, were used in each multiplex experiment which allowed robust comparisons of each sample between different multiplexes. Bridging controls enabled the characterisation of individual samples, avoiding several of the limitations observed with pooling, allowing the inter-sample differences in CLL to be observed and quantified.

The following chapter details a proteomic investigation characterising 14 individual CLL samples across two TMT 10-plex experiments compared in parallel using three healthy donor B-cell samples as bridging controls. The characterisation of these individual samples enabled the identification of confident, robust characteristics of CLL and the statistical comparisons of differential protein expression in CLL subtypes. The work completed in this chapter has also provided an opportunity to identify similarities and differences in B-cell cancer protein expression between human CLL and the CLL-like cancer which develops in Eμ-*TCL1* mice, described in **Chapter 5**.

7.2 Experimental Design and Sample Preparation for the Proteomics Characterisation of CLL

The heterogeneity of CLL meant good experimental design and sample selection was critical to identifying potential subtype-specific and subtype-independent protein signatures and characteristics (**Figure 7.1**). While several characteristics were considered, it was concluded that CLL cases carrying a mutation in either the *NOTCH1* (N=5) or *SF3B1* (N=5) genes compared to CLL cases without a mutation in either gene, termed 'WT' (N=4), would provide a meaningful sample set for analysis. Amongst these 14 samples was also a distribution of several other characteristics, including; IGHV mutation status (7 M-CLL, 7 U-CLL), trisomy 12 (n=5), $CD38^{+}$ (3 with >95%), 13q14 deletion (8 ++, 5 +- and 1 --), and gender (3 female, 11 male). While fresh CLL PBMCs were considered to be ideal, frozen samples were used for practical purposes, as

these allowed for the use of highly characterised samples at known stages to be evaluated. To accommodate for the use of frozen samples, the healthy donor PBMCs were isolated and frozen in an identical way, using DMSO to maintain viability. B cells were then isolated from defrosted, healthy donor and CLL patient-derived PBMCs using negative selection without the depletion of CD43, a protein with potential expression on B-cell cancers. B cells were washed extensively to minimise contaminating debris and adherent extracellular protein. $CD19^+$ cell purity was validated with flow cytometry and B-cell pellets were snap frozen. Lysates were prepared from these pellets by rigorous solubilisation and 100 µg of protein aliquoted for each CLL lysate and 200 µg of protein aliquoted for each healthy donor B-cell (HD) sample for proteolysis and isobaric labelling. CLL sample subtypes were distributed equally across the two 10-plexes to minimise any differences between the two rounds of 2D LC-MS/MS or any differences in quantitative analysis. For the HD bridging controls, to minimise variability, pools of 200 µg of protein were proteolysed, labelled and then bifurcated with 100 µg of peptides allocated to each 10-plex pool.

The two pools of 10 labelled HD and CLL peptides, denoted 10-plex A and 10-plex B, were separately subjected to peptide prefractionation to reduce sample complexity. Across the approximately 100 peaks of the two chromatograms, the reproducibility was extremely high, with the average absolute deviation of peak elution times being less than 5 seconds for 89 of the common fractionation points, over 120 minute gradients. The top 25 peaks, common across both 10-plexes, were analysed individually, whereas the remaining, lower-content fractions were pooled by orthogonal concatenation, pairing from the extremes of chromatographic retention times. This resulted in 60 fractions or fraction pools for each 10-plex which were subjected to approximately 200 hours of LC-MS per 10-plex.

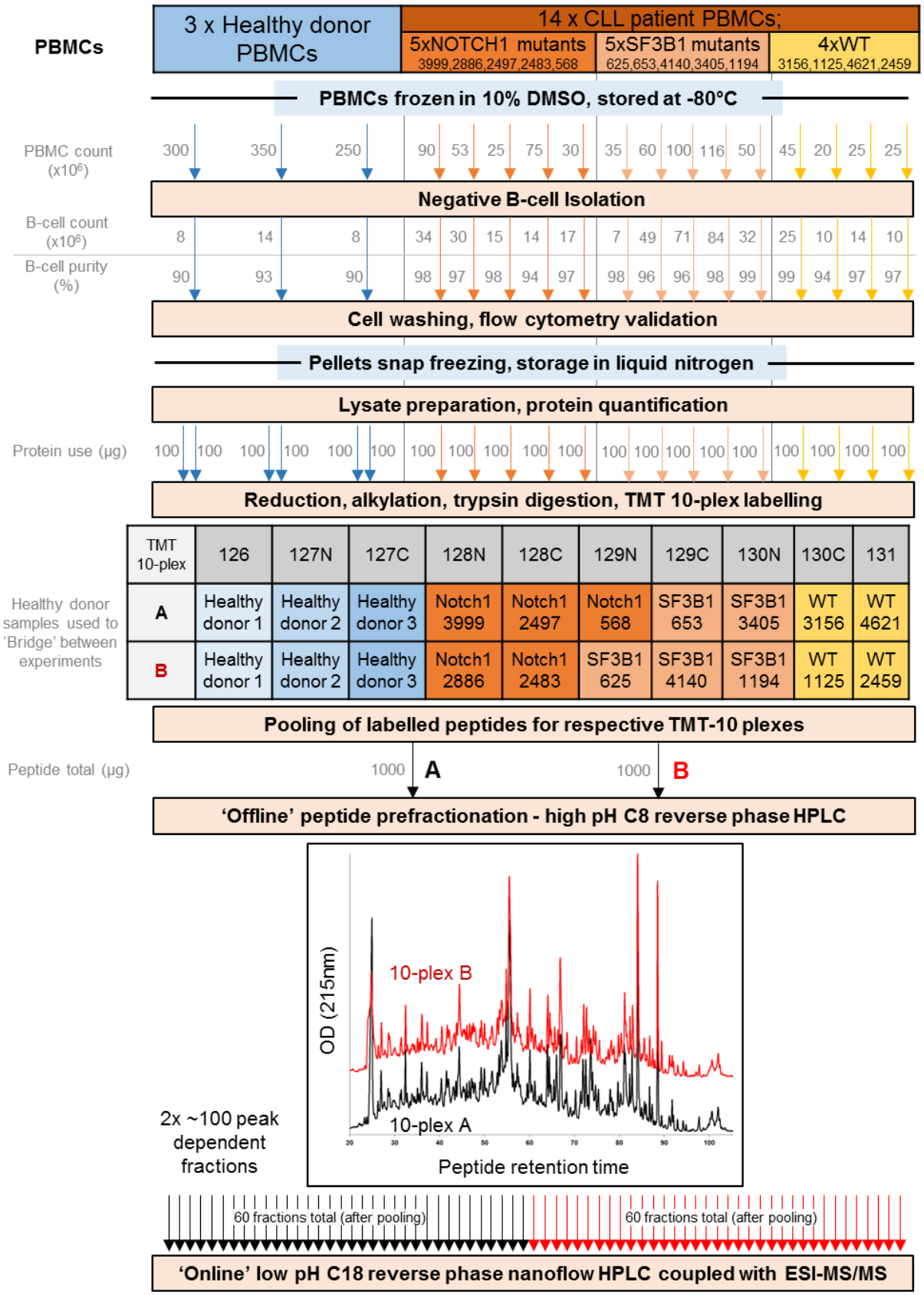

PBMCs
3 x Healthy donor PBMCs
14 x CLL patient PBMCs;
5xNOTCH1 mutants
3999,2886,2497,2483,568
5xSF3B1 mutants
625,653,4140,3405,1194
4xWT
3156,1125,4621,2459
PBMCs frozen in 10% DMSO, stored at -80°C
PBMC count (x10⁶)
300 350 250 90 53 25 75 30 35 60 100 116 50 45 20 25 25
Negative B-cell Isolation
B-cell count (x10⁶)
8 14 8 34 30 15 14 17 7 49 71 84 32 25 10 14 10
B-cell purity (%)
90 93 90 98 97 98 94 97 98 96 96 98 99 99 94 97 97
Cell washing, flow cytometry validation
Pellets snap freezing, storage in liquid nitrogen
Lysate preparation, protein quantification
Protein use (µg)
100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100
Reduction, alkylation, trypsin digestion, TMT 10-plex labelling
Healthy donor samples used to 'Bridge' between experiments
TMT 10-plex | 126 | 127N | 127C | 128N | 128C | 129N | 129C | 130N | 130C | 131
A | Healthy donor 1 | Healthy donor 2 | Healthy donor 3 | Notch1 3999 | Notch1 2497 | Notch1 568 | SF3B1 653 | SF3B1 3405 | WT 3156 | WT 4621
B | Healthy donor 1 | Healthy donor 2 | Healthy donor 3 | Notch1 2886 | Notch1 2483 | SF3B1 625 | SF3B1 4140 | SF3B1 1194 | WT 1125 | WT 2459
Pooling of labelled peptides for respective TMT-10 plexes
Peptide total (µg)
1000 A
1000 B
'Offline' peptide prefractionation - high pH C8 reverse phase HPLC
OD (215nm)
10-plex B
10-plex A
Peptide retention time
2x ~100 peak dependent fractions
60 fractions total (after pooling)
60 fractions total (after pooling)
'Online' low pH C18 reverse phase nanoflow HPLC coupled with ESI-MS/MS

Figure 7.1. 2D-LC MS/MS quantitative proteomic characterisation of 14 individual CLL samples and healthy donor B-cell controls. Peripheral blood mononuclear cells (PBMCs) from healthy donors and CLL patients presenting with *NOTCH1*-mutant status (N=5), *SF3B1*-mutant status (N=5) and wildtype *NOTCH1* and *SF3B1* status (N=4) were isolated from CLL-patient derived blood using density gradient centrifugation. PBMCs were frozen in foetal calf serum containing 10% DMSO to maintain cell viability. PBMCs were defrosted and subjected to negative B-cell isolation using an antibody cocktail (EasySep) without CD43 depletion. B-cell lysates were prepared by trituration and sonication. 100 µg of each CLL lysate and two 100µg aliquots of the healthy donor B cells lysates were reduced, alkylated and proteolysed with trypsin. CLL peptides were distributed across two TMT 10-plex label sets. Three identical aliquots of healthy donor B-cell peptides were used in each 10-plex as a means of inter-experimental comparison, termed 'bridging controls'. Samples were then pooled to their respective 10-plex experiments. Labelled peptide pools were subjected to peptide prefractionation generating around 100 fractions per 10-plex. The low-concentration fractions were concatenated at the extremes of the chromatogram to give around 60 fractions, per 10-plex, for LC-MS analysis. Each peptide fraction was subject to chromatography, in line with electrospray ionisation, MS detection and data-dependent MS/MS characterisation.

7.3 PROTEIN IDENTIFICATION AND RELATIVE QUANTIFICATION IN CLL RELATIVE TO HEALTHY DONOR B CELLS

The MS characterisation of peptides from 10-plex A and 10-plex B were subjected to separate target decoy searches to allow independent quantification of proteins within the respective 10-plexes (**Figure 7.2**). Fractions were excluded from the analysis if they contained a low quantity of peptide information or quantifications, leaving 55 and 53 fractions suitable for target decoy searching for 10-plex A and 10-plex B, respectively. These were subjected to the same two-stage, stringent and relaxed, target decoy searches described in **Chapters 5 and 6** and **Section 2.19.6**. For 10-plex A and 10-plex B, respectively, 405,176 and 348,788 PSMs were identified ($q<0.01$), amongst which 195,997 and 171,547 had associated TMT reporter regions. These PSMs identified 8,112 and 8,382 proteins, of which 6,686 and 7,358 were fully quantified for 10-plex A and 10-plex B, respectively, giving a total of 6150 proteins commonly identified in both experiments. Ratios were calculated for each of the 14 CLL samples relative to each of the 3 HD samples. For each set of 3 ratios indicating the extent of differential expression between a CLL sample and the 3 HD controls, only the minimally deviated (from no fold change) of the three ratios was kept. This had the effect of reducing both technical and biological noise, by only stating an observation of up or downregulation when all three ratios were indicative of such a trend, in a

more stringent and effective way than averages. This also contributed to reducing technical noise between the two experiments allowing a more accurate inter-experimental quantitation.

An initial analysis of the data revealed that several platelet proteins were present in the HD samples, which did not appear in the CLL samples, giving a false impression of protein downregulation in CLL. To address this contamination, 194 proteins with negative regulation, observed amongst the top 1000 platelet proteins by copy number [471], were excluded from further analyses. Additionally the WT CLL sample 4621, appeared as an outlier and was therefore excluded from further analyses. For proteins being subjected to analyses comparing subtypes of CLL, 133 proteins exhibiting an FDR-corrected significant difference between the two TMT 10-plex experiments, A and B, were excluded.

5956 proteins were subsequently interrogated for the significance, magnitude and consistency of differential expression relative to HD controls, of the 13 CLL samples. As described in **Chapters 5** and **6**, a regulation score and an FDR-corrected one sample, two tailed T-test were calculated as complimentary measures of differential expression. Given the far larger number of samples, compared with that of previous analyses, providing greater statistical power to conclude lower fold-changes, a regulation score cut off defining differential expression was set at 0.25 (detailed in **Sections 2.19.7**). This quantitative analysis determined that 819 proteins (13.8%) were significantly overexpressed ($RS>0.25$, $p<0.05$) and 746 (12.5%) were significantly underexpressed in CLL ($RS<-0.25$, $p<0.05$), relative to all HD controls.

7.4 Reproducible Protein Identification and Quantitation in Independent Isobaric-Labelled Shotgun Proteomics Experiments

Given the independent nature of 10-plexes A and B, the technical reproducibility across the two experiments was assessed to support the reliability of any biological findings in the dataset (**Figure 7.3**). Firstly, the number of commonly identified and quantified proteins between experiment A and B were demonstrated by Venn diagram (**Figure 7.3A**). In experiment A and B, a total of 8694 proteins were identified ($q<0.01$) and of these 90% (7800) were identified in both proteomes, 96% and 93% of A and B, respectively. Due to an alternative, less sensitive means of detection using HCD, rather than CID and HCD, the extent of quantitation was not as reproducible as identifications. 78% (6150) of the 7894 total proteins quantified over the two experiments were quantified in both; 92% of A and 84% of B.

The number of peptides and PSMs, attributable to the 7800 commonly identified proteins, were compared between 10-plexes A and B, with an R^2 of >0.95 observed in both cases (**Figure 7.3B**). The $\log_2$ (ratios) between the common, HD bridging controls within each experiment was also compared between the two experiments (**Figure 7.3C**). This highlighted, alongside

substantial protein dysregulation in HD3, a reproducible pattern of differential protein expression, which, while differentially affected by ratio compression, gave at least an R^2 value of 0.65, and 0.75 where more pronounced differential expression was observed in HD3. While this correlation was quite low, very few proteins were observed with the opposite direction of regulation. Of HD1 and HD2, the most consistent of these controls, the proteins observed most differentially regulated were MHC proteins; genes with a high level of inter-individual allelic differences, indicating reproducible quantitation of absolute presents or absence of polymorphic peptides.

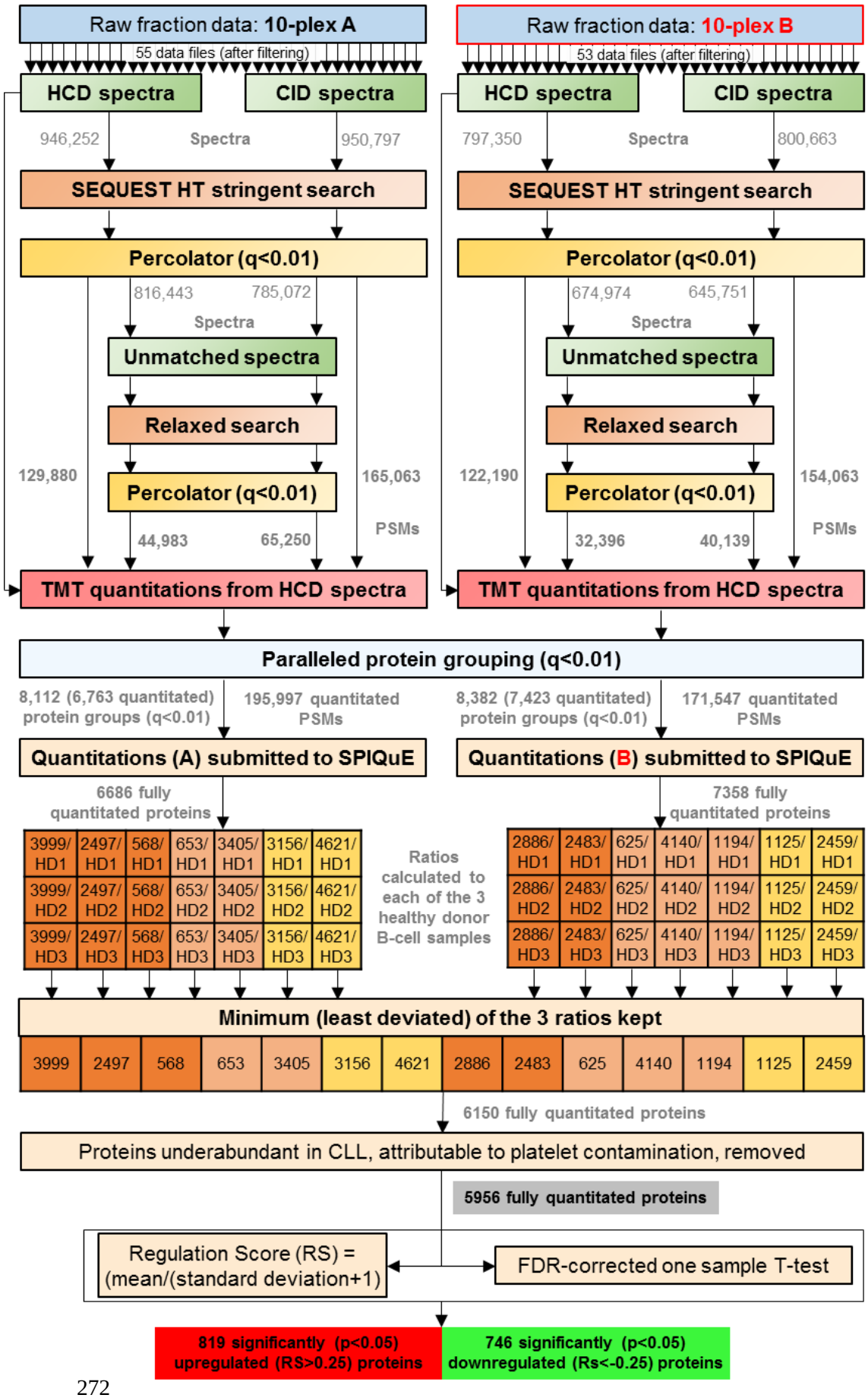

Raw fraction data: 10-plex A
55 data files (after filtering)
HCD spectra
CID spectra
946,252
Spectra
950,797
SEQUEST HT stringent search
Percolator (q<0.01)
816,443
785,072
Spectra
Unmatched spectra
Relaxed search
129,880
Percolator (q<0.01)
165,063
PSMs
44,983
65,250
TMT quantitations from HCD spectra
Raw fraction data: 10-plex B
53 data files (after filtering)
HCD spectra
CID spectra
797,350
Spectra
800,663
SEQUEST HT stringent search
Percolator (q<0.01)
674,974
645,751
Spectra
Unmatched spectra
Relaxed search
122,190
Percolator (q<0.01)
154,063
PSMs
32,396
40,139
TMT quantitations from HCD spectra
Paralleled protein grouping (q<0.01)
8,112 (6,763 quantitated) protein groups (q<0.01)
195,997 quantitated PSMs
8,382 (7,423 quantitated) protein groups (q<0.01)
171,547 quantitated PSMs
Quantitations (A) submitted to SPIQuE
Quantitations (B) submitted to SPIQuE
6686 fully quantitated proteins
7358 fully quantitated proteins
3999/HD1 2497/HD1 568/HD1 653/HD1 3405/HD1 3156/HD1 4621/HD1
3999/HD2 2497/HD2 568/HD2 653/HD2 3405/HD2 3156/HD2 4621/HD2
3999/HD3 2497/HD3 568/HD3 653/HD3 3405/HD3 3156/HD3 4621/HD3
Ratios calculated to each of the 3 healthy donor B-cell samples
2886/HD1 2483/HD1 625/HD1 4140/HD1 1194/HD1 1125/HD1 2459/HD1
2886/HD2 2483/HD2 625/HD2 4140/HD2 1194/HD2 1125/HD2 2459/HD2
2886/HD3 2483/HD3 625/HD3 4140/HD3 1194/HD3 1125/HD3 2459/HD3
Minimum (least deviated) of the 3 ratios kept
3999 2497 568 653 3405 3156 4621 2886 2483 625 4140 1194 1125 2459
6150 fully quantitated proteins
Proteins underabundant in CLL, attributable to platelet contamination, removed
5956 fully quantitated proteins
Regulation Score (RS) = (mean/(standard deviation+1)
FDR-corrected one sample T-test
819 significantly (p<0.05) upregulated (RS>0.25) proteins
746 significantly (p<0.05) downregulated (Rs<-0.25) proteins

Figure 7.2. Protein identification and quantification of CLL proteomes from two TMT 10-plex 2D LC-MS/MS experiments. Fractions respective to experiment A and B were processed using two stage target decoy searches with Proteome Discoverer 1.4. Spectra were initially sorted based on fragmentation type, either CID or HCD and subjected to an initial search for peptides spectrum matches using stringent settings; allowing for no missed cleavages and no dynamic modifications. Spectra failing to match to a peptide with a percolator-determined false discovery rate of $q<0.01$ were subjected to a second, relaxed target decoy search allowing for missed cleavages and dynamic modifications. For the purposes of consistent protein grouping, reports for 10-plexes A and B were opened in tandem, while keeping quantitations separate. The PSM quantitations and features with the protein group assignments were exported separately for each 10-plex for analysis with SPIQuE. Ratios were calculated for each CLL sample relative to each of the three healthy donor B-cell controls. To adjust for variability from the controls and to keep only the most consistent observations specific to CLL only the least deviated ratio was kept for subsequent analyses. A platelet contamination in the healthy controls was observed and therefore proteins were filtered based on previous platelet proteomics. As a means of distinguishing the most confident, upregulated proteins across all samples 13 CLL $\log_2$ (ratios), with the exception of the clear outlier 4621, were analysed to determine a 'regulation score' (RS); a measure of the magnitude and consistency of regulation across multiple samples defined as the ratio of the mean $\log_2$ (ratio) to the SD + 1. A one sample, false discovery rate corrected T-test was performed to determine the significance of protein expression across the 13 samples.

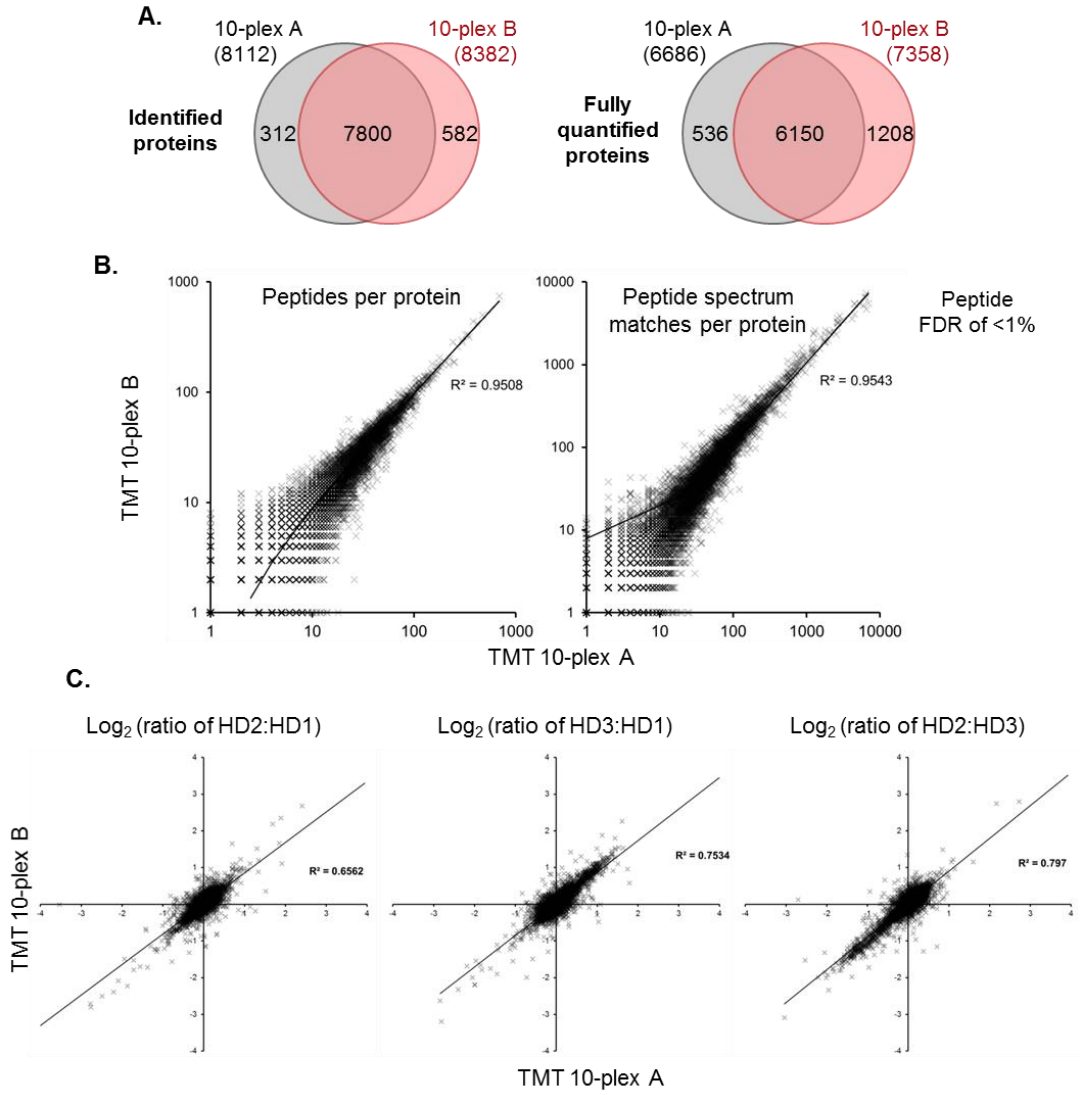


Figure 7.3. Evaluation of the qualitative and quantitative reproducibility between 10-plex A and B. A. The numbers of identified and quantified proteins determined in experiments A and B and the commonly determined numbers for each of these. **B.** The respective numbers of PSMs and peptides for each of the 7800 commonly identified proteins across experiments A and B. **C.** The reproducibility of protein quantifications for the healthy donor (HD) bridging controls characterised in each experiment. Log_2 (ratios) of the 6150 commonly quantified proteins comparing HD1, HD2 and HD3 determined by SPIQuE analysis of quantitations derived from 10-plexes A and B.

7.5 TOPOLOGICAL ANALYSIS OF QUANTITATIVE PROTEOMICS IDENTIFIES A SUBTYPE-INDEPENDENT CLL SIGNATURE

The 5956 proteins fully quantitated, as outlined in **Figure 7.2**, were subjected to a series of analyses to identify possible topological trends in differential protein expression (**Figure 7.4**).

Firstly, **Figure 7.4A** outlined the known characteristics of the 14 CLL samples subjected to quantitative proteomics, including the experiment, A or B, to which each was assigned. The order of the samples in this table was determined by the cluster analysis of the $\log_2$ (ratios) by Cluster 3.0 and Euclidian distance, of the 14 CLL samples (**Figure 7.4B**). This analysis highlighted two major clusters and the outlier, '4621', which was subsequently excluded from all further analyses. No single feature could successfully separate these two clusters, however trisomy 12 was present exclusively in the right-hand cluster with 5/8 samples bearing an additional copy of chr12. Clustering also highlighted highly consistent regions containing several proteins with over or underexpression in all samples compared to HD samples. There were also regions which demonstrated heterogeneity. However, the majority of proteins were either unregulated and upregulated, or unregulated and downregulated, with few having different directions of regulation, relative to HD expression, for the same protein. Just 20 proteins were observed with a more than 2-fold change both up and down for at least one CLL sample relative to HD samples.

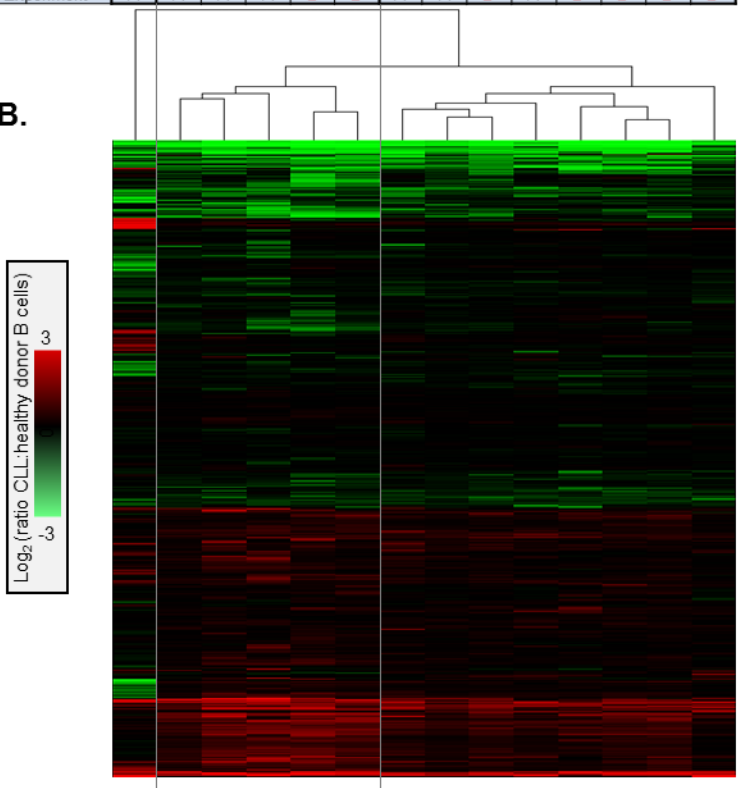
Figure 7.4C demonstrated the distribution of protein $\log_2$ (ratios) for each CLL sample relative to HD samples, highlighting the median and 1^{st}, 25^{th}, 75^{th} and 99^{th} percentiles as well as outliers beyond the 1^{st} and 99^{th} percentiles. This demonstrated a consistent extent of protein deviation across the samples, without subtype or batch effect influence, with no significant difference observed between the standard deviations ($p=0.81$), median values ($p=0.26$), 99^{th} percentile ($p=0.69$) or 1^{st} percentile ($p=0.21$) between samples in 10-plexes A and B. The average standard deviation of all 13 protein $\log_2$ (ratios) was 0.46, ranging from 0.57 for sample '625' to 0.36 for '3999'.

To visualise patterns within the dataset and emphasise sample differences, principal component analysis was employed (**Figure 7.5**). The same PCA plot was visualised from two angles; the left, with the third component axis coming out of the page and; the right, a 90° rotation around axis 1, with the second component axis coming out of the page. The same pairs of rotated plots were overlaid with annotations for CLL subtypes; mutations of *SF3B1* or *NOTCH1*, IGHV mutations status, trisomy 12 status, $CD38^+$ status and 13q14 deletion. While no separation was observed for *SF3B1*-mutant samples, *NOTCH1*-mutant samples partly clustered behind the plane of axis 1 and 3. U-CLL samples were predominantly clustered behind the plane of axis 1 and 2, with a single outlier. Despite this, the M-CLL cases clustered in the centre of the plot. Trisomy 12 cases clustered closely towards the positive end of axis 1 and the three samples also positive for CD38 were at the extreme of this clustering. Heterozygous 13q14 loss was indicative of some clustering towards the centre which was focal around the sample with homozygous loss.

A.

Subtype	WT	Notch1	SF3B1	WT	SF3B1	WT	Notch1	Notch1	Notch1	SF3B1	SF3B1	SF3B1	WT	Notch1
Patient ID	4621	2497	653	3156	625	2459	568	3999	2886	3405	4140	1194	1125	2483
Gender	M	M	F	F	M	M	M	M	M	M	M	M	M	F
AAD	66	60	66	63	44	57	79	66	67	72	51	61	56	69
White cell count	nd	41	83	99	**131**	nd	nd	**100**	**271**	**121**	71	70	**123**	**281**
Zap70 (%)	nd	2	nd	**16**	10	4	nd	**15**	10	**45**	4	1	2	nd
CD38 (%)	nd	6	nd	37	10	3	nd	49	4	**99**	1	**99**	6	**99**
Chr 12	nd	++	++	++	++	++	++	+++	+++	+++	++	+++	++	+++
13q14	nd	+-	++	+-	++	++	+-	++	--	++	+-	+-	++	++
11q23	++	++	++	+-	++	+-	++	++	++	++	++	++	++	++
IgHV mutated	M	M	**U**	**U**	**U**	M	**U**	**U**	M	**U**	M	M	M	**U**
Experiment	A	A	A	A	B	B	A	A	B	A	B	B	B	B

B.



C.

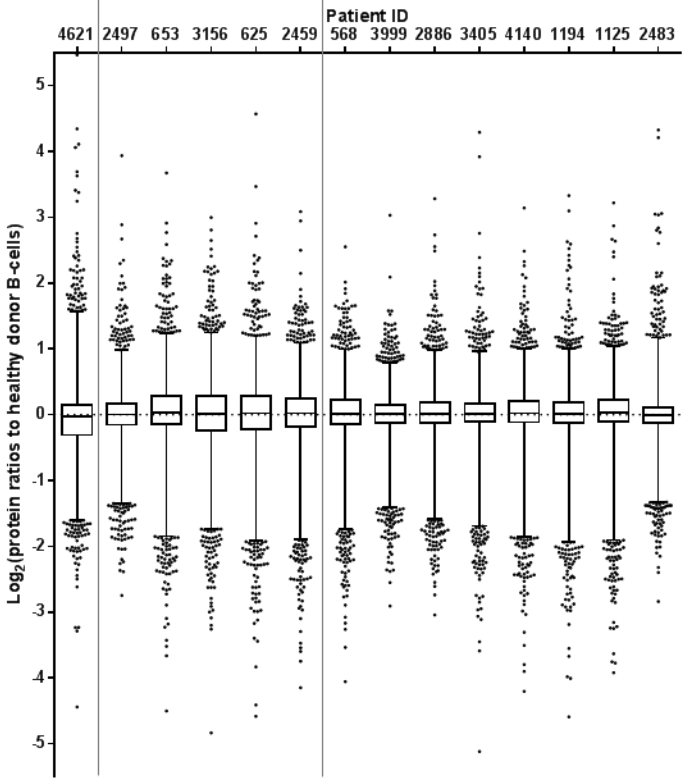


Figure 7.4. Clustering analysis of the individual CLL samples. A. CLL characteristics, patient details and subtype classifications for each of the 14 samples characterised. nd = not determined. **B.** Euclidian distance clustering of 5956 protein log_2 (ratios) for the 14 CLL samples. **C.** Distributions of 5956 protein log_2 (ratios) for the 14 CLL samples relative to HD B cells. The box represents the 25th, 50th and 75th percentiles and the whiskers, the 1st and 99th. Dots represent outliers exceeding these bounds.

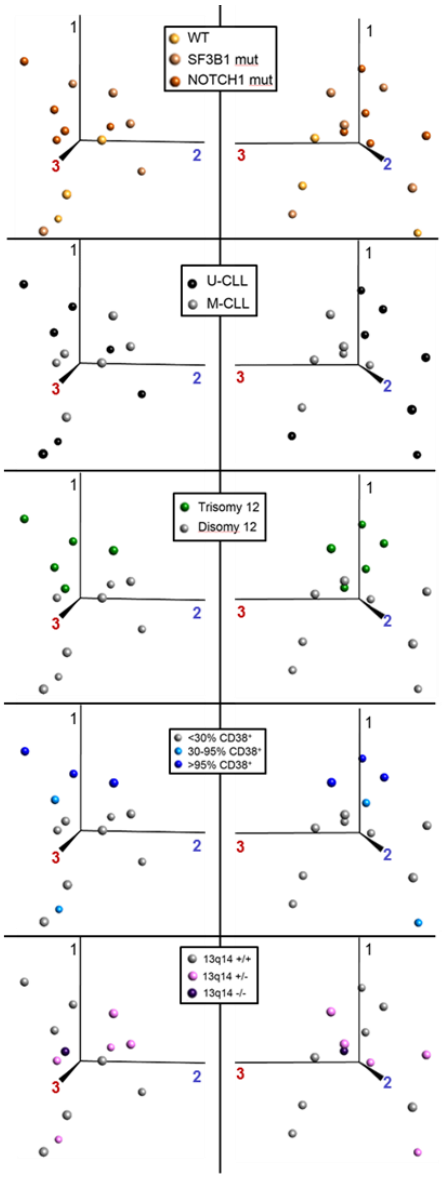


Figure 7.5. Principal component analysis of the individual CLL samples. Principal component analyses (PCA) for 13 CLL cases (excluding 4621) highlighting the three dimensional separation of features. The log2 (ratio) quantitations for the 5956 proteins were analysed using Qlucore Omics Explorer 3.1 using the default settings.

7.6 PROTEOMIC IDENTIFICATION OF ANTICIPATED PROTEIN QUANTIFICATIONS IN CLL

CLL, as discussed in **Section 1.2.3,** has been extensively characterised with several proteins consistently or variably expressed across cases. As a means of validating the quantitative proteomics findings, confirming the quality of the 2D-LC MS/MS analysis, the statistical data analysis and the successful isolation of the samples themselves, anticipated protein identifications were listed alongside their respective quantitations (**Figure 7.6**). This presented the expression across each of the 14 CLL samples for the minimum determined fold change relative to HD samples. CD5, a canonical upregulated marker of CLL, was amongst the 20 most overexpressed proteins and more than 2-fold upregulated in 11 of the 14 CLL samples, with a regulation score of 0.84 ($p=8.6x10^{-6}$). Several other reported cell surface markers were identified overexpressed ($RS>0.25$, $p<0.05$), as anticipated, relative to HD samples, including; receptor tyrosine kinase-like orphan receptor (ROR1), Suppressor of tumorigenicity 14 protein (ST14 or matriptase), T-cell differentiation antigen CD6 and low affinity Ig epsilon Fc receptor CD23. Subtype-specific regulations were also observed, with CD38 expression being consistently more than 2-fold downregulated in all but those 3 cases phenotyped with 99% CD38 expression. CD49d, a marker which correlates strongly with trisomy 12 status, was downregulated more than 2-fold, specifically in the cases without trisomy 12.

B-cell receptor components and B-cell-specific proteins were observed significantly downregulated ($RS>0.25$, $p<0.05$) as expected, included; CD22, CD79A, CD79B and B-lymphocyte antigen CD20. CD19, a B-cell specific surface protein, which was used to assess sample purity, was also identified and quantified, but without any significant differential expression between CLL and HD B cells. Intracellular proteins with anticipated and observed significant overexpression, included; the B- and T-cell receptor response regulator – tyrosine-protein phosphatase non-receptor type 22 (PTPN22), the apoptosis regulator and characteristic CLL protein BCL2, NFAT proteins, the tyrosine-protein kinase, ZAP-70, and the histone-lysine N-methyltransferase SETDB1.

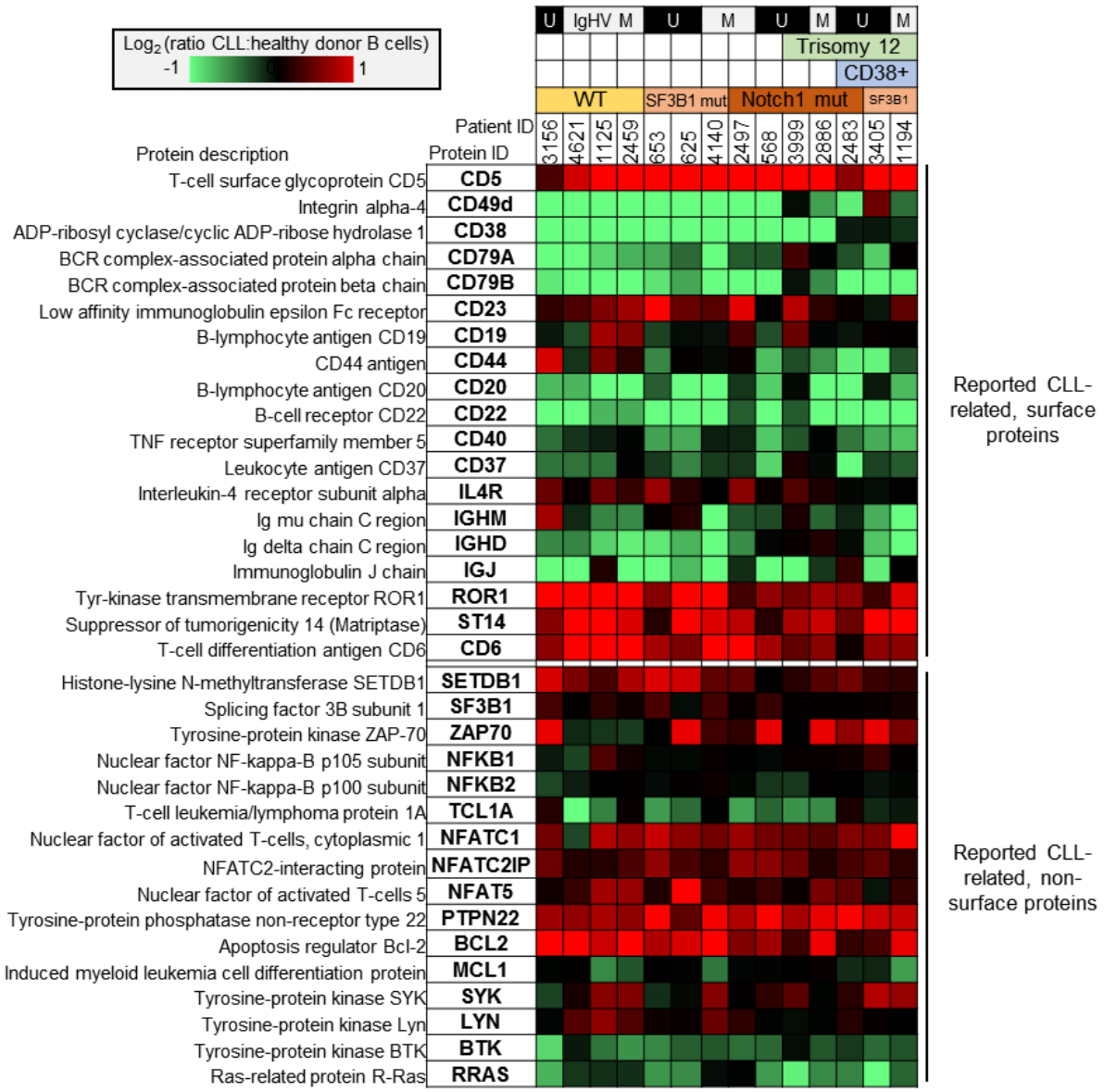


Figure 7.6. Anticipated CLL protein identification and quantification. The expression determined by quantitative proteomics relative to healthy donor B cells (upregulated – red, downregulated – green), with the respective CLL subtypes annotated. These are separated into cell surface proteins and non-surface proteins.

7.7 CONFIDENT AND CONSISTENT DIFFERENTIAL PROTEIN EXPRESSION IN CLL

With the technical reproducibility confirmed and the successful characterisation of key proteins relating to known CLL biology identified, the data were analysed to identify differential expression potentially describing novel CLL biology. The RS and FDR-corrected p-values

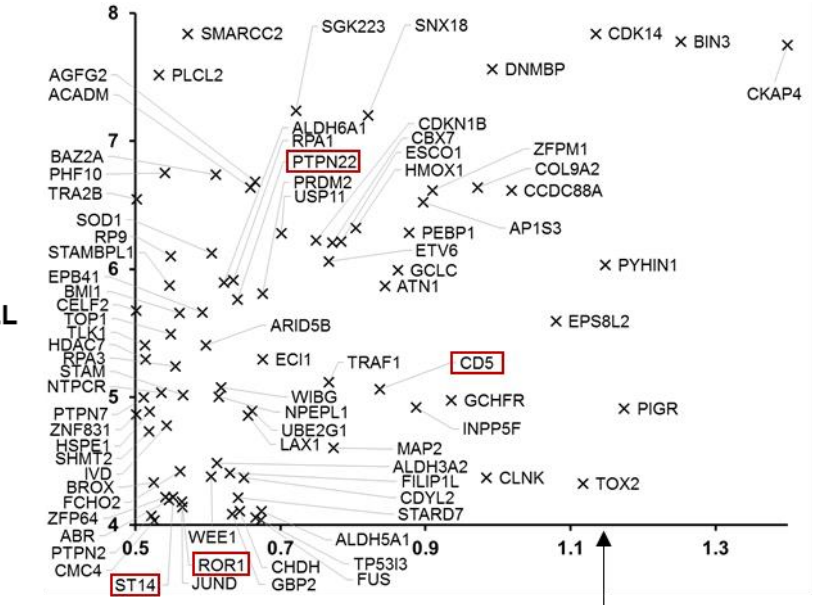
describing the magnitude and significance of differential expression, relative to HD samples, were used to produce a volcano plot (**Figure 7.7**). Of the 819 and 746 proteins with significant up and downregulation in CLL, respectively, those proteins at the extremes of this plot were annotated. The most significantly upregulated protein was cytoskeleton-associated protein 4 (CKAP4) (RS=1.4, $p=1.8x10^{-8}$), a putative receptor for antiproliferative factor (APF), which was observed with between 250% and 590% expression in all CLL samples (averaging 370%). These observations were supported by the number of identifications and quantifications of the protein; with 34 unique peptides identified from over 500 PSMs. Bridging integrator 3 (BIN3), a protein involved in actin regulation and cytokinesis, was the second most significantly overexpressed (RS=1.25, $p=1.7x10^{-8}$) in CLL, with an average of 300% expression, ranging from 200% to 400%. The third most significantly upregulated protein was cyclin-dependent kinase 14 (CDK14) (RS=1.14, $p=1.4x10^{-8}$), a regulator of cell cycle progression and proliferation, which was consistently upregulated to 200-350% expression in all 13 cases. Other proteins with both a high degree of significance and upregulation included; dynamin-binding protein (DNMBP), zinc finger protein M1 (ZFPM1), collagen alpha-2(IX) chain (COL9A2), girdin (CCDC88A), AP-1 complex subunit sigma-3 (AP1S3) and phosphatidylethanolamine-binding protein 1 (PEBP1). Strongly significant proteins ($p<10^{-7}$), with a lower regulation score ($1>RS>0.5$) included; sorting nexin-18 (SNX18), sugen tyrosine-protein kinase 223 (SGK223), SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily C member 2 (SMARCC2) and inactive phospholipase C-like protein 2 (PLCL2). Conversely, proteins with less significance ($10^{-4}>p>10^{-6}$), but a high regulation score ($RS>1$) included; pyrin and HIN domain-containing protein 1 (PYHIN1), epidermal growth factor receptor kinase substrate 8-like protein 2 (EPS8L2), polymeric Ig receptor (PIGR) and TOX high mobility group box family member 2 (TOX2).

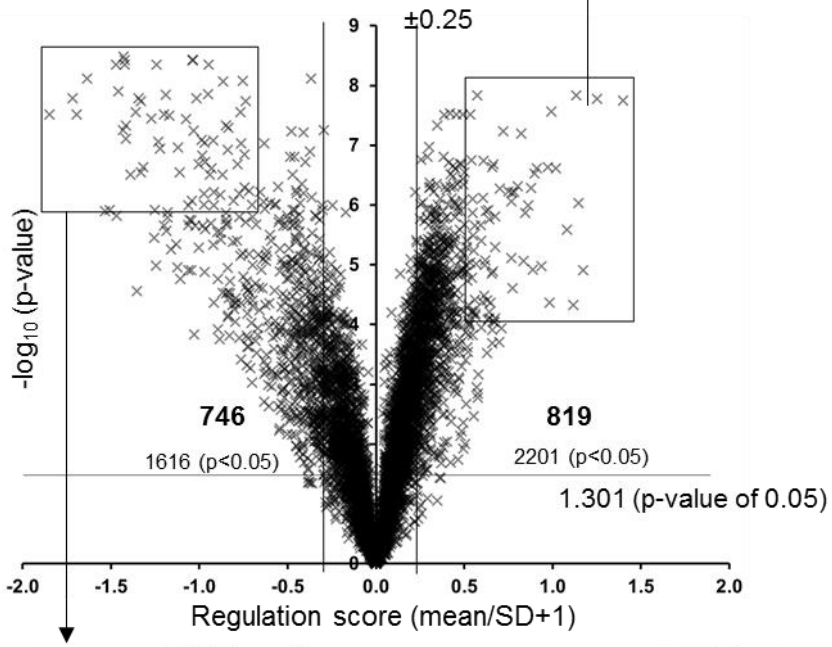
Amongst the ~80 upregulated proteins annotated in **Figure 7.7**, 9 were mitochondrial proteins; enoyl-CoA delta isomerase 1 (ECI1), StAR-related lipid transfer protein 7 (STARD7), 10 kDa heat shock protein (HSPE1), serine hydroxymethyltransferase (SHMT2) and of these, 5 were dehydrogenase enzymes; medium-chain specific acyl-CoA dehydrogenase (ACADM), choline dehydrogenase (CHDH), isovaleryl-CoA dehydrogenase (IVD), succinate-semialdehyde dehydrogenase (ALDH5A1) and methylmalonate-semialdehyde dehydrogenase [acylating] (ALDH6A1). Additionally to this, fatty acid dehydrogenase (ALDH3A2) was also amongst these proteins. Three members of the tyrosine-protein phosphatase non-receptor family – PTPN2, PTPN7 and PTPN22 were observed in this region of the volcano plot. Kinases were also frequently observed amongst these proteins, including; CDK14, SGK223, wee1-like protein kinase (WEE1) and ROR1.

Of the 746 significantly downregulated proteins, as with the upregulated proteins, approximately the top 10% were plotted and annotated. This highlighted four proteins as the most significantly downregulated relative to HD samples; kalirin (KALRN) (RS= -1.85, $p=3.1x10^{-8}$), RNA-binding protein with multiple splicing 2 (RBPMS2) (RS= -1.72, $p=1.6x10^{-8}$), aspartyl/asparaginyl beta-hydroxylase (ASPH) (RS= -1.69, $p=3.1x10^{-8}$) and DENN domain-containing protein 3 (DENND3) (RS= -1.64, $p=7.6x10^{-9}$). Kalirin, a cytoskeletal remodelling protein and guanyl-nucleotide exchange factor characterised in neurons and observed with sensitivity to splicing [472], was downregulated to an average of 15% expression over the 13 CLL samples, ranging from 8-25% expression, relative to HDs. This quantitation was derived from 10 unique peptides and 68 PSMs. The minimally characterised, but apparent smooth muscle cell-fate determinant protein, RBPMS2, was similarly consistently downregulated to approximately 30% in all CLL samples. RBPSM2 expression averaged 18% and had a maximal downregulation to just 12.5% of the expression in HD samples. ASPH, a transmembrane ER protein with tumour suppressor capacity in mice [473], was similarly underexpressed, averaging 15%, ranging from 11 to 34% expression of that of HD samples across the CLL samples. Another guanyl-nucleotide exchange factor, DENND3, a protein with a role in endosome transport and recycling [474] was observed with a slightly lower average of 21% expression, but with more consistent and significant expression with a range of 15-32% of HD samples.

Other proteins downregulated in CLL with the greatest significance were; cathepsin W (CTSW), transmembrane ligand ephrin-B1 (EFNB1), microtubule-associated protein 1A (MAP1A), uncharacterised protein KIAA0513, stonin-2 (STON2), special AT-rich sequence-binding protein 1 (SATB1), cytosolic phospholipase A2 (PLA2G4A) and chloride intracellular channel protein 3 (CLIC3). Additional more substantially downregulated proteins included; specifically androgen-regulated gene protein (SARG/C1orf116), prostacyclin receptor (PTGIR), schlafen family member 14 (SLFN14), apoptosis regulator Bcl-2-like protein 1 (BCL-X/BCL2L1), latent-transforming growth factor beta-binding protein 1 (LTBP1), monocyte differentiation antigen CD14 and c-Maf-inducing protein (CMIP).

Upregulated proteins in CLL





Downregulated proteins in CLL

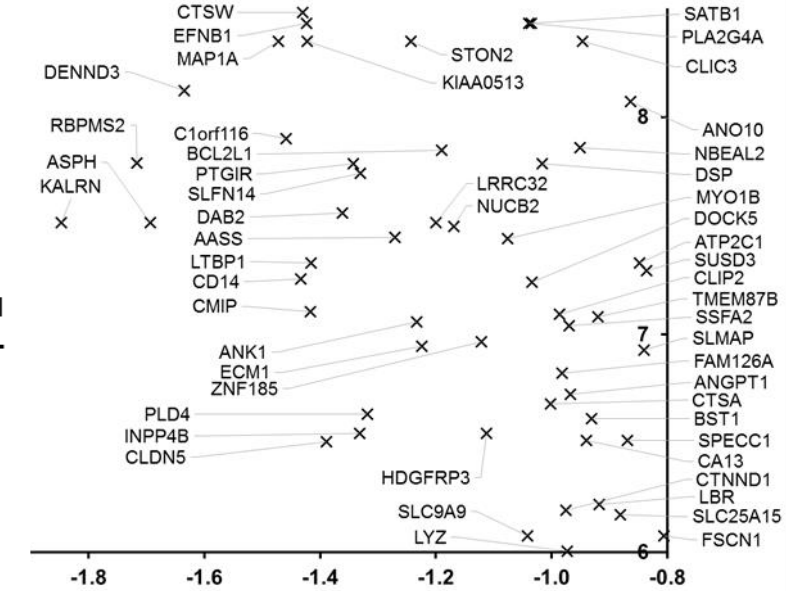


Figure 7.7. Differential protein expression in the CLL proteome. The regulation score and negative $\log_{10}$ (p-values) of protein expression relative to a $\log_2$ (ratio) of 0 were plotted for the 5956 proteins commonly identified, non-platelet proteins. Those proteins reaching significance ($p<0.05$), with a regulation score of >0.25 or <-0.25 were considered as up and downregulated, shown in the upper right and left regions, and annotated in the upper and lower panels, respectively. Proteins highlighted in red are a selection of positive controls, previously observed with upregulation in CLL.

7.8 OVEREXPRESSED CLL CELL SURFACE PROTEINS WITH POTENTIAL FOR IMMUNOTHERAPY TARGETING

In light of this confident substantial dysregulation across the proteome, the first question was that of potential clinical targets that may suggest novel ways in which CLL can be treated. The most prominent of these were cell surface targets. Understanding the cell surface protein expression in CLL presents several potential clinical benefits. CD5, for instance, allows diagnosis of CLL, CD49d expression offers a potential prognostic indicator and CD20 and CD52 present targets for immunotherapy which can treat the disease. To interrogate the proteomics findings for existing, and potentially novel, cell surface expression of clinical targets, proteins were filtered on the basis of GO term annotation as well as several other tools and databases for annotating proteins with cell surface expression, described in **Section 2.12.6**. A total of 395 quantitated proteins, 6.6% of all quantitated proteins, were determined to have consistent annotation for expression on the surface of cells and the regulation scores of these proteins were plotted against their $-\log_{10}$ (p-values) (**Figure 7.8A**). This volcano plot identified a total of 21 proteins, 5% of the membrane proteins, with significant upregulation (RS>0.25, $p<0.05$) on the surface of CLL cells, which were subsequently tabulated with a sample-specific breakdown of protein expression (**Figure 7.8B**).

Amongst the 21 upregulated surface proteins on CLL (**Figure 7.8A** and **B**) were the anticipated proteins described in **Figure 7.6**; CD5, ROR1, ST14/matriptase, CD23 and CD6; alongside CKAP4, described previously as the most upregulated protein identified in CLL (**Figure 7.7)**. Also known as climp-63, CKAP4, is reported to be a predominantly ER-localised protein, but with a propensity for cell surface expression as a receptor for APF. No previous reports appear to have demonstrated a role for CKAP4 in CLL or B-cell cancers. PIGR, an Fc receptor, with no previous links to CLL or B-cell cancers, is capable of transporting Ig across epithelial membranes by transcytosis, whereupon the extracellular region is cleaved. PIGR was the second most upregulated CLL surface protein, observed with an average upregulation of 468% of that of HD samples, ranging from 158% to 1,106%, with at least 200% expression in 12

samples, and 400% in 8 samples. 4 unique peptides, matching only the intracellular portion of PIGR, were identified. CD5 was the third most overexpressed surface protein, behind only CKAP4 and PIGR.

The fourth most upregulated cell membrane protein was lymphocyte transmembrane adaptor 1 (LAX1) (RS=0.66, $p=1.35x10^{-5}$), a negative regulator of B- and T- cell receptor signalling [475] with no previous implications or characterisations in CLL. On average, expression of LAX1 in CLL was 185% of that of HD samples, consistently above 140%, with 5 samples exceeding 200% expression. The next most upregulated protein without any previous characterisation in B-cell cancers was the glycan-binding receptor and proliferating germinal B-cell-expressed protein 'prolectin', also known as c-type lectin domain family 17, member A (CLEC17A). Prolectin was significantly upregulated (RS=0.52, $p=2.4x10^{-4}$), averaging 168% expression, with 10 samples exceeding 125%, and 7 exceeding 180% expression.

Two peripheral ion channels were observed significantly upregulated in CLL, without any previous reports; ATPase, Ca^{2+} transporting, plasma membrane 4 (ATP2B4) (RS=0.48, $p=5.7x10^{-5}$) and ATPase, Na^{+}/K^{+} transporting, beta 1 polypeptide (ATP1B1) (RS=0.36, $p=3.6x10^{-2}$). ATP2B4, a calcium exporter with a putative link to BCR-induced calcium efflux [476], was overexpressed (>125%) in all but 1 CLL sample, averaging 155% of that of HD samples. While less consistently expressed, the regulatory and potentially cell-adhesive subunit controlling Na^{+} export, ATP1B1, was observed with >200% expression in 6 samples, >140% expression in a further 4 samples and a downregulation to less than 50% in sample '653'.

In addition to the Fc receptors CD23/FCER2 and PIGR, two Fc receptor-like (FCRL) surface proteins were identified with significant upregulation in CLL. FCRL2 (CD307B), a BCR signalling inhibitor [477] previously reported in CLL [478, 479] was identified with an average expression of 145%, with 10 samples indicating some overexpression (>125%) compared with HD samples. FCRL5 (CD307E), another modulator of BCR signalling [480] with reported involvement in CLL [481], had a greater average overexpression of 160%, and 11 samples exceeding 125%.

Four further peripheral membrane proteins with previous characterisations in CLL were also identified; signalling threshold regulating transmembrane adaptor 1 (SIT1) (RS=0.38, $p=6.4x10^{-6}$), fas apoptotic inhibitory molecule 3 (FAIM3/TOSO) (RS=0.38, $p=3.3x10^{-3}$), ST6 beta-galactosamide alpha-2,6-sialyltranferase 1 (CD75) (RS=0.37, $p=1.6x10^{-4}$) and intercellular adhesion molecule 3 (ICAM3/CD50) (RS=0.25, $p=1.9x10^{-3}$). Peptides matching three minimally characterised, putative cell surface proteins; transmembrane and coiled-coil domain family 3 (TMCC3) (RS=0.46, $p=1.8x10^{-4}$), chromosome 3 open reading frame 33 (C3orf33) (RS=0.36,

$p=1.3x10^{-4}$) and chromosome 17 open reading frame 80, cell migration-inducing gene 3 protein (C17orf80) (RS=0.30, $p=1.3x10^{-5}$) ; were indicative of upregulation in CLL. Finally, neuroplastin (NPTN), an Ig-superfamily cell adhesion molecule with a widely expressed isoform (Np55) capable of activating fibroblast growth factor receptor 1 (FGFR1) [482, 483] was significantly overexpressed in CLL (RS=0.29, $p=3.5x10^{-4}$). NPTN was upregulated to 125% in all but 4 of the samples. No previous observations were apparent describing this overexpression in CLL or other B-cell cancers.

7.9 UNDEREXPRESSED CLL CELL SURFACE PROTEINS

While of less clinical significance for immunotherapy targets and diagnosis, proteins downregulated in CLL are indicative of alteration in the biology of interactions between the cancer and the immune system and the microenvironment. The downregulated surface proteins were more frequent than the upregulated proteins in CLL (**Figure 7.8A**). Approximately 1 in 3 fully quantified surface proteins (133/395) demonstrated significant underexpression (RS<-0.25, $p<0.05$) – over six times that of the upregulated proteins. Less emphasis was placed on the downregulated proteins, firstly due to the greater clinical interest in upregulated proteins, but also the potential contamination observed from platelets in the HD samples, which may extend to other cell material not accounted for by the adjustments made by filtering platelet proteins. The most significantly downregulated cell membrane proteins included the previously described (**Section 7.7**); EFNB1 (RS= -1.42, $p=3.7x10^{-9}$), PTGIR (RS= -1.34, $p=1.6x10^{-8}$) and CD14 (RS= -1.43, $p=5.5x10^{-8}$); and additionally leucine-rich repeat-containing protein 32 (LRRC32), (RS= -1.20, $p=3.1x10^{-8}$), phospholipase D4 (PLD4) (RS= -1.32, $p=2.3x10^{-7}$) and Claudin-5 (CLDN5) (RS= -1.39, $p=3.1x10^{-7}$). Overall, several trends emerged amongst the downregulated CLL surface proteins in **Figure 7.8A** such as a high frequency of proteins annotated as transporters, cell adhesion proteins, immune system regulator and transmembrane receptors.

Amongst the transporter proteins, 16 were ion transporters; Ca^{2+}-transporting ATPase type 2C member 1 (ATP2C1), anoctamin-10 (ANO10), two choline transporter-like proteins; 1 and 2 (SLC44A1 and SLC44A2), H(+)/Cl(-) exchange transporter 7 (CLCN7), inositol 1,4,5-trisphosphate receptor type 1 (ITPR1), metalloreductase STEAP3 (STEAP3), multidrug resistance-associated protein 4 (ABCC4), pannexin-1 (PANX1), probable phospholipid-transporting ATPase IF (ATP11B), protein tweety homolog 3 (TTYH3), sideroflexin-3 (SFXN3), two sodium/hydrogen exchangers (SLC9A1 and SLC9A9), solute carrier family 12 member 2 (SLC12A2) and transient receptor potential cation channel subfamily V member 2 (TRPV2). An additional 13 proteins related to vesicle mediated transport were downregulated; two secretory carrier-associated membrane proteins (SCAMP2 and SCAMP3); syntaxins; STX3

and STX17; vesicle transport proteins; GOLT1B and USE1; vesicle transport through interaction with t-SNAREs homolog 1B (VTI1B), vesicle-associated membrane protein 3 (VAMP3) atlastin-3 (ATL3), B-cell receptor-associated protein 31 (BCAP31), CD14, protein YIPF5 and sortilin-related receptor (SORL1). Amongst these transporters were also 5 proteins related to lipid transport; long-chain fatty acid transport protein 4 (SLC27A4), ATP11B, apolipoprotein B receptor (APOBR), phospholipid-transporting ATPase IA (ATP8A1) and SORL1.

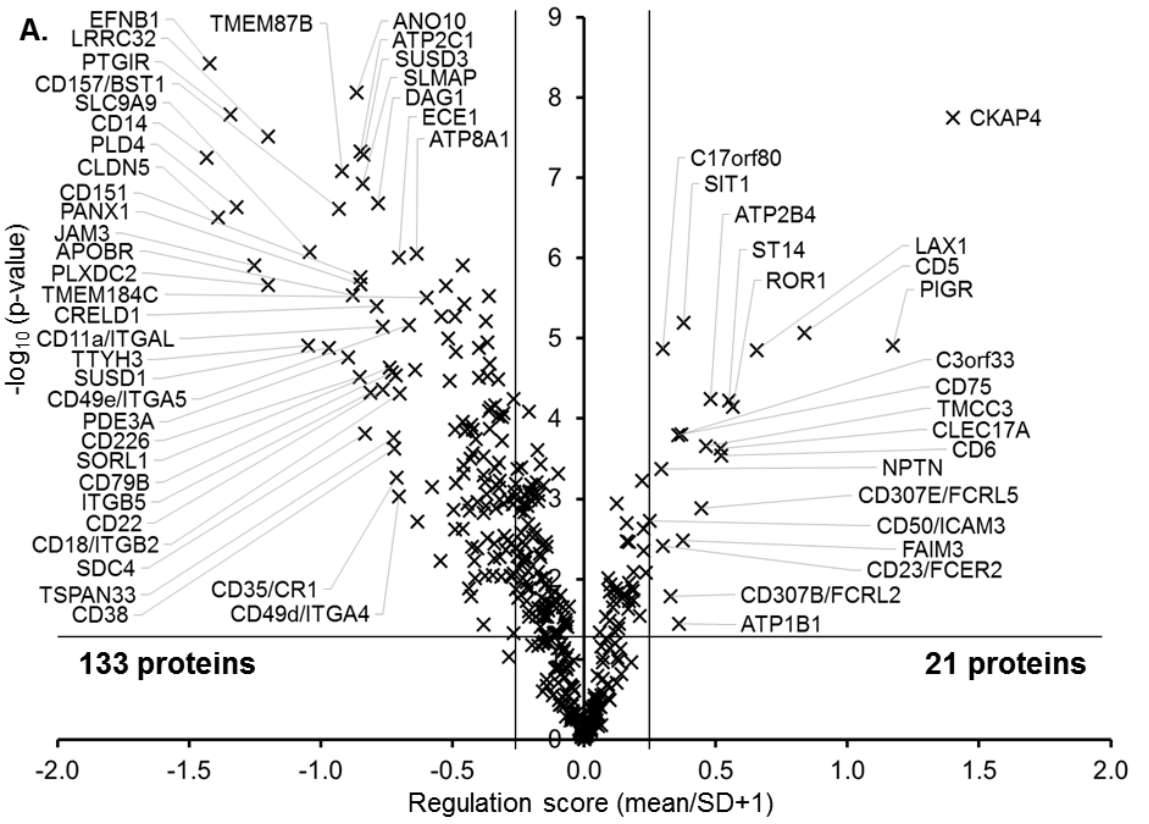
21 surface proteins involved in cell adhesion processes were determined to be downregulated in CLL, including; 7 integrins; ITGA4/CD49d, ITGA5/CD49e, ITGAL/CD11a, ITGAM/CD11b, ITGB1/CD29, ITGB2/CD18 and ITGB5; 7 non-integrin CD molecules; CD22, CD47, CD72, CD84, CD97, CD151 and CD226; and 6 other adhesion molecules; claudin 5, L-selectin (CD62L), embigin (EMB), EFNB1 and lysosome membrane protein 2 (SCARB2).

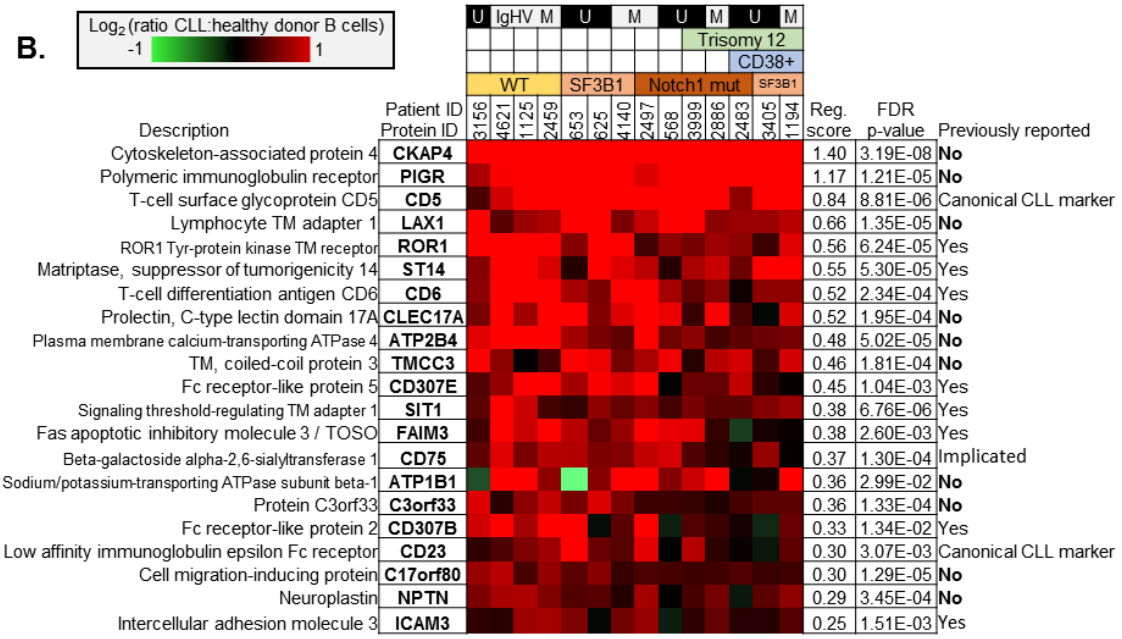
Immune system regulators were also frequently downregulated, represented by 33 of the 133 proteins, including; ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1 and 2 (CD38 and CD157); B-cell proteins; CD79A, CD79B, CD22, CD20, IGHG2 and BCAP31; 5 HLA class 2 molecules; HLA-DMB, HLA-DOB, HLA-DQB1, HLA-DRB4 and HLA-DRB1; all but ITGB5 of the above integrins, 13 further CD molecules; CD14, CD26, CD35, CD37, CD40, CD46, CD47, CD55, CD97, CD180, CD226, CD257/BAFF and CD305; and other regulators; lunapark (LNP) and EFNB1.

Approximately one in four downregulated CLL surface proteins had receptor activity (39/133), including; the CD molecules; CD14, CD20, CD35, CD38, CD40, CD46, CD47, CD72, CD79A, CD79B, CD84, CD97, CD123/IL3RA, CD180, CD226, CD305 ; all 7 integrins described above; all 5 HLA molecules described above; two receptor-type tyrosine protein phosphatases; alpha (PTPRA) and epsilon (PTPRE); additionally, receptors to; prostacyclin (PTGIR), Inositol 1,4,5-trisphosphate (ITPR1) and apolipoprotein B (APOBR).

Figure 7.8. Cell surface expression of proteins in CLL. Fully quantitated proteins were filtered on the basis of GO-annotated localisation to membranes (GO:0005886, GO:0005887, GO:0009986, GO:0031226, GO:0009897, GO:0016020 and GO:0016021) and subsequently filtered to remove any proteins also annotated with localisation to the endoplasmic reticulum (GO:0005783), mitochondria (GO:0005739), or nucleus (GO:0005634). This list was then combined with those proteins described by the MS surface atlas [368]. This resulting list was used to filter for those proteins characterised here and the remaining proteins filtered against the surfaceome database [369], a tool combining surface expression observations from several resources. In total, this process enriched for a list of 395 predominantly surface-expressed,

transmembrane proteins. **A.** Volcano plot highlighting proteins with a strong statistical significance and magnitude of regulation. **B.** The detailed results for those 21 proteins determined to be significantly upregulated (RS>0.25, p<0.05) on the surface of CLL, with the final column indicating whether or not the protein has previously been described in CLL.





7.10 UPREGULATED TARGETS OF SMALL MOLECULAR INHIBITORS IN CLL

Several approaches which have proved successful in the treatment of CLL have involved the use of small molecular inhibitors, either alone or, more frequently, in combination with immunotherapy. A key question of the proteomics data was therefore the identification of potential intracellular protein targets, that may be driving or sustaining CLL, which are vulnerable to interference with small molecular inhibitors. The inbuilt capabilities of IPA software were used to annotate proteins to identify drugable targets, and a list of upregulated proteins with potential for such interference was generated (**Figure 7.9**). BCL2, an established target of small molecular inhibitors in CLL was in this list (RS= 0.55, $p=1.5x10^{-4}$) with small molecular inhibitors such as obatoclax, navitoclax/ABT263, ABT737and ABT199 [59, 60], and the antisense oligodeoxyribonucleotide oblimersen capable of interfering with BCL2 production.

The most upregulated protein with potential for interference was collagen alpha-2 (IX) chain (COL9A2) (RS= 0.97, $p=2.3x10^{-7}$), a typically extracellular protein, forming part of the extracellular matrix, identified with 6 unique peptides and between 175% and 424% overexpression in all CLL cases, relative to HD samples. While no small molecular inhibitors were reported or annotated with IPA, a degradation enzyme collagenase clostridium histolyticum is available. The next most upregulated target of inhibition was heme oxygenase 1 (HMOX1) (RS= 0.80, $p=4.8x10^{-7}$), an enzyme which degrades free heme, thereby offering cytoprotective effects. HMOX1 was also potentially involved in mitochondrial biogenesis and previously implicated in CLL [484, 485]. HMOX1 was upregulated to >190% expression in 12 cases and was potentially a target of inhibition by tin mesoporphyrin. The second functional heme oxygenase isoform (HMOX2) was also upregulated, though to a lesser extent (RS= 0.26, $p=1.4x10^{-3}$) with 9 samples exceeding 120% expression - a novel finding in CLL. Microtubule-associated protein 2 (MAP2) was the third most upregulated potential interference target (RS= 0.77, $p=2.5x10^{-5}$), with the inhibitor estramustine potentially capable of interfering with the suggested stabilising effects of MAP2 on microtubules. MAP2 was upregulated at least 150% in all but one sample, and averaging 225% expression relative to HD samples.

Six kinases with potential for inhibition were observed upregulated in CLL, compared to HD controls; two mitogen-activated protein kinases; MAPK8/JNK1 (RS= 0.34, $p=3.0x10^{-4}$) and MAPK13 (RS= 0.24, $p=1.6x10^{-2}$); two tyrosine-protein kinase proto-oncoproteins; FGR (RS= 0.29, $p=4.8x10^{-3}$) and LCK (RS= 0.41, $p=2.8x10^{-4}$); and two proteins with a role in cell cycle progression cyclin-dependent kinase 7 (CDK7) (RS= 0.23, $p=2.9x10^{-5}$) and wee1-like protein kinase (WEE1) (RS= 0.60, $p=4.2x10^{-5}$). The upregulation of the kinase JNK1, a protein with several functions capable of promoting tumourigenesis, was potentially a target of inhibition by the cyclic depsipeptide aplidine. Aplidine has previously been used to treat leukaemias and JNK

isoforms were found to be responsible, at least in part for its mechanism of action [486, 487]. JNK1 was expressed at 125% of that in HD samples in 9 CLL samples. MAPK13, the delta isoform of the p38 MAPK family, has potential roles in regulating protein translation and microtubule remodelling in response to cytokine signalling, DNA damage and stress. The protein was upregulated >160% in 6 CLL samples. The compound talmapimod is a known small molecular inhibitor of MAPK13, though with specificity to all members of the p38 MAPK family. LCK, a kinase which phosphorylates and activates ZAP70, normally upon T-cell signalling, was upregulated >125% in 10 of the CLL samples and in 6 of these samples, >150%. Reduction of LCK activity has previously been shown to reduce BCR signalling in CLL [488]. FGR, a kinase potentially subject to inhibition by vemurafenib, was upregulated (>125%) in 8 CLL cases. FGR normally functions to direct immune response and cytoskeletal remodelling in immune cells in response to ITGB2 (downregulated) signalling [489].

The cell cycle progression promoting kinase, WEE1, was observed upregulated to >175% in 9, and to >125% in a further 4, CLL samples verses HD control. Annotation identified the WEE1-specific inhibitor MK1775, which has shown successful implementation in the treatment of several cancers, as a potential inhibitor of any effects of this WEE1 upregulation. The second cell cycle regulating kinase, CKD7, was identified as a potential inhibition target of alvocidib, with 9 cases with >120% expression. Alvocidib has previously demonstrated cytotoxicity in CLL [490].

Three HDAC enzymes were identified with some degree of upregulation in CLL; HDAC1 (RS= 0.24, $p=1.3x10^{-4}$), HDAC3 (RS= 0.31, $p=1.7x10^{-5}$) and HDAC7 (RS= 0.51, $p=5.0x10^{-6}$). All three enzymes were identified with the potential to be inhibited with compounds termed HDACi. Entinostat, for instance, a specific inhibitor of HDAC1 and HDAC3 re-established tumour suppressor expression in CLL [491]. The most upregulated of these three proteins, HDAC7, with >135% expression in all but one sample, did not have any previous findings of upregulation at the protein level. NAD-dependent protein deacylase sirtuin-5, mitochondrial (SIRT5), an enzyme with deacetylase activity, but more probably demalonylation and desuccinylation (similar-property small carbonyl motifs to acetylation), was marginally upregulated (RS= 0.24, $p=1.0x10^{-3}$). SIRT5 was identified with potential inhibition by suramin.

Another identified target of inhibition was that of succinate-semialdehyde dehydrogenase, mitochondrial (ALDH5A1), a substantially and consistently overexpressed enzyme in the data, with all samples overexpressing by at least 125% and 8 samples over 200%. ALDH5A1, frequently abbreviated to SSADH, catalyses a step in the degradation of gamma-aminobutyric acid, specifically succinate semialdehyde to succinic acid, which is reported to be non-specifically inhibited by valproic acid.

Fibroblast growth factor (FGF2) was identified significantly upregulated (RS= 0.48, $p=6.0x10^{-3}$), but with variable expression; four samples had 3.5- to 6.3- fold upregulation, and 7 samples had 1.4- to 2.3-fold upregulation. A typically extracellular signalling molecule with capabilities of inducing cell survival and proliferation, FGF2, was annotated to have inhibitors such as suradista which could reduce its activity.

Four CLL-upregulated proteins with roles in DNA replication were identified with available chemotherapeutic reagents capable of inhibiting these proteins. DNA polymerase epsilon subunit 4 (POLE4) (RS= 0.23, $p=5.5x10^{-4}$), DNA (cytosine-5-)-methyltransferase 1 (DNMT1) (RS= 0.29, $p=1.3x10^{-4}$), DNA topoisomerase I (TOP1) (RS= 0.55, $p=3.2x10^{-6}$) and ribonucleoside-diphosphate reductase subunit M2 (RRM2B) (RS= 0.31, $p=1.1x10^{-4}$) were annotated to be inhibited by nucleotide analogues such as gemcitabine, cytarabine, decitabine and 5-azacytidine and, specifically to TOP1; the alkaloids camptothecin and irinotecan.

Four other proteins had significant upregulation with a potential for inhibition; retinoid X receptor, alpha (RXRA) (RS= 0.34, $p=8.1x10^{-4}$), promyelocytic leukemia protein (PML) (RS= 0.27, $p=3.0x10^{-3}$), peptidyl-prolyl cis-trans isomerase A (PPIA) (RS= 0.38, $p=7.5x10^{-6}$) and arachidonate 5-lipoxygenase (ALOX5) (RS= 0.42, $p=2.6x10^{-5}$).

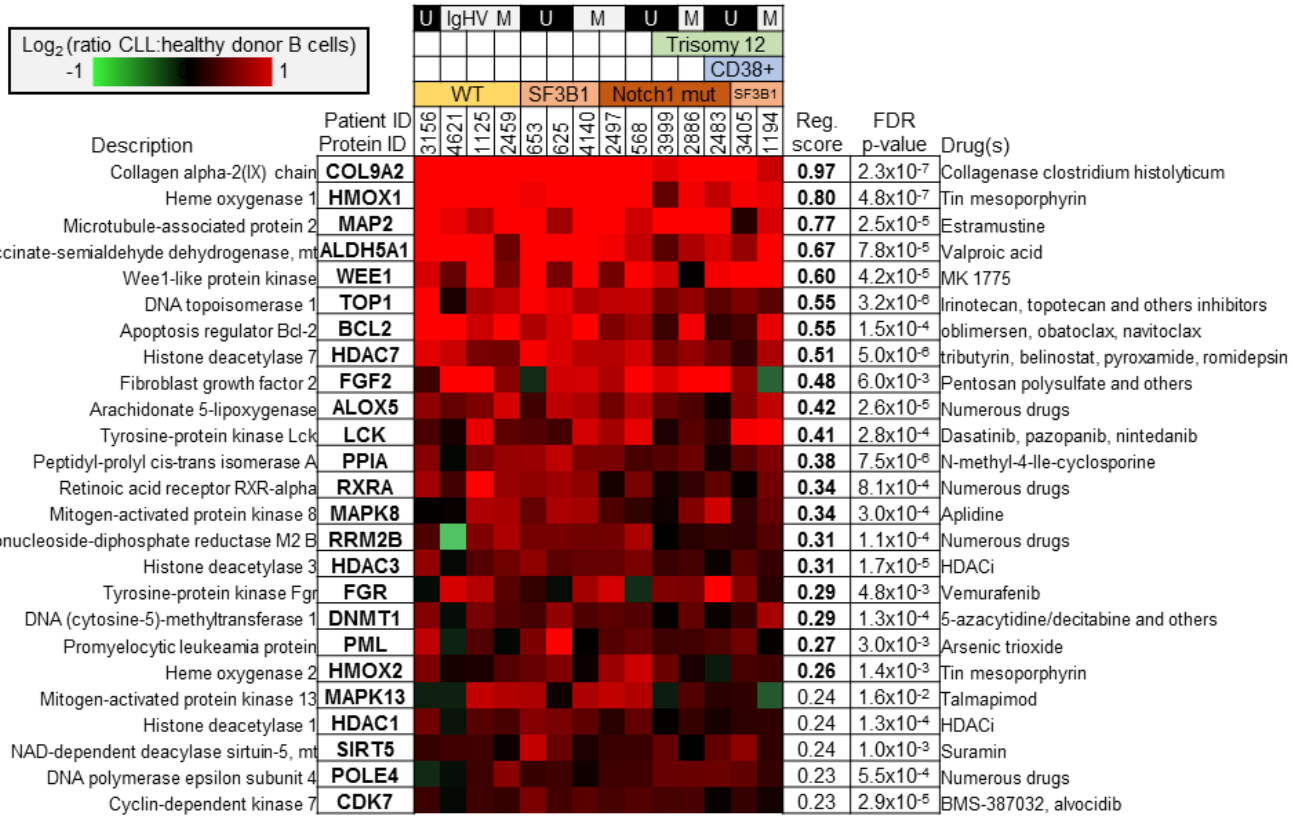


Figure 7.9. Protein targets of small molecule inhibition in CLL. Proteins upregulated in CLL with potential for interference with inhibitors. The top 25 previously described drug targets, as annotated by ingenuity pathway analysis, based on regulation score with a false discovery rate corrected p-value of <0.05.

7.11 LOCALISATION AND FUNCTIONS OF DIFFERENTIALLY EXPRESSED PROTEINS IN CLL

To gain an insight into the patterns emerging amongst the differentially regulated proteins in CLL, not observable by the analysis of single proteins, a series of bioinformatic tools were employed. Firstly, the significantly upregulated (819) or downregulated (746) proteins in CLL, as well as the list of all those fully quantitated proteins (5956), were annotated to describe the canonical localisations and functions using IPA software. The numbers of proteins annotated with each localisation for the regulated or full list of proteins were plotted against the numbers observed from all proteins to identify patterns of enrichment (**Figure 7.10A**). 3.4% (200) of all fully quantitated proteins were annotated as canonically extracellular. Of the up and downregulated proteins, 2.2% (18) and 6.6% (46) were extracellular, demonstrating under and overrepresentation, respectively. The same trend was even more pronounced at the plasma membrane (highlighted previously in **Figure 7.8**), with 30 (3.7%) and 118 (16.9%) of the over and underexpressed proteins, respectively, localised to the cell periphery compared to 419 (7%) of all the proteins. This demonstrated that over a quarter of proteins identified at the cell surface were consistently downregulated in CLL. Just under half of all proteins were cytoplasmic (2830, 47.5%) which were again, although marginally, enriched amongst the downregulated proteins (394, 56.4%) and underrepresented amongst the upregulated proteins (227, 28.2%). The most striking trend was that observed for the nucleus, where 58.3% (470) of the proteins determined to be upregulated were annotated as being canonically localised. 1 in 4 of the 1880 (31.6%) proteins identified from the nucleus were therefore determined with consistent upregulation. The trend was complimented by the observation that just 10.6% (74) of downregulated proteins were nuclear. Proteins annotated without a canonical localisation had approximately expected numbers with 10.5% (627), 7.6% (61) and 9.6% (67) of all, up and downregulated proteins respectively.

An identical analysis was performed, plotted as pie charts, for the annotations made which described canonical protein classifications (**Figure 7.10B**). Approximately half of proteins for each group were annotated with a classification and for simplicity, those without were not considered when determining percentages. Transcriptional regulators, the second most frequently identified class of proteins (552, 18%), were over and underrepresented amongst the

up and downregulated proteins in CLL, respectively. 37% (141) of those upregulated proteins were annotated as transcriptional regulators, more than a quarter of all identified transcriptional regulators. Just 22 (6%) of the downregulated proteins were transcriptional regulators. An inverse trend was observed for those proteins annotated as transporters (373, 12%), kinases (306, 10%) and peptidases (175, 6%), which were all marginally overrepresented, by percentage, amongst the underexpressed proteins; 20% (76), 13% (50) and 8% (31), of these proteins respectively. The overexpressed proteins also underrepresented these three classifications, with 28 (7%) transporters, 25 (6%) kinases and 11 (3%) peptidases. Phosphatases showed no enrichment with 4% of classified proteins for each category. Translational regulators appeared with minimal differential expression in CLL, with 3 (0.78%) and 6 (1.56%) of the 76 (2.45%) over and underexpressed. Transmembrane receptors (69, 2.2%), in agreement with the trend for plasma membrane proteins (**Figure 7.10A** and **Figure 7.8**), were substantially skewed towards downregulation (27, 7%) with more than a third of all those identified underexpressed, and just 3 (0.78%) with significant upregulation. Alongside this were ion channels of which 9 (2.3%) of the 35 (1.1%) identified were significantly downregulated, without any exhibiting upregulation. Of the 12 proteins classified as cytokines, two; interleukin-like epithelial to mesenchymal transition inducer (FAM3C/ILEI) and IK down-regulator of HLA II cytokine (RED/IK) were significantly upregulated; and two; guanine nucleotide exchange factor VAV3 and IL18 were significantly downregulated.

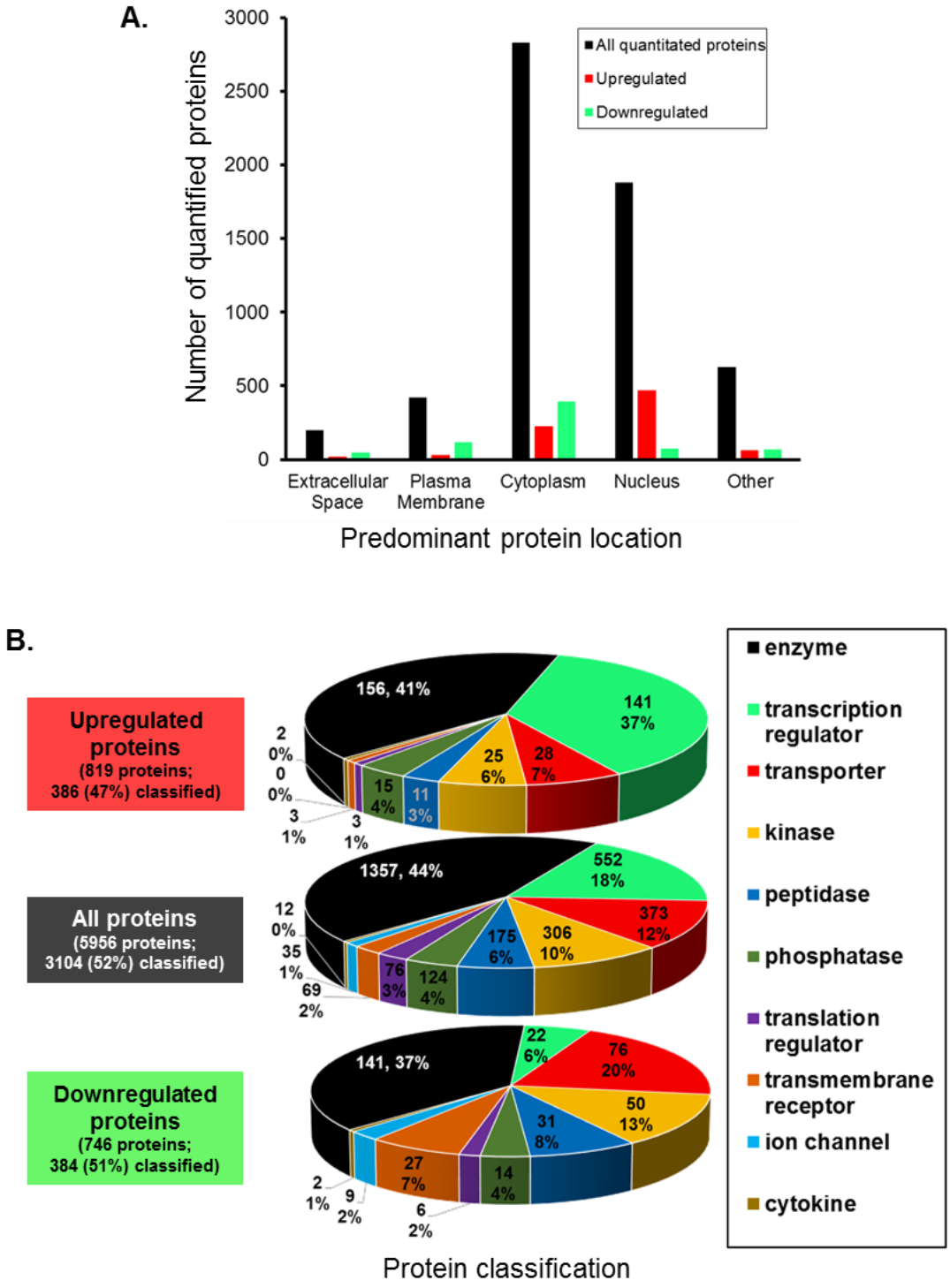


Figure 7.10. Annotation of protein localisations and subtypes for the up and downregulated proteins in CLL. A. Proteins defined as up and down regulated in CLL were categorised for their predominant protein localisation and compared with all protein identifications. **B.** As A. for the annotated functional proteins categories displayed as pie charts. Unclassified proteins are not plotted and the percentages of these proteins do not greatly differ.

7.12 GENE ONTOLOGY ENRICHMENT OF DIFFERENTIALLY EXPRESSED PROTEINS IN CLL

To gain a more detailed and specific insight into the biological trends emerging in CLL in **Figure 7.10**, GO term enrichment was employed. GO term enrichment was determined for the 819 and 746 significantly up and downregulated proteins against a background of all 5956 fully quantitated proteins with DAVID and visualised by Revigo. Significant GO term enrichment amongst the upregulated proteins in CLL, is plotted in **Figure 7.11** describing biological processes, cellular components and molecular functions.

For biological processes, several terms describing the upregulation of general metabolism, regulation and cell processes were significantly enriched; 'metabolic process (GO:0008152)' ($p=3.35x10^{-8}$), 'cellular process (GO:0009987)' ($p=2.13x10^{-5}$), 'biological regulation (GO:0065007)' ($p=4.31x10^{-2}$), 'regulation of metabolic process (GO:0019222)' ($p=1.11x10^{-10}$), 'biosynthetic process (GO:0009058)' ($p=5.63x10^{-3}$), 'regulation of cellular process (GO:0050794)' ($p=1.61x10^{-2}$), 'cellular metabolic process (GO:0044237)' ($p=8.00x10^{-11}$), 'primary metabolic process (GO:0044238)' ($p=1.49x10^{-10}$) and 'regulation of biological process (GO:0050789)' ($p=2.76x10^{-2}$). In addition to these, two processes relating to macromolecules, such as DNA, RNA and proteins were highlighted; 'macromolecule biosynthetic process (GO:0009059)' ($p=2.50x10^{-5}$) and 'macromolecule metabolic process (GO:0043170)' ($p=4.19x10^{-13}$). The two most significantly enriched terms 'nucleobase-containing compound metabolic process (GO:0006139)' ($p=3.36x10^{-38}$) and 'nitrogen compound metabolic process (GO:0006807)' ($p=2.69x10^{-40}$) both indicated that nucleic acid metabolism, linked with nitrogen metabolism, was upregulated. The terms; 'RNA processing (GO:0006396)' ($p=4.46x10^{-23}$), 'mRNA processing (GO:0006397)' ($p=1.23x10^{-34}$) and 'mRNA metabolic process (GO:0016071)' ($p=2.37x10^{-31}$); indicated that CLL cells consistently upregulated proteins involved in the synthesis of mRNA, significantly more frequently than would be expected by chance. Further to this, other terms which reached substantial significance, not shown due to overlapping term redundancy, specified that splicing and the spliceosome were a major component of this finding; 'RNA splicing (GO:0008380)' ($p=9.01x10^{-32}$) and 'mRNA splicing, via spliceosome (GO:0000398)' ($p=2.69x10^{-25}$). Processes prior to splicing were also enriched, with aspects of aspects of transcription and gene expression suggested to be upregulated; 'gene expression (GO:0010467)' ($p=1.93x10^{-24}$) and 'negative regulation of transcription, DNA-templated (GO:0045892)' ($p=3.52x10^{-7}$). The term 'chromatin organization (GO:0006325)' ($p=2.05x10^{-11}$), based on related and redundant terms, were indicative of a significant enrichment of upregulated proteins involved in histone acetylation, specifically of histone H2A and H4 ($p<0.05$) which also suggested transcriptional dysregulation. 'DNA

metabolic process (GO:0006259)' ($p=6.49x10^{-5}$), combined with the terms 'cellular response to DNA damage stimulus (GO:0006974)' ($p=6.13x10^{-3}$) and 'cell aging (GO:0007569)' ($p=3.22 x10^{-2}$), were indicative of a state induced by cell division and DNA replication.

Cellular component GO term enrichment confirmed, most significantly, the previous observation of frequently upregulated nuclear proteins in CLL ('nucleus (GO:0005634)' ($1.89x10^{-47}$)). The analysis also provided a breakdown of more specific enriched terms; 'nucleoplasm (GO:0005654)' ($p=1.81x10^{-17}$), 'nucleoplasm part (GO:0044451)' ($p=1.54x10^{-11}$), 'nuclear matrix (GO:0016363)' ($p=7.41x10^{-3}$) and 'nuclear periphery (GO:0034399)' ($p=1.85x10^{-2}$). Additional terms were in agreement with biological processes relating to mRNA processing; 'spliceosomal complex (GO:0005681)' ($p=7.62x10^{-17}$) and 'nuclear speck (GO:0016607)' ($p=2.06x10^{-8}$) – a region of the nucleus rich in splicing activity - further indicated an upregulated of splicing. Further GO terms identified significant enrichment of nuclear proteins relating to DNA and the regulation of chromosomes; 'chromosome (GO:0005694)' ($p=4.39x10^{-7}$), 'chromatin remodeling complex (GO:0016585)' ($p=5.93x10^{-3}$), 'protein-DNA complex (GO:0032993)' ($p=6.17x10^{-3}$), 'chromosomal part (GO:0044427)' ($p=5.49x10^{-5}$) and 'histone acetyltransferase complex (GO:0000123)' ($p=5.80x10^{-3}$). The terms 'small nuclear ribonucleoprotein complex (GO:0030532)' ($p=1.10x10^{-2}$), 'heterogeneous nuclear ribonucleoprotein complex (GO:0030530)' ($p=8.55x10^{-5}$) and 'ribonucleoprotein complex (GO:0030529)' ($p=6.93x10^{-4}$) were related to the overexpression of proteins localised to DNA and RNA, such as the chromosomes and spliceosome, respectively.

The most significantly enriched molecular function in CLL, 'nucleic acid binding (GO:0003676)' ($p=6.40x10^{-32}$), summarised the majority of other molecular function terms, most specifically; 'DNA binding (GO:0003677)' ($p=2.68x10^{-17}$) and 'RNA binding (GO:0003723)' ($p=3.28x10^{-15}$). DNA binding could be further subdivided into the similar but more specific terms 'sequence-specific DNA binding transcription factor activity (GO:0003700)' ($p=1.56x10^{-3}$), 'transcription regulator activity (GO:0030528)' ($p=1.75x10^{-11}$) and 'transcription factor binding (GO:0008134)' ($p=1.89x10^{-3}$) and the enrichment of several other terms further detailed these; 'transcription activator activity (GO:0016563)' ($p=1.35x10^{-3}$), 'transcription repressor activity (GO:0016564)' ($p=1.01x10^{-4}$), 'transcription coactivator activity (GO:0003713)' ($p=2.19x10^{-2}$) and 'transcription cofactor activity (GO:0003712)' ($p=6.73x10^{-4}$). The terms 'zinc ion binding (GO:0008270)' ($p=2.89x10^{-5}$) and 'transition metal ion binding (GO:0046914)' ($p=1.24x10^{-3}$) suggested the type of transcription factors, such as the 45 zinc finger proteins, which were most significantly enriched in these data.

Significantly downregulated protein GO term enrichment in CLL (**Figure 7.12**) confirmed and provided greater insight into many of the patterns described in **Figure 7.10**. For

biological processes, the term ‘localization (GO:0051179)’ was the most significantly enriched ($p=9.58x10^{-11}$). This term described several subtypes of localisation, the most significant of which was ‘vesicle-mediated transport (GO:0016192)’ ($p=4.00x10^{-6}$), which was observed alongside related terms suggesting a downregulation of proteins related to membrane vesicular movement; ‘lipid localization (GO:0010876)’ ($p=1.23x10^{-3}$), ‘lipid metabolic process (GO:0006629)’ ($p=1.87x10^{-2}$), ‘membrane invagination (GO:0010324)’ ($p=3.09x10^{-2}$) and ‘cellular component movement (GO:0006928)’ ($p=2.39x10^{-2}$). Other forms of intracellular protein and peptide movement were also suggested to be downregulated; ‘peptide transport (GO:0015833)’ ($p=1.57x10^{-2}$) and ‘protein processing (GO:0016485)’ ($p=3.86x10^{-2}$). Processes relating to whole cell localisation and adhesion of were also significantly enriched for; ‘cell migration (GO:0016477)’ ($p=2.49x10^{-2}$), ‘localization of cell (GO:0051674)’ ($p=2.46x10^{-2}$), ‘locomotion (GO:0040011)’ ($p=1.86x10^{-3}$), ‘cell adhesion (GO:0007155)’ ($p=9.52x10^{-3}$) and ‘biological adhesion (GO:0022610)’ ($p=9.52x10^{-3}$). Related to such processes were ‘cytoskeleton organization (GO:0007010)’ ($p=1.01x10^{-3}$) and ‘actin filament-based process (GO:0030029)’ ($p=8.61x10^{-4}$). CLL also demonstrated a significant trend in the downregulation of proteins relating to communication with the immune system; ‘immune response (GO:0006955)’ ($p=1.92x10^{-5}$) and ‘immune system process (GO:0002376)’ ($p=1.30x10^{-4}$); already highlighted in the trends which emerged amongst the downregulated membrane proteins (**Figure 7.8**). Other terms describing intercellular communication; ‘response to stimulus (GO:0050896)’ ($p=1.64x10^{-4}$), ‘response to external stimulus (GO:0009605)’ ($p=2.64x10^{-4}$), ‘response to wounding (GO:0009611)’ ($p=1.10x10^{-4}$), ‘regulation of response to external stimulus (GO:0032101)’ ($p=1.65x10^{-3}$), and ‘regulation of response to stimulus (GO:0048583)’ ($p=2.04x10^{-2}$); were also overrepresented. Three terms highlighted some possible specific signalling processes by which these more general processes were being downregulated; ‘integrin-mediated signaling pathway (GO:0007229)’ ($p=2.47x10^{-4}$), ‘small GTPase mediated signal transduction (GO:0007264)’ ($p=1.08x10^{-2}$) and ‘protein phosphorylation (GO:0006468)’ ($p=3.41x10^{-2}$).

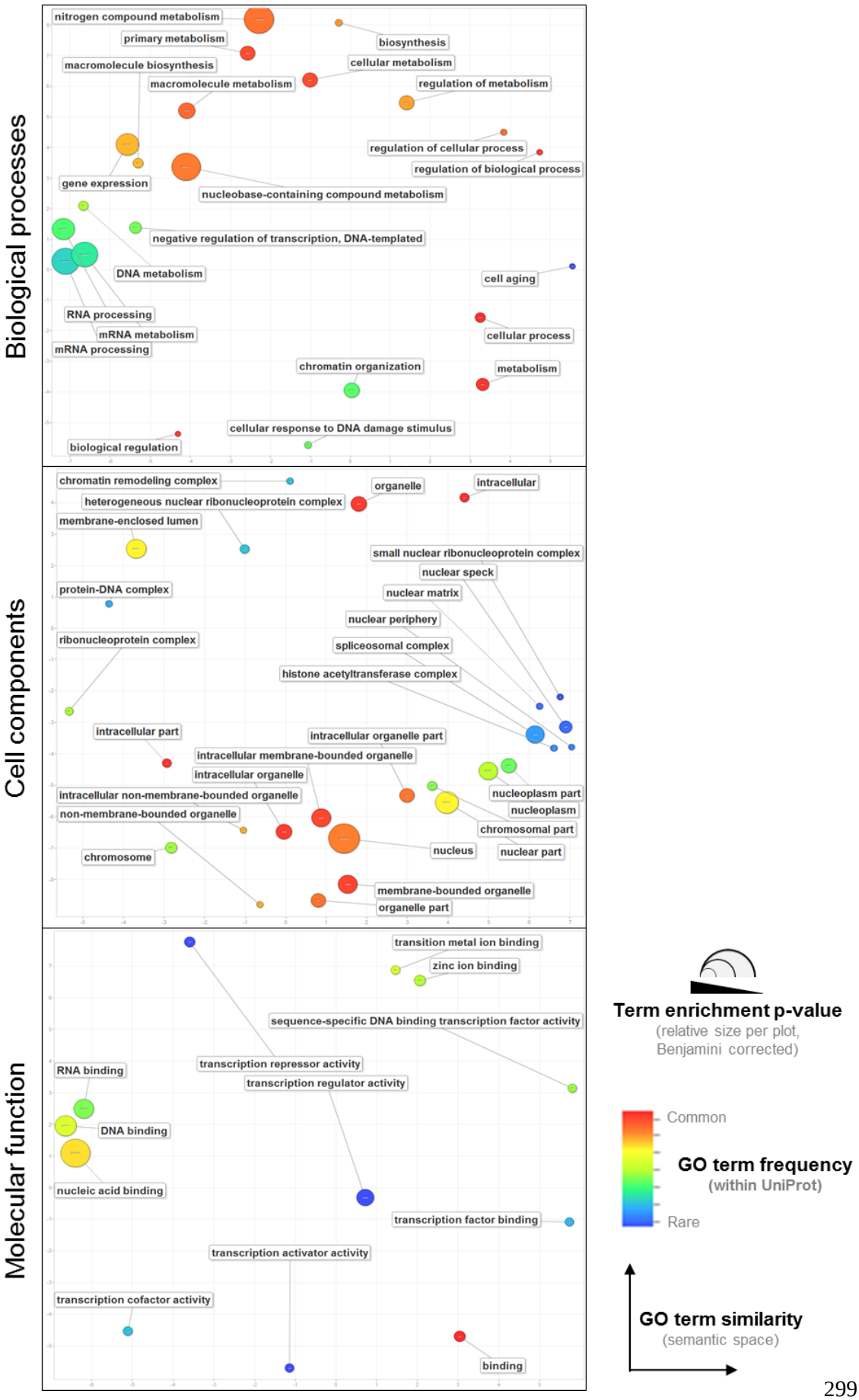
For cellular components the most prominent trend was that of membrane protein downregulation, with the terms ‘membrane (GO:0016020)’ and ‘membrane part (GO:0044425)’ by far the most significantly downregulated ($p<10^{-40}$). This analysis highlighted the localisation to different membranous regions within the cell, with the cell membrane being the most significant; ‘plasma membrane (GO:0005886)’ ($p=1.15x10^{-20}$) and ‘cell surface (GO:0009986)’ ($p=2.71x10^{-5}$). Other membranous compartments were also described with significant enrichment; ‘vacuole (GO:0005773)’ ($p=1.59x10^{-7}$), ‘lysosome (GO:0005764)’ ($p=2.32x10^{-7}$), ‘endoplasmic reticulum (GO:0005783)’ ($p=5.73x10^{-6}$), ‘golgi apparatus (GO:0005794)’

($p=6.31x10^{-6}$), 'cytoplasmic membrane-bounded vesicle (GO:0016023)' ($p=1.22x10^{-4}$) and 'vesicle (GO:0031982)' ($p=1.97x10^{-4}$). Terms also described the different associations of proteins with membranes, suggesting that protein without a covalent integration into the membrane ('extrinsic component of membrane (GO:0019898)' ($p=6.34x10^{-4}$)) has less enrichment than those with a covalent component within the membrane; 'integral component of membrane (GO:0016021)' ($p=6.23x10^{-32}$) and 'intrinsic component of membrane (GO:0031224)' ($p=2.90x10^{-33}$). This trend also held true at the cell surface where the terms 'intrinsic component of plasma membrane (GO:0031226)' and 'integral component of plasma membrane (GO:0005887)' were significant ($p<10^{-5}$), while no extrinsic term was enriched. The significantly enriched terms 'receptor complex (GO:0043235)' ($p=1.58x10^{-3}$), 'cell projection (GO:0042995)' ($p=4.80x10^{-3}$), 'cell junction (GO:0030054)' ($p=1.20x10^{-2}$) and 'basolateral plasma membrane (GO:0016323)' ($p=2.77x10^{-4}$) gave a more detailed impression of the types of cell surface proteins downregulated. Additional cellular component terms described frequent downregulation of extracellular proteins; 'extracellular region (GO:0005576)' ($p=3.07x10^{-11}$), 'extracellular space (GO:0005615)' ($p=2.71x10^{-5}$) and 'extracellular matrix (GO:0031012)' ($p=1.77x10^{-2}$); as well as a prevalence of underexpressed cytoskeletal components; 'actin cytoskeleton (GO:0015629)' ($p=1.12x10^{-3}$).

Significantly enriched molecular functions described a trend of downregulation in proteins relating to responses to external signals, with the three most significant terms in CLL being; 'molecular transducer activity (GO:0060089)' ($p=3.86x10^{-8}$), 'receptor activity (GO:0004872)' ($p=2.13x10^{-6}$), and 'signal transducer activity (GO:0004871)' ($p=3.86x10^{-8}$); in addition to the term; 'transmembrane signaling receptor activity (GO:0004888)' ($p=4.57x10^{-2}$). Transportation also demonstrated significant enrichment ('transporter activity (GO:0005215)' ($p=8.54x10^{-5}$)), predominantly describing terms related specifically to transmembrane transport; 'transmembrane transporter activity (GO:0022857)' ($p=8.58x10^{-4}$), 'active transmembrane transporter activity (GO:0022804)' ($p=5.64x10^{-3}$), 'substrate-specific transporter activity (GO:0022892)' ($p=3.53x10^{-4}$) and 'substrate-specific transmembrane transporter activity (GO:0022891)' ($p=2.73x10^{-2}$). Another previously observed trend of downregulated cytoskeletal components was observed, with the significant enrichment of terms related to cytoskeletal binding; 'cytoskeletal protein binding (GO:0008092)' ($p=3.05x10^{-6}$), 'actin binding (GO:0003779)' ($p=1.54x10^{-4}$) and 'actin filament binding (GO:0051015)' ($p=3.16x10^{-2}$). 10 protein annotated with heme binding, a tetrapyrrole molecule, were also significantly downregulated in CLL giving enrichments for the terms 'heme binding (GO:0020037)' and 'tetrapyrrole binding (GO:0046906)' of $p=6.12x10^{-3}$ and $p=1.21x10^{-2}$, respectively. 15 downregulated proteins, predominantly Ig and MHC class II molecules, gave significant

enrichment to the term ‘antigen binding (GO:0003823)’ ($p=5.36x10^{-3}$). The downregulation of 14 serine proteases gave a significant enrichment to the terms ‘serine-type peptidase activity (GO:0008236)’ ($p=2.57x10^{-2}$) and ‘serine hydrolase activity (GO:0017171)’ ($p=3.18x10^{-2}$). ‘GTPase activity (GO:0003924)’ was significantly enriched ($p=1.20x10^{-3}$) due to downregulation in 28 proteins with this function, 11 of which were Ras-related proteins. GTPase activity has pleitropic roles in many of the enriched processes described amongst the downregulated proteins, including vesicle transport, transmembrane signal transduction and protein translocation.

Figure 7.11. Gene ontology (GO) term enrichment amongst upregulated CLL proteins. Significantly enriched GO terms ($p<0.05$) for biological processes, cell components and molecular function determined by DAVID and represented using Revigo. Benjamini-corrected GO term p-values were determined using those 819 significantly upregulated ($RS>0.25$, $p<0.05$) proteins against a background of the 5956 fully quantitated proteins.


Biological processes
nitrogen compound metabolism
primary metabolism
biosynthesis
macromolecule biosynthesis
cellular metabolism
macromolecule metabolism
regulation of metabolism
regulation of cellular process
regulation of biological process
gene expression
nucleobase-containing compound metabolism
negative regulation of transcription, DNA-templated
DNA metabolism
cell aging
RNA processing
cellular process
mRNA metabolism
mRNA processing
chromatin organization
metabolism
cellular response to DNA damage stimulus
biological regulation
Cell components
chromatin remodeling complex
organelle
intracellular
heterogeneous nuclear ribonucleoprotein complex
membrane-enclosed lumen
small nuclear ribonucleoprotein complex
nuclear speck
protein-DNA complex
nuclear matrix
nuclear periphery
ribonucleoprotein complex
spliceosomal complex
histone acetyltransferase complex
intracellular part
intracellular organelle part
intracellular membrane-bounded organelle
intracellular organelle
nucleoplasm part
intracellular non-membrane-bounded organelle
nucleoplasm
non-membrane-bounded organelle
chromosomal part
nucleus
nuclear part
chromosome
membrane-bounded organelle
organelle part
Molecular function
transition metal ion binding
zinc ion binding
sequence-specific DNA binding transcription factor activity
transcription repressor activity
RNA binding
transcription regulator activity
DNA binding
nucleic acid binding
transcription factor binding
transcription activator activity
transcription cofactor activity
binding
Term enrichment p-value
(relative size per plot,
Benjamini corrected)
Common
GO term frequency
(within UniProt)
Rare
GO term similarity
(semantic space)

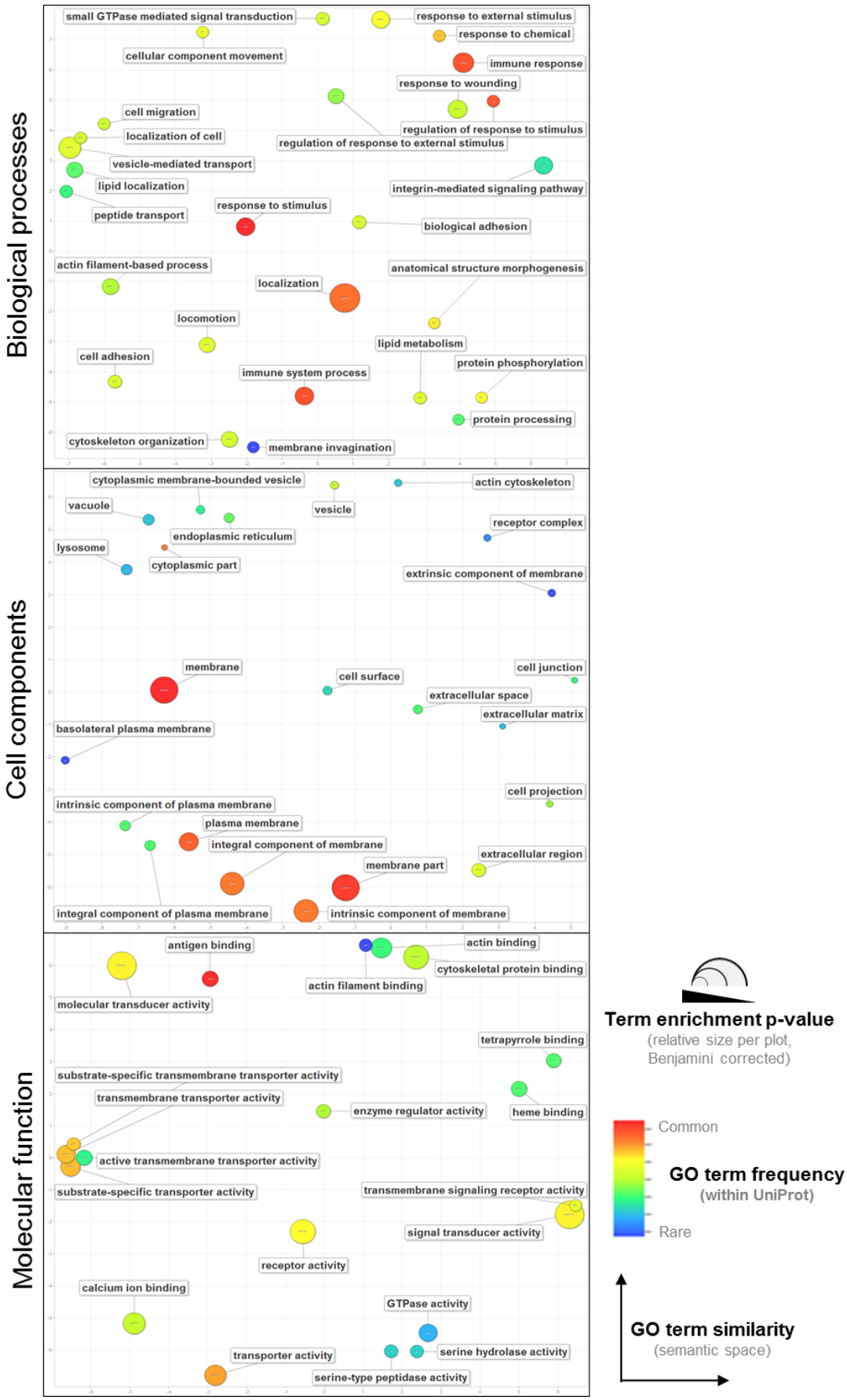

Biological processes
small GTPase mediated signal transduction
response to external stimulus
response to chemical
cellular component movement
immune response
response to wounding
cell migration
regulation of response to stimulus
localization of cell
regulation of response to external stimulus
vesicle-mediated transport
integrin-mediated signaling pathway
lipid localization
peptide transport
response to stimulus
biological adhesion
actin filament-based process
anatomical structure morphogenesis
localization
locomotion
lipid metabolism
cell adhesion
protein phosphorylation
immune system process
protein processing
cytoskeleton organization
membrane invagination
Cell components
cytoplasmic membrane-bounded vesicle
actin cytoskeleton
vacuole
vesicle
receptor complex
endoplasmic reticulum
lysosome
cytoplasmic part
extrinsic component of membrane
membrane
cell surface
cell junction
extracellular space
extracellular matrix
basolateral plasma membrane
cell projection
intrinsic component of plasma membrane
plasma membrane
integral component of membrane
extracellular region
membrane part
integral component of plasma membrane
intrinsic component of membrane
Molecular function
antigen binding
actin binding
cytoskeletal protein binding
actin filament binding
molecular transducer activity
tetrapyrrole binding
substrate-specific transmembrane transporter activity
transmembrane transporter activity
enzyme regulator activity
heme binding
active transmembrane transporter activity
substrate-specific transporter activity
transmembrane signaling receptor activity
signal transducer activity
receptor activity
calcium ion binding
GTPase activity
transporter activity
serine hydrolase activity
serine-type peptidase activity
Term enrichment p-value
(relative size per plot,
Benjamini corrected)
Common
GO term frequency
(within UniProt)
Rare
GO term similarity
(semantic space)

Figure 7.12. Gene ontology (GO) term enrichment amongst downregulated CLL proteins. As described for **Figure 7.11**, a representation of significant GO terms enrichment ($p<0.05$) for biological processes, cell components and molecular function determined by DAVID and represented using Revigo. Benjamini-corrected GO term p-values were determined using those 746 significantly downregulated ($RS< -0.25$, $p<0.05$) proteins against a background of the 5956 fully quantitated proteins.

7.13 PATHWAY ENRICHMENT OF DIFFERENTIALLY EXPRESSED PROTEINS IN CLL

GO term enrichment presented several strong trends in the up and downregulated proteins, outlining patterns of underlying biological process in CLL. In order to assess these in the context of specific proteins and groups of related proteins working towards similar functions, pathways were evaluated for enrichment.

The spliceosome was the most significantly enriched pathway amongst the differentially regulated proteins in CLL ($p=2.0x10^{-44}$). The spliceosome, a series of proteins which control the removal of introns from pre-mRNA, involves approximately 133 gene products. 111 of these were fully quantitated in CLL and amongst these 61 of the 819 significantly upregulated proteins ($RS>0.25$, $p<0.05$) were components. A further 41 of the significant proteins which didn't reach a regulation score of 0.25 ($RS>0$, $p<0.05$) also mapped to the spliceosome. These proteins were annotated onto the KEGG pathway describing the spliceosome (**Figure 7.13A**). This trend was evaluated further by overlaying all identified spliceosome proteins onto the volcano plot from **Figure 7.7** (**Figure 7.13B**). Of the 111 quantified spliceosome proteins, all but 1 had a positive regulation score ($RS>0$), and only 9 were not significant. The average regulation score for these proteins was 0.249, with an average significance of $1.16x10^{-4}$. The most significantly upregulated components were WW domain-binding protein 11 (WBP11), RNA-binding motif protein, X chromosome (RMBX), transformer-2 protein homolog beta (TRA2B) and RNA-binding protein 8A (RBM8A) ($RS>0.38$, $p<10^{-5}$). Another significantly enriched KEGG pathway amongst the upregulated proteins in CLL was 'valine, leucine and isoleucine degradation' ($p=7.2x10^{-9}$) (**Figure 7.14**). Of the 47 proteins described in this pathway, 17 were matched from the 819 significantly upregulated CLL proteins.

Pathway enrichment was also performed with IPA, analysing three sets of proteins; those with significant upregulation ($RS>0.25$, $p<0.05$), those with significant downregulation ($RS<-0.25$, $p<0.05$) and an analysis considering all the significantly differentially regulated proteins ($RS>0.25/RS<-0.25$, $p<0.05$), summarised in **Appendix A11**. In each case, a stringent search was performed allowing for only direct relationships and experimentally observed

findings in human samples. The regulation score was used for the analyses to give a robust measurement representative of the overall trend in expression in all samples.

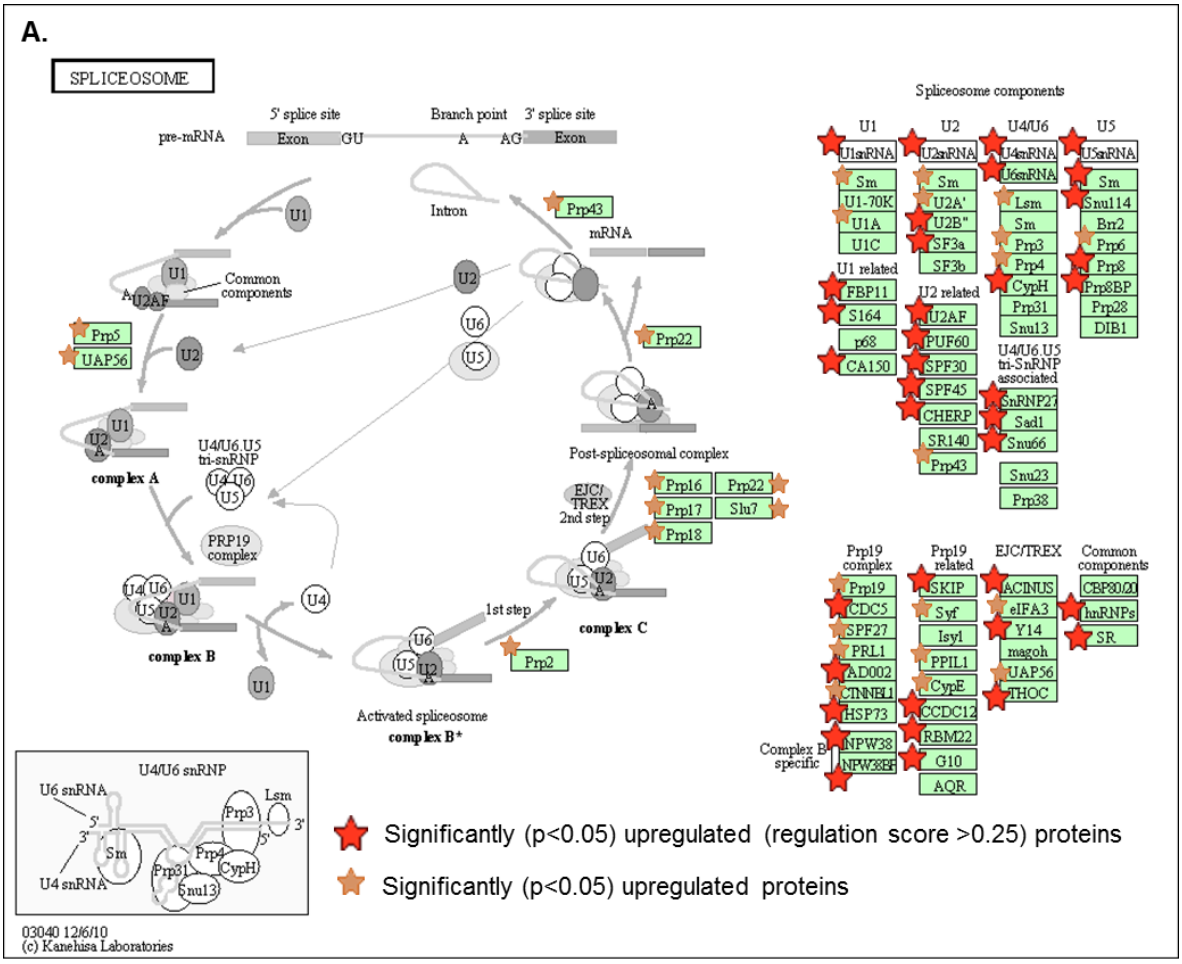
Analysis of the upregulated proteins identified the most significantly enriched canonical pathway as 'cleavage and polyadenylation of pre-mRNA' ($p=1.02x10^{-12}$). This highlighted 10 of the 819 upregulated proteins which mapped to the 12 proteins of this pathway, which were overlaid with the regulation score to summarise the expression across all CLL samples (**Figure 7.15**). To detail the expression in each of the CLL samples, bar charts were used for each protein which described the expression in the same order used in **Figure 7.8** and **7.9**, with the final bar describing the regulation score. For symbols describing multiple protein isoforms the quantitated members were displayed to one side of the pathway, denoted by a broken grey arrow and 'MB' referring to 'members'. This pathway demonstrated a consistent pattern of upregulated proteins both driving and facilitating the 3' adenylation of mRNAs.

Mechanisms involved in transcriptional regulation were also upregulated; observed in the significantly enriched pathway 'DNA methylation and transcriptional repression' ($p=1.2x10^{-3}$) (**Figure 7.16**). This pathway highlighted epigenetic regulation via DNA methylation and subsequent histone acetylation. DNMT1 involved in the transfer of existing methylation patterns to copied DNA, was consistently upregulated (RS= 0.29, $p=1.3x10^{-4}$), while DNMT3A and DNMT3B were not quantified. The most significantly upregulated protein was methyl-CpG-binding protein 2 (MECP2) (RS=0.413, $p=6.1x10^{-4}$), a protein capable of binding methylated DNA without sequence-specificity and recruiting an HDAC complex. The majority of the components of the HDAC complexes, capable of transcriptionally repressing DNA by the acetylation of histones, were also consistently upregulated.

Another significantly enriched pathway in CLL was 'nucleotide excision repair' ($8.2x10^{-5}$), to which 8 significantly upregulated proteins aligned (**Figure 7.17**). Mapping regulation scores and isoforms to this pathway revealed several other more marginally upregulated proteins involved in this process. The pathway, describing a DNA damage response, identified the replication proteins (RP) RPA1 and RPA3 as the most significantly upregulated components ($RS>0.55$, $p<10^{-5}$). RNA polymerase II components were also identified with marginal, but consistent upregulation as part of this pathway. Related to this observation, was the significant enrichment of the assembly of the RNA polymerase II complex amongst the upregulated CLL proteins ($p=5.75x10^{-6}$). Alongside the RNA polymerase components, several associated factors were also observed with generally consistent upregulation, such as 13 TBP-associated factor (TAF) isoforms and 5 general transcription factors (GTF) (**Figure 7.18**). The core component, TATA-box binding protein (TBP) was, however, consistently downregulated (RS= -0.623, $p=7.0x10^{-6}$).

A pathway significantly enriched due to a predominance of downregulated CLL proteins was that of integrin signalling ($p= 3.80x10^{-8}$) (**Figure 7.19**). This highlighted the downregulation of 26 proteins in this pathway, including the 7 integrins described in **Figure 7.8** as well as the additional, marginal downregulation of ITGB7. In addition to these downregulations, the key integrin-related signalling molecule, focal adhesion kinase (FAK) was also downregulated in CLL (RS= -0.52, $p=1.2x10^{-3}$). While the upstream regulators were downregulated, several downstream proteins exhibited some upregulation in these pathways, including; 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1 (PLCG1), mitogen-activated protein kinase 8 (JNK1/MAPK8) and Ras-related protein R-Ras2 (RRAS2).

Closely related to integrin signalling and similarly enriched amongst the proteins downregulated in CLL, was leukocyte extravasation signalling ($1.35x10^{-7}$), the process by which leukocytes adhere to and migrate through the epithelial layer (**Figure 7.20**). 25 proteins were enriched in this pathway including proteins involved in the signalling in both the migrating leukocytes and the epithelial cells. In addition to 6 integrins downregulated, other cell adhesion molecules were also underexpressed, including most significantly junctional adhesion molecule C (JAM3) (RS= -1.25, $p=1.2x10^{-6}$) and claudin-5 (CLDN5) (RS= -1.39, $p=3.1x10^{-7}$). Two components of the docking structure, however, were upregulated including the T-cell protein leukosialin (CD43/SPN) (RS=0.19, $p=1.4x10^{-2}$) and intercellular adhesion molecule 3 (ICAM3/CD50) (RS=0.25, $p=1.9x10^{-3}$).


A.
SPLICEOSOME
5' splice site
Branch point
3' splice site
pre-mRNA
Exon
GU
A
AG
Exon
Intron
mRNA
Common components
U2AF
Prp5
UAP56
Prp43
Prp22
complex A
U4/U6.U5 tri-snRNP
PRP19 complex
complex B
Activated spliceosome
complex B*
1st step
complex C
Post-spliceosomal complex
EJC/ TREX 2nd step
Prp16
Prp17
Prp18
Slu7
Prp2
Spliceosome components
U1
U2
U4/U6
U5
U1snRNA
U2snRNA
U4snRNA
U6snRNA
U5snRNA
Sm
U1-70K
U1A
U1C
U2A'
U2B"
SF3a
SF3b
Lsm
Snu114
Brr2
Prp3
Prp4
Prp6
Prp8
CypH
Prp8BP
Prp31
Prp28
Snu13
DIB1
U1 related
FBP11
S164
p68
CA150
U2 related
U2AF
PUF60
SPF30
SPF45
CHERP
SR140
Prp43
U4/U6.U5 tri-SnRNP associated
SnRNP27
Sad1
Snu66
Snu23
Prp38
Prp19 complex
Prp19
CDC5
SPF27
PRL1
AD002
CINNBL1
HSP73
NPW38
NPW38BP
Complex B specific
Prp19 related
SKIP
Syf
Isy1
PPIL1
CypE
CCDC12
RBM22
G10
AQR
EJC/TREX
ACINUS
eIFA3
Y14
magoh
UAP56
THOC
Common components
CBP80/20
hnRNPs
SR
U4/U6 snRNP
U6 snRNA
U4 snRNA
Lsm
Prp3
Sm
Prp4
Prp31
CypH
Snu13
Significantly (p<0.05) upregulated (regulation score >0.25) proteins
Significantly (p<0.05) upregulated proteins
03040 12/6/10
(c) Kanehisa Laboratories

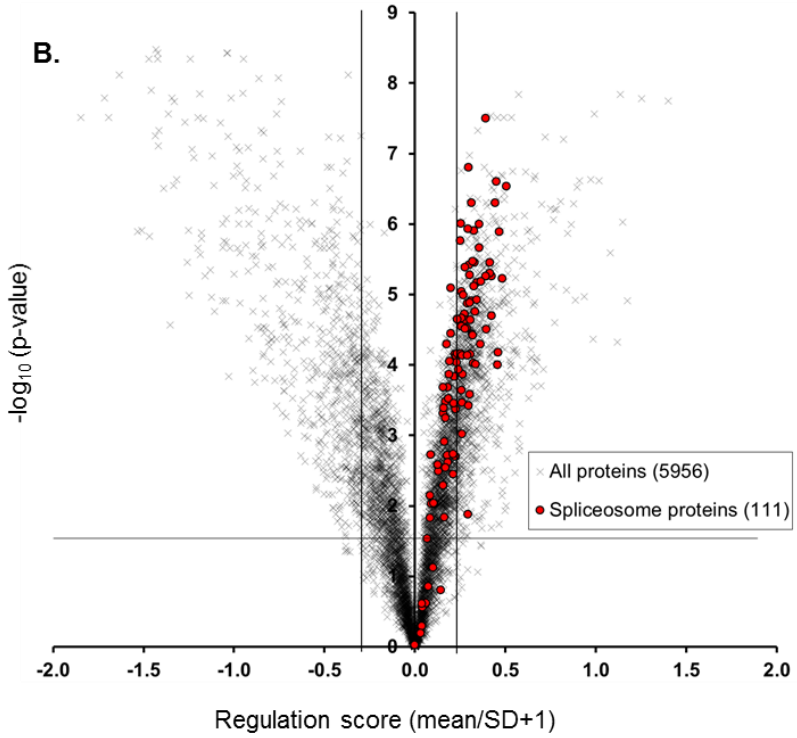

B.
-log10 (p-value)
Regulation score (mean/SD+1)
All proteins (5956)
Spliceosome proteins (111)
9
8
7
6
5
4
3
2
1
0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0

Figure 7.13. **Significantly upregulated spliceosome pathway proteins in CLL. A.** Significant enrichment of the KEGG pathway ($p=2.0x10^{-44}$) describing the components of the spliceosome overlaid with the 61/819 significantly upregulated proteins (RS>0.25, p<0.05) (red) and 41 of those with significance and a positive regulation score (RS>0, p<0.05) (orange) in CLL, relative to healthy donor B cells. **B.** The volcano plot detailed in **Figure 7.7**, overlaid with 111 of the 133 spliceosome proteins identified amongst the 5956 proteins.

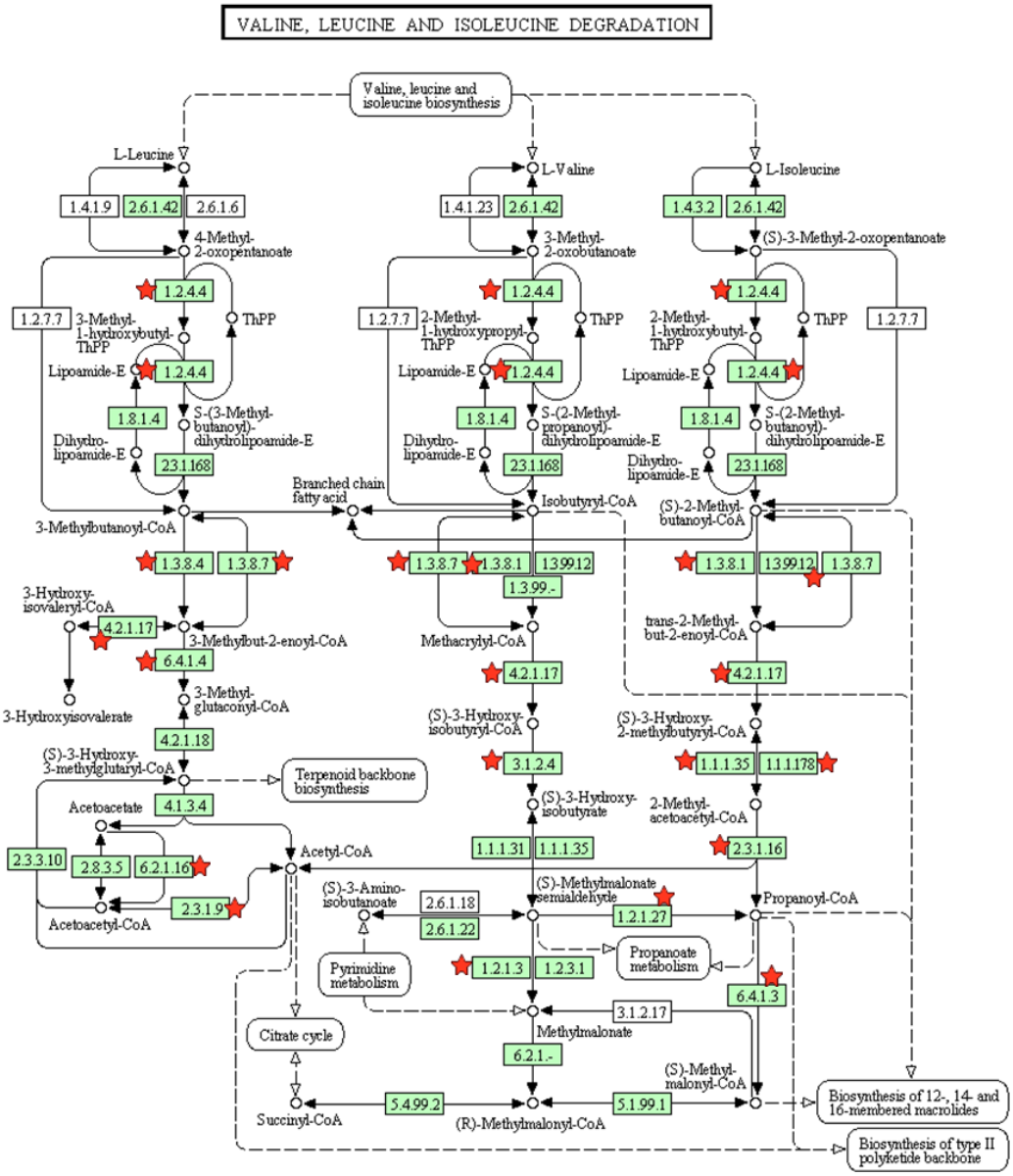


Figure 7.14. **Significantly upregulated CLL proteins mapping to the valine, leucine and isoleucine degradation pathways.** The overlaying of significantly upregulated (RS>0.25, p<0.05) CLL proteins onto the KEGG pathway describing valine, leucine and isoleucine degradation. A total of 17 of the 819 upregulated proteins mapped to the 47 proteins of the pathway, with a significant enrichment of $p=7.2x10^{-9}$.

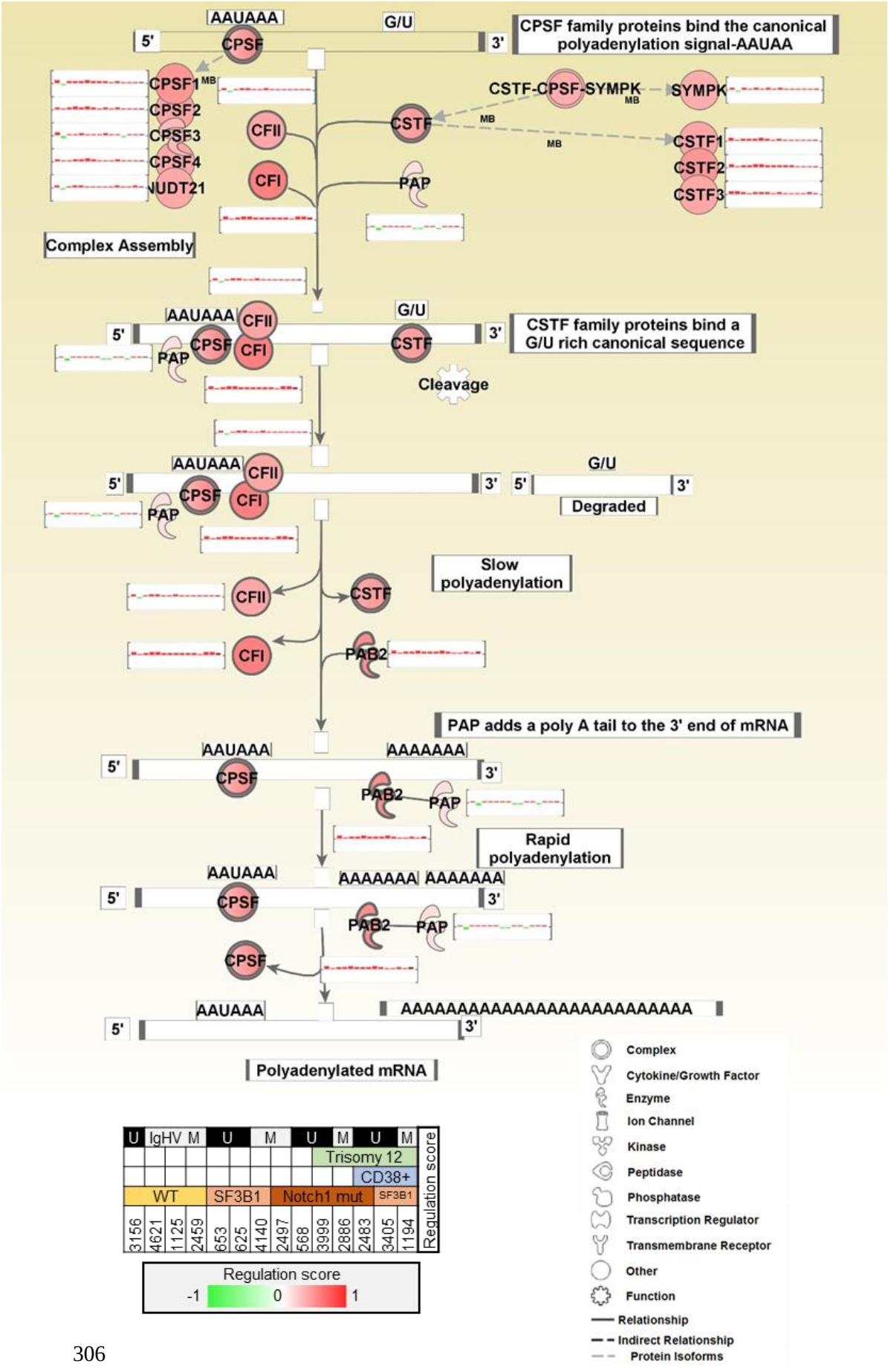
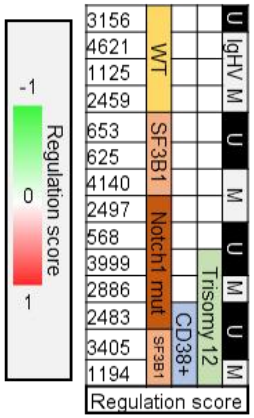
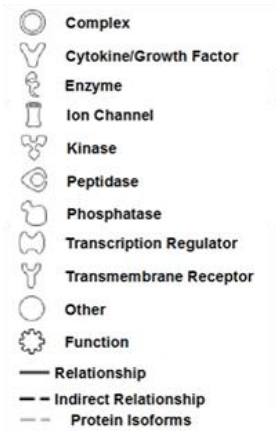

AAUAAA
G/U
5'
CPSF
3'
CPSF family proteins bind the canonical polyadenylation signal-AAUAA
CPSF1
CPSF2
CPSF3
CPSF4
UDT21
MB
CFII
CFI
CSTF
CSTF-CPSF-SYMPK
SYMPK
CSTF1
CSTF2
CSTF3
PAP
Complex Assembly
CSTF family proteins bind a G/U rich canonical sequence
Cleavage
Degraded
Slow polyadenylation
PAB2
PAP adds a poly A tail to the 3' end of mRNA
AAAAAAA
Rapid polyadenylation
AAUAAA
Polyadenylated mRNA
U
IgHV
M
Trisomy 12
CD38+
WT
SF3B1
Notch1 mut
Regulation score
3156
4621
1125
2459
653
625
4140
2497
568
3999
2886
2483
3405
1194
-1
0
1
Complex
Cytokine/Growth Factor
Enzyme
Ion Channel
Kinase
Peptidase
Phosphatase
Transcription Regulator
Transmembrane Receptor
Other
Function
Relationship
Indirect Relationship
Protein Isoforms

Figure 7.15. **RNA polyadenylation pathway in CLL.** Proteins mapping to the IPA canonical pathway 'cleavage and polyadenylation of pre-mRNA' to which 10 significantly upregulated proteins map to the 12 proteins of the pathway with an enrichment p-value of $1.02x10^{-12}$. Proteins are coloured on the basis of up or downregulation according to the regulation score. The bar charts adjacent to proteins indicate the expression across all 14 samples. Protein isoforms are indicated by broken light grey lines labelled as members (MB).


-1
Regulation score
0
1
3156
4621
1125
2459
653
625
4140
2497
568
3999
2886
2483
3405
1194
WT
SF3B1
Notch1 mut
SF3B1
CD38+
Trisomy 12
IgHV
U
M
Regulation score


Complex
Cytokine/Growth Factor
Enzyme
Ion Channel
Kinase
Peptidase
Phosphatase
Transcription Regulator
Transmembrane Receptor
Other
Function
Relationship
Indirect Relationship
Protein Isoforms

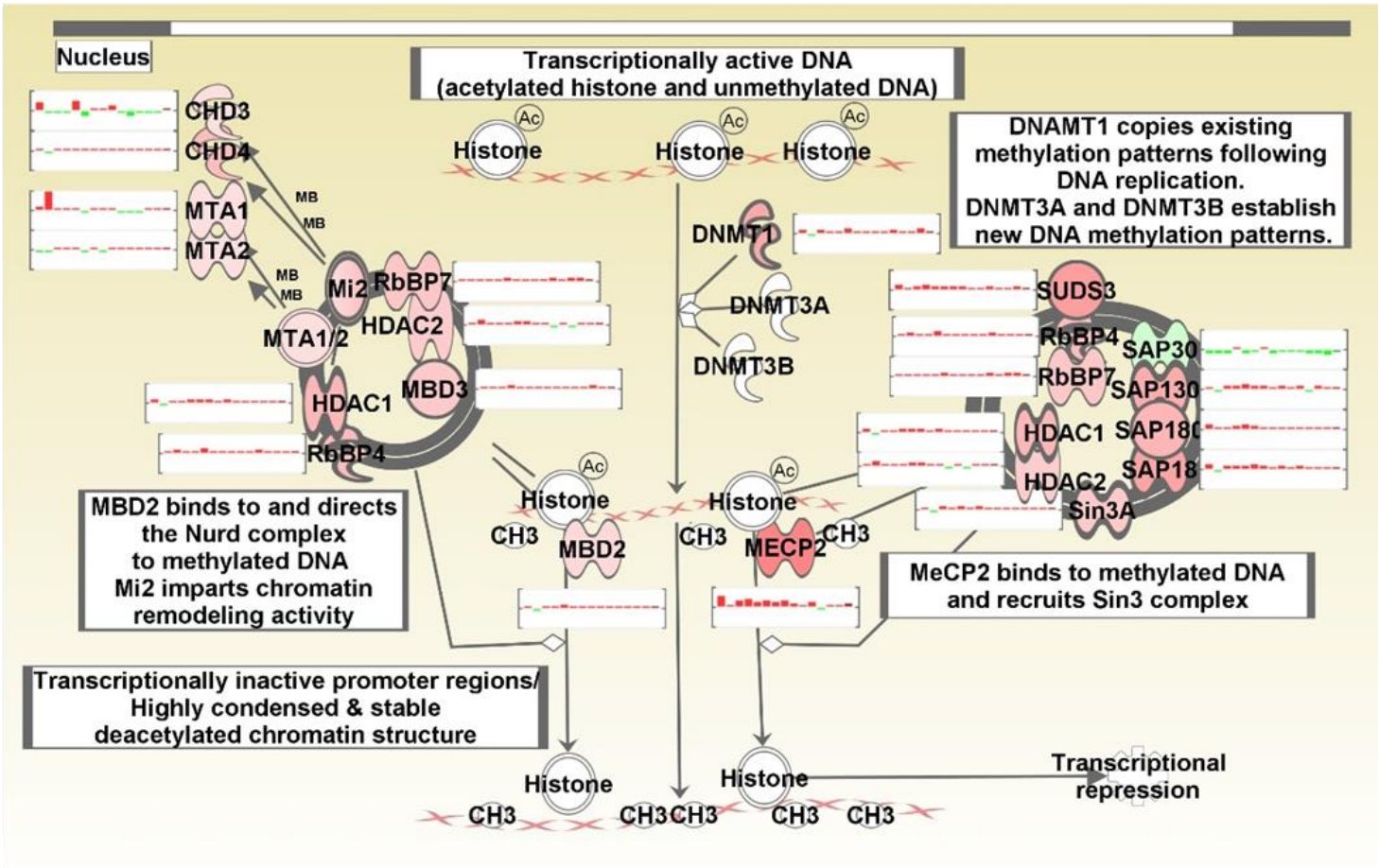

Nucleus
Transcriptionally active DNA
(acetylated histone and unmethylated DNA)
Histone
Ac
CHD3
CHD4
MTA1
MTA2
MB
Mi2
RbBP7
HDAC2
MTA1/2
HDAC1
MBD3
RbBP4
DNMT1
DNMT3A
DNMT3B
DNAMT1 copies existing methylation patterns following DNA replication.
DNMT3A and DNMT3B establish new DNA methylation patterns.
SUDS3
RbBP4
SAP30
RbBP7
SAP130
HDAC1
SAP180
HDAC2
SAP18
Sin3A
CH3
MBD2
MECP2
MBD2 binds to and directs the Nurd complex to methylated DNA
Mi2 imparts chromatin remodeling activity
MeCP2 binds to methylated DNA and recruits Sin3 complex
Transcriptionally inactive promoter regions/
Highly condensed & stable
deacetylated chromatin structure
Transcriptional repression

Figure 7.16. **Transcriptional repression in CLL.** Proteins mapping to the IPA canonical pathway 'DNA methylation and transcriptional repression' with an enrichment p-value of $1.2x10^{-3}$. Proteins are coloured on the basis of up or downregulation according to the regulation score. The bar charts adjacent to proteins indicate the expression across all 14 samples. Protein isoforms are indicated by broken light grey lines labelled as members (MB).

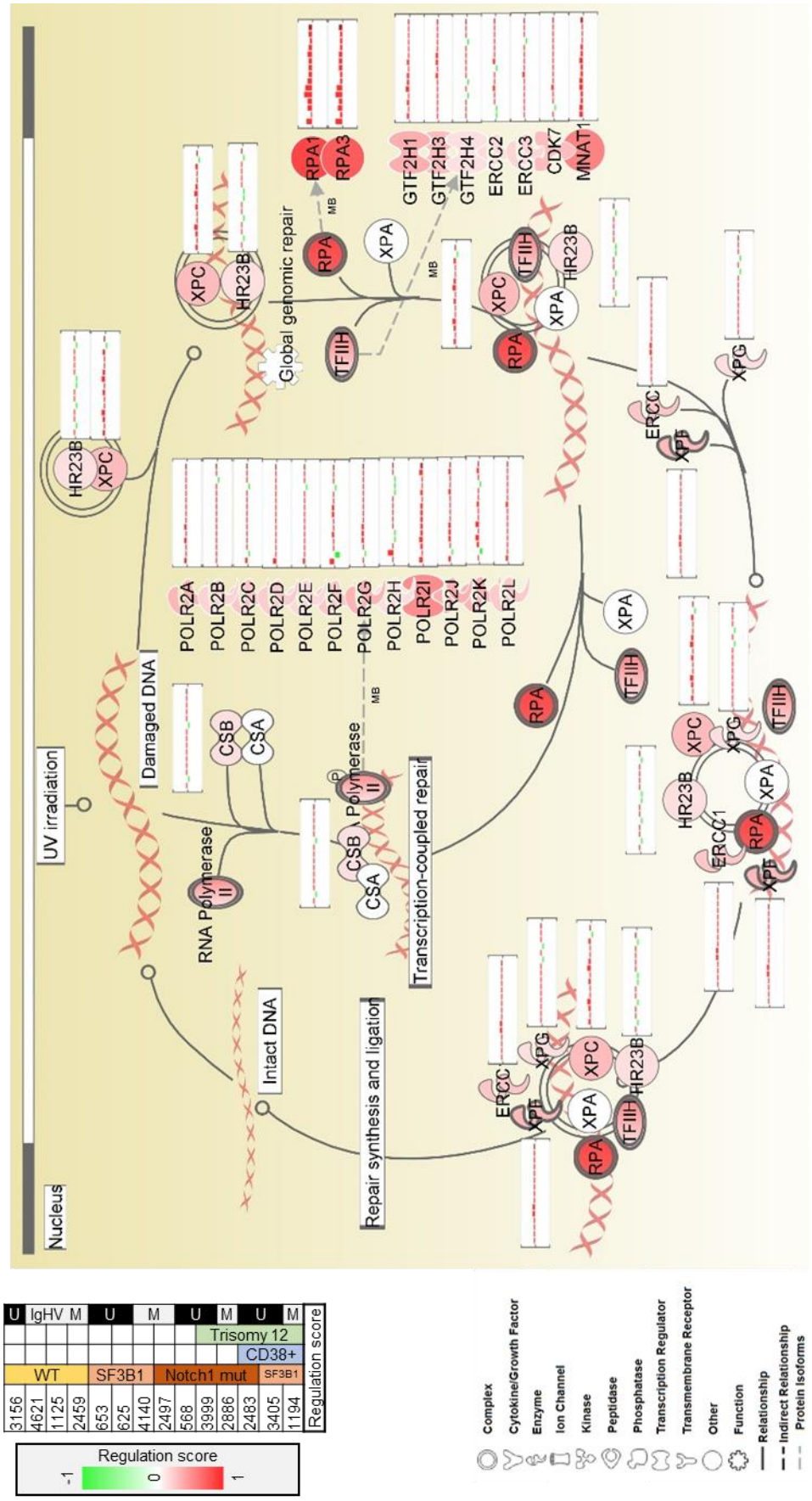

Nucleus
UV irradiation
Damaged DNA
Intact DNA
Repair synthesis and ligation
Transcription-coupled repair
Global genomic repair
RNA Polymerase II
Polymerase II
CSB
CSA
XPC
HR23B
RPA
XPA
TFIIH
XPG
ERCC
ERCC1
POLR2A
POLR2B
POLR2C
POLR2D
POLR2E
POLR2F
POLR2G
POLR2H
POLR2I
POLR2J
POLR2K
POLR2L
RPA1
RPA3
GTF2H1
GTF2H3
GTF2H4
ERCC2
ERCC3
CDK7
MNAT1
MB
P
IgHV
U
M
Trisomy 12
CD38+
WT
SF3B1
Notch1 mut
SF3B1
Regulation score
3156
4621
1125
2459
653
625
4140
2497
568
3999
2886
2483
3405
1194
Regulation score
-1
0
1
Complex
Cytokine/Growth Factor
Enzyme
Ion Channel
Kinase
Peptidase
Phosphatase
Transcription Regulator
Transmembrane Receptor
Other
Function
Relationship
Indirect Relationship
Protein Isoforms

Figure 7.17. **DNA damage repair in CLL.** Proteins mapping to the IPA canonical pathway 'nucleotide excision repair' with an enrichment p-value of $8.2x10^{-5}$. Proteins are coloured on the basis of up or downregulation according to the regulation score. The bar charts adjacent to proteins indicate the expression across all 14 samples. Protein isoforms are indicated by broken light grey lines labelled as members (MB).

Path Designer Assembly of RNA Polymerase II Complex

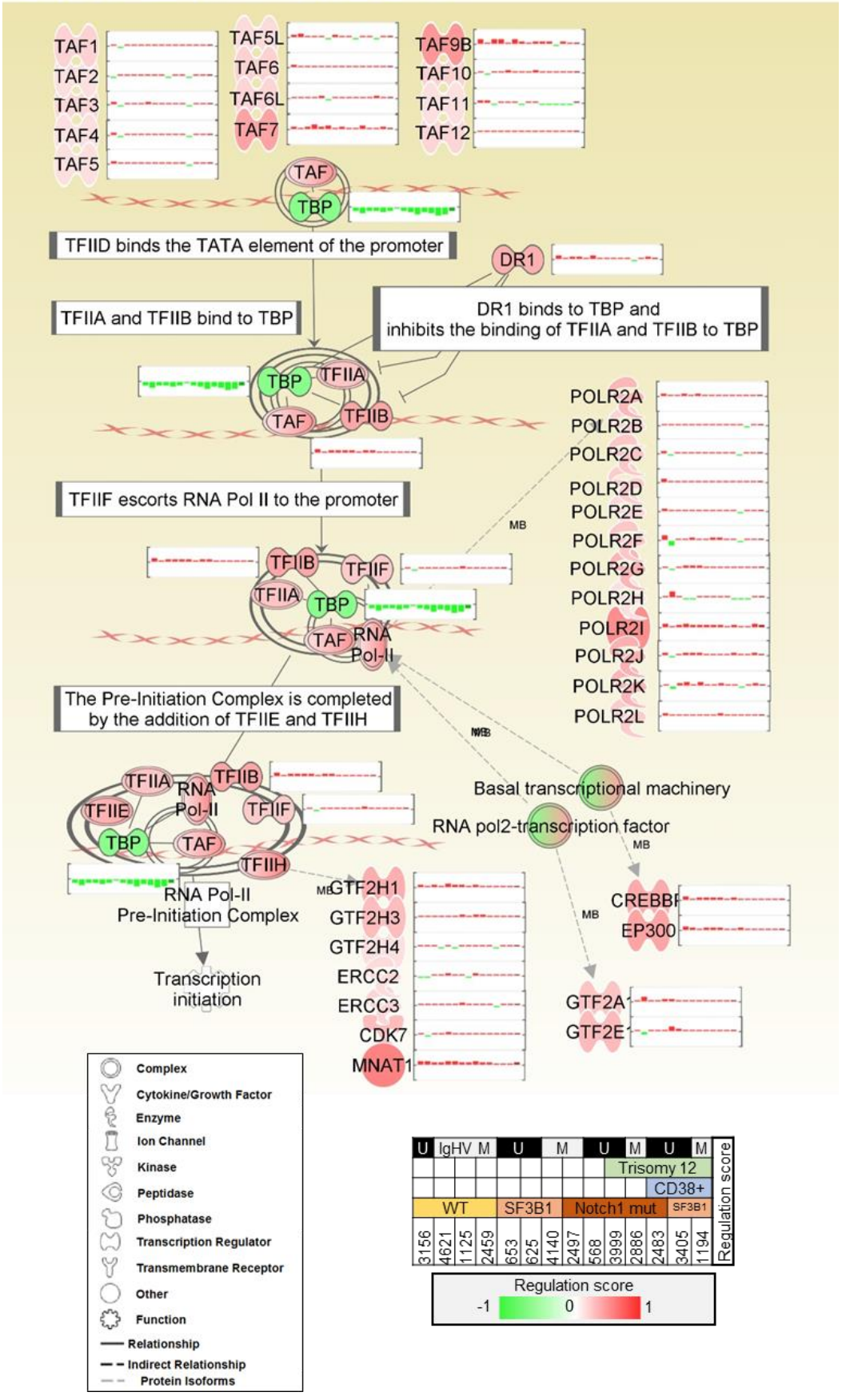


Figure 7.18. **RNA polymerase II assembly in CLL.** Proteins mapping to the IPA canonical pathway 'assembly of RNA polymerase II complex' with an enrichment p-value of $5.75x10^{-6}$. Proteins are coloured on the basis of up or downregulation according to the regulation score. The bar charts adjacent to proteins indicate the expression across all 14 samples. Protein isoforms are indicated by broken light grey lines labelled as members (MB).

Path Designer Integrin Signaling

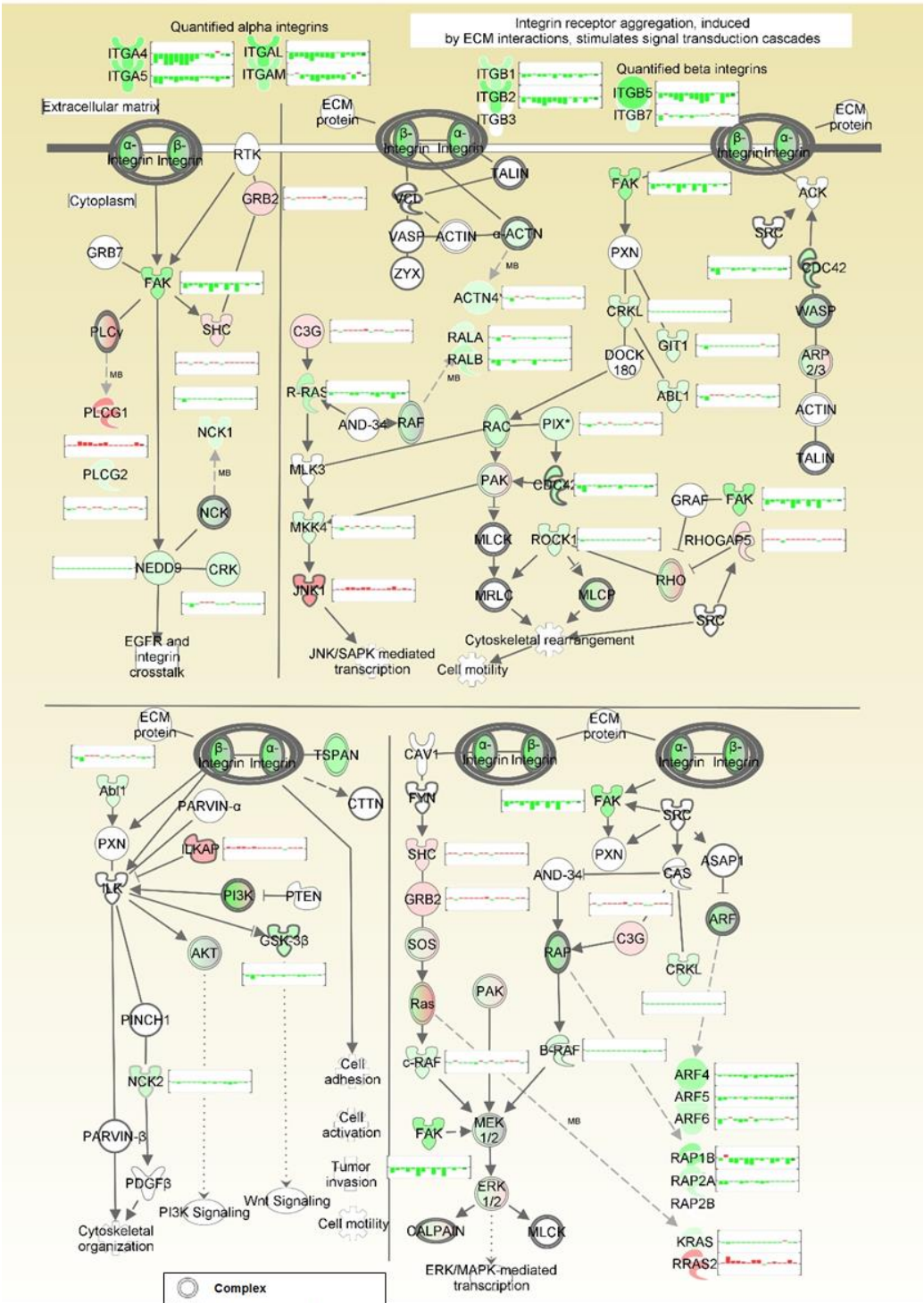


Figure 7.19. **Integrin signalling in CLL.** Proteins mapping to the IPA canonical pathway 'integrin signalling' with an enrichment p-value of $3.80x10^{-8}$. Proteins are coloured on the basis of up or downregulation according to the regulation score. The bar charts adjacent to proteins indicate the expression across all 14 samples. Protein isoforms are indicated by broken light grey lines labelled as members (MB).

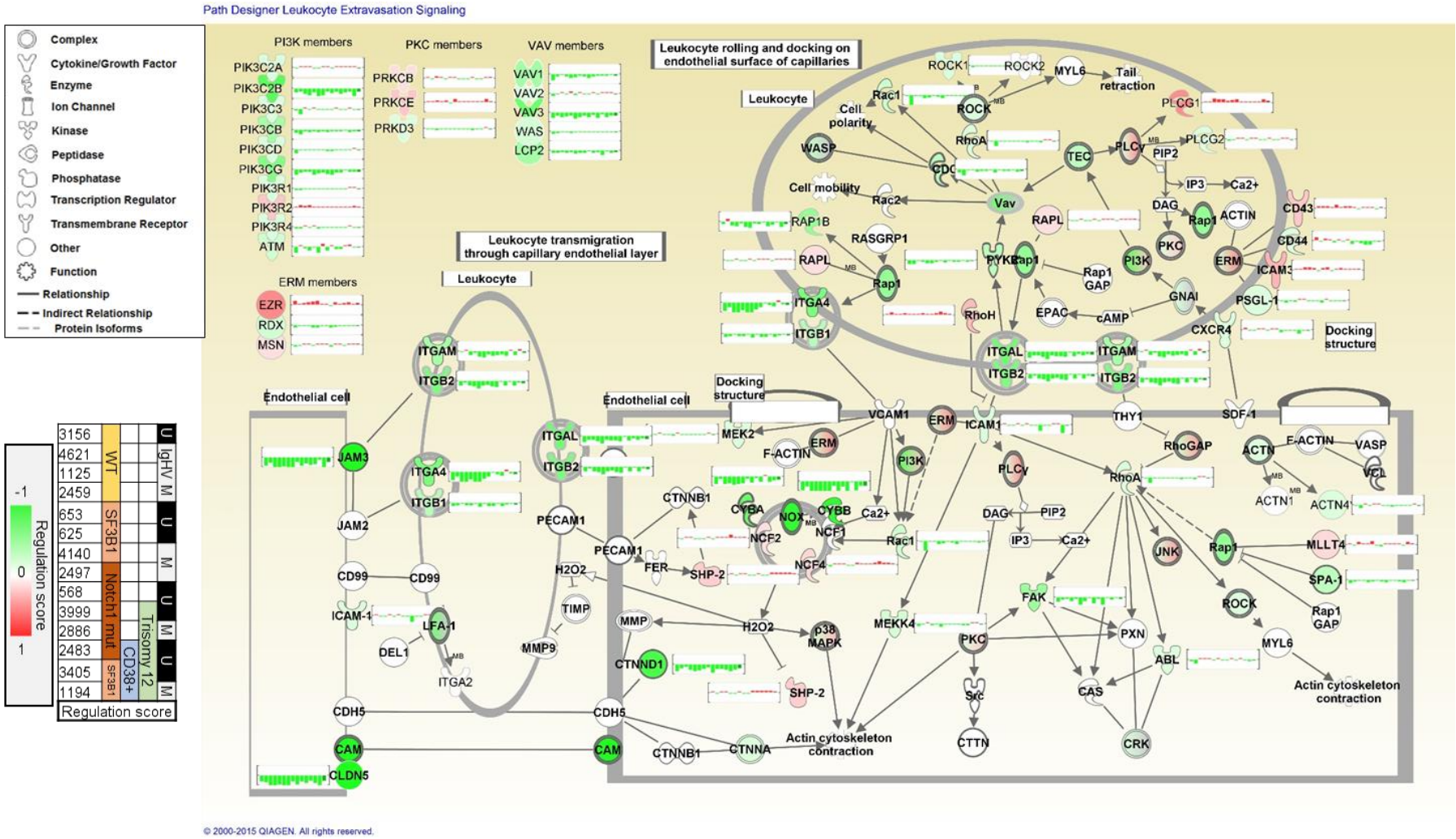

Path Designer Leukocyte Extravasation Signaling
Complex
Cytokine/Growth Factor
Enzyme
Ion Channel
Kinase
Peptidase
Phosphatase
Transcription Regulator
Transmembrane Receptor
Other
Function
Relationship
Indirect Relationship
Protein Isoforms
PI3K members
PIK3C2A
PIK3C2B
PIK3C3
PIK3CB
PIK3CD
PIK3CG
PIK3R1
PIK3R2
PIK3R4
ATM
PKC members
PRKCB
PRKCE
PRKD3
VAV members
VAV1
VAV2
VAV3
WAS
LCP2
ERM members
EZR
RDX
MSN
Leukocyte rolling and docking on endothelial surface of capillaries
Leukocyte transmigration through capillary endothelial layer
Leukocyte
Endothelial cell
Docking structure
Actin cytoskeleton contraction
Regulation score
-1
0
1
3156
4621
1125
2459
653
625
4140
2497
568
3999
2886
2483
3405
1194
WT
SF3B1
Notch1 mut
SF3B1
CD38+
Trisomy 12
IgHV M
U
M
U
M
U
M
U
M
Regulation score
© 2000-2015 QIAGEN. All rights reserved.

Figure 7.20. **Leukocyte extravasation signalling in CLL.** Proteins mapping to the IPA canonical pathway 'leukocyte extravasation signalling' with an enrichment p-value of $1.35x10^{-7}$. Proteins are coloured on the basis of up or downregulation according to the regulation score. The bar charts adjacent to proteins indicate the expression across all 14 samples. Protein isoforms are indicated by broken light grey lines labelled as members (MB).

7.14 DIFFERENTIAL EXPRESSION OF PROTEINS OF THE B-CELL RECEPTOR PATHWAY IN CLL

The up or downregulation of a protein in a pathway does not definitively indicate activation or inactivation of its function. As a general rule, especially in cancer, underexpression of a proteins can be used to infer a likelihood that the function is less critical to proliferation or survival. Overexpression can imply that the upregulation of a certain protein has provided the cancer with, for instance, a survival advantage. One of the most well studied pathways in CLL is the B-cell receptor (BCR) signalling pathway, and the simultaneous, non-biased quantitation of the components of this pathway had potential to reveal novel insight into the function of BCR signalling in CLL.

The BCR signalling pathway was the most significantly enriched pathway amongst the downregulated proteins in CLL ($p=1.1x10^{-10}$) with 28 significantly underexpressed proteins ($RS< -0.25$, $p<0.05$). In addition to these proteins, 10 significantly upregulated proteins ($RS>0.25$, $p<0.05$) also mapped to the pathway. To demonstrate the localisation of these protein regulation within the pathway and within the cell, expression values for all 14 CLL samples were overlaid onto the canonical BCR signalling pathway (**Figure 7.21**). Core components of the B-cell receptor were significantly downregulated; B-cell antigen receptor complex-associated protein alpha chain (CD79A) ($RS= -0.34$, $p=3.6x10^{-3}$) and the beta chain (CD79B) ($RS= -0.73$, $p=2.7x10^{-5}$); as well as all 9 quantitated Ig chains. CD22 was also very downregulated ($RS= -0.76$, $p=4.3x10^{-5}$). Other cell surface accessory molecules, such as CD19, CD45, low affinity Ig gamma Fc region receptor II-b (FCGR2B) were generally unregulated.

Amongst the downstream BCR signalling molecules some, inconsistent marginal regulation was observed for the majority of proteins, such as a 1.5-fold upregulation in four of the CLL samples for tyrosine-protein kinase SYK. However, proteins downstream of SYK were consistently significantly downregulated, such as Bruton tyrosine kinase (BTK) ($RS= -0.34$, $p=3.7x10^{-5}$), a protein indispensable for BCR signalling,

GRB2-associated-binding protein 2 (GAB2), a protein capable of initiating phosphatidylinositol kinase activity downstream of several transmembrane receptors, was the only significantly upregulated molecule immediately downstream of the BCR ($RS=0.28$,

$p=5.3x10^{-4}$). Phosphatidylinositol 3-kinase regulatory subunit beta (PIK3R2), the most overexpressed PIK subunit, was marginally, but significantly upregulated (RS=0.19, $p=1.4x10^{-3}$). A phosphatidylinositol phosphatase, SAC2/INPP5F, capable of reversing the effects of PIK was, however, confidently identified with 23 unique peptides and 123 PSMs, as the most significantly upregulated protein in the BCR pathway (RS=0.89, $p=1.2x10^{-5}$). Proteins downstream of phosphatidylinositol trisphosphate (PIP3) signalling were, again, generally downregulated, such as dual adapter for phosphotyrosine and 3-phosphotyrosine and 3-phosphoinositide (DAPP1) (RS= -0.41, $p=5.5x10^{-6}$) and RAC-alpha serine/threonine-protein kinase (AKT1) (RS= -0.13, $p=9.7x10^{-3}$). BTK was also annotated to be downstream of PIP3 signalling.

Four members of the mitogen-activated protein kinase (MAPK)/extracellular signal-regulated kinase (ERK) kinase kinase (MEKK) cascade were marginally upregulated; mitogen-activated protein kinase 8 (MAPK8) (RS=0.34, $p=3.0x10^{-4}$), MAPK kinase kinase 1 (MAP3K1) (RS=0.19, $p=6.7x10^{-4}$), dual specificity MAPK kinase 6 (MAP2K6) (RS=0.21, $p=2.8x10^{-3}$) and mitogen-activated protein kinase 13 (MAPK13) (RS=0.24, $p=1.6x10^{-2}$). RRAS2 was the only observed significantly upregulated upstream signalling molecule of the MEKK pathways (RS=0.33, $p=9.7x10^{-4}$). All three quantitated transcription factors downstream of the MEKK pathway were significantly upregulated histone acetyltransferases (HATs); CREB-binding protein (CREBBP) (RS=0.29, $p=4.0x10^{-5}$), histone acetyltransferase p300 (EP300) (RS=0.31, $p=1.2x10^{-5}$) and cyclic AMP-dependent transcription factor 2 (ATF2) (RS=0.38, $p=1.4x10^{-5}$).

Proteins of the NF-κB pathway were generally unregulated and the upstream regulators, mucosa-associated lymphoid tissue lymphoma translocation protein 1 (MALT1) (RS= -0.32, $p=7.9x10^{-5}$) and B-cell lymphoma/leukaemia 10 (BCL10) (RS= -0.23, $p=8.1x10^{-4}$) were downregulated. Bcl-2-like protein 1 (BCL-XL/BCL2L1), downstream of NF-κB was significantly downregulated, consistently in all cases of CLL (RS= -1.19, $p=1.4x10^{-8}$).

Other significantly overexpressed downstream transcription factors included two nuclear factor of activated T cells (NFAT) proteins; cytoplasmic isoform 1 (NFATC1) (RS=0.48, $p=1.3x10^{-6}$) and isoform 5 (NFAT5) (RS=0.24, $p=8.2x10^{-3}$); forkhead box protein O1 (FOXO1) (RS=0.25, $p=5.4x10^{-4}$) and transcription factor E2-alpha (TCF3) (RS=0.40, $p=1.2x10^{-4}$). The protein most specifically involved in the induction of B-cell commitment, alongside FOXO1 and TCF3, early B-cell factor 1 (EBF1), was significantly downregulated (RS= -0.92, $p=4.6x10^{-6}$).

Other functions of the BCR signalling pathway, such as signalling proteins involved in cytoskeletal rearrangement were generally downregulated, such as focal adhesion kinases 1 and

2 (FAK (RS= -0.52, $p=1.2x10^{-3}$) and FAK2/PTK2B (RS= -0.36, $p=1.5x10^{-4}$)) and Ras-related proteins RAP1B (RS= -0.56, $p=1.5x10^{-4}$) and RAP2A (RS= -0.28, $p=7.7x10^{-4}$).

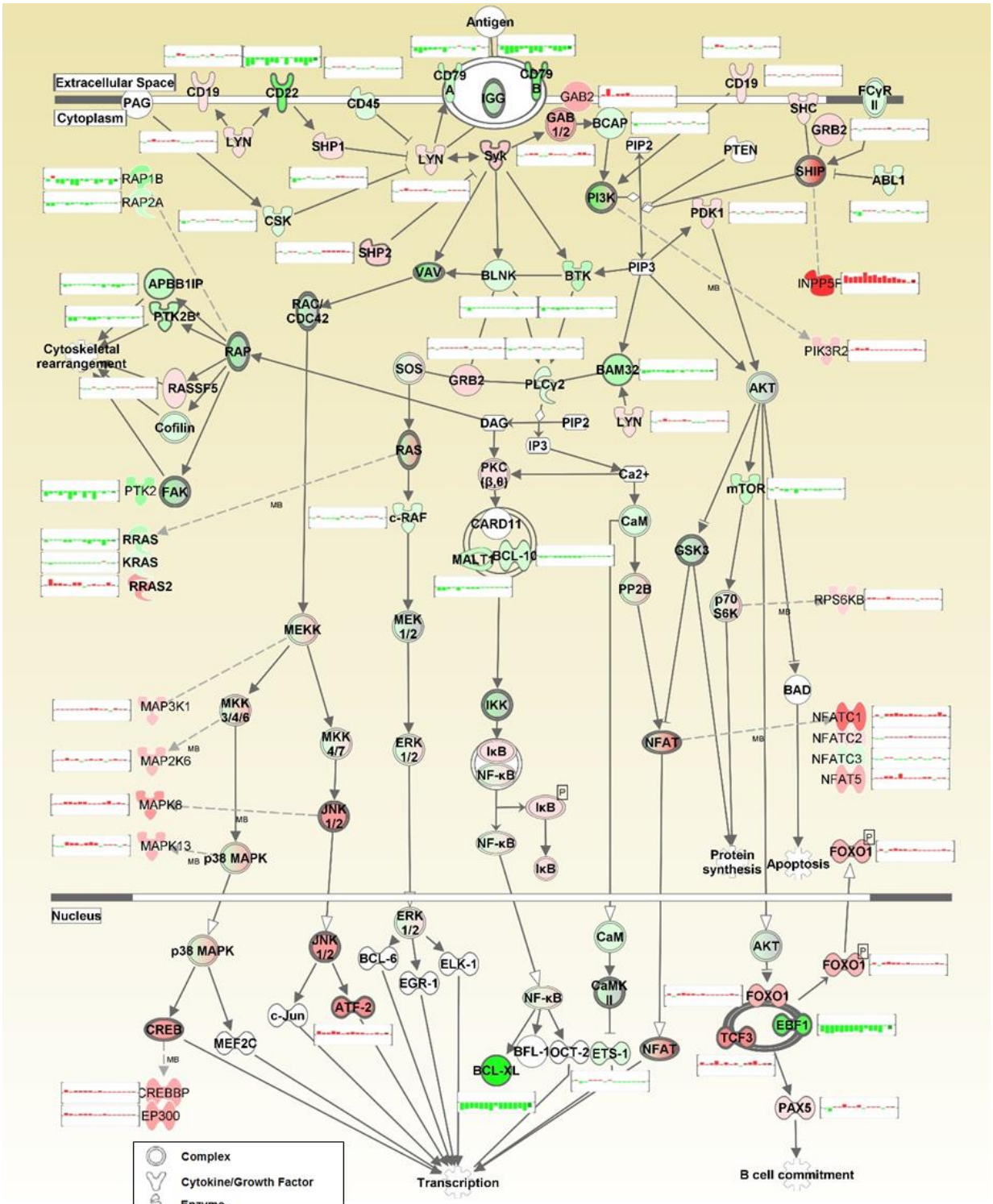

Antigen
Extracellular Space
Cytoplasm
PAG
CD19
CD22
CD45
CD79 A
IGG
CD79 B
GAB2
CD19
SHC
FCγR II
LYN
SHP1
LYN
Syk
GAB 1/2
BCAP
PIP2
PTEN
GRB2
SHIP
ABL1
RAP1B
RAP2A
CSK
PI3K
PDK1
SHP2
VAV
BLNK
BTK
PIP3
INPP5F
APBB1IP
PTK2B
RAC/ CDC42
PIK3R2
Cytoskeletal rearrangement
RAP
SOS
GRB2
PLCγ2
BAM32
AKT
RASSF5
Cofilin
DAG
PIP2
LYN
IP3
RAS
PKC (β,θ)
Ca2+
mTOR
PTK2
FAK
c-RAF
CaM
RRAS
KRAS
RRAS2
CARD11
MALT1
BCL-10
GSK3
PP2B
p70 S6K
RPS6KB
MEKK
MEK 1/2
BAD
MAP3K1
MKK 3/4/6
IKK
NFATC1
NFATC2
NFATC3
NFAT5
MAP2K6
MKK 4/7
ERK 1/2
IκB
NF-κB
NFAT
MAPK8
JNK 1/2
IκB
MAPK13
p38 MAPK
NF-κB
IκB
Protein synthesis
Apoptosis
FOXO1
Nucleus
p38 MAPK
JNK 1/2
ERK 1/2
CaM
AKT
FOXO1
BCL-6
ELK-1
EGR-1
NF-κB
CaMK II
c-Jun
ATF-2
FOXO1
EBF1
CREB
MEF2C
TCF3
BFL-1
OCT-2
ETS-1
NFAT
BCL-XL
CREBBP
EP300
PAX5
Transcription
B cell commitment
Complex
Cytokine/Growth Factor
Enzyme
Ion Channel
Kinase
Peptidase
Phosphatase
Transcription Regulator
Transmembrane Receptor
Other
Function
Relationship
Indirect Relationship
Protein Isoforms
U
IgHV
M
U
M
U
M
U
M
Trisomy 12
CD38+
WT
SF3B1
Notch1 mut
SF3B1
3156
4621
1125
2459
653
625
4140
2497
568
3999
2886
2483
3405
1194
Regulation score
Regulation score
-1
0
1

Figure 7.21. **Annotation of the BCR signalling pathway with CLL protein expression.** The significantly enriched IPA canonical pathway 'B-cell receptor signalling pathway' ($p=1.1x10^{-10}$) overlaid with proteomics-determined protein expression relative to healthy donor B cells. Proteins are coloured on the basis of up or downregulation according to the regulation score. The bar charts adjacent to proteins indicate the expression across all 14 samples. Protein isoforms are indicated by broken light grey lines labelled as members (MB).

7.15 SUBTYPE-SPECIFIC DIFFERENTIAL EXPRESSION IN CLL

Across the 14 quantitated CLL samples, while protein expression generally followed the same direction of up or downregulation (**Figure 7.4**) relative to healthy donor B cells, the extent of this regulation varied greatly across the samples. On the whole, the regulation score gave an impression of the overall regulation in CLL, however overlooked that certain CLL subtype samples may be exhibiting a vastly different pattern of expression. Additionally, given that no definitive clustering was identified amongst these samples it was of interest to better understand the specific proteins with the greatest extent of heterogeneity.

To achieve a subtype analysis, the individual $\log_2$ (ratios) of each CLL sample to each of the three HD controls were first averaged, rather than the minimum taken, as described in **Figure 7.2**. Such values could be compared to one another, rather than specifically to the HD controls (**Figure 7.22**). Proteins were again filtered to remove platelet proteins. Additionally, consideration was given to the bias that may have been introduced due to a greater extent of ratio compression for a protein's quantitation derived from either 10-plex. A T-test was therefore used to filter out 237 proteins for which expression differed significantly between 10-plex A and 10-plex B; leaving 5719 proteins. Again, sample 4621 was excluded from all analyses.

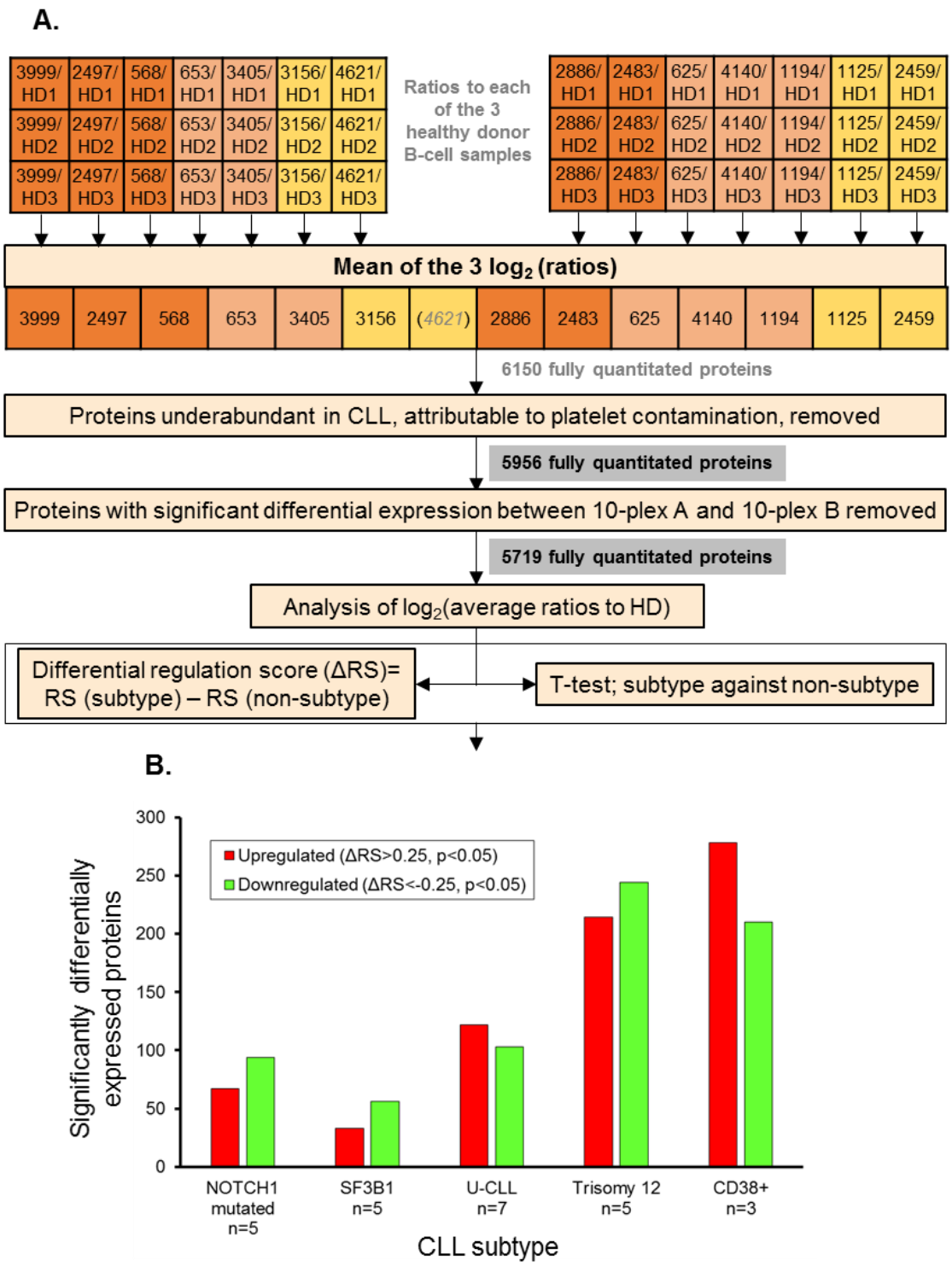
The number of samples had insufficient statistical power to confidently conclude discrete biomarkers or characteristics of each subtype. However, it was still possible to generate lists of proteins with trends indicative of subtype-specific protein expression, using the T-test without multiple correction to give an approximate, rather than a precise measure of confidence. Such an analysis was used by [352].

To evaluate the magnitude of subtype-specific expression patterns in CLL, differential regulation scores (ΔRS) were used, defined as the difference between the two RS calculated for those CLL samples fitting into each subtype. The RS of the named subtype had the RS of those samples which were not of that subtype, subtracted; e.g. for a protein overexpressed in the trisomy 12 cases compared to the non-trisomy 12 cases, the ΔRS would be positive.

This analysis was performed to identify subtype-specific expression for trisomy 12, $CD38^+$ cases, *NOTCH1*-mutated cases, *SF3B1*-mutated cases and unmutated CLL (U-CLL) cases (**Figure 7.22B**). In each case, the number of proteins with significant, subtype-specific upregulation ($\Delta RS>0.25$, $p<0.05$) and downregulation ($\Delta RS< -0.25$, $p<0.05$) was determined. For the 7 U-CLL cases, 122 proteins were significantly overexpressed relative to M-CLL, while 103 had significantly lower expression. The 5 *NOTCH1*-mutated cases had 67 over and 94 underexpressed proteins and the 5 *SF3B1*-mutated CLL had 33 over and 56 underexpressed proteins. The strongest signatures were observed for trisomy 12 cases and $CD38^+$ cases, with 458 and 478 proteins differentially regulated, respectively. However, it should be noted that all three $CD38^+$ cases also had trisomy 12.

When considering subtype-specific expression, it was also of interest to note how each protein was differentially regulated compared with the HD B-cell controls. Proteins were therefore annotated to highlight whether the regulation score was greater than or less than 0.25, relative to the HD samples, specifically for those samples with the named subtype.

Figure 7.22. **Determination of differential expression between CLL subtypes. A.** The workflow describing the determination of significant differential expression between CLL subtypes. Using the SPIQuE-determined $\log_2$ (ratios) relative to the three HD controls, mean $\log_2$ (ratios) for each CLL sample were calculated. Proteins were filtered for those proteins described in **Figure 7.2** which represented platelet contamination in the HD samples. Proteins were also filtered to remove instances where the observed differential expression between 10-plex A and 10-plex B was statistically significant. Differential expression was then determined between CLL subtypes using a differential regulation score and an uncorrected T-test. **B.** The number of proteins reaching significant subtype-specific expression ($\Delta RS>0.25$, $p<0.05$ or $\Delta RS< -0.25$, $p<0.05$) for each subtype comparison.


A.
3999/ HD1
2497/ HD1
568/ HD1
653/ HD1
3405/ HD1
3156/ HD1
4621/ HD1
3999/ HD2
2497/ HD2
568/ HD2
653/ HD2
3405/ HD2
3156/ HD2
4621/ HD2
3999/ HD3
2497/ HD3
568/ HD3
653/ HD3
3405/ HD3
3156/ HD3
4621/ HD3
Ratios to each of the 3 healthy donor B-cell samples
2886/ HD1
2483/ HD1
625/ HD1
4140/ HD1
1194/ HD1
1125/ HD1
2459/ HD1
2886/ HD2
2483/ HD2
625/ HD2
4140/ HD2
1194/ HD2
1125/ HD2
2459/ HD2
2886/ HD3
2483/ HD3
625/ HD3
4140/ HD3
1194/ HD3
1125/ HD3
2459/ HD3
Mean of the 3 log2 (ratios)
3999
2497
568
653
3405
3156
(4621)
2886
2483
625
4140
1194
1125
2459
6150 fully quantitated proteins
Proteins underabundant in CLL, attributable to platelet contamination, removed
5956 fully quantitated proteins
Proteins with significant differential expression between 10-plex A and 10-plex B removed
5719 fully quantitated proteins
Analysis of log2(average ratios to HD)
Differential regulation score (ΔRS)= RS (subtype) – RS (non-subtype)
T-test; subtype against non-subtype
B.
Upregulated (ΔRS>0.25, p<0.05)
Downregulated (ΔRS<-0.25, p<0.05)
Significantly differentially expressed proteins
300
250
200
150
100
50
0
NOTCH1 mutated n=5
SF3B1 n=5
U-CLL n=7
Trisomy 12 n=5
CD38+ n=3
CLL subtype

7.16 U-CLL-SPECIFIC PROTEIN EXPRESSION

Ribonuclease P protein subunit p25 (RPP25), an RNAse with a role in tRNA maturation, was the most upregulated protein in U-CLL, relative to M-CLL and to HD controls (**Figure 7.23**). Another protein, serine/threonine-protein kinase DCLK2, was upregulated, compared with HD samples, in all but two M-CLL samples, and demonstrated substantially more upregulation in U-CLL than M-CLL. The most non-overlapping signature of U-CLL was the mitochondrial carnitine/acylcarnitine carrier (SLC25A20) with an expression, relative to HD samples, of between 45% and 65% in M-CLL and 70% and 100% in U-CLL. In addition to SLC25A20, 7 other ion channels had greater expression in U-CLL than M-CLL, with solute carrier family 12 member 2 (SLC12A2), calcium-binding mitochondrial carrier protein Aralar2 (SLC25A13), adenosine 3'-phospho 5'-phosphosulfate transporter 1 (SLC35B2), zinc transporter ZIP11 (SLC39A11) and voltage-dependent anion-selective channel proteins (VDACs); VDAC1, VDAC2 and VDAC3. Of these; all three VDACs, SLC25A13 and SLC25A20 are mitochondrial transporters. Both IgM and IgE were detected with higher expression in U-CLL samples compared to M-CLL samples.

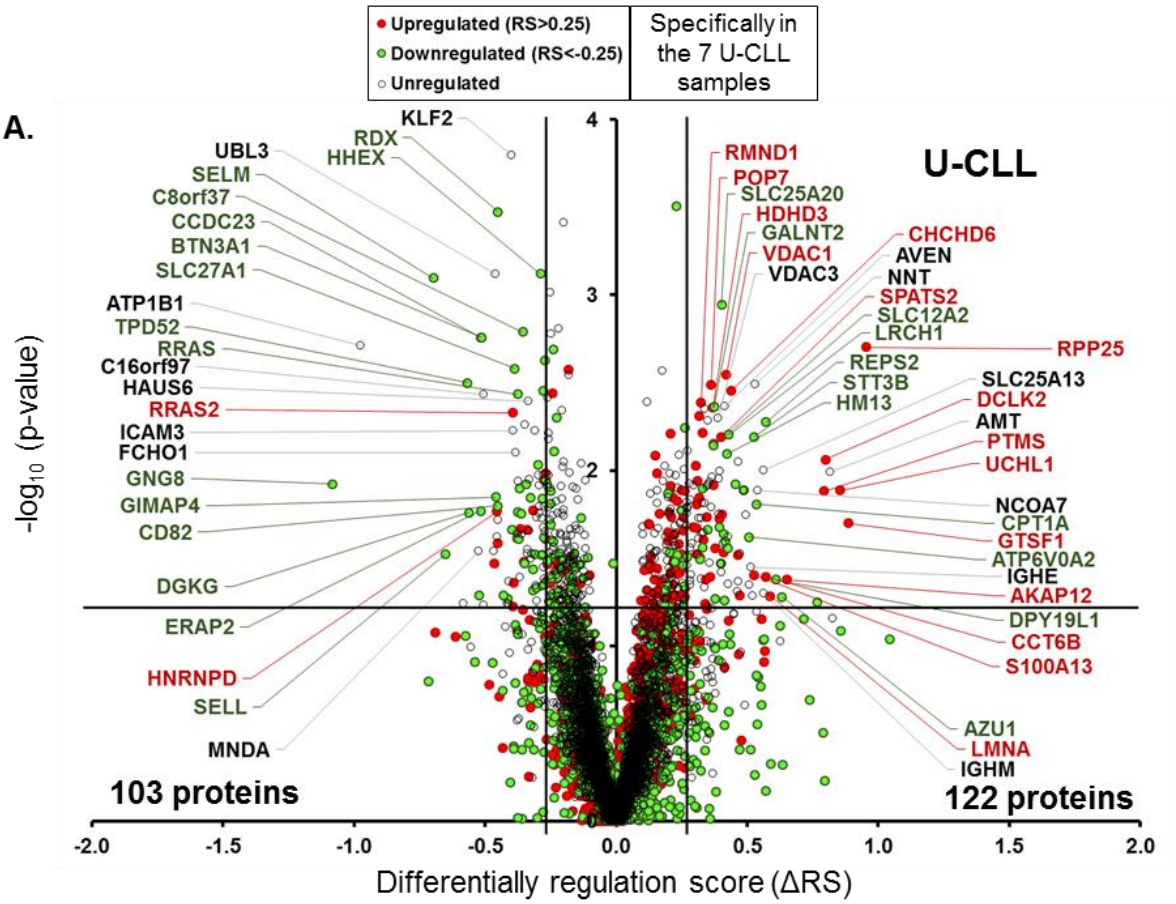
Altered glycosylation was represented by 6 U-CLL-dominant proteins including; oligosaccharyl transferase subunit STT3B, probable C-mannosyltransferase DPY19L1, beta-1,3-galactosyltransferase 6 (B3GALT6), polypeptide N-acetylgalactosaminyltransferase 2 (GALNT2), dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit (DDOST), magnesium transporter protein 1 (MAGT1). Two carbohydrate kinases, potentially extending these post-translational modification were upregulated in U-CLL; 6-phosphofructokinase type C (PFKP) and sphingosine kinase 2 (SPHK2). Acylation also showed a strong trend amongst these proteins, with carnitine O-palmitoyltransferases 1, liver isoform (CPT1A) and 2, mitochondrial (CPT2) both overexpressed in U-CLL, alongside apolipoprotein E (APOE), sialic acid synthase (NANS) and elongation of very long chain fatty acids protein 1 (ELOVL1). MICOS (mitochondrial contact site and cristae organizing system complex) subunits MIC19, MIC25 and MIC60 also demonstrated a trend of similar proteins with increased.

GO term enrichment revealed that 34 out of 122 proteins (28%) significantly overexpressed in U-CLL were annotated as mitochondrial, with an enrichment p-value of $4.5x10^{-8}$.

Amongst the proteins with significantly less expression in U-CLL, were, most notably; the previously described upregulated cell surface (**Figure 7.8**) Na^+/K^+ transporting ATPase subunit beta-1 (ATP1B1), selenoprotein M (SELM) - a putative thiol-disulfide oxidoreductase, the transcription factor Krueppel-like factor 2 (KLF2) and an adaptor of actin to the cell

membrane radixin (RDX). Other CLL-upregulated surface proteins with lower expression on U-CLL were ICAM3 and CD6. Additionally ICAM2, CD40, CD74 and CD82, downregulated across CLL relative to HD controls, were more downregulated in U-CLL cases. A trend was observed relating to the downregulation of Ras protein signalling in U-CLL, with Ras-related protein (RRAS) more downregulated and RRAS2 less upregulated in U-CLL samples. Guanine nucleotide-binding protein subunit alpha-13 (GNA13), son of sevenless homolog 1 (SOS1) and 14-3-3 protein beta/alpha (YWHAB) were also related to Ras signalling, while underexpressed in U-CLL, relative to M-CLL. Three proteins annotated with a role in B-cell activation were amongst the U-CLL-underexpressed proteins; CD40, B-cell scaffold protein with ankyrin repeats 1 (BANK1) and tumour protein D52. The theta, and more substantially gamma, subunits of the diacylglycerol kinase enzyme were marginally upregulated in M-CLL, while unregulated or downregulated in U-CLL. Other U-CLL-downregulated kinases included protein kinase C epsilon (PRKCE), bifunctional polynucleotide phosphatase/kinase (PNKP), casein kinase I isoform gamma-1 (CSNK1G1) beta-adrenergic receptor kinase 2 (ADRBK2) and G protein-coupled receptor kinase 5 (GRK5). Hyaluronan binding was also more prevalent in M-CLL, with hyaluronan-binding protein 2 (HABP2) and intracellular hyaluronan-binding protein 4 (HABP4) both less upregulated in U-CLL. Three proteins related to antigen presentation; HLA class II histocompatibility antigen gamma chain (CD74), endoplasmic reticulum aminopeptidase 2 (ERAP2) and TAP-binding protein tapasin (TAPBP) tended to have a higher expression in M-CLL. Cell projection appeared downregulated in U-CLL with 8 proteins annotated with such a role; ena/VASP-like protein (EVL), catenin delta 1 (CTNND1), centrosomal protein 290kDa (CEP290), girdin (GRDN), ICAM2, GNA13, metastasis suppressor protein 1 (MTSS1) and RDX.

A.



B.

Protein	Symbol	Unique Peptides	PSMs	M-CLL							U-CLL							ΔRS	p-value
				4621	2459	1125	4140	2497	2886	1194	3405	2483	3999	568	625	653	3156		
Ribonuclease P protein subunit p25	RPP25	4	15	-0.54	-0.75	-1.35	-1.36	-0.48	0.59	-1.11	0.84	1.25	0.45	-0.33	1.33	1.31	1.04	0.95	0.0020
Serine/threonine-protein kinase DCLK2	DCLK2	17	70	1.58	0.57	1.04	1.55	0.14	1.60	0.06	1.53	1.34	1.78	2.17	2.18	2.86	1.77	0.80	0.0087
RalBP1-associated Eps domain-containing 2	REPS2	5	25	-1.25	-3.02	-2.14	-1.81	-1.46	-1.31	-2.10	-0.26	-0.58	-0.98	-0.92	-1.00	-1.08	-1.37	0.57	0.0053
Calcium-binding mitochondrial carrier Aralar2	SLC25A13	5	63	-0.53	-1.62	-1.25	-1.40	-0.52	-1.15	-0.79	-0.21	0.02	-0.24	0.16	-1.35	-0.46	-0.40	0.56	0.0100
NAD(P) transhydrogenase, mitochondrial	NNT	38	402	-0.11	-0.64	-0.57	-0.92	-0.56	-0.43	-0.86	-0.55	0.20	-0.27	0.49	-0.29	-0.08	0.27	0.53	0.0032
Oligosaccharyl transferase subunit STT3B	STT3B	6	47	1.35	-1.25	-1.18	-0.98	-1.14	-1.40	-0.95	-0.51	-0.38	-1.02	-0.57	-1.04	-0.73	-0.05	0.53	0.0064
Glutaredoxin-like protein C5orf63	C5orf63	1	5	-0.52	-0.86	-0.85	-0.87	-0.09	-0.20	-0.88	-0.71	0.16	0.01	0.32	0.07	0.04	0.21	0.47	0.0090
MICOS complex subunit MIC25	CHCHD6	4	25	0.23	-0.23	0.00	-0.48	0.13	0.21	-0.02	0.19	0.10	0.43	0.71	0.60	0.52	0.97	0.44	0.0035
Leu-rich repeat and calponin domain-containing 1	LRCH1	10	43	-1.07	-1.71	-1.49	-1.28	-1.21	-1.09	-0.88	-0.36	-0.49	-0.71	-0.53	-0.87	-0.87	-1.23	0.43	0.0063
Minor histocompatibility antigen H13	HM13	6	34	0.26	-0.85	-1.22	-1.30	-0.64	-1.34	-1.01	-0.76	-0.62	-0.76	-0.27	-0.59	-0.79	0.04	0.42	0.0081
Voltage-dependent anion-selective channel 1	VDAC1	11	290	0.73	-0.43	0.07	-0.01	-0.13	0.33	-0.09	0.25	0.40	0.24	0.64	0.34	0.69	0.68	0.42	0.0028
Cell death regulator Aven	AVEN	1	4	-0.34	-0.58	-0.30	0.09	-0.37	-0.03	-0.14	0.71	0.55	0.14	0.52	-0.05	0.09	0.18	0.42	0.0043
Mitochondrial carnitine/acylcarnitine carrier	SLC25A20	11	89	0.03	-1.15	-0.83	-0.93	-0.68	-0.61	-0.66	-0.23	-0.02	-0.53	-0.41	-0.51	-0.16	-0.40	0.40	0.0011
Spermatogenesis-associated serine-rich 2	SPATS2	2	3	0.34	0.11	0.18	0.07	0.81	0.33	0.30	0.70	0.89	0.75	0.58	0.45	1.02	1.03	0.40	0.0065
Voltage-dependent anion-selective channel 3	VDAC3	7	161	0.17	-1.17	-0.83	-0.80	-0.60	-0.36	-0.76	-0.35	-0.46	-0.31	-0.21	-0.48	0.18	-0.12	0.39	0.0050
Protein C8orf37	C8orf37	2	5	0.54	-0.17	-0.07	0.00	-0.33	-0.31	-0.47	-0.69	-0.61	-0.38	-0.62	-0.64	-0.75	-0.59	-0.36	0.0016
Ras-related protein R-Ras	RRAS	6	38	-0.61	-0.49	-0.50	-0.36	-0.42	-0.85	-0.70	-1.82	-0.82	-1.35	-1.00	-0.86	-0.99	-1.11	-0.38	0.0037
FCH domain only protein 1	FCHO1	16	121	-0.12	0.47	0.14	-0.01	0.28	0.02	0.69	-0.23	-0.63	-0.27	-0.33	-0.02	0.19	-0.26	-0.38	0.0079
Long-chain fatty acid transport protein 1	SLC27A1	3	9	1.12	0.02	0.22	0.07	0.22	-0.20	-0.03	-0.47	-0.14	-0.53	-0.46	-0.92	-0.46	-0.09	-0.39	0.0026
Intercellular adhesion molecule 3	ICAM3	11	102	0.30	0.79	0.77	0.59	0.74	0.54	0.47	0.29	0.29	0.32	-0.03	-0.21	0.64	0.32	-0.39	0.0059
Ras-related protein R-Ras2	RRAS2	6	79	1.66	0.75	0.75	0.97	1.09	0.49	0.84	0.48	0.45	0.58	-0.04	0.06	0.66	0.26	-0.40	0.0047
Krueppel-like factor 2	KLF2	3	18	0.32	0.78	0.68	0.43	0.67	0.73	0.60	0.22	-0.11	0.14	0.29	0.31	0.37	0.23	-0.40	0.0002
Radixin	RDX	16	952	-0.43	-0.12	-0.05	-0.09	-0.13	-0.44	0.00	-0.29	-0.56	-0.61	-0.89	-0.80	-0.87	-0.93	-0.45	0.0003
Ubiquitin-like protein 3	UBL3	1	6	0.00	0.57	0.57	0.58	0.17	0.40	0.42	-0.28	-0.21	0.06	-0.40	0.35	0.00	-0.15	-0.46	0.0008
Uncharacterized protein C16orf97	C16orf97	1	30	0.80	0.80	0.57	0.81	0.40	0.32	0.83	-0.55	0.04	-0.18	-0.13	0.52	0.40	-0.10	-0.51	0.0037
Butyrophilin subfamily 3 member A1	BTN3A1	3	32	0.91	0.33	0.03	0.12	0.15	-0.41	-0.33	-0.74	-0.81	-0.63	-0.77	-0.37	-0.87	-0.33	-0.51	0.0017
Coiled-coil domain-containing protein 23	CCDC23	3	8	-1.03	0.38	0.49	-0.06	0.19	0.17	0.38	-0.17	-0.61	0.19	-0.40	-0.93	-0.54	-0.38	-0.52	0.0017
Tumor protein D52	TPD52	9	137	-1.06	0.40	0.04	-0.39	0.29	-0.36	-0.38	-0.92	-0.60	-0.41	-1.37	-0.41	-1.03	-1.35	-0.57	0.0032
Selenoprotein M	SELM	2	7	-0.42	0.94	0.98	0.41	0.00	0.29	0.44	-0.62	-1.02	-0.06	-0.33	-0.45	-0.53	-0.06	-0.70	0.0008
Na+/K+ transporting ATPase subunit beta-1	ATP1B1	2	5	2.20	0.78	1.50	1.21	2.12	2.05	2.57	0.04	-0.01	0.91	0.65	0.76	-1.32	-0.38	-0.98	0.0019

Figure 7.23. **Subtype specific protein expression comparing U-CLL with M-CLL samples. A.** Volcano plot describing the differential regulation score (ΔRS) plotted against the $-\log_{10}$ (p-value) of differential regulation between U-CLL and M-CLL samples. The plot highlights those proteins for which the regulation score calculated for the 7 U-CLL samples was >0.25 (red) and < -0.25 (green). **B.** The $\log_2$ (ratios) for the top 15 significantly overexpressed and top 15 underexpressed proteins in U-CLL, detailing the sample expression, the ΔRS and the uncorrected p-value estimating the probability of subtype specific expression between the 7 U-CLL samples and 6 M-CLL samples (4621 was excluded from the analysis as an outlier).

7.17 *NOTCH1*-MUTANT CLL-SPECIFIC PROTEIN EXPRESSION

NOTCH1-mutated cases demonstrated a strong trend of increasing the expression of proteins otherwise significantly downregulated in CLL with 28 of the 67 proteins having a regulation of less than -0.25 across all CLL samples (**Figure 7.24**). This was emphasised by the protein with the greatest degree of specificity to *NOTCH1*-mutated cases; the actin-binding stabiliser tropomyosin beta chain (TPM2). In cases without *NOTCH1* mutation, TPM2 expression varied from 1.5% to 5.8%, while *NOTCH1*-mutated cases expressed between 6.5% and 27.2% that of HD controls. Another cytoskeletal protein, microtubule-associating protein MAP1A was also less downregulated in cases with mutant *NOTCH1* (22.8-30.3%) compared to other cases (16.7-20.4%). Similarly regulated was the cytoskeletal reorganisation/migration regulator – serine/threonine-protein kinase MRCK beta (CDC42BPB).The T-cell signalling protein C-Maf-inducing protein (CMIP), and SARG/C1orf116 also followed this trend of reduced downregulation in samples with mutations of *NOTCH1*.

The most specifically upregulated proteins in *NOTCH1*-mutant cases was the DNA damage response regulator and histone H4 modifier - E3 ubiquitin-protein ligase DTX3L. Poly [ADP-ribose] polymerase 9 (PARP9) also demonstrated *NOTCH1I*-specific upregulation – a functional binding partner of DTX3L required for ubiquitination activity, for which the expression between the two proteins correlated with an R^2 value of 0.74. While the third partner in this complex, PARP1, was unregulated across all cases, expression of an alternative PARP, PARP14, correlated to DTX3L with an R^2 value of 0.78. PARP14, which independently is an IL4-responsive survival factor in B cells 19147789, has been shown to form a complex with PARP9 and DTX3L inducing several oncogenic phenotypes in prostate cancer [492]. Another noteworthy upregulation with a higher prevalence in *NOTCH1*-mutated samples was that of CKAP4, the protein exhibiting the greatest upregulation in CLL, relative to HD samples. In cases with *NOTCH1* mutations, CKAP4 expression varied from 390% to 624%, while in other samples it was between 260 and 420%.

Two Ig; Ig alpha-1 chain C region (IGHA1) and Ig delta chain C region (IGHD); and the BCR component CD79A had higher expression in *NOTCH1*-mutant CLL cases. Other immune regulators which followed this trend included; complement component 1, q subcomponent binding protein (C1QBP), granzyme A (GZMA), guanylate binding protein 1, interferon-inducible, 67kDa (GBP1), interferon, gamma-inducible protein 30 (IFI30), tafazzin (TAZ) and toll-like receptor 7 (TLR7).

Carboxylic acid metabolism appeared upregulated in CLL cases with a mutation in *NOTCH1* with 6 proteins with greater expression; 2-oxoglutarate dehydrogenase complex component E2 (DLST), acyl-coenzyme A thioesterase 8 (ACOT8), long-chain-fatty-acid--CoA ligase 5 (ACSL5), fructose-1,6-bisphosphatase 1 (FBP1), bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, mitochondrial (MTHFD2), serine hydroxymethyltransferase, mitochondrial (SHMT2).

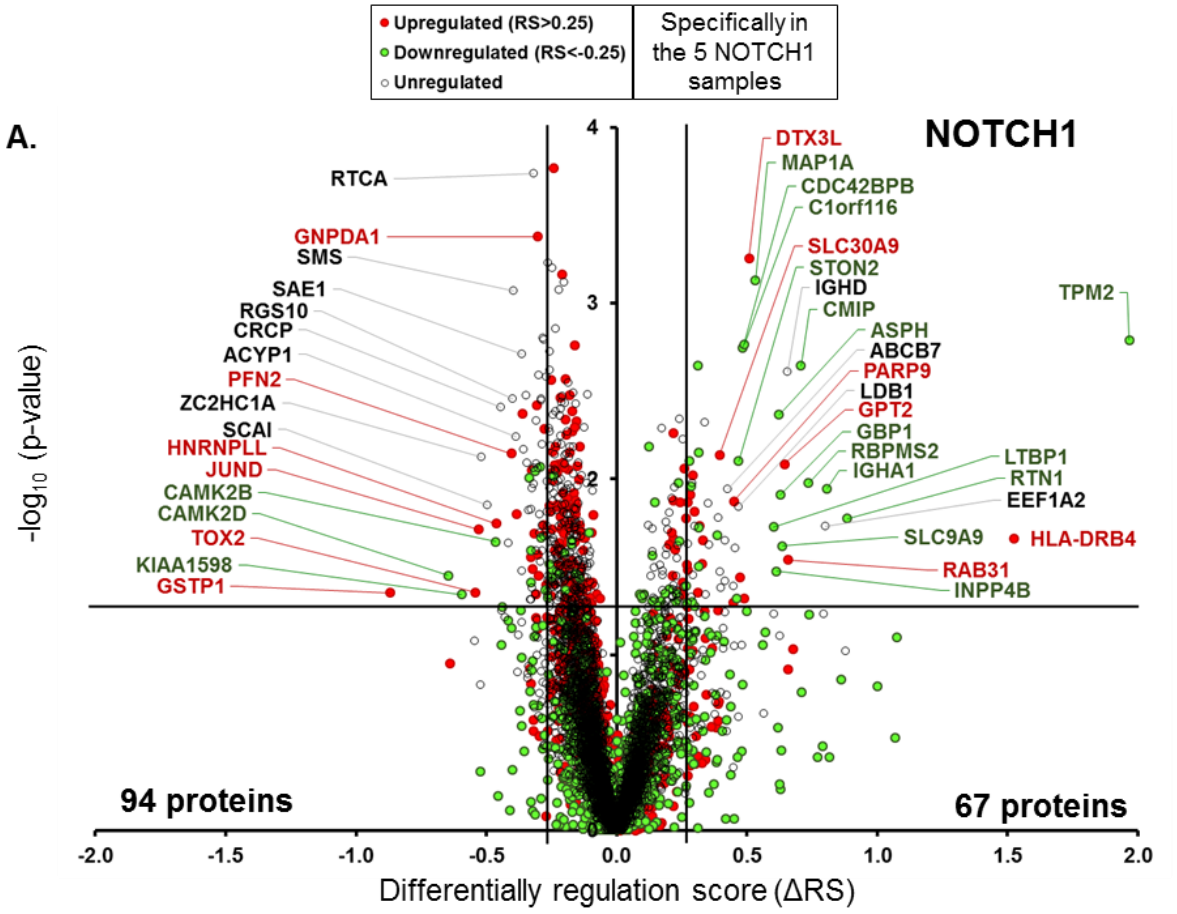
Amongst the 94 proteins relatively downregulated in the presence of *NOTCH1* mutation, 54 were proteins which would have otherwise been upregulated (CLL RS of >0.25), relative to HD controls. The majority of these therefore appeared unregulated, relative to HD samples, in the *NOTCH1*-mutant samples. RNA 3'-terminal phosphate cyclase (RTCA) was an example of this trend, with between 109 and 120% expression in *NOTCH1*-mutant cases, compared with 125 to 175% in cases with wildtype NOTCH1. Other proteins following this trend included, most notably; spermine synthase (SMS), SUMO-activating enzyme subunit 1 (SAE1), regulator of G-protein signalling 10 (RGS10), DNA-directed RNA polymerase III subunit RPC9 (CRCP) and zinc finger C2HC domain-containing protein 1A (ZC2HC1A). Proteins were also present that maintained their upregulation (*NOTCH1*-mutant sample RS of >0.25); glucosamine-6-phosphate isomerase 1 (GNPDA1), profilin-2 (PFN2), heterogeneous nuclear ribonucleoprotein L-like (HNRNPLL) and transcription factor JUND. Two calcium/calmodulin-dependent protein kinase type II subunits (CAMK2), CAMK2B and CAMK2D, exhibited downregulation in *NOTCH1*-mutant relative to both HD samples and wildtype NOTCH1 samples.

Shootin-1 (KIAA1598) demonstrated downregulation in *NOTCH1*-mutant samples, relative to samples with WT *NOTCH1*. This was also the only *NOTCH1*-mutant sample-downregulated protein to demonstrate significant downregulation when considering all CLL samples (CLL RS of < -0.25, $p<0.05$). Other noteworthy proteins appearing in this list were BCL2, two alcohol dehydrogenase enzymes (ADH5 and AKR1A1) and an aldose reductase (AKR1B1), two annexins ANX7 and ANX11,

The predominant trend amongst the proteins underexpressed in *NOTCH1*-mutant CLL samples was that of gene expression and nucleotide metabolism. In addition to RTCA, CRCP,

ZC2HC1A, HNRNPLL and JUND, 18 of the 94 proteins were annotated with a role in transcriptional regulation. Proteins canonically localised to the nucleus were also significantly overrepresented ($p=1.9x10^{-7}$).

A.



B.

Protein	Symbol	Unique Peptides	PSMs	WT				SF3B1 mut					Notch1 mut					ΔRS	p-value
				4621	3156	1125	2459	653	625	4140	3405	1194	2497	568	3999	2886	2483		
Tropomyosin beta chain	TPM2	1	77	-5.39	-5.78	-4.11	-5.04	-5.45	-5.48	-5.10	-6.07	-5.49	-2.06	-3.54	-2.11	-3.94	-1.88	1.97	0.0016
C-Maf-inducing protein	CMIP	6	33	-2.08	-2.57	-2.84	-2.84	-2.55	-2.79	-2.80	-2.18	-2.86	-1.76	-2.35	-1.93	-2.01	-1.52	0.71	0.0023
Ig delta chain C region	IGHD	4	81	-0.78	-0.78	-1.27	-1.39	-1.16	-1.16	-2.62	-1.03	-2.90	-0.80	-0.27	0.07	0.00	-0.41	0.66	0.0025
Alanine aminotransferase 2	GPT2	4	17	0.97	0.37	0.91	0.52	1.89	0.78	0.81	0.67	0.31	2.38	1.38	1.77	1.15	2.05	0.64	0.0082
Aspartyl/asparaginyl beta-hydroxylase	ASPH	8	26	-1.40	-3.05	-3.23	-3.01	-2.57	-3.05	-2.96	-2.53	-3.26	-2.62	-1.74	-2.22	-2.29	-2.39	0.62	0.0043
Microtubule-associated protein 1A	MAP1A	21	70	-1.68	-2.45	-2.29	-2.35	-2.41	-2.40	-2.29	-2.58	-2.43	-1.83	-1.94	-1.96	-2.14	-1.72	0.53	0.0007
E3 ubiquitin-protein ligase DTX3L	DTX3L	19	114	0.39	-0.04	0.25	-0.08	0.02	0.48	-0.14	-0.05	-0.42	0.89	0.57	0.64	0.48	0.43	0.51	0.0006
Specifically androgen-regulated gene protein	C1orf116	7	38	-2.10	-2.33	-2.81	-2.80	-2.84	-2.79	-2.77	-3.01	-2.90	-2.09	-2.51	-2.12	-2.40	-2.01	0.49	0.0017
Serine/threonine-protein kinase MRCK beta	CDC42BPB	15	64	-0.92	-1.91	-2.09	-2.18	-1.44	-2.15	-2.27	-1.32	-1.58	-0.96	-1.50	-1.24	-1.13	-0.78	0.48	0.0018
Stonin-2	STON2	4	19	-1.51	-2.01	-2.10	-2.17	-2.04	-2.03	-2.19	-2.07	-2.11	-1.59	-1.78	-1.84	-1.95	-1.56	0.47	0.0079
Zinc transporter 9	SLC30A9	5	22	0.63	0.33	0.29	-0.07	-0.42	-0.45	-0.27	-0.16	-0.71	0.37	0.32	0.42	0.10	0.25	0.40	0.0073
Uncharacterized protein C7orf43	C7orf43	6	32	0.37	-0.25	0.17	0.00	-0.51	-0.07	-0.19	0.08	-0.29	0.21	0.00	0.35	0.33	0.48	0.34	0.0048
Long-chain-fatty-acid--CoA ligase 5	ACSL5	9	75	-0.04	-1.17	-0.74	-1.00	-0.97	-1.31	-0.42	-1.03	-0.70	-0.50	-0.30	-0.67	-0.66	-0.26	0.32	0.0071
Pannexin-1	PANX1	3	16	-1.21	-1.99	-1.51	-1.70	-1.49	-1.76	-1.69	-1.40	-2.42	-1.32	-1.37	-0.90	-1.15	-1.17	0.31	0.0023
Stomatin-like protein 2, mitochondrial	STOML2	14	254	0.60	-0.12	-0.03	0.02	0.27	0.33	-0.18	-0.15	0.09	0.36	0.36	0.27	0.20	0.67	0.29	0.0096
RNA 3'-terminal phosphate cyclase	RTCA	14	126	0.00	0.44	0.75	0.60	0.49	0.81	0.73	0.33	0.52	0.26	0.22	0.17	0.12	0.18	-0.32	0.0002
MHC class II regulatory factor RFX1	RFX1	10	49	0.29	0.64	0.42	0.72	0.46	0.59	0.46	0.33	0.52	0.09	0.26	0.19	0.39	-0.10	-0.32	0.0087
Choline dehydrogenase, mitochondrial	CHDH	15	90	1.34	1.41	1.07	1.33	1.70	2.00	1.22	0.61	0.70	0.63	0.99	0.44	0.63	0.51	-0.33	0.0089
Aldose reductase	AKR1B1	11	247	0.01	0.33	0.33	0.25	-0.19	0.08	0.25	-0.04	-0.18	-0.37	-0.53	0.01	-0.18	-0.44	-0.33	0.0099
Cactin	CACTIN	7	25	-0.65	0.55	0.47	0.19	0.42	0.63	0.56	0.29	0.82	0.09	-0.28	0.27	0.13	0.20	-0.34	0.0081
Runt-related transcription factor 1 (Fragment)	RUNX1	1	6	0.14	0.39	0.39	0.42	0.32	0.44	0.44	-0.04	0.37	-0.26	-0.15	0.17	0.08	-0.16	-0.35	0.0034
Alpha-ketoglutarate-dependent dioxygenase FTO	FTO	16	93	-0.29	0.67	0.45	0.39	0.58	0.63	0.47	0.22	0.44	0.05	0.26	0.30	-0.14	-0.10	-0.36	0.0068
Protein DEK	DEK	17	762	-0.21	1.05	0.85	1.01	0.66	0.93	0.70	0.31	0.65	0.51	0.28	0.32	0.44	-0.01	-0.36	0.0042
SUMO-activating enzyme subunit 1	SAE1	19	374	-0.37	0.68	0.59	0.77	0.69	0.92	0.61	0.50	0.78	0.25	0.47	0.25	0.42	0.06	-0.36	0.0020
Acylphosphatase-1	ACYP1	5	64	-0.17	0.75	0.47	0.51	0.66	1.26	0.55	0.16	0.07	0.17	0.20	-0.05	0.07	-0.27	-0.39	0.0058
Spermine synthase	SMS	16	158	0.25	0.20	0.67	0.42	0.22	0.35	0.32	-0.03	0.76	-0.02	-0.09	-0.05	-0.21	-0.21	-0.40	0.0008
Regulator of G-protein signaling 10	RGS10	6	141	-0.23	0.65	0.78	0.51	0.76	0.81	0.58	0.47	0.44	0.21	0.36	0.37	0.03	-0.08	-0.40	0.0035
Profilin-2	PFN2	2	42	0.70	2.22	1.36	1.54	1.43	2.16	1.65	2.09	0.50	0.58	1.04	0.69	1.24	0.56	-0.40	0.0072
DNA-directed RNA polymerase III subunit RPC9	CRCP	3	11	-0.50	0.39	0.25	0.38	0.52	0.70	0.24	0.43	0.37	-0.15	-0.38	-0.05	0.24	-0.21	-0.45	0.0039
Zinc finger C2HC domain-containing protein 1A	ZC2HC1A	9	33	0.19	1.24	0.95	1.05	0.78	0.90	0.88	0.58	0.00	0.34	0.46	0.18	-0.17	-0.39	-0.52	0.0075

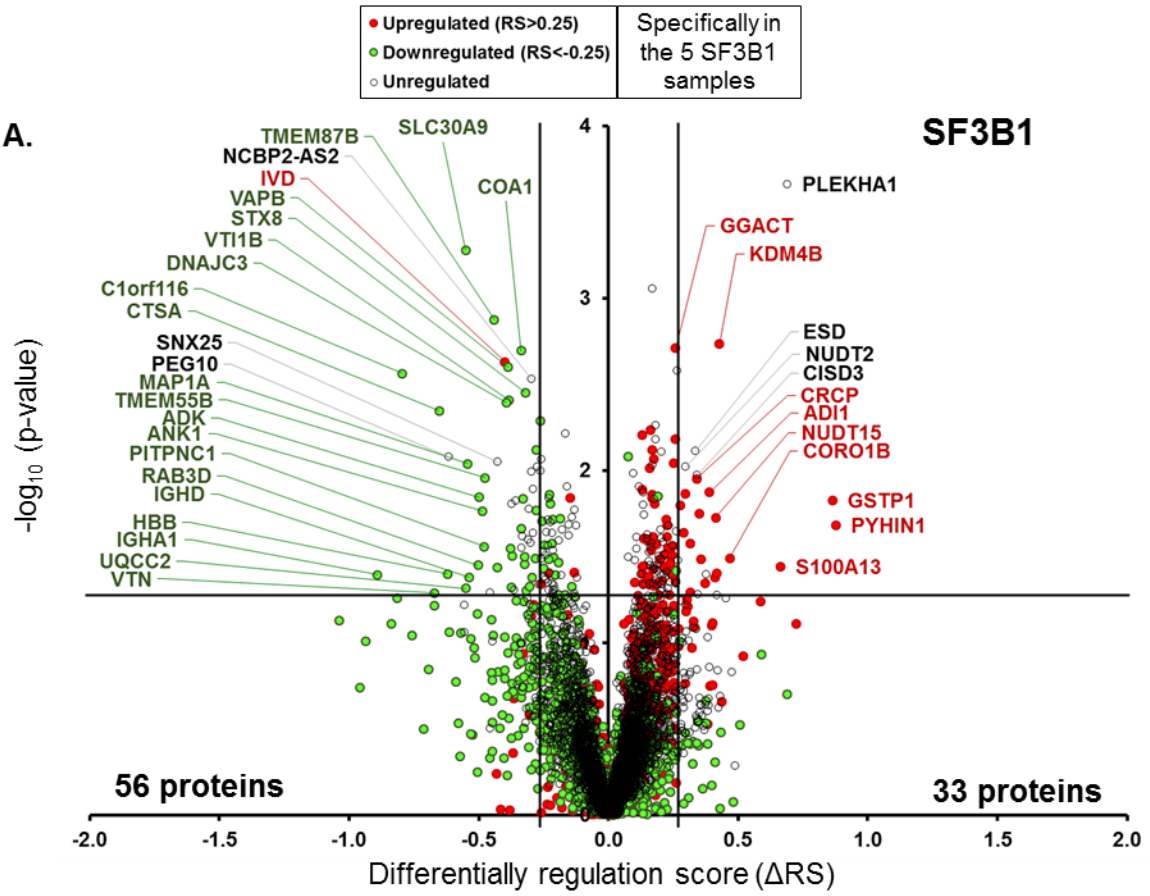
Figure 7.24. **Subtype specific protein expression in *NOTCH1*-mutant CLL. A.** Volcano plot describing the differential regulation score (ΔRS) plotted against the $-\log_{10}$ (p-value) of differential regulation between samples bearing mutations of the *NOTCH1* gene and samples with wildtype *NOTCH1*. The plot highlights those proteins for which the regulation score calculated for the 5 *NOTCH1*-mutant samples was >0.25 (red) and < -0.25 (green). **B.** The $\log_2$ (ratios) for the top 15 significantly overexpressed and top 15 underexpressed proteins in *NOTCH1*-mutant CLL, detailing the sample expression, the ΔRS and the uncorrected p-value estimating the probability of subtype-specific expression between the 5 *NOTCH1*-mutant samples and 8 wildtype *NOTCH1*-mutant samples (4621 was excluded from the analysis as an outlier).

7.18 *SF3B1*-MUTANT CLL-SPECIFIC PROTEIN EXPRESSION

SF3B1-mutant CLL samples had the smallest degree of subtype specific regulation amongst the subtypes analysed. However, the observed trend was one of amplification, with the proteins specifically upregulated in *SF3B1*-mutants (33) exhibiting upregulation relative to HD samples, and the downregulated proteins (56) being downregulated relative to the HD samples. Pleckstrin homology domain-containing family A member 1 (PLEKHA1) was specifically overexpressed relative to wildtype *SF3B1* samples in which expression ranged from 33% to 57% of that of HD samples, whereas in *SF3B1*-mutant samples expression was between 72% and 108%. Also observed in this list was pleckstrin homology domain-containing family G member 1 (PLEKHG1) with a generally greater expression in samples bearing an *SF3B1* mutation (ΔRS=0.41). Gamma-glutamylaminecyclotransferase (GGACT) had the most specific upregulated expression in *SF3B1*-mutants with 140 to 170% of HD control expression compared to 110 to 143% expression in samples wildtype for *SF3B1*. Another observation amongst the upregulated *SF3B1*-mutant samples included a generally higher expression of CD5 (250-540%) than in the other samples (135-300%).

Proteins downregulated in samples with an SF3B1 mutation relative to both HD samples and *SF3B1*-wildtype samples included; zinc transporter 9 (SLC30A9), SARG/C1orf116, lysosomal protective protein (CTSA) and transmembrane protein 87B (TMEM87B). Just two proteins exhibited upregulation relative to HD samples (*SF3B1*-mutated sample RS>0.25) that was less than that of the other samples (ΔRS< -0.25); Ribonuclease UK114 (HRSP12) and isovaleryl-CoA dehydrogenase, mitochondrial (IVD). Membranous proteins were well-represented in this list with 33 out of 56 with this annotation, which included localisation to lysosomes (6) and the ER (12) and 7 proteins functionally related to vesicle mediated transport.

A.



B.

Protein	Symbol	Unique Peptides	PSMs	WT				Notch1 mut					SF3B1 mut					ΔRS	p-value
				4621	3156	1125	2459	2497	568	3999	2886	2483	653	625	4140	3405	1194		
Pleckstrin homology domain-containing A1	PLEKHA1	4	16	-0.38	-1.19	-1.12	-1.12	-1.62	-1.10	-0.83	-0.89	-0.82	-0.04	0.11	-0.45	-0.47	-0.21	0.69	0.0002
Lysine-specific demethylase 4B	KDM4B	9	72	0.03	-0.24	0.46	0.14	0.20	0.21	0.14	0.19	-0.42	0.54	0.50	0.64	0.39	0.69	0.43	0.0018
S-formylglutathione hydrolase	ESD	12	195	-0.68	0.09	0.12	-0.59	-0.31	-0.06	-0.20	-0.36	-0.47	0.15	0.49	-0.04	0.14	0.20	0.34	0.0077
Bis(5'-nucleosyl)-tetraphosphatase	NUDT2	6	15	0.16	-0.34	0.20	-0.40	-0.30	-0.38	-0.07	-0.56	-0.55	0.23	0.12	-0.12	-0.10	0.21	0.30	0.0095
F-box only protein 30	FBXO30	2	9	-0.40	-0.03	-0.19	-0.19	0.06	-0.35	-0.53	-0.36	-0.40	0.08	0.03	0.01	0.03	0.15	0.27	0.0026
Gamma-glutamylaminecyclotransferase	GGACT	3	21	0.28	0.26	0.52	0.52	0.35	0.44	0.42	0.14	0.23	0.66	0.63	0.63	0.48	0.76	0.26	0.0020
Protein arginine N-methyltransferase 1	PRMT1	16	156	0.15	0.42	0.71	0.62	0.27	0.26	0.43	0.53	0.57	0.75	0.91	0.86	0.57	0.71	0.26	0.0066
Ribonuclease P protein subunit p30	RPP30	6	21	-0.13	0.01	0.01	-0.24	-0.14	0.15	0.23	0.15	0.35	0.50	0.36	0.16	0.41	0.30	0.25	0.0091
DNA polymerase beta	POLB	14	102	-0.29	-0.26	-0.04	0.15	-0.18	0.13	0.09	-0.03	-0.14	0.12	0.34	0.09	0.16	0.25	0.20	0.0079
Surfeit locus protein 4	SURF4	2	14	-2.08	-0.38	-0.32	-0.36	-0.49	-0.31	-0.62	-0.37	-0.60	-0.10	-0.25	-0.14	-0.28	-0.33	0.18	0.0055
Programmed cell death protein 6	PDCD6	8	85	-0.34	-0.16	-0.09	-0.02	-0.19	-0.39	-0.12	-0.26	0.03	0.10	-0.07	0.12	0.00	0.12	0.18	0.0066
NEDD8-activating E1 regulatory subunit	NAE1	12	103	-0.37	0.34	0.20	0.09	0.03	0.08	0.26	0.21	0.13	0.44	0.36	0.27	0.25	0.48	0.18	0.0086
Glycogen synthase kinase-3 alpha	GSK3A	6	87	-0.37	-0.31	0.03	0.03	-0.12	-0.28	-0.11	-0.13	-0.17	0.03	-0.06	0.08	0.06	0.19	0.17	0.0093
Nuclear RNA export factor 1	NXF1	18	84	-0.41	0.18	0.19	0.23	-0.04	0.03	-0.10	0.13	0.17	0.30	0.23	0.34	0.17	0.36	0.17	0.0076
Phospholipase A-2-activating protein	PLAA	21	139	-0.12	-0.23	-0.04	-0.18	-0.27	-0.19	-0.17	-0.01	-0.13	0.02	0.13	0.02	-0.01	0.00	0.17	0.0009
Uncharacterized protein NCBP2-AS2	NCBP2-AS2	2	6	0.32	0.31	0.16	0.25	0.48	0.63	0.11	0.46	0.49	0.03	0.13	0.07	-0.24	0.07	-0.30	0.0029
Syntaxin-8	STX8	4	32	-0.24	-0.06	-0.19	0.00	0.00	-0.01	-0.03	-0.23	-0.39	-0.43	-0.31	-0.41	-0.57	-0.74	-0.32	0.0035
Cytochrome c oxidase assembly factor 1 homolog	COA1	4	22	-0.30	0.05	0.04	0.11	-0.27	0.06	-0.06	-0.27	-0.08	-0.69	-0.47	-0.40	-0.36	-0.28	-0.34	0.0020
Vesicle transport through interaction with t-SNAREs homolog 1B	VTI1B	5	53	-0.22	-0.21	-0.43	-0.19	-0.37	-0.28	-0.19	-0.52	-0.58	-0.88	-0.61	-0.76	-0.74	-1.18	-0.38	0.0039
Vesicle-associated membrane protein-associated protein B/C	VAPB	11	86	-0.28	0.45	-0.12	0.29	0.01	0.31	-0.16	-0.09	0.00	-0.48	-0.07	-0.48	-0.32	-0.52	-0.39	0.0025
DnaJ homolog subfamily C member 3	DNAJC3	12	57	-1.11	-1.01	-1.40	-1.31	-0.92	-1.09	-1.37	-0.84	-0.92	-1.28	-1.47	-1.47	-1.42	-1.59	-0.39	0.0040
Isovaleryl-CoA dehydrogenase, mitochondrial	IVD	16	228	0.95	0.73	1.11	1.00	0.72	1.15	0.95	0.75	0.86	0.73	0.21	0.51	0.41	0.38	-0.40	0.0023
Sorting nexin-25	SNX25	3	11	-0.81	0.36	0.25	0.28	0.60	0.53	0.40	-0.03	0.02	-0.52	-0.40	-0.13	0.21	-0.29	-0.43	0.0089
Transmembrane protein 87B	TMEM87B	2	10	-0.68	-0.85	-1.40	-1.23	-0.92	-1.01	-1.37	-1.31	-1.26	-1.63	-1.38	-1.59	-1.59	-1.53	-0.44	0.0013
Microtubule-associated protein 1A	MAP1A	21	70	-1.68	-2.45	-2.29	-2.35	-1.83	-1.94	-1.96	-2.14	-1.72	-2.41	-2.40	-2.29	-2.58	-2.43	-0.54	0.0092
Zinc transporter 9	SLC30A9	5	22	0.63	0.33	0.29	-0.07	0.37	0.32	0.42	0.10	0.25	-0.42	-0.45	-0.27	-0.16	-0.71	-0.55	0.0005
Retrotransposon-derived protein PEG10	PEG10	4	10	0.40	1.42	1.14	1.17	0.82	1.37	0.66	1.08	0.61	0.05	0.60	0.78	-0.07	-0.11	-0.61	0.0084
Lysosomal protective protein	CTSA	7	71	-1.34	-1.08	-1.69	-1.84	-1.48	-1.89	-1.44	-1.27	-1.12	-1.98	-1.83	-1.87	-2.03	-1.90	-0.65	0.0045
Specifically androgen-regulated gene protein	C1orf116	7	38	-2.10	-2.33	-2.81	-2.80	-2.09	-2.51	-2.12	-2.40	-2.01	-2.84	-2.79	-2.77	-3.01	-2.90	-0.79	0.0027
Tropomyosin beta chain	TPM2	1	77	-5.39	-5.78	-4.11	-5.04	-2.06	-3.54	-2.11	-3.94	-1.88	-5.45	-5.48	-5.10	-6.07	-5.49	-2.64	0.0062

Figure 7.25. **Subtype specific protein expression in *SF3B1*-mutant CLL. A.** Volcano plot describing the differential regulation score (ΔRS) plotted against the $-\log_{10}$ (p-value) of differential regulation between samples bearing mutations of the *SF3B1* gene and samples with wildtype *SF3B1*. The plot highlights those proteins for which the regulation score calculated for the 5 *SF3B1*-mutant samples was >0.25 (red) and < -0.25 (green). **B.** The $\log_2$ (ratios) for the top 15 significantly overexpressed and top 15 underexpressed proteins in *SF3B1*-mutant CLL, detailing the sample expression, the ΔRS and the uncorrected p-value estimating the probability of subtype-specific expression between the 5 *SF3B1*-mutant samples and 8 wildtype *SF3B1*-mutated samples (4621 was excluded from the analysis as an outlier).

7.19 CD38$^+$ CLL-SPECIFIC PROTEIN EXPRESSION

CD38, ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1, is a transmembrane marker which correlates with immune cell activation. When expressed on CLL, CD38 is associated with a more aggressive phenotype. For the three samples, 2483, 3405 and 1194 immunophenotyped as 99% CD38$^+$, CD38 expression was determined by proteomics to be the most significantly upregulated ($p=2.4x10^{-6}$) in these cases with 96%, 76% and 91% expression, respectively, on average, relative to the healthy donor B-cell samples (**Figure 7.26**). The remaining samples, immunophenotyped with less than 50% CD38$^+$ expression, were detected with significant downregulation to between 13% and 35% expression compared with the HD controls.

A marker of cell proliferation, KI67/MKI67, was observed with an even more significant specific (than CD38) expression in the CD38$^+$ samples ($p=1.0x10^{-6}$) (103-110% of HD expression) while downregulated in the remaining samples (35-70% of HD expression). The correlation between KI67 and CD38 expression had an R^2 value of 0.79. A third protein which fit this pattern of expression was the key cell cycle control protein cyclin-dependent kinase 1 (CDK1), also correlating with CD38 expression ($R^2=0.75$). Condensin complex subunit 3 (NCAPG), alongside subunit 1 (NCAPD2), a complex which condenses chromosomes for mitosis, and the uncharacterised coiled-coil domain-containing protein 102A (CCDC102A) exhibited downregulation in all but the CD38$^+$ samples.

Several groups of proteins had higher protein expression specifically in CD38$^+$ samples, including; 21 ribosomal proteins, 7 clathrin-associated adaptor protein complex subunits (AP1G2, AP1S2, AP3B1, AP3D1, AP3S1, AP5S1 and AP5Z1), 6 E3 ubiquitin ligases (RNF123, RNF), 6 integrator complex subunits (INTS1, INTS2, INTS4, INTS5, INTS6 and INTS8), 6 proteosome components (PSMB6, PSMB7, PSMD11, PSMD14, PSMD9, PSME4), 5 exportin subunits (XPO1, XPO2, XPO4, XPO5 and XPOT), 5 importin subunits (IPO4, IPO5, IPO7, IPO9 and IPO13), 3 dedicator of cytokinesis proteins (DOCK2, DOCK7 and DOCK11), 3

inositol 1,4,5-trisphosphate receptors (ITPR1, ITPR2 and ITPR3), 3 N-alpha-acetyltransferases (NAA10, NAA16 and NAA25), 3 ubiquitin carboxyl-terminal hydrolases (USP5, USP15 and USP34) and 3 unconventional myosins (MYO1D, MYO1F and MYO9B). Other noteworthy proteins included DNA replication licensing factors MCM4 and MCM7, retinoblastoma-associated protein (RB1), serine-protein kinase ATM, structural maintenance of chromosomes proteins SMC2 and SMC4, and the RAF proto-oncogene.

GO term enrichment of the 278 $CD38^+$ CLL-specific proteins identified strong trends relating to extracellular matrix focal adhesion ($p=1.38x10^{-5}$), vesicles ($p=2.83x10^{-5}$), nucleolus ($p=4.42x10^{-4}$) and nuclear pores ($p=4.17x10^{-4}$).

The most $CD38^+$-specifically underexpressed protein was adenlyate kinase 1 (AK1) an enzyme capable of transferring phosphate groups from ATP to AMP, which exhibited, relative to HD control expression, 44-62% expression in $CD38^+$ cases compared with 80-410% in other CLL samples. Another protein with sizable downregulation in $CD38^+$ cases was F-box only protein 2 (FBXO2), a protein with a role in the ER degradation pathway ubiquitinating unfolded proteins targeting them for degradation. A potential chaperone with unfolded protein binding, DnaJ homolog subfamily B member 4 (DNAJB4), was observed alongside FBXO2.

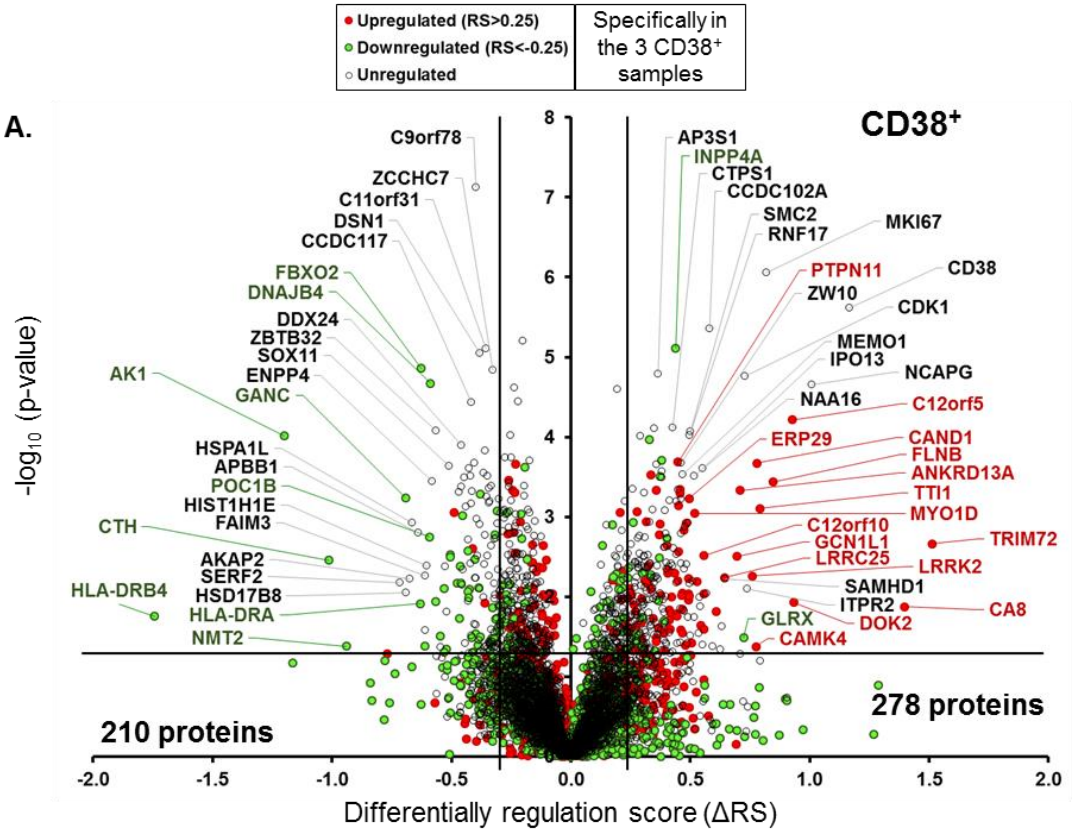
Cell surface proteins specifically underexpressed in $CD38^+$ cases, included CD40, CD44, CD46, IL4RA, FAIM3, ICAM2 and TGF-beta receptor type-2 (TGFBR2). Other underexpressed protein included; 6 MHC class II molecules, 4 histones (H1.2, H1.4, H4 and H2A) and 2 high mobility group nucleosome-binding domain-containing proteins (HMGN3 and HMGN4), 4 histone modifying enzymes; histone deacetylase complex subunits SAP18 and SAP30 and histone-lysine N-methyltransferases NSD3 and SETD2, 3 subunits of 17-beta-hydroxysteroid dehydrogenase (HSD17B4, HSD17B8 and HSD17B11), 3 pre-mRNA splicing factors (CWC25, ISY1 and WTAP) and 3 of the 4 components of the MIS12/MIND kinetochore subcomplex (DSN1, NSL1 and MIS12).

RNA binding, specifically poly(A) RNA binding was significantly overrepresented amongst the $CD38^+$-specific, underexpressed proteins ($p=1.65x10^{-5}$). Additionally, proteins relating to chromatin organisation, regulation of lymphocyte activation, regulation of mRNA splicing, via spliceosome were overrepresented.

Figure 7.26. **Subtype-specific protein expression in $CD38^+$ CLL samples. A.** Volcano plot describing the differential regulation score (ΔRS) plotted against the $-\log_{10}$ (p-value) of differential regulation between $CD38^+$ CLL samples (>99%) and $CD38^-$ CLL samples (<50% $CD38^+$). The plot highlights those proteins for which the regulation score calculated for the 3 $CD38^+$ samples was >0.25 (red) and < -0.25 (green). **B.** The $\log_2$ (ratios) for the top 15

significantly overexpressed and top 15 underexpressed proteins in CD38^{+} CLL, detailing the sample expression, the ΔRS and the uncorrected p-value estimating the probability of subtype-specific expression between the 3 CD38^{+} samples and 10 CD38^{-} CLL samples (4621 was excluded from the analysis as an outlier).

A.



B.

Protein	Symbol	Unique Peptides	PSMs	4621	3156	1125	2459	853	625	4140	2497	568	3999	2886	CD38 2483	CD38 3405	CD38 1194	ΔRS	p-value
ADP-ribosyl cyclase 1	CD38	4	17	-1.34	-1.55	-2.88	-2.24	-1.98	-1.96	-2.23	-1.62	-1.37	-1.54	-1.90	-0.05	-0.39	-0.14	1.17	0.000002
Condensin complex subunit 3	NCAPG	3	17	-1.48	-2.97	-2.73	-2.48	-1.58	-2.56	-1.03	-2.08	-0.83	-1.34	-1.43	-0.06	-0.24	0.04	1.01	0.000022
Fructose-2,6-bisphosphatase TIGAR	C12orf5	6	45	-0.12	-0.57	-0.01	0.14	-0.47	0.00	0.54	-0.27	0.22	0.90	0.48	1.11	0.98	1.17	0.93	0.000061
Filamin-B	FLNB	86	1174	-1.36	-0.82	-1.48	-1.74	-0.65	-0.62	-0.89	-0.91	0.17	-1.12	0.58	0.66	0.45	0.32	0.85	0.000364
Antigen KI-67	MKI67	15	46	-0.36	-1.40	-1.50	-1.01	-0.99	-0.99	-1.01	-1.09	-0.49	-0.70	-0.66	0.04	0.15	0.04	0.82	0.000001
TELO2-interacting protein 1 homolog	TTI1	3	16	-0.79	-1.79	-0.52	-0.95	-1.39	-1.39	0.22	-0.95	0.25	-0.33	-0.57	0.68	0.27	0.33	0.79	0.000786
Cullin-associated NEDD8-dissociated protein 1	CAND1	39	466	-0.67	-1.04	-0.56	-1.36	-0.36	-1.61	0.28	-1.22	-0.09	-0.88	-0.68	0.22	0.33	0.49	0.78	0.000212
Cyclin-dependent kinase 1	CDK1	3	56	-0.71	-1.10	-1.56	-1.75	-0.71	-1.80	-0.97	-1.05	-0.53	-0.84	-1.04	-0.17	-0.01	-0.02	0.73	0.000017
Ankyrin repeat domain-containing protein 13A	ANKRD13A	15	140	0.00	-0.77	0.49	0.17	-0.23	0.00	0.08	0.28	-0.20	0.64	0.60	0.84	0.77	0.88	0.71	0.000468
Coiled-coil domain-containing protein 102A	CCDC102A	3	12	-0.40	-1.41	-0.97	-1.14	-1.04	-1.20	-1.27	-0.55	-1.15	-0.64	-1.00	-0.26	-0.32	-0.18	0.58	0.000004
Importin-13	IPO13	7	18	-0.16	-1.14	-0.14	-0.62	-0.41	-0.71	0.16	-0.76	-0.20	-0.69	-0.16	0.14	0.25	0.30	0.55	0.000244
Unconventional myosin-Id	MYO1D	6	28	0.11	-1.13	-0.42	-0.65	-0.27	-0.26	-0.23	-0.16	0.51	-0.69	-0.43	0.20	0.27	0.37	0.52	0.000908
N-alpha-acetyltransferase 16, NatA auxiliary	NAA16	6	85	-0.53	-0.98	0.22	-0.55	-0.80	-0.65	-0.08	-0.40	-0.07	-0.30	-0.39	0.18	0.23	0.30	0.52	0.000308
Structural maintenance of chromosomes 2	SMC2	13	45	-0.44	-0.75	-0.59	-0.63	-0.51	-0.66	-0.37	-0.55	-0.16	-0.17	-0.73	0.08	0.18	0.00	0.50	0.000085
RING finger protein 17	RNF17	1	8	0.23	-0.51	-0.34	-0.54	-0.91	-0.66	-0.08	-1.03	-0.95	-0.97	-0.49	0.01	0.02	-0.02	0.50	0.000096
Prenylcysteine oxidase 1	PCYOX1	11	66	1.43	1.17	0.61	0.96	0.64	0.69	0.72	0.33	0.62	0.07	0.36	0.03	-0.01	0.12	-0.43	0.000246
Lactoylglutathione lyase	GLO1	12	395	0.24	1.39	0.75	1.30	1.01	1.68	1.00	0.18	1.09	0.52	0.80	0.29	0.37	0.14	-0.44	0.000754
Transcription factor E3	TFE3	4	21	0.48	0.33	0.09	0.41	0.49	0.32	0.17	0.36	0.20	-0.14	0.15	-0.17	-0.35	-0.28	-0.44	0.000321
Echinoderm microtubule-associated-like 2	EML2	11	93	-0.41	0.68	0.00	0.10	0.57	0.58	0.19	0.08	0.24	0.02	-0.12	-0.41	-0.31	-0.18	-0.45	0.000958
ATP-dependent RNA helicase DDX24	DDX24	31	230	0.26	0.61	0.56	0.44	0.98	0.39	0.81	0.45	1.29	0.36	0.36	-0.03	0.03	0.07	-0.46	0.000128
Centrin-3	CETN3	4	14	0.30	0.91	1.27	1.60	0.60	0.94	1.00	0.62	0.59	0.22	0.91	0.21	0.07	0.27	-0.46	0.000424
Centriole, cilia and spindle-associated protein	CCSAP	1	9	1.13	0.90	0.69	1.46	0.98	1.05	0.81	1.05	0.44	0.64	1.00	0.17	0.22	0.41	-0.47	0.000493
Cx9C motif-containing protein 4	CMC4	4	17	1.48	1.59	0.93	0.96	1.46	1.16	0.74	1.21	1.04	0.86	1.06	0.39	0.60	0.33	-0.49	0.000883
Transcription factor SOX-11	SOX11	1	11	0.51	0.77	0.60	0.81	0.26	0.74	0.61	0.32	0.34	0.16	0.75	-0.17	0.04	-0.12	-0.51	0.000229
Zinc finger and BTB domain-containing 32	ZBTB32	2	6	0.39	0.74	0.94	0.71	1.83	0.82	0.61	1.37	1.45	0.78	0.87	0.06	0.24	0.22	-0.56	0.000084
Bis(5'-adenosyl)-triphosphatase ENPP4	ENPP4	6	42	1.22	0.75	0.88	0.85	0.43	0.56	0.25	0.52	0.27	0.70	0.31	-0.05	-0.11	-0.29	-0.58	0.000358
DnaJ homolog subfamily B member 4	DNAJB4	5	16	-1.05	0.78	0.28	0.27	0.08	0.39	0.18	0.20	0.15	-0.17	0.05	-0.47	-0.35	-0.51	-0.59	0.000021
F-box only protein 2	FBXO2	2	8	0.20	0.19	0.71	0.64	0.20	0.43	-0.06	0.63	-0.09	0.35	0.52	-0.35	-0.45	-0.32	-0.63	0.000014
Neutral alpha-glucosidase C	GANC	10	51	-0.65	0.41	-0.52	0.83	-0.28	0.51	-0.18	0.81	0.48	-0.21	-0.20	-0.56	-0.61	-0.74	-0.69	0.000572
Adenylate kinase isoenzyme 1	AK1	8	111	-0.91	1.72	0.57	1.19	2.03	1.14	1.29	-0.03	0.78	-0.27	0.14	-0.82	-0.69	-1.20	-1.20	0.000097

7.20 TRISOMY 12 CLL-SPECIFIC PROTEIN EXPRESSION

The acquisition of third copy of chromosome 12, termed trisomy 12, is observed in around 15-20% of CLL cases and results in a more aggressive phenotype. A marker that frequently correlates with trisomy 12 is integrin alpha 4, known as CD49d, a cell surface receptor with the potential to initiate binding to the extracellular matrix and induce intracellular signalling. Indeed, the protein demonstrating the greatest degree of specificity in trisomy 12 CLL cases was CD49d, with 47-145% expression of that of HD controls, compared with 16-40% in disomy 12 cases (**Figure 7.27**). Integrin beta 7, a co-receptor required for the binding of CD49d to fibronectin, vascular CAM 1 (VCAM1) or mucosal addressin CAM 1 (MADCAM1), was also specifically overexpressed in trisomy 12 cases (96-135% of HD) compared to disomy 12 cases (45-85% of HD). A third integrin, alpha-L (ITGAL), while still downregulated in trisomy 12 cases (40-80% of HD), generally had higher expression than that of disomy 12 cases (30-45% of HD).

Ubiquitin carboxyl-terminal hydrolase 5 (UPS5) a chr. 12 protein for which loss increases p53 accumulation [493] was specifically overexpressed (135-140% of HD) in trisomy 12 cases. With a nearly identical expression pattern was another chr.12 protein tyrosine-protein phosphatase non-receptor type 11 (PTPN11 or SHP2) (140-145% of HD), a cell surface to nucleus signal transducer, which has been shown to correlate with telomerase activity [494]. Both proteins had minimal regulation (88-119%) compared with HD controls in disomy 12 samples. IQ motif and SEC7 domain-containing protein 1 (IQSEC1), which when downregulated amplified cell spreading and accumulation of cell surface integrin expression [495], demonstrated 70-100% expression in trisomy 12 samples, whereas in disomy 12, was downregulated to 30-50% of HD control expression.

Additional proteins with trisomy 12-specific upregulation, relative to both HD samples and disomy 12 samples, included; TP53-induced glycolysis and apoptosis regulator (TIGAR) – a protein offering protection against and p53-induced reactive oxygen species-related apoptosis [496], TP53-regulated inhibitor of apoptosis 1 (TRIAP1), an inhibitor or caspase 9 activation [497] and extended synaptotagmin-1 (ESYT1) a Ca^{2+}-responsive ER glycerophospholipid binding/transport protein which plays a role in anchoring the ER to the plasma membrane [498]. All three of these proteins, TIGAR, TRIAP1 and ESYT1 are encoded by chr. 12.

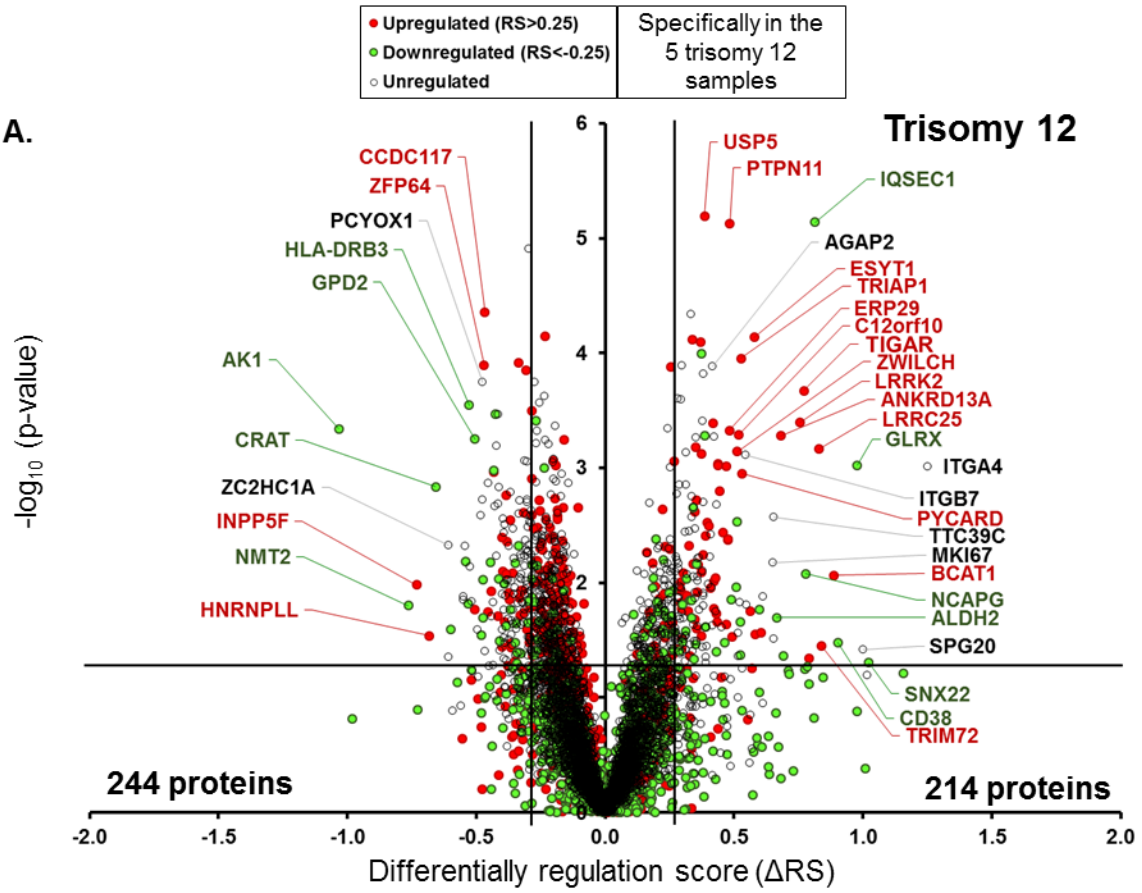
Two cell surface proteins were specifically upregulated in trisomy 12, relative to both disomy 12 samples ($\Delta RS>0.25$) and HD samples ($RS>0.25$); Leucine-rich repeat-containing protein 25 (LRRC25) and receptor-type tyrosine-protein phosphatase O (PTPRO). Three proteins related to B-cell activation had greater expression in trisomy 12 CLL cases; proto-

oncogene VAV1, PTPN6 and phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform (PIK3CG). PTPN6 and PTPRO are also both encoded by chr. 12.

Amongst the 214 trisomy 12-upregulated proteins, 72, 106 and 36 proteins were upregulated (RS>0.25), unregulated (0.25>RS> -0.25) and downregulated (RS< -0.25) amongst the 5 trisomy 12 samples, relative to the HD samples, respectively. The proteins were analysed for chromosome enrichment using DAVID and the three lists were overlaid onto the human genome using the Ensembl genome browser (**Figure 7.28**). Of these three sub-lists of proteins, 42/72 (58.3%) upregulated proteins, 37/106 (34.9%) unregulated proteins and 3/36 (8.3%) downregulated proteins mapped to chromosome 12; a total of 82 proteins, with an enrichment p-value of $2.3x10^{-50}$. A comparative enrichment analysis of the 819 proteins upregulated across all CLL samples also revealed a significant enrichment, with 70/819 proteins (8.5%) mapping to chr. 12 ($p=3.2x10^{-5}$).

For the 244 trisomy 12-specific downregulated proteins, a predominant trend was one of the downregulation of otherwise upregulated proteins, with 175/244 (72%) giving a regulation score of >0.25 when calculating from the 8 disomy 12 samples. There was also a trend of overlapping proteins observed specific to $CD38^+$ samples, due to the co-occurrence of the subtypes; such as CCDC117 and AK1. While also significant in $CD38^+$ CLL, zinc finger protein 64 (ZFP64) underexpression was more significantly specific to trisomy 12 cases, with less upregulation, 120-150%, compared with 175-300% in disomy 12 cases, comparing to that of HD control samples. ZFP64 acts as a potential transcription factor upregulated in response to toll-like receptor signalling, inducing several cytokines. Prenylcysteine oxidase 1 (PCYOX1), a protein with a role in degrading prenylated proteins, was unregulated in trisomy 12 cases (100-125% of HD) while upregulated in disomy 12 samples (125-225% of HD). INPP5F, described in **Figure 7.21** as the most upregulated protein potentially inhibiting aspects of the BCR signalling pathway, demonstrated a lower expression amongst trisomy 12 (110-255% of HD), relative to disomy 12, samples (255-510% of HD). Cell surface proteins underexpressed in trisomy 12 cases, relative to disomy 12 cases included; CD6, FAIM3, CD75, CD43, FCRL3/CD307c, CD166, CXCR4, CD44 and CD55. 74 of the 244 trisomy 12-underexpressed proteins were annotated with nucleic acid binding ($p=5.5x10^{-5}$), 24 of which were RNA binding ($p=5.7x10^{-3}$) and 50 were DNA binding ($p=1.2x10^{-2}$). 8 proteins were spliceosome components.

A.



B.

Protein	Symbol	Unique Peptides	PSMs	4621	3156	1125	2459	653	625	4140	2497	568	3999	2886	2483	3405	1194	ΔRS	p-value
Integrin alpha-4	ITGA4	11	107	-1.27	-2.11	-2.61	-2.66	-2.06	-2.48	-2.57	-1.84	-1.31	-0.13	-0.77	-1.08	0.54	-0.58	1.25	0.00097
Glutaredoxin-1	GLRX	5	70	-1.70	-2.98	-2.97	-2.94	-3.27	-2.83	-2.73	-1.94	-2.98	-1.47	-2.00	-1.53	-1.87	-0.82	0.98	0.00094
Leucine-rich repeat-containing protein 25	LRRC25	2	14	-0.46	-0.70	-0.13	-0.70	-0.87	-0.10	0.34	0.50	-0.11	0.77	0.90	0.81	0.73	0.59	0.83	0.00068
IQ motif and SEC7 domain-containing protein 1	IQSEC1	11	97	-1.42	-1.32	-1.76	-1.46	-1.03	-1.54	-1.34	-1.07	-1.56	-0.06	-0.48	-0.24	-0.49	-0.48	0.81	0.00001
Fructose-2,6-bisphosphatase TIGAR/C12orf5	TIGAR	6	45	-0.12	-0.57	-0.01	0.14	-0.47	0.00	0.54	-0.27	0.22	0.90	0.48	1.11	0.98	1.17	0.77	0.00021
Leu-rich repeat serine/threonine kinase 2	LRRK2	20	75	-0.31	-1.10	-0.61	-0.47	-0.23	-0.77	-0.01	-1.01	-0.47	0.18	0.17	0.26	0.83	0.73	0.75	0.00040
Ankyrin repeat domain-containing protein 13A	ANKRD13A	15	140	0.00	-0.77	0.49	0.17	-0.23	0.00	0.08	0.28	-0.20	0.64	0.60	0.84	0.77	0.88	0.68	0.00053
Extended synaptotagmin-1	ESYT1	40	670	0.52	0.32	0.02	0.21	-0.22	0.19	0.01	0.11	0.50	0.75	0.72	0.67	0.80	0.71	0.58	0.00007
Integrin beta-7	ITGB7	7	43	0.36	-1.17	-0.54	-0.35	-0.77	-0.25	-0.75	-0.36	-0.28	-0.01	0.44	-0.06	0.02	0.35	0.54	0.00078
TP53-regulated inhibitor of apoptosis 1	TRIAP1	3	31	0.27	0.01	0.17	0.40	0.63	0.62	0.16	0.37	0.40	0.86	1.01	0.93	0.77	0.84	0.53	0.00011
UPF0160 protein MYG1, mitochondrial	C12orf10	7	71	0.45	-0.06	0.01	0.05	0.09	0.18	0.19	-0.13	0.21	0.49	0.56	0.94	0.63	0.81	0.52	0.00051
Protein zwilch homolog	ZWILCH	3	7	0.67	-0.48	-0.02	-0.38	0.23	-0.51	0.20	-0.50	-0.06	0.39	0.30	0.49	0.29	0.67	0.51	0.00072
Tyr-protein phosphatase non-receptor 11	PTPN11	26	197	-0.12	0.01	0.24	0.11	-0.18	-0.05	0.12	-0.08	-0.08	0.48	0.47	0.53	0.54	0.55	0.48	0.00001
Endoplasmic reticulum resident protein 29	ERP29	13	500	-0.26	0.44	-0.05	0.10	0.30	0.57	-0.09	0.43	0.02	0.57	0.70	0.70	0.81	0.87	0.48	0.00047
Elongation factor Ts, mitochondrial	TSFM	11	84	0.31	-0.04	0.35	0.34	0.69	0.34	0.55	0.22	0.51	0.90	1.19	1.06	0.79	0.70	0.47	0.00097
Exosome complex component RRP4	EXOSC2	12	69	0.69	0.68	0.60	0.61	0.62	0.66	0.70	0.35	0.76	0.43	0.27	0.35	0.26	0.18	-0.29	0.00032
UPF0696 protein C11orf68	C11orf68	8	77	0.17	0.61	0.59	0.50	0.59	0.50	0.50	0.41	0.33	0.27	0.17	0.10	0.16	0.16	-0.30	0.00001
Major centromere autoantigen B	CENPB	12	97	0.84	0.78	0.62	0.61	0.36	0.49	0.44	0.40	0.45	0.17	0.19	0.06	0.36	0.08	-0.30	0.00088
DNA topoisomerase 1	TOP1	37	520	0.17	1.15	0.74	0.80	1.04	0.95	0.76	0.80	0.81	0.50	0.64	0.42	0.53	0.43	-0.31	0.00014
Zinc finger CCHC domain-containing protein 7	ZCCHC7	4	12	0.28	0.44	0.52	0.48	0.51	0.74	0.57	0.54	0.50	0.29	0.35	0.07	0.16	0.13	-0.31	0.00088
Mesoderm induction early response protein 1	MIER1	10	45	0.33	0.97	0.78	0.97	0.78	0.86	0.62	0.78	0.71	0.34	0.59	0.38	0.47	0.36	-0.34	0.00012
Vesicle-fusing ATPase	NSF	42	609	0.31	0.86	0.96	1.00	0.53	0.56	0.83	0.47	1.10	0.34	0.37	0.19	0.04	0.32	-0.41	0.00034
Estradiol 17-beta-dehydrogenase 11	HSD17B11	9	50	0.35	0.60	0.39	1.01	0.53	0.41	0.54	0.73	0.08	-0.10	0.16	0.11	-0.14	-0.02	-0.42	0.00065
Glycosaminoglycan xylosylkinase	FAM20B	5	20	0.56	0.36	-0.08	0.00	0.22	-0.19	0.07	-0.19	-0.22	-0.67	-0.46	-0.36	-0.56	-0.38	-0.43	0.00034
Coiled-coil domain-containing protein 117	CCDC117	5	18	0.92	0.87	0.81	0.97	1.21	1.02	0.87	0.56	0.96	0.42	0.52	0.21	0.27	0.27	-0.47	0.00004
Zinc finger protein 64 homolog	ZFP64	2	10	-0.11	1.24	0.86	1.01	1.55	1.04	1.00	0.80	1.11	0.53	0.61	0.25	0.36	0.55	-0.47	0.00013
Prenylcysteine oxidase 1	PCYOX1	11	66	1.43	1.17	0.61	0.96	0.64	0.69	0.72	0.33	0.62	0.07	0.36	0.03	-0.01	0.12	-0.47	0.00018
Glycerol-3-phosphate dehydrogenase, mt	GPD2	37	341	0.65	0.22	0.02	0.36	0.80	0.59	0.06	0.31	0.13	-0.50	-0.35	-0.40	-0.31	0.00	-0.50	0.00056
HLA class II antigen, DR beta 3 chain	HLA-DRB3	3	149	0.13	-0.07	0.12	0.55	0.60	0.21	0.40	-0.21	0.23	-0.27	-0.54	-0.36	-0.33	-0.45	-0.53	0.00028
Adenylate kinase isoenzyme 1	AK1	8	111	-0.91	1.72	0.57	1.19	2.03	1.14	1.29	-0.03	0.78	-0.27	0.14	-0.82	-0.69	-1.20	-1.03	0.00046

Figure 7.27. **Subtype specific protein expression in trisomy 12 CLL samples. A.** Volcano plot describing the differential regulation score (ΔRS) plotted against the $-\log_{10}$ (p-value) of differential regulation between trisomy 12 and disomy 12 samples. The plot highlights those proteins for which the regulation score calculated for the samples was >0.25 (red) and < -0.25 (green). **B.** The $\log_2$ (ratios) for the top 15 significantly overexpressed and top 15 underexpressed proteins in trisomy 12, detailing the sample expression, the ΔRS and the uncorrected p-value estimating the probability of subtype-specific expression between the 5 trisomy 12 samples and 8 disomy 12 samples (4621 was excluded from the analysis as an outlier).

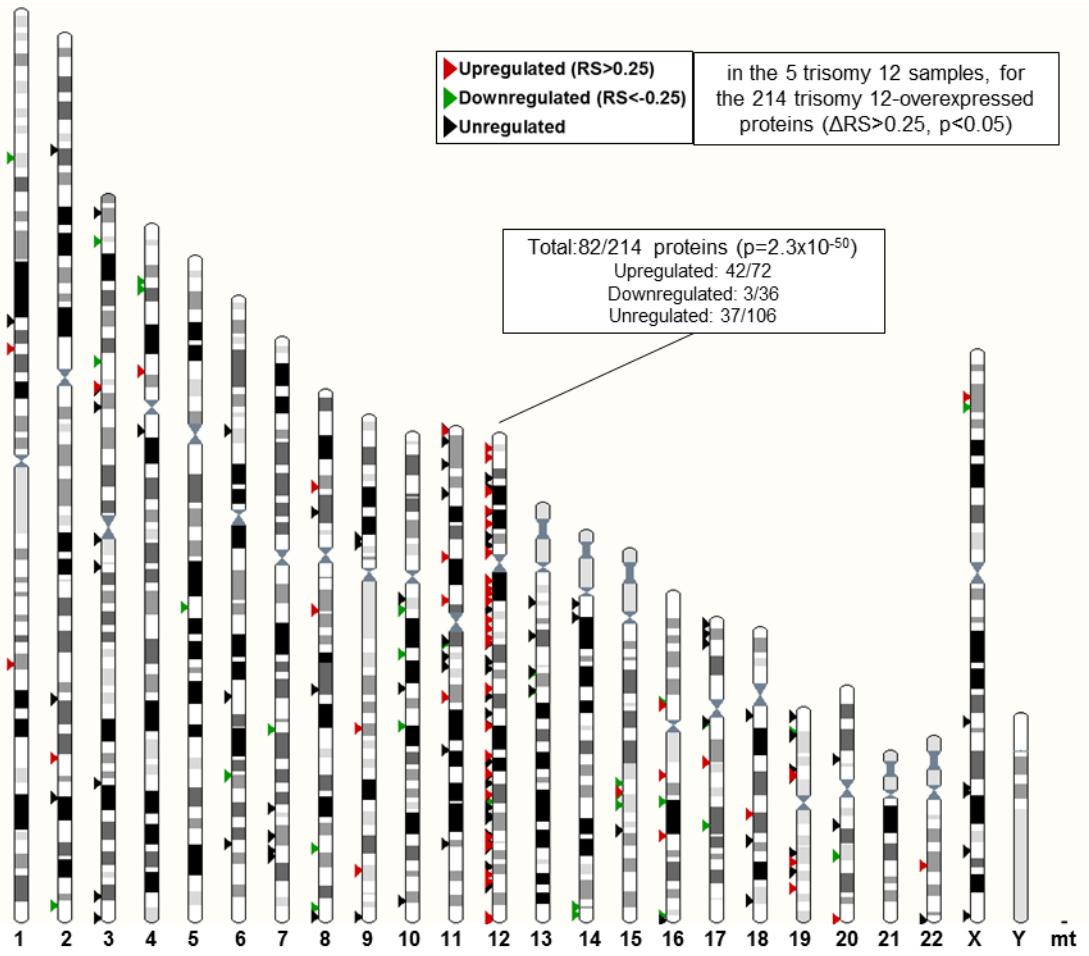


Figure 7.28. **Chromosome mapping of trisomy 12 CLL-specific proteins.** Chromosome mapping of the 214 proteins with subtype-specific overexpression in trisomy 12 cases (ΔRS>0.25, p<0.05), separated into those proteins demonstrating upregulation (RS>0.25, red), downregulation (< -0.25, green) and no regulation (0.25>RS> -0.25, black) calculated for the 5 trisomy 12 samples relative to HD samples. The mapping of these proteins to chromosome 12 is described.

7.21 INTER-PROTEOME COMPARISON BETWEEN HUMAN CLL AND CANCERS OF THE Eμ-*TCL1* MOUSE MODEL

The Eμ-*TCL1* mouse has been proposed as potential model of CLL, mimicking such features as the slow progression of a $CD5^+$ leukaemia and the development of tumour populations in the bone marrow, spleen and lymph nodes. The comparison of the quantitative CLL proteome described in this chapter with the Eμ-*TCL1* proteome described in **Chapter 5**, presented an opportunity to identify similarities and differences emerging between the two cancer types at the level of individual orthologous protein expression. While this comparison was between splenic B cells in mice and circulatory B cells in humans, it was hoped that the strongest biological trends would still be apparent. Any emerging signatures would potentially be indicative of proteins which played a pivotal role in $CD5^+$ B-cell cancers, based on the evolutionary conservation of B-cell biology, capable of facilitating neoplasia development and malignancy.

The UniProtKB entry names were used to align between the human and murine proteomes, which determined that a total of 4720 proteins were commonly identified in both experiments (**Figure 7.29A**). To determine whether or not the alignment of the two proteomes provided an accurate inter-proteome comparison, the number of PSMs, a very approximate measure of the abundance of protein within a proteome, for each protein was plotted (**Figure 7.29B**). This identified a correlation between the two proteomes with an R^2 value of 0.598, with the majority of proteins being detected with no more than a 10-fold difference in their approximate abundance between orthologs.

To define the number of commonly regulated proteins, the threshold described in **Section 5.7** and **Section 7.3**, of an RS exceeding ±0.5 or ±0.25, for Eμ-*TCL1* and CLL, respectively, with an FDR-corrected p-value of less than 0.05 was used (**Figure 7.29C**). From the orthologs common to both proteomes, this identified 664 and 656 upregulated proteins for Eμ-*TCL1* and CLL, respectively. A total of 65 overexpressed proteins were common to both CLL and the Eμ-*TCL1* tumour cells; 9.8% of the Eμ-*TCL1*-upregulated proteins and 9.9% of the CLL-upregulated proteins. 35 proteins exhibited common downregulation; 23.3% of the 150 Eμ-*TCL1*-downregulated proteins and 6.5% of the 537 proteins underexpressed in CLL.

The regulation scores for the 4720 proteins identified with comparable orthologs were plotted to demonstrate the correlation between CLL and Eμ-*TCL1* tumours (**Figure 7.30**). This identified no overall correlation between the two cancers, with an R^2 value of less than 0.01. Correlations between CLL subtypes and the Eμ-*TCL1* tumours, or against the Eμ-*myc* tumours, did not improve upon this correlation coefficient. Despite this lack of an overall trend, the

comparison of CLL to Eμ-*TCL1* quantitations identified several proteins with consistent expression patterns. The most significantly upregulated protein identified in CLL, the anti-proliferative factor receptor CKAP4, (RS=1.4, $p=1.8x10^{-8}$) also demonstrated significant upregulation in Eμ-*TCL1* tumours (RS=1.07, p=0.02), with more than 250 PSMs and 25 unique peptides in both proteomes. Tyrosine-protein phosphatase non-receptor type 22 (PTPN22) has been shown to negatively regulate T-cell receptor signalling, by the dephosphorylation of proteins which may also regulate BCR signalling in CLL, such as ZAP70 [499]. PTPN22 was the ninth most upregulated, commonly identified ortholog in Eμ-*TCL1* (RS=1.68, p=0.004) and also exhibited significant upregulation in CLL (RS=0.64, $p=1.7x10^{-6}$). CD5, a defining marker of both tumours was detected with similar overexpression in CLL (RS=0.84, $p=8.6x10^{-6}$) and Eμ-*TCL1* tumours (RS=0.98, p=0.04).

Two highly upregulated proteins in CLL, the cell cycle control protein cyclin-dependent kinase 14 (CDK14) (RS=1.14, $p=1.4x10^{-8}$) and zinc finger protein ZFPM1 (RS=0.91, $p=2.4x10^{-7}$), were also upregulated in Eμ-*TCL1* tumours but to a lesser extent (RS>0.5, p<0.05). A cluster of five proteins was observed with substantial upregulation in Eμ-*TCL1* tumours (RS>1) and consistent upregulation in CLL (RS>0.4); the cell cycle regulator WEE1, the B-cell survival regulator TNF receptor-associated factor 3 (TRAF3), the NOTCH signalling repressor ataxin-1 (ATXN1), the minimally characterised 629 kDa structural protein neuroblast differentiation-associated protein AHNAK and the uncharacterised protein GRAM domain-containing protein 1B (GRAM1B). Two uncharacterised proteins arf-GAP domain and FG repeat-containing protein 2 (AGFG2) and UPF0544 protein C5orf45 were identified with consistent and equivalent upregulation in both cancers.

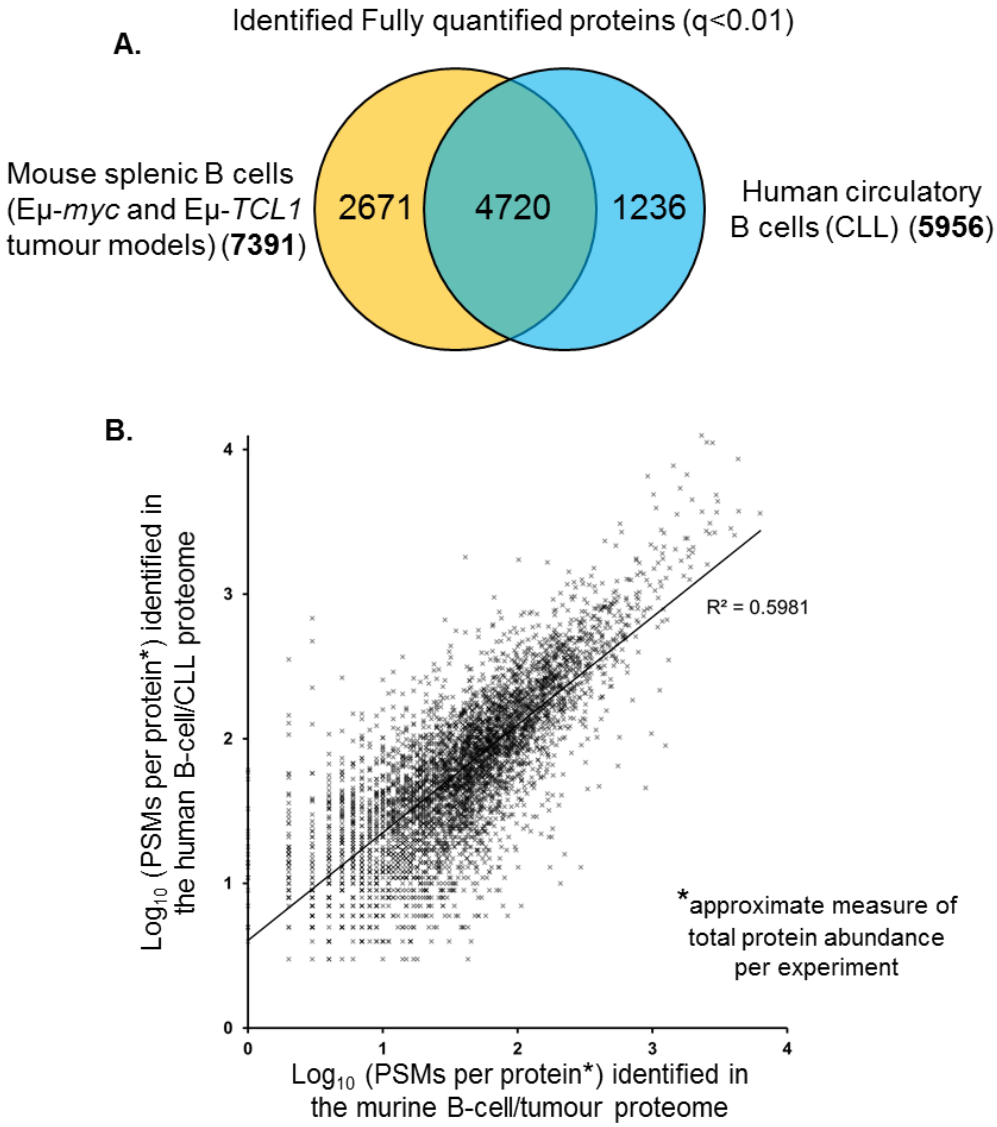
Amongst the full list of 65 proteins determined to be significantly upregulated in CLL and the Eμ-*TCL1* tumours, 19 proteins mapped to the mitochondria, a 5-fold enrichment ($p=5.6x10^{-7}$). 11 of these proteins were enriched (14-fold) to the matrix and lumen ($p=5.3x10^{-7}$). Other noteworthy proteins included the upregulation of two DNA replication licensing factors MCM3 and MCM5, a second PTPN protein involved in lymphocyte signalling, PTPN7, the minimally studied anti-apoptotic protein BCL7C, the BCR-inducible transcription regulator NFATC1 and another transcription regulator, reported to affect B-cell differentiation and Ig expression TCF3. 28 of these 65 proteins (43%) also had significant upregulation in the Eμ-*myc* tumours (RS>0.5, p<0.05).

One of the most downregulated CLL proteins, DENND3 (RS= -1.64, $p=7.6x10^{-9}$), a regulator of endosome transport and recycling, was also significantly downregulated in Eμ-*TCL1* tumours (RS= -1.37, p=0.02). Ig delta chain C region (IGHD/IgD), the predominant surface Ig expressed on peripheral B cells, also demonstrated downregulation in both cancers,

with a far more prominent underexpression in Eμ-*TCL1* tumours (RS= -2.46, p=0.01) than CLL (RS= -0.42, p=0.01). Another commonly downregulated Ig protein was Ig alpha-1 chain C region (IGHA1), which demonstrated greater underexpression in CLL than in Eμ-*TCL1* tumours. The protein with the most consistent downregulation in both tumours was special AT-rich sequence-binding protein 1 (SATB1), a global gene regulator (RS< -1). Four proteins also demonstrating similar downregulation in both cancers, included; sortilin-related receptor (SORL1), CD38, CD22 and fascin (FSCN1) (RS<-0.7). Two interferon-induced proteins with tetratricopeptide repeats (IFITs) were identified amongst the proteins downregulated in CLL and Eμ-*TCL1* tumours, IFIT2 and IFIT3 (RS< -0.5), typically proteins involved in antiviral response. Two B-cell CLL/lymphoma proteins BCL2L1 and BCL7A, were also in this list. The list also included 6 immune regulators, CD38, CD40, CD47, CD55, SAM domain and HD domain 1 (SAMHD1) and dipeptidyl-peptidase 4 (DPP4) and 6 cell adhesion molecules CD22, CD47, CD72, CD84, ITGB5 and selectin L (SELL).

The 35 proteins significantly downregulated in CLL and Eμ-*TCL1* tumours were significantly enriched for membrane proteins ($p=7.35x10^{-5}$), with 26 (74%) of the proteins annotated as such. Another trend amongst these proteins was the observation that 28 (80%) also demonstrated significant downregulation in the Eμ-*myc* tumours.

Figure 7.29. **Comparison of murine and human B-cell cancer proteomes. A.** The common and uniquely identified protein orthologs in the fully quantified murine and human B-cell cancer proteomes. **B.** The number of PSMs (plotted as $\log_{10}$ (PSM number)) for each protein for the orthologs common to both proteomes, as an approximate indicated of protein abundance. **C.** The overlap of common orthologs with significant up or downregulation in CLL and Eμ-*TCL1* tumours.


A.
Identified Fully quantified proteins (q<0.01)
Mouse splenic B cells (Eμ-myc and Eμ-TCL1 tumour models) (7391)
2671
4720
1236
Human circulatory B cells (CLL) (5956)
B.
Log10 (PSMs per protein*) identified in the human B-cell/CLL proteome
Log10 (PSMs per protein*) identified in the murine B-cell/tumour proteome
R² = 0.5981
*approximate measure of total protein abundance per experiment
0
1
2
3
4
C.
Significantly upregulated proteins (p<0.05)
Eμ-TCL1 tumours (RS>0.5) (664)
CLL (RS>0.25) (656)
599
65
591
Significantly downregulated proteins (p<0.05)
Eμ-TCL1 tumours (RS<-0.5) (150)
CLL (RS<-0.25) (537)
115
35
502

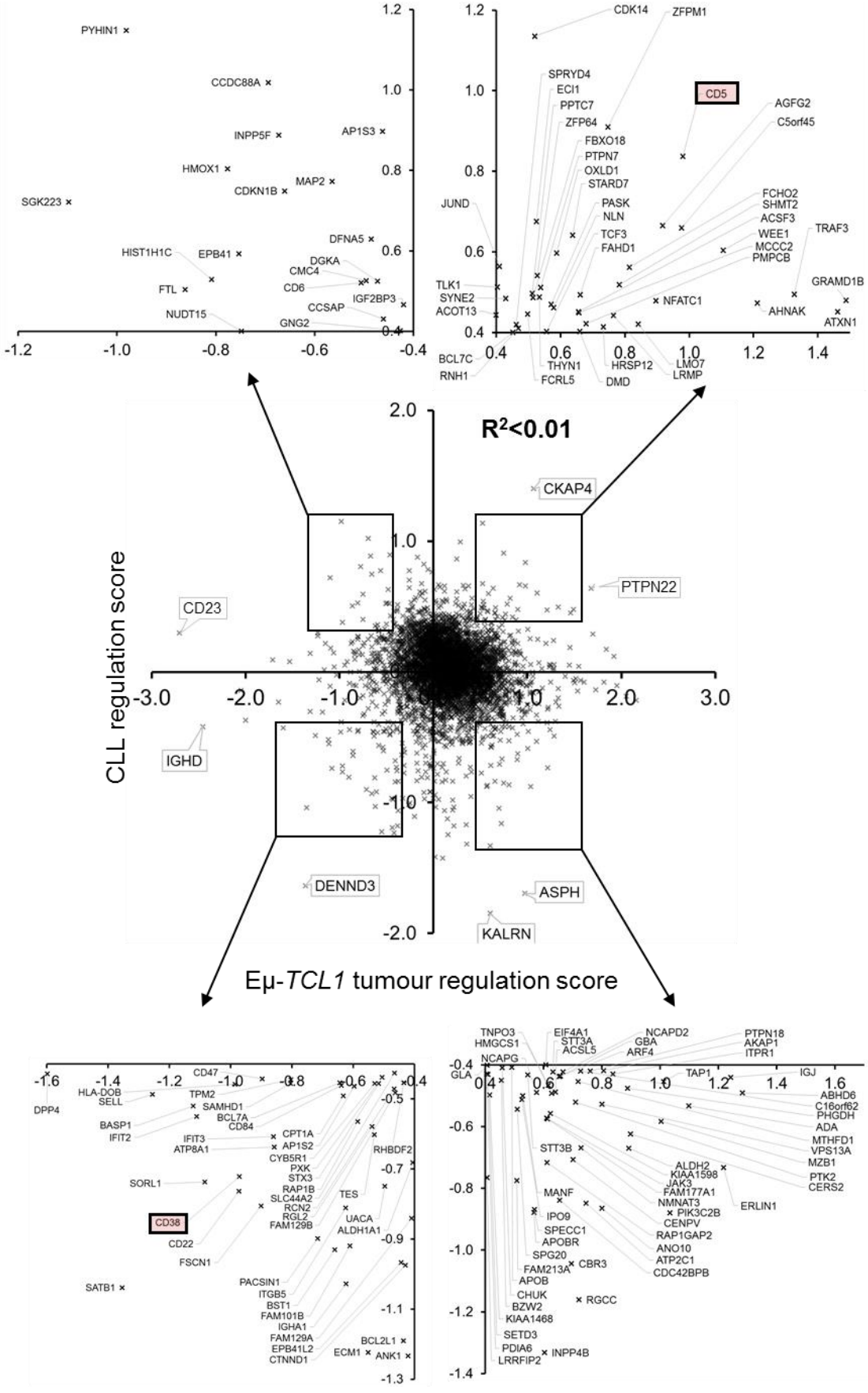

R²<0.01
CKAP4
PTPN22
CD23
IGHD
DENND3
ASPH
KALRN
CD5
CD38
CLL regulation score
Eμ-TCL1 tumour regulation score

Figure 7.30. **Quantitative comparison of CLL and Eμ-*TCL1* tumour proteomes.** Comparison of the regulation scores relative to healthy, non-tumour B cells determined for CLL and the Eμ-*TCL1* tumours. The four correlating or inversely correlating quadrants are expanded and plotted above and below.

7.22 CHAPTER DISCUSSION

In light of the successful applications of proteomics, detailed in **Chapters 3-6**, this chapter aimed to reproduce this quality of results for the characterisation of primary human CLL samples. Compared with other proteomics analyses of CLL, this characterisation has exceeded all other studies in the number of protein identifications and quantitations, therefore potentially offering the most comprehensive phenotyping of CLL produced to date.

The resulting CLL proteome fully quantified fewer proteins overall, compared with the B-cell proteome (**Chapter 5**). This was partly due to the lesser extent of fractionation, and peptide pooling needed to accommodate two full 2D-LC MS/MS 10-plex experiments in a finite amount of MS time. The lower protein quantification number is also attributable to the greater stringency implicated by the requirement that proteins be identified and quantified in the two discrete shotgun proteomics experiments (**Figure 7.3A**). It can therefore be concluded that a more carefully optimised 2D-LC MS/MS characterisation utilising more extensive fractionation and more MS time holds the potential to further increase the depth of characterisation described here.

The 2D-LC MS/MS characterisations from the two parallel 10-plex experiments were successful from a technical perspective, demonstrating highly reproducible RP LC traces (**Figure 7.1**), a substantial degree of overlap between proteomes (**Figure 7.3A**); which extended to the numbers of PSMs and peptides describing each protein (**Figure 7.3B**) and the reproducibility of protein quantitations between 10-plex A and B (**Figure 7.3C**). While the R^2 value was lower for the comparison between the $\log_2$ (ratios of HD2:HD1) of 10-plex A and B (**Figure 7.3C**), compared to HD3 comparisons, this was partly due to the high degree of similarity of the two B-cell samples. With fewer ratios above the threshold of noise, the R^2 value was less able to reflect a correlation. Those proteins with substantial fold changes were observed to be reproducible. Interestingly, the proteins exhibiting differential expression between HD1 and HD2 were HLA proteins, which are known for allelic inter-individual differences [500]. This detection therefore, in part, served as a positive control where this substantial differential expression was representative of the discrete presence or absence of polymorphic peptides in the samples. These values also suggested the threshold of noise for the vast majority of protein ratios

at around a $\log_2$ (ratio) of 0.3 (a fold change of 1.23). More extensive protein expression differences were observed in HD3 than the other HD samples which highlighted a more reproducible pattern of differential protein abundance resulting in more accurate R^2 values. The exclusion of HD3 from the analysis was considered, however, it was decided that in combination with the minimum (least deviated) value approach that this would help to preserve only the most CLL-specific signatures. No substantial pattern emerged to suggest a cause of the differences observed for HD3.

The reproducibility of the HD bridging control ratios and protein identifications confirmed the potential of inter-proteome quantitative comparison. While a typical approach to comparing between the two 10-plexes would have been to formulate a single average to all three HD controls, for each CLL sample, this caused a clustering between the two experiments when performing hierarchical clustering. This demonstrated that the proteome-specific differences, such as variable effects of ratio compression were apparent within the quantitations. By selecting the minimum deviated ratio, of those derived from the three HD samples, this experimental clustering was abolished (**Figure 7.4**). This approach also offered a means of conserving only the most reproducible results, providing confidence that each presented $\log_2$ (ratio) was consistently representative of all three ratios, not a single outlier point causing an increased average. This also reduced the effects of quantitative background noise such as that observed in **Figure 7.3C**.

The success of the quantitative proteome characterisation was also reinforced by the observation of several expected biological characteristics amongst the CLL samples. Positive controls such as CD5 and BCL2 upregulation were observed highlighting the successful comparison between CLL and healthy donor B-cell samples (**Figure 7.6**). Anticipated inter-sample differences confirmed the capability of the results to profile individual samples and conclude subtype specific differences, such as the specific expression of CD38 in the 99% $CD38^+$ CLL cases and CD49d expression in the trisomy 12 CLL cases (**Figure 7.6**). A further observation was that of the highly specific detection of two Y-chromosome-encoded proteins exclusively in male CLL cases. These observations, in combination with several other anticipated findings detailed throughout this chapter, confirm that the sample procurement and processing, the 2D-LC MS/MS proteomics analysis, the data processing and the quantitative analyses adopted have provided an accurate representation of the phenotypes of CLL, relative to both the HD controls and to the other CLL samples and subtypes, from which further hypotheses about CLL biology can be drawn.

The initial analysis of the data also revealed some limitations to the approach such as the observation of platelet contamination. The contamination, however, was restricted to the HD

control samples which only limited conclusions that could be drawn relating to the proteins downregulated in CLL. The contamination was filtered from the analysis using a previously characterised platelet proteome [471], which eliminated many of the proteins artefactually appearing as substantially 'downregulated'. However, proteins such as CD14, a marker of other immune cells remained, suggesting the possibility of marginal contamination from other PBMCs. These contaminations were no doubt related to the vastly increased volume of blood required to isolate equivalent numbers of B cells from healthy donors compared to CLL patients. In addition to this, thrombocytopenia, describing a lower number of platelets, is also characteristic of CLL further exacerbating this disparity. Because of this, conclusions were drawn more cautiously in relation to the proteins determined to be downregulated.

Another anomaly in the data was that of CLL sample '4621' identified by the outlying results observed in the hierarchical clustering (**Figure 7.4B**) and ratio distribution (**Figure 7.4C**). The clustering of this sample was separated from all other CLL samples suggesting that either a substantial technical or biological difference was present. No trends were immediately apparent explaining this outlier, however sample handling differences, such as a greater contamination from non-CLL cells may explain this observation. Sample '4621' could also possibly be representative of a different disease. CLL sample '4621' was therefore excluded from any further analyses, but the results for this sample were presented wherever possible.

It should also be noted that the circulatory B cells used as a non-cancer control were not ideal. A pan-B-cell isolation from each donor may have highlighted protein expression differences that were attributable to the B-cell type from which CLL arises from, rather than specifically as a result of transformation to cancer, such as CD5. However, pan-B-cell samples provided the best available option given the low percentage of B-cell subtype cells that can be isolated from PBMCs, comparable to the circulatory CLL cells. Additionally, uncertainty remains as to the specific CLL cell of origin; and therefore most ideal healthy control. Despite this limitation, any therapeutic targets identified pertaining to the CLL cell of origin, would still offers a selective CLL target, which would avoid the elimination of the majority of healthy B cells.

One of the most striking topological observations in the data analysis was that, despite variation of subtypes and CLL sample origin (**Figure 7.4A**), a strong trend emerged from hierarchical clustering demonstrating a consistent direction of regulation for the majority of proteins in CLL, relative to HD samples (**Figure 7.4B**). While many proteins demonstrated heterogeneity, these variable proteins each appeared either upregulated and unregulated, or

downregulated and unregulated across the CLL samples, relative to the HD controls. This suggests that a CLL-specific protein signature exists that commonly and consistently defines the potential of up or downregulation of each protein compared to healthy B cells. Homogeneity, similar to this, has also been consistently observed in those studies evaluating CLL at the gene expression level [71, 72]. While several factors, such as CLL subtype, appear to influence the protein expression profile, it appears that these subtypes may be either amplifying or suppressing characteristics of B cells to promote a CLL phenotype. It also suggests that alternative pathways and mechanisms may exist to achieve the same acquisition of cancer traits through convergent evolution, which has also been highlighted at the genetic level in CLL [501]. This is consistent with the principle that CLL arises without a single causative mutation, such as myc translocations in BL, and can exist without the canonical genomic or genetic aberrations frequently observed in CLL. Overall, this finding of a subtype-independent CLL signature disputes the notion that different subtypes of CLL constitute distinct diseases [502].

The proteomics experiments performed in this chapter provided a successful means of characterising surface-expressed transmembrane proteins in CLL. The study identified several novel candidates with potential for investigation as immunotherapy targets (**Figure 7.8B**). The validity of these novel targets is reinforced by the previously characterised cell surface proteins within the list and is additionally reinforced by the independent observation and quantitation in both TMT 10-plex experiments.

CKAP4 was not only the most CLL-upregulated protein with annotated cell surface expression, but was also the single most CLL-upregulated protein identified in this proteome (RS=1.4, $p=1.8x10^{-8}$). The additional high-frequency of unique peptides (34) and PSMs (501) substantially reinforced this observation. This finding is novel at the protein level and only a single study has discussed upregulation of CKAP4 mRNA [503]. Also known as CLIMP-63, the cell surface expression of CKAP4 is well documented in relation to its function as a receptor for molecules including; tissue plasminogen activator (tPA) [504], surfactant protein A (SP-A) [505] and APF [506].

The protein is also recognised to be predominantly localised to the rough ER as a microtubule-binding protein with a role in ER sheet thickness and additional roles in regulating the miRNA processing protein ‘dicer’, translocon assembly and ribosome binding [507-510].

APF treatment of a bladder cancer cell line resulted in reduced proliferation, attributable to substantially reduced phosphorylation of AKT and GSK3β and an increased expression of p53

[511]. CKAP4 has additionally been shown to reduce proliferation by the suppression of epithelial growth factor receptor (EGFR) signalling [512].

The novel observation of substantial CKAP4 protein upregulation in CLL, suggests that the targeting of this pathway might offer a means of reducing the proliferation of CLL. The observation that CKAP4 is also upregulated in the tumours of the Eμ-*TCL1* mouse, but not Eμ-*myc*, reinforces a mechanistic role in $CD5^+$ B-cell tumours as well as providing a means of evaluating this hypothesis *in vivo*. The exact role and potential for immunotherapy are difficult to conclude however, given that CKAP4 cellular localisation and functions appear variably reported. It does however appear likely that CKAP4 may function in a growth suppression role, potentially as a negative feedback response to increased proliferation. Evaluation of CKAP4 localisation in CLL and the response to various stimuli will be critical in confirming any potential clinical benefits that can be exploited from CKAP4 overabundance in CLL.

Amongst the other novel cell surface proteins identified with significant upregulation in CLL was a trend relating to BCR function or B-cell biology. The minimally studied protein LAX1 was shown to have an inhibitory role in B- and T-cell signalling in lymphocytes [475] and ATP2B4 may have a role in BCR-induced calcium efflux [476]. Prolectin, also known as CLEC17A, is a germinal centre B-cell-expressed protein whose expression correlates with proliferation [513]. Despite this protein being a marker of B-cell proliferation, no previous observations have been made as to its upregulation in B-cell cancers. Prolectin also has a putative role in BCR signalling, demonstrated to have a rapid and substantial association with BLNK upon BCR stimulation [514]. Prolectin, LAX1 and ATP2B4 not only suggest novel specific target of immunotherapy, but also present a means of interfering with and better understanding BCR signalling in CLL.

Of the 4 Fc receptors identified as upregulated on CLL, FCER2/CD23, FCRL2/CD307B and FCRL5/CD307E, were previously reported findings, while PIGR was a novel observation. PIGR is an Fc receptor for IgA and IgM which facilitates transcytosis of antibodies, typically through epithelial cells. Upon completion of transport, the extracellular portion, known as the secretory component, is cleaved to release the antibody, leaving a 35 amino acid portion of the extracellular domain [515-517]. While PIGR upregulation has been implicated in a number of solid tumours, [518-523], it has not been characterised in any leukaemias or lymphomas, suggesting this finding is novel. Interrogation of the peptide level data suggested that only peptides matching to the intracellular portion of PIGR were identified. This suggested that the secretory component of PIGR is cleaved, at least sufficiently for any matching peptides to fall below the threshold of detection. The implication of such a cleavage would be that the transcytosis action of PIGR is functional; an unusual observation given that leukaemia cells do

not have either a basolateral or apical surface for antibodies to be transcytosed between. Regardless of the cleavage and functionality, validation of PIGR as a cell surface marker of CLL could present a clinical tool, even if just the 35 amino acid portion remains on the surface of cells. Further investigation is also warranted into the function of PIGR on CLL.

The proteomics experiment performed in this chapter also provided a range of proteins upregulated in CLL, indicative of cancer-promoting mechanisms which could be targeted by inhibitors (**Figure 7.9**). Again, this list included novel targets, corroborated by previously identified targets, such as BCL2 and LCK.

The upregulation of both HMOX proteins (HMOX1 and HMOX2) suggested, in addition to potential therapeutic targets, a critical requirement for an increase in free heme degradation in CLL metabolism. This is consistent with the observation that CLL exhibits a high level of oxidative stress [524], and that free heme accumulation can induce apoptosis. It was previously demonstrated that HMOX1 was the most upregulated gene in response to a thioredoxin reductase-induced increase in reactive oxygen species, and inhibition of HMOX1 synergistically increased apoptosis [525]. However, this, and other similar analyses have failed to consider HMOX2. These finding suggest that any attempts at inhibiting HMOX1 should be attempted with compounds which also inhibit HMOX2, such as tin mesoporphyrin which may otherwise act to circumvent any specific means of HMOX1 inhibition. They also reinforce the notion that inhibition of the HMOX proteins has therapeutic potential in CLL, which is yet to be fully explored.

Another trend amongst these targets of inhibition was that of overexpressed HDACs. HDACi have been trialled in CLL, however have not been adopted as a means of treatment, due to poor response and tolerability [526-528]. The use of pan-HDACi such as SAHA have likely contributed to these limitations, with fewer off-target effects. More targeted therapies promise greater efficacy such as the HDACi, entinostat, which has a high degree of specificity to HDAC1 and HDAC3, both identified upregulated in these results. Entinostat induced several proapoptotic effects on CLL [529]. To date, only mRNA level gene expression data has been published relating to global HDAC expression in CLL, which also identified HDAC7 as the most upregulated HDAC, relative to HD controls [527]. Overall, this finding of HDAC7 protein overexpression constitutes a novel observation in CLL and suggests that targeted strategies to HDAC7 may have unexplored clinical potential with fewer off-target effects. No HDAC7-specific inhibitors appear to have been successfully characterised however.

Global analysis of the trends exhibited amongst the differentially regulated proteins in CLL highlighted a substantial pattern indicating the overexpression of nuclear proteins (**Figure 7.10A**). This trend could be more specifically accounted for by proteins annotated as transcriptional regulators (**Figure 7.10B**) and GO term enrichment highlighted an equivalent overrepresentation of transcriptional repressors and activators (**Figure 7.11**). Pathway enrichment analysis detailed specific means by which transcriptional regulators were differentially expressed in CLL, describing the upregulation of transcriptional repression by epigenetic means and transcriptional initiation by the pre-initiation complex (**Figure 7.16** and **Figure 7.18**).

Another substantial trend, pertaining to this nuclear protein enrichment, was that of the global upregulation of the spliceosomal complex (**Figure 7.11**). Subsequent pathway analysis identified the vast majority of the spliceosome components exhibiting overexpression in CLL (**Figure 7.13**). In addition to spliceosome upregulation, the pathway describing mRNA polyadenylation, and subsequent stabilisation was also enriched (**Figure 7.15**)

The combined global analysis of protein upregulation demonstrates that CLL presents substantially altered transcriptional and post-transcriptional processes compared with HD controls. This is in agreement with many of the transcriptome-level analyses conducted characterising CLL, describing major alterations to mRNA expression [71, 72, 105, 530, 531].

The role of aberrant splicing in CLL is highlighted by frequent mutations to the *SF3B1* gene [96]. Recent deep transcriptome analysis confirmed that CLL demonstrates vast dysregulation in splicing patterns, relative to HD B cells [531]. The proteomics results detailing spliceosome protein overexpression (**Figure 7.13**) therefore offered an explanation as to how CLL cells induce these aberrations. This novel observation of global spliceosome upregulation strongly suggests that methods of interfering with the spliceosome reducing overall splicing could offer a means of treating CLL.

Trends relating to downregulated proteins (**Figure 7.10A**) further highlighted the overall underexpression of cell surface proteins seen in **Figure 7.8**, confirming the patterns seen in transmembrane receptor and ion channel downregulation (**Figure 7.10B**). **Figure 7.12** detailed this reduced expression of proteins relating to cellular migration, immune system interactions, cell surface expression and signal transduction. These findings were consistent with previous observations suggesting that CLL in the circulation have impaired transmigration capabilities [532]. The proteomics data presented in **Figure 7.19** and **Figure 7.20,** describing the downregulation of integrin signalling and leukocyte transmigration, therefore details several specific mechanisms which may explain the circulatory phenotype of this subpopulation of CLL

cells. To more accurately conclude these mechanisms, however, comparisons would need to be made between CLL cells and B cells isolated from the lymph nodes, bone marrow and spleen.

Substantial dysregulation was observed amongst the proteins of the BCR signalling pathway, and while the majority of proteins appeared downregulated, specific downstream signalling molecules were upregulated (**Figure 7.21**). Overexpression of downstream transcription regulators suggested a rewiring of the BCR pathway enabling CLL cells to modulate or potentially circumvent their dependency on BCR complex-induced upstream signalling. This is in agreement with the observation that proliferation in CLL occurs independently of external antigen-induced signalling [94]. This could also be described as the adoption of an anergic phenotype similar to that induced in B cells after antigen engagement, with a downregulation of surface IgM, and lack of BCR responsiveness.

NFAT transcription factor activity, observed upregulated in CLL (**Figure 7.21**) is proposed to contribute to the establishment of an anergic phenotype [533]. Additionally the inositol phosphatase INPP5D/SHIP1 has been shown to act as a negative regulator of BCR signalling and is activated in anergic B cells [534]. While INPP5D did not demonstrate differential regulation, a lesser-studied inositol phosphatase of the INPP5 family, INPP5F/SAC2, was confidently observed with significant overexpression in CLL, a previously unreported finding. Further investigation into the role of INPP5F in BCR signalling and its overall effects on CLL are warranted, given the confidence and consistency of these findings and its relatively poor characterisation to date. This finding is also partly counterintuitive, given that INPP5F is suggested to reverse the effects of PI3K signalling, thereby inhibiting the phosphorylation of AKT and upregulation of pro-survival/proliferation pathways. INPP5F is therefore poised to play a key role in the establishment of aberrant anergy and subsequently the undifferentiated B-cell phenotype observed in CLL.

While the overall trend was downregulation of upstream BCR proteins, in many cases this was only marginal or variable, which is consistent with the retained functionality of these BCR signalling pathways in most but not all cases of CLL, most frequently in U-CLL [535, 536]. Despite this, few BCR signalling pathway proteins demonstrated any differential expression between U-CLL and M-CLL. An exception was, however, IgM itself (**Figure 2.23**), which was overexpressed in U-CLL, relative to M-CLL. This is consistent with previous findings and the observation of greater antigen sensitivity in U-CLL [536].

RRAS2 was significantly upregulated in M-CLL, relative to both U-CLL and HD controls. This was of particular interest given that RRAS2/TC21 has been shown to be critical

for B-cell development and survival [537]. This suggests a key difference in the B-cell signalling pathways between U-CLL and M-CLL, with RRAS2 potentially supplementing survival signals in the more anergic M-CLL phenotype. This finding identifies RRAS2 signalling as a potential therapeutic target in M-CLL cases.

Despite the relatively large number of samples characterised, the evaluation of the data for subtype-specific trends was generally insufficient to confidently draw conclusions for individual proteins. However, trends amongst the data revealed several strong patterns of protein dysregulation suggesting unique characteristics of each CLL subtype. Overall analyses revealed that CD38 expression and trisomy 12 status produced the most distinct phenotype in CLL, which while overlapping, still highlighted discrete differences (**Figure 7.22**).

U-CLL presented a strong trend of mitochondrial protein upregulation compared with that of M-CLL. This included both mitochondrial transmembrane transporters and mitochondrial structural proteins. This potentially indicated that mitochondria were more abundant or more functional in U-CLL. Given that mitochondrial stress is far greater in CLL, than normal B cells [485], and that U-CLL adopts a more aggressive form of the disease, it follows that mitochondrial components may be upregulated in response to greater BCR signalling and proliferation in U-CLL.

Glycosylation appeared upregulated in U-CLL relative to M-CLL, which is consistent with previous observations of increased mannosylation of surface IgM in U-CLL [538]. These data therefore suggests specific proteins may be responsible for this phenomenon. The amplified ability of U-CLL to potentially upregulate glycosylation may have implications when considering immunotherapy as aberrant glycosylation may interfere with antibody binding to target antigens, such as that described for ROR1 in CLL [539].

U-CLL also demonstrated downregulation of antigen presentation and cell surface proteins, possibly reflecting a greater need for immune evasion, give the more aggressive nature of the disease, and reduced requirement of any pro-survival signals from transmembrane receptors.

The most prominent trend amongst the *NOTCH1*-mutant CLL samples was that of the subtype-specific overexpression of immune system-related proteins such as CD79a, Ig, granzyme A, CMIP and TLR7 (**Figure 7.24**). *NOTCH1*-mutant-specific overexpression of DTX3L and PARP9 suggested an upregulation of a DNA damage repair pathway. However, the

correlative expression of PARP14, combined with the potential of DTX3L and PARP9 to form a cell-survival promoting complex [492], suggested an alternative or additional pathway might be present. PARP14 has been shown to be a binding partner of STAT6, capable of protecting against apoptosis in IL4-treated B cells [540].

Together, these findings indicate that *NOTCH1* mutation and constitutive activation may induce a reorganisation of immunoregulatory signalling. This agrees with the altered immune cell lineages induced by the overexpression and deletion of the *NOTCH1* gene in mice [541]. The co-expression of the DTX3L-PARP9-PARP14 complex suggest *NOTCH1*-mutant CLL samples may have a greater responsiveness to IL4 and other immune-related stimuli.

Transcriptional regulation demonstrated a general trend in downregulation in *NOTCH1*-mutant samples, suggesting some specific mechanisms behind the alterations to gene and protein expression induced by NOTCH1 signalling.

SF3B1 mutation induced the least discrete protein expression profile amongst the CLL subtypes and few overall trends emerged amongst these proteins (**Figure 7.25**). This is surprising considering the adverse outcomes associated with *SF3B1* mutation and the frequency with which it is observed [542, 543]. It has been suggested that the acquisition of *SF3B1* mutations typically occurs at an advanced stage of CLL progression [544]. This implies that the effect of mutant *SF3B1* may be more subtle and specific than initiating mutations, while still sufficient to provide a selective advantage, potentially explaining this minimal signature.

The two most specifically upregulated protein in *SF3B1*-mutant CLL, compared to both HD samples and CLL WT for *SF3B1* were histone demethylase KDM4B/JMJD2B and arginine methyltransferase 1 (PRMT1), both proteins with a role in epigenetic regulation of gene expression [545, 546]. KDM4B has an additional role in the DNA damage response, offering a survival advantage in several cancers, especially in the context of cytotoxic therapies [547]. KDM4B has been shown to enhance proliferation and epithelial to mesenchymal transition in gastric cancers via its epigenetic mechanisms and potentially promotes genomic instability [545, 548-550]. A KDM4 inhibitor, NCDM-32B, has been shown to effectively reduce cell viability in breast cancer [551], and may therefore offer a therapeutic option with regards to *SF3B1*-mutant CLL cases.

PRMT1 expression has been shown to promote resistance to apoptosis in breast cancer [552] and has also been shown to have a role in epithelial to mesenchymal transition in hepatocellular carcinoma and breast cancer [553-555]. Arginine methylation by PRMT1 has been shown to enhance BCR signalling [556], suggesting a further mechanism by which *SF3B1*

mutation may drive CLL. Inhibitors of PRMT1 have been identified with antiproliferative effects in a variety of cancers [557].

Neither KDM4B nor PRMT1 have previously been implicated in CLL and these novel findings suggest mechanisms behind the observations of genomic instability, epigenetic dysregulation and poor clinical outcome in *SF3B1*-mutated CLL [558].

$CD38^+$ (99%) CLL cases demonstrated the most dysregulated subtype-specific profile consistent with the observation that CD38 expression correlates with a more proliferative disease. The correlation between CD38 and proliferation marker KI67 expression in CLL, demonstrating this, has previously been established [122] and both findings act as positive controls for this subtype analysis (**Figure 7.26)**.

The analysis also confirmed the proliferative nature of $CD38^+$ CLL by highlighting the subtype-specific expression of cell cycle promoting proteins. Importantly, these proteins were discretely expressed in the three $CD38^+$ CLL cases, and not in the 2 trisomy 12 CLL cases, with less than 50% $CD38^+$ expression.

The three $CD38^+$ CLL cases appeared more metabolically active with a greater expression of ribosome proteins, nuclear translocation proteins, a greater activity of vesicle movement and greater cellular adhesion capabilities. A large number of these proteins had expression that while greater than the non-$CD38^+$ CLL cases, was equivalent to HD B cells. The same was true for the majority of non-$CD38^+$ CLL downregulated proteins. This suggests that maintaining CD38 expression in CLL actually produces a phenotype more closely resembling healthy B cells; partly highlighted by a reduced topological expression pattern (**Figure 7.4)**. This poses the interesting question of what causes low CD38 expression in some CLL cases, especially given the proliferative advantage provided by CD38. Given that $CD38^+$ CLL cases do not express more CD38 than that of the healthy B cells, it seems possible that CD38 expression is not a characteristic which CLL can selectively upregulate. This may suggest an activated $CD38^+$ CLL precursor in these cases, or that further CD38 expression is ineffectual.

The observation of substantial, specific downregulation of adenylate kinase in $CD38^+$ cases suggested a mechanism between two metabolic regulators. AK1, however, reduces cellular ATP by transferring a phosphate to AMP forming 2 ADP molecules. Given that CD38 function is inhibited by ATP, this process would seem preferential to CD38 function. While inconclusive, this suggests that AK1 function, may represent a critical difference between $CD38^+$ and $CD38^-$ CLL metabolisms induced by CD38 expression, targetable by therapeutics.

The analysis of proteins downregulated in $CD38^+$ CLL cases compared to non-$CD38^+$ CLL cases suggested several mechanisms which appeared, as with the overexpressed proteins, with expression more similar to healthy, circulatory B cells in the $CD38^+$ CLL cases. These included lower expression of transmembrane receptors, MHC proteins, lymphocyte regulation, chromatin reorganisation and most interestingly; spliceosomal components. This implies that $CD38^+$ CLL cases show less dependency on spliceosome overexpression. Overall, the protein expression profile observed in $CD38^+$ CLL cases suggests that CD38 imposes a more proliferative phenotype with more features similar to healthy B cells, while still presenting a clear, subtype-independent CLL phenotype.

The subtype-specific overexpression of CD49d (ITGA4), ITGB7 and ITGAL (CD11a) provided positive controls for trisomy 12 CLL cases [89, 559] (**Figure 7.27**). IQSEC1/BRAG2/GEP100, has been shown to have various roles in promoting integrin expression and cancer cell migration [560-562]. Trisomy 12-specific overexpression of IQSEC1 strongly suggests a novel candidate for this increased surface expression of integrins and the lymph node-homing phenotype in trisomy 12 CLL cases.

Other novel protein findings highlighted a trend of trisomy 12-overexpressed proteins with roles in oncogenic mechanisms, such as telomerase transduction, p53 inhibition, B-cell activation and apoptosis inhibitors. Interestingly the majority of these proteins were encoded by chromosome 12, suggesting several candidates for specific evolutionary advantages posed by the acquisition of a third copy of chr12 in CLL. Chromosome analysis identified a strong enrichment of chr12 proteins specifically upregulated in trisomy 12 CLL cases relative to both disomy 12 CLL and HD controls (**Figure 7.28**). This confirms the principle that the third copy of chr12 promotes the upregulation of chr12 proteins and provides some explanation as to the advantages presented by the acquisition of chr12; a previously unexplained phenomenon.

The analysis comparing between CLL and the Eμ-*TCL1* mouse models, was conducted to evaluate any trends which emerged suggesting common mechanism in the human and mouse diseases. While key proteins such as CD5 correlated, the vast majority of proteins demonstrated no similarities between mouse and human $CD5^+$ B-cell cancers (**Figure 7.30**). While the overall protein abundances, measured very approximately using PSM numbers showed some extent of correlation (**Figure 7.29**), this was not reproduced in the observed quantitations for these proteins relative to their respective, non-tumour controls. It should first be noted that this analysis may not have effectively aligned proteins due to differences in protein grouping,

resulting from the issues of protein inference from peptides. It is also likely that differences are amplified given the isolation of cancer samples from different niches; peripheral blood for CLL and spleen for Eμ-*TCL1*-derived tumours.

Given the evolutionary differences, it is additionally possible that many proteins have marginally different roles exchanged between common isoforms, amplified in the context of cancer. HDAC and phosphatase expression, for instance, were common dysregulation in both tumours, however different isoforms were specifically upregulated, while downregulated in the opposing tumour type. The role of INPP5F, upregulated in CLL, while downregulated in Eμ-*TCL1* tumours may be explained by the role of INPP5B, upregulated in Eμ-*TCL1* tumours, while downregulated in CLL. This might be supported by similarities in GO terms, however, some of the strongest trends in CLL, such as the global spliceosome upregulation, were not seen in the Eμ-*TCL1* tumours. Overall, however, this suggests that at the protein level, CLL and the tumours emerging from the overexpression of TCL1 in mouse B cells are very different entities.

In summary, this chapter highlights the potential of the adopted proteomics approaches in the characterisation of human CLL samples to achieve disease specific, subtype specific and sample specific profiling of relative protein expression. The experiment, which is potentially the most comprehensive characterisation of CLL phenotype to date, has yielded both anticipated and novel findings pertaining to therapeutic targets as well suggestions of biological mechanisms driving CLL. It can therefore be concluded that the investigation of the results produced in this characterisation and the further proteomics characterisation of CLL promises to provide novel insight into CLL treatment options and a greater understanding of CLL biology.

8.0 FINAL DISCUSSION AND FUTURE DIRECTIONS

Studies characterising B-cell cancers prior to these investigations were limited by technological capabilities or only focused on the analysis of cultured, immortalised cell line material, not truly representative of an *in vivo* B-cell cancer phenotype [345, 352]. The application of some of the latest advances in MS proteomics to carefully procured primary B-cell cancer samples and controls has produced three data-rich characterisations detailing novel and anticipated biological and clinical findings. The combined characterisation of plasma and B cells in the Eμ-*myc* and Eμ-*TCL1* mouse B-cell cancers models provides one of the most comprehensive phenotypic characterisations of cancer models produced to date. The quantitative proteomics profiling of CLL has presented the most detailed description of a primary human B-cell cancer phenotype to date.

The data produced characterising FFE-isolated Eμ-*myc* tumour cells (**Chapter 3**) provided a starting point for demonstrating that quantitative MS proteomics could provide data which could be validated successfully using Western blotting. While the principle of FFE was not entirely compatible with proteomics given the low yields and extensive handling time of cells, it allowed a representative sample to be characterised which generally provided accurate validation. This suggested that the proteomics approaches combining isobaric tags and 2D-LC MS/MS gave a successful, albeit minimal characterisation. Given the limitations of the analysis,

it was concluded that further analysis, building upon the lessons learnt from this experimental design offered an opportunity to extend the depth of proteome characterisation.

Studies, including the characterisation detailed in **Chapter 3**, highlighted the issue posed by co-isolation in the compression of ratios when using isobaric tags for relative quantitative proteomics. Arguably ratio compression poses the most substantial disadvantage to iTRAQ and TMT quantitative proteomics. It was therefore considered that a means of predicting ratio compression at a PSM level might offer the opportunity to correct for this limitation. The approach described in **Chapter 4**, weighting isobaric tag quantitations on the basis of PSM features proved widely and consistently successful in reducing ratio compression in multiple data sets. It was therefore concluded that it would prove applicable to the biological characterisation presented in this thesis offering a substantial improvement to the majority of ratios.

To expand upon the work analysed in **Chapter 3**, alterations were made to the original experimental design, such as the additional characterisation of the Eμ-*TCL1* B-cell cancer model alongside tumours derived from Eμ-*myc* mice **(Chapter 5)**. Several improvements were offered by technical advances which proved highly successful, identifying over 9000 proteins and quantitatively profiling 7391. The quantitative results were equally successful with 36% of these proteins demonstrating regulation for the Eμ-*myc* tumours. Analysis of these data revealed anticipated protein identification and quantification, such as myc and TCL1, which acted as positive controls within the experiment. The findings of proteins upregulated in both tumours collectively provided a consistent trend of cancer-related proteins, even suggesting proteins with clear putative oncogenic functions, without any prior findings in the literature. Together, this suggests that these data may constitute a description of several novel cancer promoting mechanisms, which were identified alongside many known tumourigenic pathways. The analysis also proved successful in the identification of targets of small molecular inhibitors, both specific to each tumour and applicable to both tumours. The combination of novel upregulated oncogenic proteins and inhibition targets present numerous possibilities with regards to the application of small molecular inhibitors in the context of these, and potentially other cancers. These data also highlights inhibition candidates which, if equivalently observed in human cancers, provides confirmation of the potential of these models as a preclinical platform for compound testing.

The cell surface phenotypes of each tumour were also successfully captured, describing over 200 differentially expressed transmembrane proteins. Amongst these were both novel and anticipated proteins, suggesting several possibilities for further investigation as immunotherapy targets in B-cell cancers. One particularly interesting trend which emerged from the membrane proteins was that of the Eμ-*TCL1* tumour-specific upregulation of both subunits required for

signal transduction in response to IL5. In combination with the consistent findings of upregulation at the PSM level, this was strongly indicative of either a selective pressure for cells expressing, or an aberrant upregulation of, the IL5R. Evaluation of the expression of IL5RA confirmed this upregulation on Eμ-*TCL1* tumours and to understand any potential role of this IL5R upregulation, responses to IL5 were evaluated *in vitro*. After initial culturing with IL5 demonstrated a significant dose-dependent expansion of cell density, individual aspects of cell proliferation and survival were interrogated. Overall, this revealed that IL5 had a dose-dependent effect on promoting cell viability, cell cycle progression and proliferation. This finding was consistent with several previous findings of IL5-resposiveness in mouse lymphomas, however was a novel observation explaining the possible origin of Eμ-*TCL1* tumours [444, 445, 449, 450]. It seems very likely that the species-specific IL5R-expressing B-cell population in mice [448] are conservatively amplified by the overexpression of TCL1. A likely mechanism behind this is the amplification of AKT signalling by TCL1 inducing proliferation via molecules such as myc [211, 212]. A possible means of testing this hypothesis would be to treat 6-week old WT and non-tumourous Eμ-*TCL1* B cells with IL5 and observe the proliferative capacity of the $IL5R^+$ population in the Eμ-*TCL1* B cells. The inhibition of this signalling by PI3K inhibitors would further reinforce the role of TCL1 in the PI3K-AKT signalling pathway. Phosphorylated GSK3β and myc expression could also be evaluated to demonstrate any signalling downstream of AKT.

While this population of $IL5R^+$ B cells does not appear in humans, the function of the TCL1 in the AKT pathway may be interchangeable with other cytokine receptors. IL4R, for instance, exerts an analogous pro-survival role in human B cells, and has additionally been shown to activate the PI3K/AKT pathway in B-cell lymphoma [563]. This could offer an explanation of the effects of TCL1 correlating with more aggressive forms of CLL. It would be of interest to investigate any correlation between TCL1 protein expression and IL4 sensitivity in CLL.

While it will be important to ascertain the specific *in vivo* function of IL5 in Eμ-*TCL1* tumours, an additional benefit of these findings is the suggestion that culturing Eμ-*TCL1* tumours with IL5 *in vitro* will potentially aid the evaluation of their biology.

The characterisation of Eμ-*TCL1* and Eμ-*myc* tumours and controls, discussed above, was performed on pooled samples to accommodate the samples in a single isobaric labelling experiment. Ideally, this work would be expanded upon by the evaluation of individual samples to better understand the heterogeneity of these tumours, potentially identifying, for instance why certain Eμ-*TCL1* and Eμ-*myc* tumours progress more aggressively than others. There is also potential behind evaluating other mouse models of CLL to identify parallels and differences to

help conclude the model which most resembles CLL at the molecular level. Models such as the APRIL transgenic mouse [564], the ROR1 x TCL1 transgenic mouse [235], the Eμ-miR-29 transgenic mouse [236], SV40 T antigen-expressing IgH.TEμ mice 19332766 and IL5-overexpressing models [447] all warrant further investigation using proteomics methodologies.

Overall, given that IL5 responsiveness appears as a critical mechanism driving the Eμ-*TCL1* tumours, a trait not reproduced in CLL, it was concluded that proteomics of primary human CLL samples would be far more suitable for the purposes of biological characterisation and clinical target discovery. The comparison between CLL and the Eμ-*TCL1* mouse models, conducted in **Chapter 7**, later furthered this suggestion that at the protein level, CLL and the tumours emerging from the overexpression of TCL1 in mouse B cells are different entities.

The characterisation of plasma samples (**Chapter 6**) paralleling the analysis of B-cell tumours and controls described in **Chapter 5** offered substantial potential in understanding the systemic effects of B-cell tumourigenesis in each tumour model. The implementation of SuPrE-SEC for the enrichment of the low molecular weight sub-proteome presented an effective and streamlined strategy in the identification and quantification of protein signatures. While limited by some reproducibility issues, these were minimised by technical replicates and the use of the RS and statistical tests.

The terminal plasma proteome characterisation revealed an extensive pattern of cell lysis products, which proved traceable to the B-cell tumours themselves, based on PSM numbers. The cell lysis signature was most predominantly observed for the Eμ-*myc* tumours; an expected observation given the high rate of apoptosis in the model [182]. While the Eμ-*TCL1* tumours also had a strong signature of tumour lysis in their plasma, an additional Eμ-*TCL1* tumour plasma-dominant signature emerged, which suggested a more subtle mechanism of plasma proteins becoming upregulated. These included several proteins indicative of an immune response to tumours, tumour-secreted proteins and proteins with signatures suggestive of shedding from the cell surface of tumours, which were also indicative of possible tumour biomarkers. The previous observations of immune aberrations such as an increase in circulatory T cells may offer some explanation to this [224]. Biological inferences could also be made between the B-cell tumours and plasma, for instance, with a common trend of hyaluronan binding in both proteomes suggesting a possible mechanism of tumour-induced microenvironment interactions [433, 434].

This plasma characterisation is reinforced and cross-validated by many of the observations made in the B-cell tumour proteomics experiments. Despite these observation in

both proteomes, however, validation of the differential proteins abundances is required in individual samples. Moreover, functional validation will help better understand the applicability of the mechanisms identified.

The characterisation of pre-terminal Eμ-*TCL1* plasma samples also proved successful in the identification of terminally-overabundant protein signatures at an early stage of the disease. However, it must be noted that the proteins of this signature predominantly presented low numbers of PSM and small differential abundances, suggesting that the concentration in the whole, pre-terminal Eμ-*TCL1* plasma, prior to enrichment would have been very low. This both highlights the difficulties, but suggests the potential behind, the plasma proteomics approaches outlined in this analysis. To conclude meaning from such subtle signatures it will be important to evaluate many individual samples, to ensure that differential protein abundances are consistently detectable.

Overall, the analysis of tumour and pre-tumour plasma demonstrates biomarkers attributable to cell lysis, immune reaction, shedding, secretion and possibly microenvironment interactions, which were differentially present in the two tumour types, in relation to tumour aggression and phenotype. While biomarker discovery is less applicable to CLL, as the circulatory leukaemia cells provide a non-invasive means of diagnosis and prognosis, the analysis presents a proof-of-principle study that suggests applicability to the identification of cancer biomarkers for human solid tumour samples.

For such human experiments, this analysis highlights the need for rigorously controlled sample procurement, the use of individual sample characterisation and, where possible, the proteomics characterisation of paralleled tumour tissue samples. The combination of higher initial volumes of plasma, higher resolution size exclusion chromatography and greater sensitivity provided by the latest MS instrumentation almost certainly has the potential to provide a robust method of plasma characterisation.

After the observation of a lack of a response to IL5 and expression of IL5RA observed in CLL, like that seen with tumours of the Eμ-*TCL1* mouse (**Chapter 5**), it became apparent that to make clinically applicable observations, CLL itself, would need to be characterised. Given the consistent success of the other proteomes described in this investigation, an experiment was designed which would characterise individual, primary CLL samples (**Chapter 7**). This incorporated the use of bridging controls into two TMT 10-plexes, extending the comparative capacity of isobaric-labelled experiments, beyond those described in **Chapters 3-6**. The resulting 2D-LC MS/MS characterisation resulted in the most comprehensive proteomics

profiling of CLL to date. The accuracy of this characterisation was corroborated by both technical reproducibility and anticipated protein expression of proteins such as CD5, BCL2, CD49d and CD38. While limitations such as inter-experimental differences and platelet contamination existed, these were addressed with a conservative use of ratios and the proteomics-based filtering of platelet proteins, respectively.

The study successfully suggested several potential clinical targets of inhibition or immunotherapy. Several known targets of small molecular inhibitors were profiled suggesting that CLL may be targetable with several compounds, such as pan-HMOX inhibitors and HDAC7-specific inhibitors. The novel protein-level observation of HDAC7 over-expression, while corroborated by mRNA overexpression data [490], offers a good candidate for validation by WB.

Cell surface proteins were successfully profiled, suggesting novel and anticipated candidates, in approximately equal proportions. CKAP4 presents a very promising candidate given its high confidence finding, but will require further investigation to better understand the localisation and precise role of this protein. Culturing CLL in the presence of any extracellular ligands of CKAP4 may induce measureable alterations in phenotype, such as the reduction of AKT phosphorylation seen previously with APF treatment of bladder cancers [511]. If the effects are recapitulated, APF, or more specific artificial ligands to CKAP4, could present a highly effective means of reducing CLL proliferation. Given its ubiquity, targeting CKAP4 may stabilise progressive CLL cases without applying aggressive therapeutics which can lead to a more aggressive, treatment-resistance disease. Additionally, the observation of similar CKAP4 overexpression in Eμ-*TCL1* tumours, provides a promising pre-clinical platform for any evaluation of CKAP4 function and targeting.

Prolectin/CLEC17A suggests a very promising target of immunotherapy, with expression limited to subsets of B cells. Validation of the cell surface expression of prolectin on a wider cohort of CLL will be required to conclude this potential. The observation of prolectin association with the BCR complex upon signalling, suggests an interesting avenue for investigation to better understand BCR signalling in CLL. The same is true for LAX1 and ATP2B4. PIGR also warrants validation and further investigation to identify why circulatory cancer cells upregulate an Fc receptor, PIGR, which typically functions in transcytosis across endothelial membranes.

Pathway analysis highlighted the consistent upregulation of transcriptional and post-transcriptional processes in CLL. The upregulation of almost every component of the spliceosome, may constitute an important finding in understanding previous observations of aberrant splicing. It will be important to identify whether overabundance of the spliceosome and subsequent overactive splicing is responsible for aberrant splicing observed in CLL.

Downregulated pathways highlighted a strong trend of downregulation amongst cell surface proteins indicative of immune evasion and a dedifferentiated B-cell state. Additionally, pathway analysis identified an apparent rewiring of the BCR pathway; suggesting INPP5F as a strong candidate for upstream induction of anergy and/or BCR signalling modulation in CLL. INPP5F biology is minimally characterised, suggesting a need for further investigation in both CLL and healthy B cells to understand any role it may have in BCR signalling.

The subtype-specific trends observed in the CLL cases examined were lower than originally anticipated, given the observation of phenotypic and clinical differences between these CLL subtypes. This analysis was, however, limited by low sample numbers and each evaluated CLL subtype did identify a distinct signature suggesting altered biological trends. U-CLL presented a strong trend of mitochondrial protein upregulation, glycosylation upregulation and antigen presentation and cell surface protein downregulation. *NOTCH1*-mutant CLL upregulated a trend of immune-related proteins and downregulated aspects of transcriptional regulation. The *SF3B1*-mutant CLL signature was minimal but highlighted specific epigenetic regulators with oncogenic potential, which could additionally be targeted with inhibitors. $CD38^{+}$ CLL cases presented a profile indicative of cells with a more active metabolism and cell cycle, in addition to more closely resembling the phenotype of healthy B cells. Trisomy 12 CLL cases identified an expected signature of upregulated chr12 proteins which suggested some specific oncogenic mechanisms such as p53 inhibition.

Overall, however, the quantitative proteomics results revealed a strong subtype-independent signature of CLL that was far greater than any subtype-specific trend of protein expression. This suggested that despite the several subtypes and heterogeneous nature of CLL the overall phenotype was indicative of a generally homogeneous disease.

To more confidently conclude this homogeneity amongst CLL cases, and to better understand the subtype-specific differences, a far greater number and range of samples will need to be analysed. Additionally a greater number of control samples would also need to be characterised, to understand the signatures arising from B-cell subtypes and inter-individual variations. More comprehensive profiling of subtypes would allow more confident conclusions

relating to how different subtypes should be clinically managed and treated. An example of this is the *SF3B1*-mutant-specific overexpression of PRMT1 and KDM4B, which may offer subtype-specific targets of inhibition.

The proteomics methodologies presented here also promise applicability to several other biological questions in the context of CLL. Characterisation of the protein expression changes which emerge in MBL and the further changes which are critical to the transformation to CLL, could offer a means of defining an early prognosis. The characterisation of B-cell subtypes in combination with MBL could additionally suggest strong phenotypic evidence for the elusive CLL cell of origin. The proteomics changes related to treatment-resistant CLL could also offer specific strategies to increase the chances of avoiding drug resistance and treating it when it occurs. Finally, it will be of interest to apply these methods to better understand how global protein expression changes in CLL cells localised to different biological niches, such as lymph nodes, bone marrow and the spleen.

In conclusion, the application of isobaric labels in 2D-LC MS/MS proteomics has provided a successful means of characterising the global protein expression in the context of B-cell cancers. This investigation presents unparalleled analyses of primary B-cell malignancies and plasma, demonstrating both expected and novel findings. These include the identification of putative targets of immunotherapy and targets of small molecular inhibitors. In addition to suggesting several novel hypotheses, the results offers some explanation to unanswered questions in B-cell cancer biology, such as advantages presented by trisomy 12 and a mechanism behind aberrant splicing in CLL.

Finally, these characterisations have highlighted the potential behind isobaric-labelled 2D-LC MS/MS proteomics. This success not only has implications for deriving further clinical impacts to CLL prognosis, diagnosis and treatment, but suggests applicability to other cancers, diseases and biological questions. With the field of proteomics continually improving characterisation quality, the applications in this investigation suggest that proteomics has the potential to become a revolutionary tool in many area of cancer biology and beyond.

9.0 REFERENCES

1. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
2. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
3. Mitelman, F., B. Johansson, and F. Mertens, *The impact of translocations and gene fusions on cancer causation.* Nat Rev Cancer, 2007. **7**(4): p. 233-45.
4. Rabbitts, T.H., *Commonality but diversity in cancer gene fusions.* Cell, 2009. **137**(3): p. 391-5.
5. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome.* Nature, 2009. **458**(7239): p. 719-24.
6. Doll, R., *The Pierre Denoix Memorial Lecture: nature and nurture in the control of cancer.* Eur J Cancer, 1999. **35**(1): p. 16-23.
7. Fletcher, O. and R.S. Houlston, *Architecture of inherited susceptibility to common cancer.* Nat Rev Cancer, 2010. **10**(5): p. 353-61.
8. Finkel, T., M. Serrano, and M.A. Blasco, *The common biology of cancer and ageing.* Nature, 2007. **448**(7155): p. 767-74.
9. Sieber, O.M., K. Heinimann, and I.P. Tomlinson, *Genomic instability--the engine of tumorigenesis?* Nat Rev Cancer, 2003. **3**(9): p. 701-8.
10. Jones, P.A. and S.B. Baylin, *The epigenomics of cancer.* Cell, 2007. **128**(4): p. 683-92.
11. Vogelstein, B. and K.W. Kinzler, *Cancer genes and the pathways they control.* Nat Med, 2004. **10**(8): p. 789-99.
12. Hoeijmakers, J.H., *Genome maintenance mechanisms for preventing cancer.* Nature, 2001. **411**(6835): p. 366-74.
13. Greaves, M. and C.C. Maley, *Clonal evolution in cancer.* Nature, 2012. **481**(7381): p. 306-13.

14. Raff, M.C., *Social controls on cell survival and cell death.* Nature, 1992. **356**(6368): p. 397-400.
15. Massague, J., *G1 cell-cycle control and cancer.* Nature, 2004. **432**(7015): p. 298-306.
16. Siegel, R., D. Naishadham, and A. Jemal, *Cancer statistics, 2013.* CA Cancer J Clin, 2013. **63**(1): p. 11-30.
17. Sidman, C.L., et al., *Multiple mechanisms of tumorigenesis in E mu-myc transgenic mice.* Cancer Res, 1993. **53**(7): p. 1665-9.
18. Lowe, S.W., E. Cepero, and G. Evan, *Intrinsic tumour suppression.* Nature, 2004. **432**(7015): p. 307-15.
19. White, R.L., *Tumor suppressing pathways.* Cell, 1998. **92**(5): p. 591-2.
20. Robanus-Maandag, E., et al., *p107 is a suppressor of retinoblastoma development in pRb-deficient mice.* Genes Dev, 1998. **12**(11): p. 1599-609.
21. Donehower, L.A., et al., *Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours.* Nature, 1992. **356**(6366): p. 215-21.
22. Visvader, J.E., *Cells of origin in cancer.* Nature, 2011. **469**(7330): p. 314-22.
23. Fialkow, P.J., et al., *Chronic myelocytic leukemia. Origin of some lymphocytes from leukemic stem cells.* J Clin Invest, 1978. **62**(4): p. 815-23.
24. Campo, E., et al., *The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications.* Blood, 2011. **117**(19): p. 5019-32.
25. Harris, N.L., et al., *The World Health Organization classification of neoplastic diseases of the haematopoietic and lymphoid tissues: Report of the Clinical Advisory Committee Meeting, Airlie House, Virginia, November 1997.* Histopathology, 2000. **36**(1): p. 69-86.
26. Kuppers, R., *Mechanisms of B-cell lymphoma pathogenesis.* Nat Rev Cancer, 2005. **5**(4): p. 251-62.
27. Cooper, M.D. and M.N. Alder, *The evolution of adaptive immune systems.* Cell, 2006. **124**(4): p. 815-22.
28. LeBien, T.W. and T.F. Tedder, *B lymphocytes: how they develop and function.* Blood, 2008. **112**(5): p. 1570-80.
29. Zhang, S. and T.J. Kipps, *The pathogenesis of chronic lymphocytic leukemia.* Annu Rev Pathol, 2014. **9**: p. 103-18.
30. Iversen, U., et al., *Cell kinetics of African cases of Burkitt lymphoma. A preliminary report.* Eur J Cancer, 1972. **8**(3): p. 305-8.
31. Molyneux, E.M., et al., *Burkitt's lymphoma.* Lancet, 2012. **379**(9822): p. 1234-44.
32. van den Bosch, C.A., *Is endemic Burkitt's lymphoma an alliance between three infections and a tumour promoter?* Lancet Oncol, 2004. **5**(12): p. 738-46.
33. Miles, R.R., S. Arnold, and M.S. Cairo, *Risk factors and treatment of childhood and adolescent Burkitt lymphoma/leukaemia.* Br J Haematol, 2012. **156**(6): p. 730-43.
34. Taub, R., et al., *Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells.* Proc Natl Acad Sci U S A, 1982. **79**(24): p. 7837-41.
35. Bernheim, A., R. Berger, and G. Lenoir, *Cytogenetic studies on African Burkitt's lymphoma cell lines: t(8;14), t(2;8) and t(8;22) translocations.* Cancer Genet Cytogenet, 1981. **3**(4): p. 307-15.
36. Kuppers, R., *B cells under influence: transformation of B cells by Epstein-Barr virus.* Nat Rev Immunol, 2003. **3**(10): p. 801-12.
37. Thorley-Lawson, D.A. and A. Gross, *Persistence of the Epstein-Barr virus and the origins of associated lymphomas.* N Engl J Med, 2004. **350**(13): p. 1328-37.
38. Brady, G., G.J. Macarthur, and P.J. Farrell, *Epstein-Barr virus and Burkitt lymphoma.* Postgrad Med J, 2008. **84**(993): p. 372-7.

39. Neri, A., et al., *Epstein-Barr virus infection precedes clonal expansion in Burkitt's and acquired immunodeficiency syndrome-associated lymphoma.* Blood, 1991. **77**(5): p. 1092-5.
40. Wilson, J.B., J.L. Bell, and A.J. Levine, *Expression of Epstein-Barr virus nuclear antigen-1 induces B cell neoplasia in transgenic mice.* EMBO J, 1996. **15**(12): p. 3117-26.
41. Lane, H.C., et al., *Abnormalities of B-cell activation and immunoregulation in patients with the acquired immunodeficiency syndrome.* N Engl J Med, 1983. **309**(8): p. 453-8.
42. Donati, D., et al., *Increased B cell survival and preferential activation of the memory compartment by a malaria polyclonal B cell activator.* J Immunol, 2006. **177**(5): p. 3035-44.
43. http://www.cancerresearchuk.org/.
44. Binet, J.L., et al., *A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis.* Cancer, 1981. **48**(1): p. 198-206.
45. Rai, K.R. and A. Sawitsky, *A review of the prognostic role of cytogenetic, phenotypic, morphologic, and immune function characteristics in chronic lymphocytic leukemia.* Blood Cells, 1987. **12**(2): p. 327-38.
46. Weiss, N.S., *Geographical variation in the incidence of the leukemias and lymphomas.* Natl Cancer Inst Monogr, 1979(53): p. 139-42.
47. Hallek, M., et al., *Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines.* Blood, 2008. **111**(12): p. 5446-56.
48. Hallek, M., et al., *Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines.* Blood, 2008. **111**(12): p. 5446-56.
49. Nabhan, C. and S.T. Rosen, *Chronic lymphocytic leukemia: a clinical review.* JAMA, 2014. **312**(21): p. 2265-76.
50. Rawstron, A.C., et al., *Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia.* N Engl J Med, 2008. **359**(6): p. 575-83.
51. Mowery, Y.M. and M.C. Lanasa, *Clinical aspects of monoclonal B-cell lymphocytosis.* Cancer Control, 2012. **19**(1): p. 8-17.
52. Molica, S., et al., *Monoclonal B-cell lymphocytosis: a reappraisal of its clinical implications.* Leuk Lymphoma, 2012. **53**(9): p. 1660-5.
53. Marti, G.E., et al., *Diagnostic criteria for monoclonal B-cell lymphocytosis.* Br J Haematol, 2005. **130**(3): p. 325-32.
54. Rai, K.R., et al., *Clinical staging of chronic lymphocytic leukemia.* Blood, 1975. **46**(2): p. 219-34.
55. Hallek, M., et al., *Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial.* Lancet, 2010. **376**(9747): p. 1164-74.
56. Brown, J.R., *The treatment of relapsed refractory chronic lymphocytic leukemia.* Hematology Am Soc Hematol Educ Program, 2011. **2011**: p. 110-8.
57. Wiestner, A., *Targeting B-Cell receptor signaling for anticancer therapy: the Bruton's tyrosine kinase inhibitor ibrutinib induces impressive responses in B-cell malignancies.* J Clin Oncol, 2013. **31**(1): p. 128-30.
58. Shah, A. and A. Mangaonkar, *Idelalisib: A Novel PI3Kdelta Inhibitor for Chronic Lymphocytic Leukemia.* Ann Pharmacother, 2015.
59. Souers, A.J., et al., *ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets.* Nat Med, 2013. **19**(2): p. 202-8.

60. Roberts, A.W., et al., *Phase 1 study of the safety, pharmacokinetics, and antitumour activity of the BCL2 inhibitor navitoclax in combination with rituximab in patients with relapsed or refractory CD20(+) lymphoid malignancies.* Br J Haematol, 2015. **170**(5): p. 669-78.
61. Campas, C., et al., *Bcl-2 inhibitors induce apoptosis in chronic lymphocytic leukemia cells.* Exp Hematol, 2006. **34**(12): p. 1663-9.
62. Dighiero, G., et al., *B-cell chronic lymphocytic leukemia: present status and future directions. French Cooperative Group on CLL.* Blood, 1991. **78**(8): p. 1901-14.
63. Dighiero, G., *Unsolved issues in CLL biology and management.* Leukemia, 2003. **17**(12): p. 2385-91.
64. Lanasa, M.C., *Novel insights into the biology of CLL.* Hematology Am Soc Hematol Educ Program, 2010. **2010**: p. 70-6.
65. Messmer, B.T., et al., *In vivo measurements document the dynamic cellular kinetics of chronic lymphocytic leukemia B cells.* J Clin Invest, 2005. **115**(3): p. 755-64.
66. Burger, J.A., M. Burger, and T.J. Kipps, *Chronic lymphocytic leukemia B cells express functional CXCR4 chemokine receptors that mediate spontaneous migration beneath bone marrow stromal cells.* Blood, 1999. **94**(11): p. 3658-67.
67. Burger, J.A., et al., *Blood-derived nurse-like cells protect chronic lymphocytic leukemia B cells from spontaneous apoptosis through stromal cell-derived factor-1.* Blood, 2000. **96**(8): p. 2655-63.
68. Redondo-Munoz, J., et al., *Alpha4beta1 integrin and 190-kDa CD44v constitute a cell surface docking complex for gelatinase B/MMP-9 in chronic leukemic but not in normal B cells.* Blood, 2008. **112**(1): p. 169-78.
69. Almasri, N.M., et al., *Reduced expression of CD20 antigen as a characteristic marker for chronic lymphocytic leukemia.* Am J Hematol, 1992. **40**(4): p. 259-63.
70. Damle, R.N., et al., *Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia.* Blood, 1999. **94**(6): p. 1840-7.
71. Rosenwald, A., et al., *Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia.* J Exp Med, 2001. **194**(11): p. 1639-47.
72. Klein, U., et al., *Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells.* J Exp Med, 2001. **194**(11): p. 1625-38.
73. Chiorazzi, N. and M. Ferrarini, *Cellular origin(s) of chronic lymphocytic leukemia: cautionary notes and additional considerations and possibilities.* Blood, 2011. **117**(6): p. 1781-91.
74. Seifert, M., et al., *Cellular origin and pathophysiology of chronic lymphocytic leukemia.* J Exp Med, 2012. **209**(12): p. 2183-98.
75. Chiorazzi, N., K.R. Rai, and M. Ferrarini, *Chronic lymphocytic leukemia.* N Engl J Med, 2005. **352**(8): p. 804-15.
76. Chiorazzi, N. and M. Ferrarini, *B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor.* Annu Rev Immunol, 2003. **21**: p. 841-94.
77. Messmer, B.T., et al., *Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia.* J Exp Med, 2004. **200**(4): p. 519-25.
78. Nolz, J.C., et al., *ZAP-70 is expressed by a subset of normal human B-lymphocytes displaying an activated phenotype.* Leukemia, 2005. **19**(6): p. 1018-24.
79. Oscier, D.G., *Cytogenetic and molecular abnormalities in chronic lymphocytic leukaemia.* Blood Rev, 1994. **8**(2): p. 88-97.

80. Stilgenbauer, S., P. Lichter, and H. Dohner, *Genetic features of B-cell chronic lymphocytic leukemia.* Rev Clin Exp Hematol, 2000. **4**(1): p. 48-72.
81. Juliusson, G., et al., *Prognostic subgroups in B-cell chronic lymphocytic leukemia defined by specific chromosomal abnormalities.* N Engl J Med, 1990. **323**(11): p. 720-4.
82. Dohner, H., et al., *Genomic aberrations and survival in chronic lymphocytic leukemia.* N Engl J Med, 2000. **343**(26): p. 1910-6.
83. Cimmino, A., et al., *miR-15 and miR-16 induce apoptosis by targeting BCL2.* Proc Natl Acad Sci U S A, 2005. **102**(39): p. 13944-9.
84. Rossi, D., et al., *Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia.* Blood, 2013. **121**(8): p. 1403-12.
85. Marasca, R., et al., *Clinical heterogeneity of de novo 11q deletion chronic lymphocytic leukaemia: prognostic relevance of extent of 11q deleted nuclei inside leukemic clone.* Hematol Oncol, 2013. **31**(2): p. 88-95.
86. Puiggros, A., G. Blanco, and B. Espinet, *Genetic abnormalities in chronic lymphocytic leukemia: where we are and where we go.* Biomed Res Int, 2014. **2014**: p. 435983.
87. Gaidano, G., R. Foa, and R. Dalla-Favera, *Molecular pathogenesis of chronic lymphocytic leukemia.* J Clin Invest, 2012. **122**(10): p. 3432-8.
88. Riches, J.C., et al., *Trisomy 12 chronic lymphocytic leukemia cells exhibit upregulation of integrin signaling that is modulated by NOTCH1 mutations.* Blood, 2014. **123**(26): p. 4101-10.
89. Zucchetto, A., et al., *CD49d is overexpressed by trisomy 12 chronic lymphocytic leukemia cells: evidence for a methylation-dependent regulation mechanism.* Blood, 2013. **122**(19): p. 3317-21.
90. Jeromin, S., et al., *SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients.* Leukemia, 2014. **28**(1): p. 108-17.
91. Rossi, D., et al., *The prognostic value of TP53 mutations in chronic lymphocytic leukemia is independent of Del17p13: implications for overall survival and chemorefractoriness.* Clin Cancer Res, 2009. **15**(3): p. 995-1004.
92. Brutsch, R., et al., *Integrin cytoplasmic domain-associated protein-1 attenuates sprouting angiogenesis.* Circ Res, 2010. **107**(5): p. 592-601.
93. Fabbri, G., et al., *Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation.* J Exp Med, 2011. **208**(7): p. 1389-401.
94. Rossi, D., et al., *Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia.* Blood, 2012. **119**(2): p. 521-9.
95. Rossi, D., et al., *Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness.* Blood, 2011. **118**(26): p. 6904-8.
96. Quesada, V., et al., *Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia.* Nat Genet, 2012. **44**(1): p. 47-52.
97. Ouillette, P., et al., *Incidence and clinical implications of ATM aberrations in chronic lymphocytic leukemia.* Genes Chromosomes Cancer, 2012. **51**(12): p. 1125-32.
98. Rossi, D., et al., *Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia.* Blood, 2012. **119**(12): p. 2854-62.
99. Ngo, V.N., et al., *Oncogenically active MYD88 mutations in human lymphoma.* Nature, 2011. **470**(7332): p. 115-9.
100. Stevenson, F.K., et al., *B-cell receptor signaling in chronic lymphocytic leukemia.* Blood, 2011. **118**(16): p. 4313-20.
101. Hamblin, T.J., et al., *Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia.* Blood, 1999. **94**(6): p. 1848-54.

102. Guarini, A., et al., *BCR ligation induced by IgM stimulation results in gene expression and functional changes only in IgV H unmutated chronic lymphocytic leukemia (CLL) cells.* Blood, 2008. **112**(3): p. 782-92.
103. Catera, R., et al., *Chronic lymphocytic leukemia cells recognize conserved epitopes associated with apoptosis and oxidation.* Mol Med, 2008. **14**(11-12): p. 665-74.
104. Chu, C.C., et al., *Chronic lymphocytic leukemia antibodies with a common stereotypic rearrangement recognize nonmuscle myosin heavy chain IIA.* Blood, 2008. **112**(13): p. 5122-9.
105. Duhren-von Minden, M., et al., *Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling.* Nature, 2012. **489**(7415): p. 309-12.
106. Landgren, O., et al., *Respiratory tract infections and subsequent risk of chronic lymphocytic leukemia.* Blood, 2007. **109**(5): p. 2198-201.
107. Kostareli, E., et al., *Molecular evidence for EBV and CMV persistence in a subset of patients with chronic lymphocytic leukemia expressing stereotyped IGHV4-34 B-cell receptors.* Leukemia, 2009. **23**(5): p. 919-24.
108. Holler, C., et al., *PKCbeta is essential for the development of chronic lymphocytic leukemia in the TCL1 transgenic mouse model: validation of PKCbeta as a therapeutic target in chronic lymphocytic leukemia.* Blood, 2009. **113**(12): p. 2791-2794.
109. Thompson, A.A., et al., *Aberrations of the B-cell receptor B29 (CD79b) gene in chronic lymphocytic leukemia.* Blood, 1997. **90**(4): p. 1387-94.
110. Vuillier, F., et al., *Lower levels of surface B-cell-receptor expression in chronic lymphocytic leukemia are associated with glycosylation and folding defects of the mu and CD79a chains.* Blood, 2005. **105**(7): p. 2933-40.
111. Chen, L., et al., *ZAP-70 enhances IgM signaling independent of its kinase activity in chronic lymphocytic leukemia.* Blood, 2008. **111**(5): p. 2685-92.
112. Chen, L., et al., *Expression of ZAP-70 is associated with increased B-cell receptor signaling in chronic lymphocytic leukemia.* Blood, 2002. **100**(13): p. 4609-14.
113. Chen, L., et al., *ZAP-70 directly enhances IgM signaling in chronic lymphocytic leukemia.* Blood, 2005. **105**(5): p. 2036-41.
114. Pede, V., et al., *Expression of ZAP70 in chronic lymphocytic leukaemia activates NF-kappaB signalling.* Br J Haematol, 2013. **163**(5): p. 621-30.
115. Antony, P., et al., *B cell receptor directs the activation of NFAT and NF-kappaB via distinct molecular mechanisms.* Exp Cell Res, 2003. **291**(1): p. 11-24.
116. Furman, R.R., et al., *Modulation of NF-kappa B activity and apoptosis in chronic lymphocytic leukemia B cells.* J Immunol, 2000. **164**(4): p. 2200-6.
117. Hewamana, S., et al., *The NF-kappaB subunit Rel A is associated with in vitro survival and clinical disease progression in chronic lymphocytic leukemia and represents a promising therapeutic target.* Blood, 2008. **111**(9): p. 4681-9.
118. Zaninoni, A., et al., *Cytokine modulation of nuclear factor-kappaB activity in B-chronic lymphocytic leukemia.* Exp Hematol, 2003. **31**(3): p. 185-90.
119. Horie, R., et al., *DHMEQ, a new NF-kappaB inhibitor, induces apoptosis and enhances fludarabine effects on chronic lymphocytic leukemia cells.* Leukemia, 2006. **20**(5): p. 800-6.
120. Vaisitti, T., et al., *The enzymatic activities of CD38 enhance CLL growth and trafficking: implications for therapeutic targeting.* Leukemia, 2015. **29**(2): p. 356-68.
121. Bulian, P., et al., *CD49d is the strongest flow cytometry-based predictor of overall survival in chronic lymphocytic leukemia.* J Clin Oncol, 2014. **32**(9): p. 897-904.
122. Damle, R.N., et al., *CD38 expression labels an activated subset within chronic lymphocytic leukemia clones enriched in proliferating B cells.* Blood, 2007. **110**(9): p. 3352-9.

123. Funaro, A., et al., *Human CD38 is associated to distinct molecules which mediate transmembrane signaling in different lineages.* Eur J Immunol, 1993. **23**(10): p. 2407-11.
124. Deaglio, S., et al., *CD38 is a signaling molecule in B-cell chronic lymphocytic leukemia cells.* Blood, 2003. **102**(6): p. 2146-55.
125. Till, K.J., et al., *The chemokine receptor CCR7 and alpha4 integrin are important for migration of chronic lymphocytic leukemia cells into lymph nodes.* Blood, 2002. **99**(8): p. 2977-84.
126. Till, K.J., et al., *CLL, but not normal, B cells are dependent on autocrine VEGF and alpha4beta1 integrin for chemokine-induced motility on and through endothelium.* Blood, 2005. **105**(12): p. 4813-9.
127. de la Fuente, M.T., et al., *Fibronectin interaction with alpha4beta1 integrin prevents apoptosis in B cell chronic lymphocytic leukemia: correlation with Bcl-2 and Bax.* Leukemia, 1999. **13**(2): p. 266-74.
128. Gattei, V., et al., *Relevance of CD49d protein expression as overall survival and progressive disease prognosticator in chronic lymphocytic leukemia.* Blood, 2008. **111**(2): p. 865-73.
129. Tsujimoto, Y., et al., *Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation.* Science, 1984. **226**(4678): p. 1097-9.
130. Mertens, D., et al., *Allelic silencing at the tumor-suppressor locus 13q14.3 suggests an epigenetic tumor-suppressor mechanism.* Proc Natl Acad Sci U S A, 2006. **103**(20): p. 7741-6.
131. Sampath, D., et al., *Histone deacetylases mediate the silencing of miR-15a, miR-16, and miR-29b in chronic lymphocytic leukemia.* Blood, 2012. **119**(5): p. 1162-72.
132. Otake, Y., et al., *Overexpression of nucleolin in chronic lymphocytic leukemia cells induces stabilization of bcl2 mRNA.* Blood, 2007. **109**(7): p. 3069-75.
133. Pepper, C., et al., *Mcl-1 expression has in vitro and in vivo significance in chronic lymphocytic leukemia and is associated with other poor prognostic markers.* Blood, 2008. **112**(9): p. 3807-17.
134. Gottardi, D., et al., *In leukaemic CD5+ B cells the expression of BCL-2 gene family is shifted toward protection from apoptosis.* Br J Haematol, 1996. **94**(4): p. 612-8.
135. Baskar, S., et al., *Unique cell surface expression of receptor tyrosine kinase ROR1 in human B-cell chronic lymphocytic leukemia.* Clin Cancer Res, 2008. **14**(2): p. 396-404.
136. Barna, G., et al., *ROR1 expression is not a unique marker of CLL.* Hematol Oncol, 2011. **29**(1): p. 17-21.
137. Fukuda, T., et al., *Antisera induced by infusions of autologous Ad-CD154-leukemia B cells identify ROR1 as an oncofetal antigen and receptor for Wnt5a.* Proc Natl Acad Sci U S A, 2008. **105**(8): p. 3047-52.
138. Janovska, P., et al., *Autocrine signaling by Wnt-5a deregulates chemotaxis of leukemic cells and predicts clinical outcome in chronic lymphocytic leukemia.* Clin Cancer Res, 2015.
139. Daneshmanesh, A.H., et al., *Monoclonal antibodies against ROR1 induce apoptosis of chronic lymphocytic leukemia (CLL) cells.* Leukemia, 2012. **26**(6): p. 1348-55.
140. Hudecek, M., et al., *Receptor affinity and extracellular domain modifications affect tumor recognition by ROR1-specific chimeric antigen receptor T cells.* Clin Cancer Res, 2013. **19**(12): p. 3153-64.
141. Sipkins, D.A., et al., *In vivo imaging of specialized bone marrow endothelial microdomains for tumour engraftment.* Nature, 2005. **435**(7044): p. 969-73.
142. Burger, J.A., et al., *The microenvironment in mature B-cell malignancies: a target for new treatment strategies.* Blood, 2009. **114**(16): p. 3367-75.

143. Herling, M., et al., *TCL1 shows a regulated expression pattern in chronic lymphocytic leukemia that correlates with molecular subtypes and proliferative state.* Leukemia, 2006. **20**(2): p. 280-5.
144. Bichi, R., et al., *Human chronic lymphocytic leukemia modeled in mouse by targeted TCL1 expression.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 6955-60.
145. Johnson, A.J., et al., *Characterization of the TCL-1 transgenic mouse as a preclinical drug development tool for human chronic lymphocytic leukemia.* Blood, 2006. **108**(4): p. 1334-8.
146. Ablain, J., et al., *How animal models of leukaemias have already benefited patients.* Mol Oncol, 2013. **7**(2): p. 224-31.
147. Cespedes, M.V., et al., *Mouse models in oncogenesis and cancer therapy.* Clin Transl Oncol, 2006. **8**(5): p. 318-29.
148. Frese, K.K. and D.A. Tuveson, *Maximizing mouse cancer models.* Nat Rev Cancer, 2007. **7**(9): p. 645-58.
149. Becher, O.J. and E.C. Holland, *Genetically engineered models have advantages over xenografts for preclinical studies.* Cancer Res, 2006. **66**(7): p. 3355-8, discussion 3358-9.
150. Rangarajan, A. and R.A. Weinberg, *Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice.* Nat Rev Cancer, 2003. **3**(12): p. 952-9.
151. Ames, B.N., M.K. Shigenaga, and T.M. Hagen, *Oxidants, antioxidants, and the degenerative diseases of aging.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 7915-22.
152. Jonkers, J., et al., *Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer.* Nat Genet, 2001. **29**(4): p. 418-25.
153. Maddison, K. and A.R. Clarke, *New approaches for modelling cancer mechanisms in the mouse.* J Pathol, 2005. **205**(2): p. 181-93.
154. Politi, K. and W. Pao, *How genetically engineered mouse tumor models provide insights into human cancers.* J Clin Oncol, 2011. **29**(16): p. 2273-81.
155. Neel, B.G., et al., *Avian leukosis virus-induced tumors have common proviral integration sites and synthesize discrete new RNAs: oncogenesis by promoter insertion.* Cell, 1981. **23**(2): p. 323-34.
156. Vennstrom, B., et al., *Isolation and characterization of c-myc, a cellular homolog of the oncogene (v-myc) of avian myelocytomatosis virus strain 29.* J Virol, 1982. **42**(3): p. 773-9.
157. Meyer, N. and L.Z. Penn, *Reflecting on 25 years with MYC.* Nat Rev Cancer, 2008. **8**(12): p. 976-90.
158. Eilers, M., et al., *Chimaeras of myc oncoprotein and steroid receptors cause hormone-dependent transformation of cells.* Nature, 1989. **340**(6228): p. 66-8.
159. Armelin, H.A., et al., *Functional role for c-myc in mitogenic response to platelet-derived growth factor.* Nature, 1984. **310**(5979): p. 655-60.
160. Blackwell, T.K., et al., *Sequence-specific DNA binding by the c-Myc protein.* Science, 1990. **250**(4984): p. 1149-51.
161. Nair, S.K. and S.K. Burley, *X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors.* Cell, 2003. **112**(2): p. 193-205.
162. Zeller, K.I., et al., *An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets.* Genome Biol, 2003. **4**(10): p. R69.
163. Kelly, K., et al., *Cell-specific regulation of the c-myc gene by lymphocyte mitogens and platelet-derived growth factor.* Cell, 1983. **35**(3 Pt 2): p. 603-10.

164. Klinakis, A., et al., *Myc is a Notch1 transcriptional target and a requisite for Notch1-induced mammary tumorigenesis in mice.* Proc Natl Acad Sci U S A, 2006. **103**(24): p. 9262-7.
165. Dang, C.V., *MYC on the path to cancer.* Cell, 2012. **149**(1): p. 22-35.
166. Freytag, S.O., *Enforced expression of the c-myc oncogene inhibits cell differentiation by precluding entry into a distinct predifferentiation state in G0/G1.* Mol Cell Biol, 1988. **8**(4): p. 1614-24.
167. Dang, C.V., et al., *The c-Myc target gene network.* Semin Cancer Biol, 2006. **16**(4): p. 253-64.
168. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
169. Murre, C., P.S. McCaw, and D. Baltimore, *A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins.* Cell, 1989. **56**(5): p. 777-83.
170. Amati, B., et al., *Oncogenic activity of the c-Myc protein requires dimerization with Max.* Cell, 1993. **72**(2): p. 233-45.
171. Grant, P.A., et al., *The ATM-related cofactor Tra1 is a component of the purified SAGA complex.* Mol Cell, 1998. **2**(6): p. 863-7.
172. Amati, B., et al., *Function of the c-Myc oncoprotein in chromatin remodeling and transcription.* Biochim Biophys Acta, 2001. **1471**(3): p. M135-45.
173. McMahon, B. and R.M. Hanson, *A toolkit for publishing enhanced figures.* J Appl Crystallogr, 2008. **41**(Pt 4): p. 811-814.
174. Nesbit, C.E., J.M. Tersak, and E.V. Prochownik, *MYC oncogenes and human neoplastic disease.* Oncogene, 1999. **18**(19): p. 3004-16.
175. Vafa, O., et al., *c-Myc can induce DNA damage, increase reactive oxygen species, and mitigate p53 function: a mechanism for oncogene-induced genetic instability.* Mol Cell, 2002. **9**(5): p. 1031-44.
176. Harris, A.W., et al., *The E mu-myc transgenic mouse. A model for high-incidence spontaneous lymphoma and leukemia of early B cells.* J Exp Med, 1988. **167**(2): p. 353-71.
177. Alitalo, K., et al., *Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma.* Proc Natl Acad Sci U S A, 1983. **80**(6): p. 1707-11.
178. Yada, M., et al., *Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7.* EMBO J, 2004. **23**(10): p. 2116-25.
179. Zindy, F., et al., *Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization.* Genes Dev, 1998. **12**(15): p. 2424-33.
180. Honda, R. and H. Yasuda, *Association of p19(ARF) with Mdm2 inhibits ubiquitin ligase activity of Mdm2 for tumor suppressor p53.* EMBO J, 1999. **18**(1): p. 22-7.
181. Kamijo, T., et al., *Functional and physical interactions of the ARF tumor suppressor with p53 and Mdm2.* Proc Natl Acad Sci U S A, 1998. **95**(14): p. 8292-7.
182. Eischen, C.M., et al., *Disruption of the ARF-Mdm2-p53 tumor suppressor pathway in Myc-induced lymphomagenesis.* Genes Dev, 1999. **13**(20): p. 2658-69.
183. Poe, J.C., et al., *A c-Myc and surface CD19 signaling amplification loop promotes B cell lymphoma development and progression in mice.* J Immunol, 2012. **189**(5): p. 2318-25.
184. Chung, E.Y., et al., *CD19 is a major B cell receptor-independent activator of MYC-driven B-lymphomagenesis.* J Clin Invest, 2012. **122**(6): p. 2257-66.
185. Felsher, D.W. and J.M. Bishop, *Transient excess of MYC activity can elicit genomic instability and tumorigenesis.* Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3940-4.

186. Ray, S., et al., *MYC can induce DNA breaks in vivo and in vitro independent of reactive oxygen species.* Cancer Res, 2006. **66**(13): p. 6598-605.
187. Li, Q. and C.V. Dang, *c-Myc overexpression uncouples DNA replication from mitosis.* Mol Cell Biol, 1999. **19**(8): p. 5339-51.
188. Adams, J.M., et al., *The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice.* Nature, 1985. **318**(6046): p. 533-8.
189. Mori, S., et al., *Utilization of pathway signatures to reveal distinct types of B lymphoma in the Emicro-myc model and human diffuse large B-cell lymphoma.* Cancer Res, 2008. **68**(20): p. 8525-34.
190. Zhu, D., et al., *Deregulated expression of the Myc cellular oncogene drives development of mouse "Burkitt-like" lymphomas from naive B cells.* Blood, 2005. **105**(5): p. 2135-7.
191. Langdon, W.Y., et al., *The c-myc oncogene perturbs B lymphocyte development in E-mu-myc transgenic mice.* Cell, 1986. **47**(1): p. 11-8.
192. Alexander, W.S., J.W. Schrader, and J.M. Adams, *Expression of the c-myc oncogene under control of an immunoglobulin enhancer in E mu-myc transgenic mice.* Mol Cell Biol, 1987. **7**(4): p. 1436-44.
193. Langdon, W.Y., A.W. Harris, and S. Cory, *Acceleration of B-lymphoid tumorigenesis in E mu-myc transgenic mice by v-H-ras and v-raf but not v-abl.* Oncogene Res, 1989. **4**(4): p. 253-8.
194. Petock, J.M., et al., *Crystal structures of Tcl1 family oncoproteins and their conserved surface features.* ScientificWorldJournal, 2002. **2**: p. 1876-84.
195. Takizawa, J., et al., *Expression of the TCL1 gene at 14q32 in B-cell malignancies but not in adult T-cell leukemia.* Jpn J Cancer Res, 1998. **89**(7): p. 712-8.
196. Said, J.W., et al., *TCL1 oncogene expression in B cell subsets from lymphoid hyperplasia and distinct classes of B cell lymphoma.* Lab Invest, 2001. **81**(4): p. 555-64.
197. Kang, S.M., et al., *Impaired T- and B-cell development in Tcl1-deficient mice.* Blood, 2005. **105**(3): p. 1288-94.
198. Narducci, M.G., et al., *TCL1 participates in early embryonic development and is overexpressed in human seminomas.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11712-7.
199. Russo, G., et al., *Molecular analysis of a t(14;14) translocation in leukemic T-cells of an ataxia telangiectasia patient.* Proc Natl Acad Sci U S A, 1989. **86**(2): p. 602-6.
200. Narducci, M.G., et al., *TCL1 is overexpressed in patients affected by adult T-cell leukemias.* Cancer Res, 1997. **57**(24): p. 5452-6.
201. Sakashita, et al., *Amplification of the TCL1 flanking region at 14q32.1 with no TCL1 gene transcription in a patient with peripheral T cell lymphoma.* Leukemia, 1998. **12**(6): p. 970-1.
202. Takizawa, J. and M. Seto, *The TCL1 oncogene is not overexpressed in patients with adult T cell leukemia.* Leukemia, 1999. **13**(2): p. 314.
203. Narducci, M.G., et al., *Regulation of TCL1 expression in B- and T-cell lymphomas and reactive lymphoid tissues.* Cancer Res, 2000. **60**(8): p. 2095-100.
204. Virgilio, L., et al., *Deregulated expression of TCL1 causes T cell leukemia in mice.* Proc Natl Acad Sci U S A, 1998. **95**(7): p. 3885-9.
205. Croce, C.M., *Role of TCL1 and ALL1 in human leukemias and development.* Cancer Res, 1999. **59**(7 Suppl): p. 1778s-1783s.
206. Petock, J.M., et al., *Structure of murine Tcl1 at 2.5 A resolution and implications for the TCL oncogene family.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 11): p. 1545-51.
207. Hoh, F., et al., *Crystal structure of p14TCL1, an oncogene product involved in T-cell prolymphocytic leukemia, reveals a novel beta-barrel topology.* Structure, 1998. **6**(2): p. 147-55.

208. Yuille, M.R., et al., *TCL1 is activated by chromosomal rearrangement or by hypomethylation.* Genes Chromosomes Cancer, 2001. **30**(4): p. 336-41.
209. Gaudio, E., et al., *Heat shock protein 70 regulates Tcl1 expression in leukemia and lymphomas.* Blood, 2013. **121**(2): p. 351-9.
210. Balatti, V., et al., *TCL1 targeting miR-3676 is codeleted with tumor protein p53 in chronic lymphocytic leukemia.* Proc Natl Acad Sci U S A, 2015. **112**(7): p. 2169-74.
211. Pekarsky, Y., et al., *Tcl1 enhances Akt kinase activity and mediates its nuclear translocation.* Proc Natl Acad Sci U S A, 2000. **97**(7): p. 3028-33.
212. Laine, J., et al., *The protooncogene TCL1 is an Akt kinase coactivator.* Mol Cell, 2000. **6**(2): p. 395-407.
213. Laine, J., et al., *Differential regulation of Akt kinase isoforms by the members of the TCL1 oncogene family.* J Biol Chem, 2002. **277**(5): p. 3743-51.
214. Suzuki, A., et al., *Critical roles of Pten in B cell homeostasis and immunoglobulin class switch recombination.* J Exp Med, 2003. **197**(5): p. 657-67.
215. Pekarsky, Y., et al., *Animal models for chronic lymphocytic leukemia.* J Cell Biochem, 2007. **100**(5): p. 1109-18.
216. Pekarsky, Y., et al., *Tcl1 functions as a transcriptional regulator and is directly involved in the pathogenesis of CLL.* Proc Natl Acad Sci U S A, 2008. **105**(50): p. 19643-8.
217. Ameyar, M., M. Wisniewska, and J.B. Weitzman, *A role for AP-1 in apoptosis: the case for and against.* Biochimie, 2003. **85**(8): p. 747-52.
218. Kriss, C.L., et al., *Overexpression of TCL1 activates the endoplasmic reticulum stress response: a novel mechanism of leukemic progression in mice.* Blood, 2012. **120**(5): p. 1027-38.
219. Zanesi, N., et al., *A Sleeping Beauty screen reveals NF-kB activation in CLL mouse model.* Blood, 2013.
220. Noguchi, M., et al., *Proto-oncogene TCL1: more than just a coactivator for Akt.* FASEB J, 2007. **21**(10): p. 2273-84.
221. Palamarchuk, A., et al., *Tcl1 protein functions as an inhibitor of de novo DNA methylation in B-cell chronic lymphocytic leukemia (CLL).* Proc Natl Acad Sci U S A, 2012. **109**(7): p. 2555-60.
222. Kimby, E., et al., *T lymphocyte subpopulations in chronic lymphocytic leukemia of B cell type in relation to immunoglobulin isotype(s) on the leukemic clone and to clinical features.* Eur J Haematol, 1987. **38**(3): p. 261-7.
223. Gorgun, G., et al., *E(mu)-TCL1 mice represent a model for immunotherapeutic reversal of chronic lymphocytic leukemia-induced T-cell dysfunction.* Proc Natl Acad Sci U S A, 2009. **106**(15): p. 6250-5.
224. Hofbauer, J.P., et al., *Development of CLL in the TCL1 transgenic mouse model is associated with severe skewing of the T-cell compartment homologous to human CLL.* Leukemia, 2011. **25**(9): p. 1452-8.
225. Chen, S.S., et al., *Epigenetic alterations in a murine model for chronic lymphocytic leukemia.* Cell Cycle, 2009. **8**(22): p. 3663-7.
226. Chen, S.S., et al., *Epigenetic changes during disease progression in a murine model of human chronic lymphocytic leukemia.* Proc Natl Acad Sci U S A, 2009. **106**(32): p. 13433-8.
227. Herling, M., et al., *High TCL1 levels are a marker of B-cell receptor pathway responsiveness and adverse outcome in chronic lymphocytic leukemia.* Blood, 2009. **114**(21): p. 4675-86.
228. Lucas, D.M., et al., *The novel deacetylase inhibitor AR-42 demonstrates pre-clinical activity in B-cell malignancies in vitro and in vivo.* PLoS One, 2010. **5**(6): p. e10941.

229. Zanesi, N., et al., *Effect of rapamycin on mouse chronic lymphocytic leukemia and the development of nonhematopoietic malignancies in Emu-TCL1 transgenic mice.* Cancer Res, 2006. **66**(2): p. 915-20.
230. Blunt, M.D., et al., *The PI3K/mTOR inhibitor PF-04691502 induces apoptosis and inhibits microenvironmental signaling in CLL and the Emicro-TCL1 mouse model.* Blood, 2015. **125**(26): p. 4032-41.
231. Simonetti, G., et al., *Mouse models in the study of chronic lymphocytic leukemia pathogenesis and therapy.* Blood, 2014. **124**(7): p. 1010-9.
232. Liu, J., et al., *Loss of p53 and altered miR15-a/16-1short right arrowMCL-1 pathway in CLL: insights from TCL1-Tg:p53(-/-) mouse model and primary human leukemia cells.* Leukemia, 2014. **28**(1): p. 118-28.
233. Enzler, T., et al., *Chronic lymphocytic leukemia of Emu-TCL1 transgenic mice undergoes rapid cell turnover that can be offset by extrinsic CD257 to accelerate disease progression.* Blood, 2009. **114**(20): p. 4469-76.
234. Lascano, V., et al., *Chronic lymphocytic leukemia disease progression is accelerated by APRIL-TACI interaction in the TCL1 transgenic mouse model.* Blood, 2013. **122**(24): p. 3960-3.
235. Widhopf, G.F., 2nd, et al., *ROR1 can interact with TCL1 and enhance leukemogenesis in Emu-TCL1 transgenic mice.* Proc Natl Acad Sci U S A, 2014. **111**(2): p. 793-8.
236. Santanam, U., et al., *Chronic lymphocytic leukemia modeled in mouse by targeted miR-29 expression.* Proc Natl Acad Sci U S A, 2010. **107**(27): p. 12210-5.
237. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.
238. Thieffry, D. and S. Sarkar, *Forty years under the central dogma.* Trends Biochem Sci, 1998. **23**(8): p. 312-6.
239. Hegde, P.S., I.R. White, and C. Debouck, *Interplay of transcriptomics and proteomics.* Curr Opin Biotechnol, 2003. **14**(6): p. 647-51.
240. Leroy, B., et al., *The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis.* Nucleic Acids Res, 2013. **41**(Database issue): p. D962-9.
241. Eddy, J.A., et al., *Relative expression analysis for molecular cancer diagnosis and prognosis.* Technol Cancer Res Treat, 2010. **9**(2): p. 149-59.
242. Maher, C.A., et al., *Transcriptome sequencing to detect gene fusions in cancer.* Nature, 2009. **458**(7234): p. 97-101.
243. Kozak, M., *Some thoughts about translational regulation: forward and backward glances.* J Cell Biochem, 2007. **102**(2): p. 280-90.
244. Silvera, D., S.C. Formenti, and R.J. Schneider, *Translational control in cancer.* Nat Rev Cancer, 2010. **10**(4): p. 254-66.
245. de Godoy, L.M.F., et al., *Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast.* Nature, 2008. **455**(7217): p. 1251-U60.
246. Wilkins, M.R., et al., *From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis.* Biotechnology (N Y), 1996. **14**(1): p. 61-5.
247. Cox, J. and M. Mann, *Is proteomics the new genomics?* Cell, 2007. **130**(3): p. 395-8.
248. Gygi, S.P., et al., *Correlation between protein and mRNA abundance in yeast.* Mol Cell Biol, 1999. **19**(3): p. 1720-30.
249. Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.* Nat Rev Genet, 2012. **13**(4): p. 227-32.
250. Duncan, M.W., R. Aebersold, and R.M. Caprioli, *The pros and cons of peptide-centric proteomics.* Nat Biotechnol, 2010. **28**(7): p. 659-64.

251. Altelaar, A.F., J. Munoz, and A.J. Heck, *Next-generation proteomics: towards an integrative view of proteome dynamics.* Nat Rev Genet, 2013. **14**(1): p. 35-48.
252. Gstaiger, M. and R. Aebersold, *Applying mass spectrometry-based proteomics to genetics, genomics and network biology.* Nat Rev Genet, 2009. **10**(9): p. 617-27.
253. Fagerberg, L., et al., *Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics.* Mol Cell Proteomics, 2014. **13**(2): p. 397-406.
254. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Anal Chem, 1988. **60**(20): p. 2299-301.
255. Hunt, D.F., et al., *Protein sequencing by tandem mass spectrometry.* Proc Natl Acad Sci U S A, 1986. **83**(17): p. 6233-7.
256. Wolters, D.A., M.P. Washburn, and J.R. Yates, 3rd, *An automated multidimensional protein identification technology for shotgun proteomics.* Anal Chem, 2001. **73**(23): p. 5683-90.
257. O'Farrell, P.H., *High resolution two-dimensional electrophoresis of proteins.* J Biol Chem, 1975. **250**(10): p. 4007-21.
258. Hanash, S.M., et al., *Lineage-related polypeptide markers in acute lymphoblastic leukemia detected by two-dimensional gel electrophoresis.* Proc Natl Acad Sci U S A, 1986. **83**(3): p. 807-11.
259. Shin, B.K., et al., *Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function.* J Biol Chem, 2003. **278**(9): p. 7607-16.
260. Lopez, J.L., *Two-dimensional electrophoresis in proteome expression analysis.* J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **849**(1-2): p. 190-202.
261. Gruber, K.A., et al., *Fluorometric assay of vasopressin and oxytocin: a general approach to the assay of peptides in tissues.* Proc Natl Acad Sci U S A, 1976. **73**(4): p. 1314-8.
262. Di Palma, S., et al., *Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis.* J Proteomics, 2012. **75**(13): p. 3791-813.
263. Switzar, L., M. Giera, and W.M. Niessen, *Protein digestion: an overview of the available techniques and recent developments.* J Proteome Res, 2013. **12**(3): p. 1067-77.
264. Washburn, M.P., D. Wolters, and J.R. Yates, 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotechnol, 2001. **19**(3): p. 242-7.
265. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
266. Kebarle, P. and U.H. Verkerk, *Electrospray: from ions in solution to ions in the gas phase, what we know now.* Mass Spectrom Rev, 2009. **28**(6): p. 898-917.
267. Yates, J.R., 3rd, et al., *Automated protein identification using microcolumn liquid chromatography-tandem mass spectrometry.* Methods Mol Biol, 1999. **112**: p. 553-69.
268. Hahne, H., et al., *DMSO enhances electrospray response, boosting sensitivity of proteomic experiments.* Nat Methods, 2013. **10**(10): p. 989-91.
269. Michalski, A., et al., *Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes.* Mol Cell Proteomics, 2012. **11**(3): p. O111 013698.
270. Kalli, A., et al., *Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: focus on LTQ-Orbitrap mass analyzers.* J Proteome Res, 2013. **12**(7): p. 3071-86.
271. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.* Anal Chem, 2000. **72**(6): p. 1156-62.
272. Wells, J.M. and S.A. McLuckey, *Collision-induced dissociation (CID) of peptides and proteins.* Methods Enzymol, 2005. **402**: p. 148-85.

273. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis.* Nat Methods, 2007. **4**(9): p. 709-12.
274. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomed Mass Spectrom, 1984. **11**(11): p. 601.
275. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.
276. Yates, J.R., 3rd, et al., *Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.* Anal Chem, 1995. **67**(8): p. 1426-36.
277. Pappin, D.J., P. Hojrup, and A.J. Bleasby, *Rapid identification of proteins by peptide-mass fingerprinting.* Curr Biol, 1993. **3**(6): p. 327-32.
278. Kall, L., et al., *Semi-supervised learning for peptide identification from shotgun proteomics datasets.* Nat Methods, 2007. **4**(11): p. 923-5.
279. Nesvizhskii, A.I. and R. Aebersold, *Interpretation of shotgun proteomic data: the protein inference problem.* Mol Cell Proteomics, 2005. **4**(10): p. 1419-40.
280. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry.* Anal Chem, 2003. **75**(17): p. 4646-58.
281. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present.* Anal Bioanal Chem, 2012. **404**(4): p. 939-65.
282. Old, W.M., et al., *Comparison of label-free methods for quantifying human proteins by shotgun proteomics.* Mol Cell Proteomics, 2005. **4**(10): p. 1487-502.
283. Bondarenko, P.V., D. Chelius, and T.A. Shaler, *Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry.* Anal Chem, 2002. **74**(18): p. 4741-9.
284. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review.* Anal Bioanal Chem, 2007. **389**(4): p. 1017-31.
285. Oda, Y., et al., *Accurate quantitation of protein expression and site-specific phosphorylation.* Proc Natl Acad Sci U S A, 1999. **96**(12): p. 6591-6.
286. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.
287. Ong, S.E., I. Kratchmarova, and M. Mann, *Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC).* J Proteome Res, 2003. **2**(2): p. 173-81.
288. Kruger, M., et al., *SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function.* Cell, 2008. **134**(2): p. 353-64.
289. Rose, K., et al., *A new mass-spectrometric C-terminal sequencing technique finds a similarity between gamma-interferon and alpha 2-interferon and identifies a proteolytically clipped gamma-interferon that retains full antiviral activity.* Biochem J, 1983. **215**(2): p. 273-7.
290. Miyagi, M. and K.C. Rao, *Proteolytic 18O-labeling strategies for quantitative proteomics.* Mass Spectrom Rev, 2007. **26**(1): p. 121-36.
291. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.
292. Zhang, R., et al., *Fractionation of isotopically labeled peptides in quantitative proteomics.* Anal Chem, 2001. **73**(21): p. 5142-9.
293. Schmidt, A., J. Kellermann, and F. Lottspeich, *A novel strategy for quantitative proteomics using isotope-coded protein labels.* Proteomics, 2005. **5**(1): p. 4-15.
294. Hsu, J.L., et al., *Stable-isotope dimethyl labeling for quantitative proteomics.* Anal Chem, 2003. **75**(24): p. 6843-52.

295. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal Chem, 2003. **75**(8): p. 1895-904.
296. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.
297. Choe, L., et al., *8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease.* Proteomics, 2007. **7**(20): p. 3651-60.
298. Dayon, L., et al., *Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags.* Anal Chem, 2008. **80**(8): p. 2921-31.
299. McAlister, G.C., et al., *Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses.* Anal Chem, 2012. **84**(17): p. 7469-78.
300. Werner, T., et al., *Ion coalescence of neutron encoded TMT 10-plex reporter ions.* Anal Chem, 2014. **86**(7): p. 3594-601.
301. Karp, N.A., et al., *Addressing accuracy and precision issues in iTRAQ quantitation.* Mol Cell Proteomics, 2010. **9**(9): p. 1885-97.
302. Ow, S.Y., et al., *iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly".* J Proteome Res, 2009. **8**(11): p. 5347-55.
303. McAlister, G.C., et al., *MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes.* Anal Chem, 2014. **86**(14): p. 7150-8.
304. Ow, S.Y., et al., *Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation.* Proteomics, 2011. **11**(11): p. 2341-6.
305. Savitski, M.M., et al., *Delayed fragmentation and optimized isolation width settings for improvement of protein identification and accuracy of isobaric mass tag quantification on Orbitrap-type mass spectrometers.* Anal Chem, 2011. **83**(23): p. 8959-67.
306. Shliaha, P.V., et al., *Additional Precursor Purification in Isobaric Mass Tagging Experiments by Traveling Wave Ion Mobility Separation (TWIMS).* J Proteome Res, 2014. **13**(7): p. 3360-9.
307. Ting, L., et al., *MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics.* Nat Methods, 2011. **8**(11): p. 937-40.
308. Wenger, C.D., et al., *Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging.* Nat Methods, 2011. **8**(11): p. 933-5.
309. Sandberg, A., et al., *Quantitative accuracy in mass spectrometry based proteomics of complex samples: the impact of labeling and precursor interference.* J Proteomics, 2014. **96**: p. 133-44.
310. Savitski, M.M., et al., *Measuring and managing ratio compression for accurate iTRAQ/TMT quantification.* J Proteome Res, 2013. **12**(8): p. 3586-98.
311. Li, H., et al., *Estimating Influence of Cofragmentation on Peptide Quantification and Identification in iTRAQ Experiments by Simulating Multiplexed Spectra.* J Proteome Res, 2014. **13**(7): p. 3488-97.
312. Hultin-Rosenberg, L., et al., *Defining, comparing, and improving iTRAQ quantification in mass spectrometry proteomics data.* Mol Cell Proteomics, 2013. **12**(7): p. 2021-31.
313. Hanash, S.M., S.J. Pitteri, and V.M. Faca, *Mining the plasma proteome for cancer biomarkers.* Nature, 2008. **452**(7187): p. 571-9.
314. Srivastava, S. and R.G. Srivastava, *Proteomics in the forefront of cancer biomarker discovery.* J Proteome Res, 2005. **4**(4): p. 1098-103.
315. Aebersold, R., et al., *Perspective: a program to improve protein biomarker discovery for cancer.* J Proteome Res, 2005. **4**(4): p. 1104-9.

316. Diamandis, E.P., *Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations.* Mol Cell Proteomics, 2004. **3**(4): p. 367-78.
317. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects.* Mol Cell Proteomics, 2002. **1**(11): p. 845-67.
318. Jones, K.A., et al., *Immunodepletion plasma proteomics by tripleTOF 5600 and Orbitrap elite/LTQ-Orbitrap Velos/Q exactive mass spectrometers.* J Proteome Res, 2013. **12**(10): p. 4351-65.
319. Patel, B.B., et al., *Assessment of two immunodepletion methods: off-target effects and variations in immunodepletion efficiency may confound plasma proteomics.* J Proteome Res, 2012. **11**(12): p. 5947-58.
320. Garbis, S.D., et al., *A novel multidimensional protein identification technology approach combining protein size exclusion prefractionation, peptide zwitterion-ion hydrophilic interaction chromatography, and nano-ultraperformance RP chromatography/nESI-MS2 for the in-depth analysis of the serum proteome and phosphoproteome: application to clinical sera derived from humans with benign prostate hyperplasia.* Anal Chem, 2011. **83**(3): p. 708-18.
321. Al-Daghri, N.M., et al., *Whole serum 3D LC-nESI-FTMS quantitative proteomics reveals sexual dimorphism in the milieu interieur of overweight and obese adults.* J Proteome Res, 2014. **13**(11): p. 5094-105.
322. Ochs, R.C. and A. Bagg, *Molecular genetic characterization of lymphoma: application to cytology diagnosis.* Diagn Cytopathol, 2012. **40**(6): p. 542-55.
323. Love, C., et al., *The genetic landscape of mutations in Burkitt lymphoma.* Nat Genet, 2012. **44**(12): p. 1321-5.
324. Di Lisio, L., et al., *MicroRNA signatures in B-cell lymphomas.* Blood Cancer J, 2012. **2**(2): p. e57.
325. Lenze, D., et al., *The different epidemiologic subtypes of Burkitt lymphoma share a homogenous micro RNA profile distinct from diffuse large B-cell lymphoma.* Leukemia, 2011. **25**(12): p. 1869-76.
326. Blume, C.J., et al., *p53-dependent non-coding RNA networks in chronic lymphocytic leukemia.* Leukemia, 2015.
327. Yustein, J.T., et al., *Induction of ectopic Myc target gene JAG2 augments hypoxic growth and tumorigenesis in a human B-cell model.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3534-9.
328. Xiao, W., et al., *High-throughput RNA sequencing in B-cell lymphomas.* Methods Mol Biol, 2013. **971**: p. 295-312.
329. Boyd, R.S., M.J. Dyer, and K. Cain, *Proteomic analysis of B-cell malignancies.* J Proteomics, 2010. **73**(10): p. 1804-22.
330. Ludvigsen, M., et al., *Proteomic approaches to the study of malignant lymphoma: analyses on patient samples.* Proteomics Clin Appl, 2015. **9**(1-2): p. 72-85.
331. Li, J., et al., *Proteomic characterization of primary diffuse large B-cell lymphomas in the central nervous system.* J Neurosurg, 2008. **109**(3): p. 536-46.
332. Fujii, K., et al., *Protein expression pattern distinguishes different lymphoid neoplasms.* Proteomics, 2005. **5**(16): p. 4274-86.
333. Kamper, P., et al., *Proteomic analysis identifies galectin-1 as a predictive biomarker for relapsed/refractory disease in classical Hodgkin lymphoma.* Blood, 2011. **117**(24): p. 6638-49.
334. Braoudaki, M., et al., *Protein biomarkers distinguish between high- and low-risk pediatric acute lymphoblastic leukemia in a tissue specific manner.* J Hematol Oncol, 2013. **6**: p. 52.

335. Polati, R., et al., *Tissue proteomics of splenic marginal zone lymphoma.* Electrophoresis, 2015. **36**(14): p. 1612-21.
336. Romesser, P.B., et al., *Development of a malignancy-associated proteomic signature for diffuse large B-cell lymphoma.* Am J Pathol, 2009. **175**(1): p. 25-35.
337. Gkiafi, Z. and G. Panayotou, *Comparative proteomic analysis implicates COMMD proteins as Epstein-Barr virus targets in the BL41 Burkitt's lymphoma cell line.* J Proteome Res, 2011. **10**(7): p. 2959-68.
338. Fernandez, M. and J.P. Albar, *2D DIGE for the analysis of RAMOS cells subproteomes.* Methods Mol Biol, 2012. **854**: p. 239-52.
339. Gurtler, A., et al., *The inter-individual variability outperforms the intra-individual variability of differentially expressed proteins prior and post irradiation in lymphoblastoid cell lines.* Arch Physiol Biochem, 2014. **120**(5): p. 198-207.
340. Klanova, M., et al., *Downregulation of deoxycytidine kinase in cytarabine-resistant mantle cell lymphoma cells confers cross-resistance to nucleoside analogs gemcitabine, fludarabine and cladribine, but not to other classes of anti-lymphoma agents.* Mol Cancer, 2014. **13**: p. 159.
341. Ghobrial, I.M., et al., *Proteomic analysis of mantle-cell lymphoma by protein microarray.* Blood, 2005. **105**(9): p. 3722-30.
342. Huang, X., et al., *Quantitative proteomics reveals that miR-155 regulates the PI3K-AKT pathway in diffuse large B-cell lymphoma.* Am J Pathol, 2012. **181**(1): p. 26-33.
343. Rolland, D., et al., *Global phosphoproteomic profiling reveals distinct signatures in B-cell non-Hodgkin lymphomas.* Am J Pathol, 2014. **184**(5): p. 1331-42.
344. Pighi, C., et al., *Phospho-proteomic analysis of mantle cell lymphoma cells suggests a pro-survival role of B-cell receptor signaling.* Cell Oncol (Dordr), 2011. **34**(2): p. 141-53.
345. Deeb, S.J., et al., *Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles.* Mol Cell Proteomics, 2012. **11**(5): p. 77-89.
346. Deeb, S.J., et al., *N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes.* Mol Cell Proteomics, 2014. **13**(1): p. 240-51.
347. Christopherson, R.I., et al., *Mechanisms of action of fludarabine nucleoside against human Raji lymphoma cells.* Nucleosides Nucleotides Nucleic Acids, 2014. **33**(4-6): p. 375-83.
348. Hofmann, A., et al., *Surfaceome of classical Hodgkin and non-Hodgkin lymphoma.* Proteomics Clin Appl, 2015. **9**(7-8): p. 661-70.
349. Diez, P., et al., *Integration of Proteomics and Transcriptomics Data Sets for the Analysis of a Lymphoma B-Cell Line in the Context of the Chromosome-Centric Human Proteome Project.* J Proteome Res, 2015.
350. Boyd, R.S., et al., *Protein profiling of plasma membranes defines aberrant signaling pathways in mantle cell lymphoma.* Mol Cell Proteomics, 2009. **8**(7): p. 1501-15.
351. Fujii, K., et al., *Proteomic study identified HSP 70 kDa protein 1A as a possible therapeutic target, in combination with histone deacetylase inhibitors, for lymphoid neoplasms.* J Proteomics, 2012. **75**(4): p. 1401-10.
352. Eagle, G.L., et al., *Total proteome analysis identifies migration defects as a major pathogenetic factor in immunoglobulin heavy chain variable region (IGHV)-unmutated chronic lymphocytic leukemia.* Mol Cell Proteomics, 2015. **14**(4): p. 933-45.
353. Alsagaby, S.A., et al., *Proteomics-based strategies to identify proteins relevant to chronic lymphocytic leukemia.* J Proteome Res, 2014. **13**(11): p. 5051-62.
354. Glibert, P., et al., *Quantitative proteomics to characterize specific histone H2A proteolysis in chronic lymphocytic leukemia and the myeloid THP-1 cell line.* Int J Mol Sci, 2014. **15**(6): p. 9407-21.

355. Kashuba, E., et al., *Proteomic analysis of B-cell receptor signaling in chronic lymphocytic leukaemia reveals a possible role for kininogen.* J Proteomics, 2013. **91**: p. 478-85.
356. O'Hayre, M., et al., *Elucidating the CXCL12/CXCR4 signaling network in chronic lymphocytic leukemia through phosphoproteomics analysis.* PLoS One, 2010. **5**(7): p. e11716.
357. Miguet, L., et al., *Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis.* J Proteome Res, 2009. **8**(7): p. 3346-54.
358. Barnidge, D.R., et al., *Quantitative protein expression analysis of CLL B cells from mutated and unmutated IgV(H) subgroups using acid-cleavable isotope-coded affinity tag reagents.* J Proteome Res, 2005. **4**(4): p. 1310-7.
359. Barnidge, D.R., et al., *Protein expression profiling of CLL B cells using replicate off-line strong cation exchange chromatography and LC-MS/MS.* J Chromatogr B Analyt Technol Biomed Life Sci, 2005. **819**(1): p. 33-9.
360. Cochran, D.A., et al., *Proteomic analysis of chronic lymphocytic leukemia subtypes with mutated or unmutated Ig V(H) genes.* Mol Cell Proteomics, 2003. **2**(12): p. 1331-41.
361. Perrot, A., et al., *A unique proteomic profile on surface IgM ligation in unmutated chronic lymphocytic leukemia.* Blood, 2011. **118**(4): p. e1-15.
362. Voss, T., et al., *Correlation of clinical data with proteomics profiles in 24 patients with B-cell chronic lymphocytic leukemia.* Int J Cancer, 2001. **91**(2): p. 180-6.
363. Boyd, R.S., et al., *Proteomic analysis of the cell-surface membrane in chronic lymphocytic leukemia: identification of two novel proteins, BCNP1 and MIG2B.* Leukemia, 2003. **17**(8): p. 1605-12.
364. Adeghe, A.J. and J. Cohen, *A better method for terminal bleeding of mice.* Lab Anim, 1986. **20**(1): p. 70-2.
365. Papachristou, E.K., et al., *The shotgun proteomic study of the human ThinPrep cervical smear using iTRAQ mass-tagging and 2D LC-FT-Orbitrap-MS: the detection of the human papillomavirus at the protein level.* J Proteome Res, 2013. **12**(5): p. 2078-89.
366. Levin, Y., *The role of statistical power analysis in quantitative proteomics.* Proteomics, 2011. **11**(12): p. 2565-7.
367. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
368. Bausch-Fluck, D., et al., *A mass spectrometric-derived cell surface protein atlas.* PLoS One, 2015. **10**(3): p. e0121314.
369. de Souza, J.E., et al., *SurfaceomeDB: a cancer-orientated database for genes encoding cell surface proteins.* Cancer Immun, 2012. **12**: p. 15.
370. Beroukhim, R., et al., *The landscape of somatic copy-number alteration across human cancers.* Nature, 2010. **463**(7283): p. 899-905.
371. Strasser, A., et al., *Novel primitive lymphoid tumours induced in transgenic mice by cooperation between myc and bcl-2.* Nature, 1990. **348**(6299): p. 331-3.
372. Ian Cumming, O.L., *A method of sorting cells*, in *European Patent Register*. 2011.
373. Geiger, T., et al., *Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins.* Mol Cell Proteomics, 2012. **11**(3): p. M111 014050.
374. !!! INVALID CITATION !!! {Besson, 2011 #78;Chang, 2011 #79;Voisin, 2011 #80;Geiger, 2012 #526}.
375. Steen, H. and M. Mann, *The ABC's (and XYZ's) of peptide sequencing.* Nat Rev Mol Cell Biol, 2004. **5**(9): p. 699-711.

376. Pan, C., et al., *Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions.* Mol Cell Proteomics, 2009. **8**(3): p. 443-50.
377. Colley, S.M., P.A. Tilbrook, and S.P. Klinken, *Increased transcription of the E mu-myc transgene and mRNA stabilisation produce only a modest elevation in Myc protein.* Oncogene, 1997. **14**(22): p. 2735-9.
378. Asano, S., M. Mishima, and E. Nishida, *Coronin forms a stable dimer through its C-terminal coiled coil region: an implicated role in its localization to cell periphery.* Genes Cells, 2001. **6**(3): p. 225-35.
379. Galkin, V.E., et al., *Coronin-1A stabilizes F-actin by bridging adjacent actin protomers and stapling opposite strands of the actin filament.* J Mol Biol, 2008. **376**(3): p. 607-13.
380. Mueller, P., et al., *Regulation of T cell survival through coronin-1-mediated generation of inositol-1,4,5-trisphosphate and calcium mobilization after T cell receptor triggering.* Nat Immunol, 2008. **9**(4): p. 424-31.
381. Foger, N., et al., *Requirement for coronin 1 in T lymphocyte trafficking and cellular homeostasis.* Science, 2006. **313**(5788): p. 839-42.
382. Pollard, T.D. and G.G. Borisy, *Cellular motility driven by assembly and disassembly of actin filaments.* Cell, 2003. **112**(4): p. 453-65.
383. Haraldsson, M.K., et al., *The lupus-related Lmb3 locus contains a disease-suppressing Coronin-1A gene mutation.* Immunity, 2008. **28**(1): p. 40-51.
384. Honore, B., S. Buus, and M.H. Claesson, *Identification of differentially expressed proteins in spontaneous thymic lymphomas from knockout mice with deletion of p53.* Proteome Sci, 2008. **6**: p. 18.
385. Sreedhar, A.S., et al., *Hsp90 isoforms: functions, expression and clinical importance.* FEBS Lett, 2004. **562**(1-3): p. 11-5.
386. DeSouza, L.V., et al., *Endometrial carcinoma biomarker discovery and verification using differentially tagged clinical samples with multidimensional liquid chromatography and tandem mass spectrometry.* Mol Cell Proteomics, 2007. **6**(7): p. 1170-82.
387. DeSouza, L.V., et al., *Absolute quantification of potential cancer markers in clinical tissue homogenates using multiple reaction monitoring on a hybrid triple quadrupole/linear ion trap tandem mass spectrometer.* Anal Chem, 2009. **81**(9): p. 3462-70.
388. Koopman, G., et al., *Annexin V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis.* Blood, 1994. **84**(5): p. 1415-20.
389. Juin, P., et al., *c-Myc-induced sensitization to apoptosis is mediated through cytochrome c release.* Genes Dev, 1999. **13**(11): p. 1367-81.
390. Johansson, H.J., et al., *Retinoic acid receptor alpha is associated with tamoxifen resistance in breast cancer.* Nat Commun, 2013. **4**: p. 2175.
391. Sheng, Q., et al., *Preprocessing Significantly Improves the Peptide/Protein Identification Sensitivity of High-resolution Isobarically Labeled Tandem Mass Spectrometry Data.* Mol Cell Proteomics, 2015. **14**(2): p. 405-17.
392. Yan, X.J., et al., *B cell receptors in TCL1 transgenic mice resemble those of aggressive, treatment-resistant human chronic lymphocytic leukemia.* Proc Natl Acad Sci U S A, 2006. **103**(31): p. 11713-8.
393. Rickert, R.C., *New insights into pre-BCR and BCR signalling with relevance to B cell malignancies.* Nat Rev Immunol, 2013. **13**(8): p. 578-91.
394. Post, S.M., et al., *p53-dependent senescence delays Emu-myc-induced B-cell lymphomagenesis.* Oncogene, 2010. **29**(9): p. 1260-9.
395. Kon, N., et al., *Inactivation of arf-bp1 induces p53 activation and diabetic phenotypes in mice.* J Biol Chem, 2012. **287**(7): p. 5102-11.

396. Qiu, F., et al., *Arginine starvation impairs mitochondrial respiratory function in ASS1-deficient breast cancer cells.* Sci Signal, 2014. **7**(319): p. ra31.
397. Miraki-Moud, F., et al., *Arginine deprivation using pegylated arginine deiminase has activity against primary acute myeloid leukemia cells in vivo.* Blood, 2015. **125**(26): p. 4060-8.
398. Shan, Y.S., et al., *Argininosuccinate synthetase 1 suppression and arginine restriction inhibit cell migration in gastric cancer cell lines.* Sci Rep, 2015. **5**: p. 9783.
399. Changou, C.A., et al., *Arginine starvation-associated atypical cellular death involves mitochondrial dysfunction, nuclear DNA leakage, and chromatin autophagy.* Proc Natl Acad Sci U S A, 2014. **111**(39): p. 14147-52.
400. Tan, G.S., et al., *Novel proteomic biomarker panel for prediction of aggressive metastatic hepatocellular carcinoma relapse in surgically resectable patients.* J Proteome Res, 2014. **13**(11): p. 4833-46.
401. Shan, Y.S., et al., *Increased expression of argininosuccinate synthetase protein predicts poor prognosis in human gastric cancer.* Oncol Rep, 2015. **33**(1): p. 49-57.
402. Tsai, W.B., et al., *Resistance to arginine deiminase treatment in melanoma cells is associated with induced argininosuccinate synthetase expression involving c-Myc/HIF-1alpha/Sp4.* Mol Cancer Ther, 2009. **8**(12): p. 3223-33.
403. Warburg, O., *On the origin of cancer cells.* Science, 1956. **123**(3191): p. 309-14.
404. Solana, R., et al., *MHC class I antigen expression is inversely related with tumor malignancy and ras oncogene product (p21ras) levels in human breast tumors.* Invasion Metastasis, 1992. **12**(3-4): p. 210-7.
405. Paulson, K.G., et al., *Downregulation of MHC-I expression is prevalent but reversible in Merkel cell carcinoma.* Cancer Immunol Res, 2014. **2**(11): p. 1071-9.
406. Friedmann-Morvinski, D. and I.M. Verma, *Dedifferentiation and reprogramming: origins of cancer stem cells.* EMBO Rep, 2014. **15**(3): p. 244-53.
407. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
408. Iacovelli, S., et al., *Two types of BCR interactions are positively selected during leukemia development in the Emu-TCL1 transgenic mouse model of CLL.* Blood, 2015. **125**(10): p. 1578-88.
409. Cheung-Ong, K., G. Giaever, and C. Nislow, *DNA-damaging agents in cancer chemotherapy: serendipity and chemical biology.* Chem Biol, 2013. **20**(5): p. 648-59.
410. Bouwman, P. and J. Jonkers, *The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance.* Nat Rev Cancer, 2012. **12**(9): p. 587-98.
411. Rai, K.R., et al., *Fludarabine compared with chlorambucil as primary therapy for chronic lymphocytic leukemia.* N Engl J Med, 2000. **343**(24): p. 1750-7.
412. Wu, J., et al., *The up-regulation of histone deacetylase 8 promotes proliferation and inhibits apoptosis in hepatocellular carcinoma.* Dig Dis Sci, 2013. **58**(12): p. 3545-53.
413. Lee, H., et al., *Histone deacetylase 8 safeguards the human ever-shorter telomeres 1B (hEST1B) protein from ubiquitin-mediated degradation.* Mol Cell Biol, 2006. **26**(14): p. 5259-69.
414. Qian, Y., et al., *DEC1 coordinates with HDAC8 to differentially regulate TAp73 and DeltaNp73 expression.* PLoS One, 2014. **9**(1): p. e84015.
415. Yan, W., et al., *Histone deacetylase inhibitors suppress mutant p53 transcription via histone deacetylase 8.* Oncogene, 2013. **32**(5): p. 599-609.
416. Gao, S.M., et al., *Histone deacetylases inhibitor sodium butyrate inhibits JAK2/STAT signaling through upregulation of SOCS1 and SOCS3 mediated by HDAC8 inhibition in myeloproliferative neoplasms.* Exp Hematol, 2013. **41**(3): p. 261-70 e4.

417. Hu, E., et al., *Cloning and characterization of a novel human class I histone deacetylase that functions as a transcription repressor.* J Biol Chem, 2000. **275**(20): p. 15254-64.
418. Choudhary, C., et al., *Lysine acetylation targets protein complexes and co-regulates major cellular functions.* Science, 2009. **325**(5942): p. 834-40.
419. Tang, W., et al., *Discovery of histone deacetylase 8 selective inhibitors.* Bioorg Med Chem Lett, 2011. **21**(9): p. 2601-5.
420. Krennhrubec, K., et al., *Design and evaluation of 'Linkerless' hydroxamic acids as selective HDAC8 inhibitors.* Bioorg Med Chem Lett, 2007. **17**(10): p. 2874-8.
421. Balasubramanian, S., et al., *A novel histone deacetylase 8 (HDAC8)-specific inhibitor PCI-34051 induces apoptosis in T-cell lymphomas.* Leukemia, 2008. **22**(5): p. 1026-34.
422. Macheda, M.L., S. Rogers, and J.D. Best, *Molecular and cellular regulation of glucose transporter (GLUT) proteins in cancer.* J Cell Physiol, 2005. **202**(3): p. 654-62.
423. Polanski, R., et al., *Activity of the monocarboxylate transporter 1 inhibitor AZD3965 in small cell lung cancer.* Clin Cancer Res, 2014. **20**(4): p. 926-37.
424. Diers, A.R., et al., *Pyruvate fuels mitochondrial respiration and proliferation of breast cancer cells: effect of monocarboxylate transporter inhibition.* Biochem J, 2012. **444**(3): p. 561-71.
425. Karunakaran, S., et al., *SLC6A14 (ATB0,+) protein, a highly concentrative and broad specific amino acid transporter, is a novel and effective drug target for treatment of estrogen receptor-positive breast cancer.* J Biol Chem, 2011. **286**(36): p. 31830-8.
426. Babu, E., et al., *Deletion of the amino acid transporter Slc6a14 suppresses tumour growth in spontaneous mouse models of breast cancer.* Biochem J, 2015. **469**(1): p. 17-23.
427. Imai, H., et al., *Inhibition of L-type amino acid transporter 1 has antitumor activity in non-small cell lung cancer.* Anticancer Res, 2010. **30**(12): p. 4819-28.
428. Franklin, R.B., et al., *ZIP14 zinc transporter downregulation and zinc depletion in the development and progression of hepatocellular cancer.* J Gastrointest Cancer, 2012. **43**(2): p. 249-57.
429. Zou, J., et al., *hZIP1 zinc transporter down-regulation in prostate cancer involves the overexpression of ras responsive element binding protein-1 (RREB-1).* Prostate, 2011. **71**(14): p. 1518-24.
430. Franklin, R.B., et al., *hZIP1 zinc uptake transporter down regulation and zinc depletion in prostate cancer.* Mol Cancer, 2005. **4**: p. 32.
431. Henshall, S.M., et al., *Expression of the zinc transporter ZnT4 is decreased in the progression from early prostate disease to invasive prostate cancer.* Oncogene, 2003. **22**(38): p. 6005-12.
432. Jin, J., et al., *Knockdown of zinc transporter ZIP5 (SLC39A5) expression significantly inhibits human esophageal cancer progression.* Oncol Rep, 2015. **34**(3): p. 1431-9.
433. Wu, L., et al., *Zinc transporter genes and urological cancers: integrated analysis suggests a role for ZIP11 in bladder cancer.* Tumour Biol, 2015.
434. Li, M., et al., *Aberrant expression of zinc transporter ZIP4 (SLC39A4) significantly contributes to human pancreatic cancer pathogenesis and progression.* Proc Natl Acad Sci U S A, 2007. **104**(47): p. 18636-41.
435. Zhang, Y., et al., *ZIP4 regulates pancreatic cancer cell growth by activating IL-6/STAT3 pathway through zinc finger transcription factor CREB.* Clin Cancer Res, 2010. **16**(5): p. 1423-30.
436. Tabarkiewicz, J. and K. Giannopoulos, *Definition of a target for immunotherapy and results of the first Peptide vaccination study in chronic lymphocytic leukemia.* Transplant Proc, 2010. **42**(8): p. 3293-6.

437. Snauwaert, S., et al., *RHAMM/HMMR (CD168) is not an ideal target antigen for immunotherapy of acute myeloid leukemia.* Haematologica, 2012. **97**(10): p. 1539-47.
438. Gurski, L.A., et al., *Hyaluronan (HA) interacting proteins RHAMM and hyaluronidase impact prostate cancer cell behavior and invadopodia formation in 3D HA-based hydrogels.* PLoS One, 2012. **7**(11): p. e50075.
439. Augustin, F., et al., *Receptor for hyaluronic acid-mediated motility (RHAMM, CD168) expression is prognostically important in both nodal negative and nodal positive large cell lung cancer.* J Clin Pathol, 2015. **68**(5): p. 368-73.
440. Ishigami, S., et al., *Prognostic impact of CD168 expression in gastric cancer.* BMC Cancer, 2011. **11**: p. 106.
441. Nagel, S., et al., *Coexpression of CD44 variant isoforms and receptor for hyaluronic acid-mediated motility (RHAMM, CD168) is an International Prognostic Index and C-MYC gene status-independent predictor of poor outcome in diffuse large B-cell lymphomas.* Exp Hematol, 2010. **38**(1): p. 38-45.
442. Gust, K.M., et al., *RHAMM (CD168) is overexpressed at the protein level and may constitute an immunogenic antigen in advanced prostate cancer disease.* Neoplasia, 2009. **11**(9): p. 956-63.
443. Slavin, S. and S. Strober, *Spontaneous murine B-cell leukaemia.* Nature, 1978. **272**(5654): p. 624-6.
444. O'Garra, A., et al., *The BCL1 B lymphoma responds to IL-4, IL-5, and GM-CSF.* Cell Immunol, 1989. **123**(1): p. 189-200.
445. Lohoff, M., F. Sommer, and M. Rollinghoff, *Suppressive effect of interferon-gamma on the BCL1 cell-dependent interleukin 5 bioassay.* Eur J Immunol, 1989. **19**(7): p. 1327-9.
446. Lasky, J.L. and G.J. Thorbecke, *Characterization and growth factor requirements of SJL lymphomas. II. Interleukin 5 dependence of the in vitro cell line, cRCS-X, and influence of other cytokines.* Eur J Immunol, 1989. **19**(2): p. 365-71.
447. Wen, X., et al., *Transgene-mediated hyper-expression of IL-5 inhibits autoimmune disease but increases the risk of B cell chronic lymphocytic leukemia in a model of murine lupus.* Eur J Immunol, 2004. **34**(10): p. 2740-9.
448. Rolink, A.G., et al., *Characterization of the interleukin 5-reactive splenic B cell population.* Eur J Immunol, 1990. **20**(9): p. 1949-56.
449. Wetzel, G.D., *Interleukin 5 regulation of peritoneal Ly-1 B lymphocyte proliferation, differentiation and autoantibody secretion.* Eur J Immunol, 1989. **19**(9): p. 1701-7.
450. Sanderson, C.J., *Interleukin-5: an eosinophil growth and activation factor.* Dev Biol Stand, 1988. **69**: p. 23-9.
451. Kopf, M., et al., *IL-5-deficient mice have a developmental defect in CD5+ B-1 cells and lack eosinophilia but have normal antibody and cytotoxic T cell responses.* Immunity, 1996. **4**(1): p. 15-24.
452. Sato, S., et al., *IL-5 receptor-mediated tyrosine phosphorylation of SH2/SH3-containing proteins and activation of Bruton's tyrosine and Janus 2 kinases.* J Exp Med, 1994. **180**(6): p. 2101-11.
453. Coffer, P.J., et al., *Analysis of signal transduction pathways in human eosinophils activated by chemoattractants and the T-helper 2-derived cytokines interleukin-4 and interleukin-5.* Blood, 1998. **91**(7): p. 2547-57.
454. Adachi, T. and R. Alam, *The mechanism of IL-5 signal transduction.* Am J Physiol, 1998. **275**(3 Pt 1): p. C623-33.
455. Mainou-Fowler, T. and A.G. Prentice, *Modulation of apoptosis with cytokines in B-cell chronic lymphocytic leukaemia.* Leuk Lymphoma, 1996. **21**(5-6): p. 369-77.

456. Clutterbuck, E., et al., *Recombinant human interleukin 5 is an eosinophil differentiation factor but has no activity in standard human B cell growth factor assays.* Eur J Immunol, 1987. **17**(12): p. 1743-50.
457. Keshishian, H., et al., *Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury.* Mol Cell Proteomics, 2015. **14**(9): p. 2375-93.
458. van Winden, A.W., et al., *Serum degradome markers for the detection of breast cancer.* J Proteome Res, 2010. **9**(8): p. 3781-8.
459. Karczmarski, J., et al., *Pre-analytical-related variability influencing serum peptide profiles demonstrated in a mass spectrometry-based search for colorectal and prostate cancer biomarkers.* Acta Biochim Pol, 2013. **60**(3): p. 417-25.
460. Yang, J., et al., *Identification of potential serum proteomic biomarkers for clear cell renal cell carcinoma.* PLoS One, 2014. **9**(11): p. e111364.
461. Davidson, M.B., et al., *Pathophysiology, clinical consequences, and treatment of tumor lysis syndrome.* Am J Med, 2004. **116**(8): p. 546-54.
462. Wossmann, W., et al., *Incidence of tumor lysis syndrome in children with advanced stage Burkitt's lymphoma/leukemia before and after introduction of prophylactic use of urate oxidase.* Ann Hematol, 2003. **82**(3): p. 160-5.
463. Otrock, Z.K., H.A. Hatoum, and Z.M. Salem, *Acute tumor lysis syndrome after rituximab administration in Burkitt's lymphoma.* Intern Emerg Med, 2008. **3**(2): p. 161-3.
464. Ludwig, A.K. and B. Giebel, *Exosomes: small vesicles participating in intercellular communication.* Int J Biochem Cell Biol, 2012. **44**(1): p. 11-5.
465. Shimizu, K., et al., *Integrity of intracellular domain of Notch ligand is indispensable for cleavage required for release of the Notch2 intracellular domain.* EMBO J, 2002. **21**(3): p. 294-302.
466. Inlay, M.A., et al., *Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development.* Genes Dev, 2009. **23**(20): p. 2376-81.
467. Flach, H., et al., *Mzb1 protein regulates calcium homeostasis, antibody secretion, and integrin activation in innate-like B cells.* Immunity, 2010. **33**(5): p. 723-35.
468. Herold, T., et al., *High expression of MZB1 predicts adverse prognosis in chronic lymphocytic leukemia, follicular lymphoma and diffuse large B-cell lymphoma and is associated with a unique gene expression signature.* Leuk Lymphoma, 2013. **54**(8): p. 1652-7.
469. Oing, C., et al., *Aberrant DNA hypermethylation of the ITIH5 tumor suppressor gene in acute myeloid leukemia.* Clin Epigenetics, 2011. **2**(2): p. 419-23.
470. Hamm, A., et al., *Frequent expression loss of Inter-alpha-trypsin inhibitor heavy chain (ITIH) genes in multiple human solid tumors: a systematic expression analysis.* BMC Cancer, 2008. **8**: p. 25.
471. Burkhart, J.M., et al., *The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways.* Blood, 2012. **120**(15): p. e73-82.
472. Kawai, T., H. Sanjo, and S. Akira, *Duet is a novel serine/threonine kinase with Dbl-Homology (DH) and Pleckstrin-Homology (PH) domains.* Gene, 1999. **227**(2): p. 249-55.
473. Dinchuk, J.E., et al., *Absence of post-translational aspartyl beta-hydroxylation of epidermal growth factor domains in mice leads to developmental defects and an increased incidence of intestinal neoplasia.* J Biol Chem, 2002. **277**(15): p. 12970-7.
474. Yoshimura, S., et al., *Family-wide characterization of the DENN domain Rab GDP-GTP exchange factors.* J Cell Biol, 2010. **191**(2): p. 367-81.

475. Zhu, M., et al., *Molecular cloning of a novel gene encoding a membrane-associated adaptor protein (LAX) in lymphocyte signaling.* J Biol Chem, 2002. **277**(48): p. 46151-8.
476. Chen, J., et al., *CD22 attenuates calcium signaling by potentiating plasma membrane calcium-ATPase activity.* Nat Immunol, 2004. **5**(6): p. 651-7.
477. Jackson, T.A., et al., *FcR-like 2 Inhibition of B cell receptor-mediated activation of B cells.* J Immunol, 2010. **185**(12): p. 7405-12.
478. Nuckel, H., et al., *FCRL2 mRNA expression is inversely associated with clinical progression in chronic lymphocytic leukemia.* Eur J Haematol, 2009. **83**(6): p. 541-9.
479. Li, F.J., et al., *FCRL2 expression predicts IGHV mutation status and clinical progression in chronic lymphocytic leukemia.* Blood, 2008. **112**(1): p. 179-87.
480. Zhu, Z., et al., *FCRL5 exerts binary and compartment-specific influence on innate-like B-cell receptor signaling.* Proc Natl Acad Sci U S A, 2013. **110**(14): p. E1282-90.
481. Kazemi, T., et al., *Fc receptor-like 1-5 molecules are similarly expressed in progressive and indolent clinical subtypes of B-cell chronic lymphocytic leukemia.* Int J Cancer, 2008. **123**(9): p. 2113-9.
482. Owczarek, S. and V. Berezin, *Neuroplastin: cell adhesion molecule and signaling receptor.* Int J Biochem Cell Biol, 2012. **44**(1): p. 1-5.
483. Owczarek, S., et al., *Neuroplastin-55 binds to and signals through the fibroblast growth factor receptor.* FASEB J, 2010. **24**(4): p. 1139-50.
484. Wu, R.P., et al., *Nrf2 responses and the therapeutic selectivity of electrophilic compounds in chronic lymphocytic leukemia.* Proc Natl Acad Sci U S A, 2010. **107**(16): p. 7479-84.
485. Jitschin, R., et al., *Mitochondrial metabolism contributes to oxidative stress and reveals therapeutic targets in chronic lymphocytic leukemia.* Blood, 2014. **123**(17): p. 2663-72.
486. Broggini, M., et al., *Aplidine, a new anticancer agent of marine origin, inhibits vascular endothelial growth factor (VEGF) secretion and blocks VEGF-VEGFR-1 (flt-1) autocrine loop in human leukemia cells MOLT-4.* Leukemia, 2003. **17**(1): p. 52-9.
487. Cuadrado, A., et al., *JNK activation is critical for Aplidin-induced apoptosis.* Oncogene, 2004. **23**(27): p. 4673-80.
488. McCaig, A.M., et al., *Dasatinib inhibits B cell receptor signalling in chronic lymphocytic leukaemia but novel combination approaches are required to overcome additional pro-survival microenvironmental signals.* Br J Haematol, 2011. **153**(2): p. 199-211.
489. Berton, G., et al., *Beta 2 integrin-dependent protein tyrosine phosphorylation and activation of the FGR protein tyrosine kinase in human neutrophils.* J Cell Biol, 1994. **126**(4): p. 1111-21.
490. Lanasa, M.C., et al., *Final results of EFC6663: a multicenter, international, phase 2 study of alvocidib for patients with fludarabine-refractory chronic lymphocytic leukemia.* Leuk Res, 2015. **39**(5): p. 495-500.
491. Jordaan, G., W. Liao, and S. Sharma, *E-cadherin gene re-expression in chronic lymphocytic leukemia cells by HDAC inhibitors.* BMC Cancer, 2013. **13**: p. 88.
492. Bachmann, S.B., et al., *DTX3L and ARTD9 inhibit IRF1 expression and mediate in cooperation with ARTD8 survival and proliferation of metastatic prostate cancer cells.* Mol Cancer, 2014. **13**: p. 125.
493. Dayal, S., et al., *Suppression of the deubiquitinating enzyme USP5 causes the accumulation of unanchored polyubiquitin and the activation of p53.* J Biol Chem, 2009. **284**(8): p. 5030-41.
494. Jakob, S., et al., *Nuclear protein tyrosine phosphatase Shp-2 is one important negative regulator of nuclear export of telomerase reverse transcriptase.* J Biol Chem, 2008. **283**(48): p. 33155-61.

495. Dunphy, J.L., et al., *The Arf6 GEF GEP100/BRAG2 regulates cell adhesion by controlling endocytosis of beta1 integrins.* Curr Biol, 2006. **16**(3): p. 315-20.
496. Bensaad, K., et al., *TIGAR, a p53-inducible regulator of glycolysis and apoptosis.* Cell, 2006. **126**(1): p. 107-20.
497. Potting, C., et al., *TRIAP1/PRELI complexes prevent apoptosis by mediating intramitochondrial transport of phosphatidic acid.* Cell Metab, 2013. **18**(2): p. 287-95.
498. Giordano, F., et al., *PI(4,5)P(2)-dependent and Ca(2+)-regulated ER-PM interactions mediated by the extended synaptotagmins.* Cell, 2013. **153**(7): p. 1494-509.
499. Wu, J., et al., *Identification of substrates of human protein-tyrosine phosphatase PTPN22.* J Biol Chem, 2006. **281**(16): p. 11002-10.
500. Geraghty, D.E., et al., *Mapping HLA for single nucleotide polymorphisms.* Rev Immunogenet, 1999. **1**(2): p. 231-8.
501. Ojha, J., et al., *Deep sequencing identifies genetic heterogeneity and recurrent convergent evolution in chronic lymphocytic leukemia.* Blood, 2015. **125**(3): p. 492-8.
502. Hamblin, T., *Chronic lymphocytic leukaemia: one disease or two?* Ann Hematol, 2002. **81**(6): p. 299-303.
503. McCarthy, B.A., et al., *A seven-gene expression panel distinguishing clonal expansions of pre-leukemic and chronic lymphocytic leukemia B cells from normal B lymphocytes.* Immunol Res, 2015.
504. Razzaq, T.M., et al., *Functional regulation of tissue plasminogen activator on the surface of vascular smooth muscle cells by the type-II transmembrane protein p63 (CKAP4).* J Biol Chem, 2003. **278**(43): p. 42679-85.
505. Gupta, N., et al., *Identification and characterization of p63 (CKAP4/ERGIC-63/CLIMP-63), a surfactant protein A binding protein, on type II pneumocytes.* Am J Physiol Lung Cell Mol Physiol, 2006. **291**(3): p. L436-46.
506. Conrads, T.P., et al., *CKAP4/p63 is a receptor for the frizzled-8 protein-related antiproliferative factor from interstitial cystitis patients.* J Biol Chem, 2006. **281**(49): p. 37836-43.
507. Nikonov, A.V., et al., *Climp-63-mediated binding of microtubules to the ER affects the lateral mobility of translocon complexes.* J Cell Sci, 2007. **120**(Pt 13): p. 2248-58.
508. Dejgaard, K., et al., *Organization of the Sec61 translocon, studied by high resolution native electrophoresis.* J Proteome Res, 2010. **9**(4): p. 1763-71.
509. Pepin, G., M.P. Perron, and P. Provost, *Regulation of human Dicer by the resident ER membrane protein CLIMP-63.* Nucleic Acids Res, 2012. **40**(22): p. 11603-17.
510. Schweizer, A., et al., *Reassessment of the subcellular localization of p63.* J Cell Sci, 1995. **108 (Pt 6)**: p. 2477-85.
511. Shahjee, H.M., et al., *Antiproliferative factor decreases Akt phosphorylation and alters gene expression via CKAP4 in T24 bladder carcinoma cells.* J Exp Clin Cancer Res, 2010. **29**: p. 160.
512. Li, S.X., et al., *CKAP4 inhibited growth and metastasis of hepatocellular carcinoma through regulating EGFR signaling.* Tumour Biol, 2014. **35**(8): p. 7999-8005.
513. Graham, S.A., et al., *Prolectin, a glycan-binding receptor on dividing B cells in germinal centers.* J Biol Chem, 2009. **284**(27): p. 18537-44.
514. Oellerich, T., et al., *The B-cell antigen receptor signals through a preformed transducer module of SLP65 and CIN85.* EMBO J, 2011. **30**(17): p. 3620-34.
515. Kaetzel, C.S., et al., *The polymeric immunoglobulin receptor (secretory component) mediates transport of immune complexes across epithelial cells: a local defense function for IgA.* Proc Natl Acad Sci U S A, 1991. **88**(19): p. 8796-800.
516. Song, W., et al., *Stimulation of transcytosis of the polymeric immunoglobulin receptor by dimeric IgA.* Proc Natl Acad Sci U S A, 1994. **91**(1): p. 163-6.

517. Eiffert, H., et al., *Determination of the molecular structure of the human free secretory component.* Biol Chem Hoppe Seyler, 1991. **372**(2): p. 119-28.
518. Park, J., et al., *Discovery and Validation of Biomarkers That Distinguish Mucinous and Nonmucinous Pancreatic Cysts.* Cancer Res, 2015. **75**(16): p. 3227-35.
519. Niu, H., K. Wang, and Y. Wang, *Polymeric immunoglobulin receptor expression is predictive of poor prognosis in glioma patients.* Int J Clin Exp Med, 2014. **7**(8): p. 2185-90.
520. Liu, F., et al., *COLORECTAL Polymeric Immunoglobulin Receptor Expression is Correlated with Hepatic Metastasis and Poor Prognosis in Colon Carcinoma Patients with Hepatic Metastasis.* Hepatogastroenterology, 2014. **61**(131): p. 652-9.
521. Wang, X., et al., *Polymeric immunoglobulin receptor expression is correlated with poor prognosis in patients with osteosarcoma.* Mol Med Rep, 2014. **9**(6): p. 2105-10.
522. Fristedt, R., et al., *Expression and prognostic significance of the polymeric immunoglobulin receptor in esophageal and gastric adenocarcinoma.* J Transl Med, 2014. **12**: p. 83.
523. Ai, J., et al., *The role of polymeric immunoglobulin receptor in inflammation-induced tumor metastasis of human hepatocellular carcinoma.* J Natl Cancer Inst, 2011. **103**(22): p. 1696-712.
524. Zhou, Y., et al., *Free radical stress in chronic lymphocytic leukemia cells and its role in cellular sensitivity to ROS-generating anticancer agents.* Blood, 2003. **101**(10): p. 4098-104.
525. Fiskus, W., et al., *Auranofin induces lethal oxidative and endoplasmic reticulum stress and exerts potent preclinical activity against chronic lymphocytic leukemia.* Cancer Res, 2014. **74**(9): p. 2520-32.
526. Garcia-Manero, G., et al., *Phase 1 study of the histone deacetylase inhibitor vorinostat (suberoylanilide hydroxamic acid [SAHA]) in patients with advanced leukemias and myelodysplastic syndromes.* Blood, 2008. **111**(3): p. 1060-6.
527. Van Damme, M., et al., *HDAC isoenzyme expression is deregulated in chronic lymphocytic leukemia B-cells and has a complex prognostic significance.* Epigenetics, 2012. **7**(12): p. 1403-12.
528. Byrd, J.C., et al., *A phase 1 and pharmacodynamic study of depsipeptide (FK228) in chronic lymphocytic leukemia and acute myeloid leukemia.* Blood, 2005. **105**(3): p. 959-67.
529. Lucas, D.M., et al., *The histone deacetylase inhibitor MS-275 induces caspase-dependent apoptosis in B-cell chronic lymphocytic leukemia cells.* Leukemia, 2004. **18**(7): p. 1207-14.
530. Fabris, S., et al., *Molecular and transcriptional characterization of 17p loss in B-cell chronic lymphocytic leukemia.* Genes Chromosomes Cancer, 2008. **47**(9): p. 781-93.
531. Ferreira, P.G., et al., *Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia.* Genome Res, 2014. **24**(2): p. 212-26.
532. Hartmann, T.N., et al., *Circulating B-cell chronic lymphocytic leukemia cells display impaired migration to lymph nodes and bone marrow.* Cancer Res, 2009. **69**(7): p. 3121-30.
533. Barrington, R.A., et al., *Involvement of NFAT1 in B cell self-tolerance.* J Immunol, 2006. **177**(3): p. 1510-5.
534. Akerlund, J., A. Getahun, and J.C. Cambier, *B cell expression of the SH2-containing inositol 5-phosphatase (SHIP-1) is required to establish anergy to high affinity, proteinacious autoantigens.* J Autoimmun, 2015. **62**: p. 45-54.

535. Lanham, S., et al., *Differential signaling via surface IgM is associated with VH gene mutational status and CD38 expression in chronic lymphocytic leukemia.* Blood, 2003. **101**(3): p. 1087-93.
536. Mockridge, C.I., et al., *Reversible anergy of sIgM-mediated signaling in the two subsets of CLL defined by VH-gene mutational status.* Blood, 2007. **109**(10): p. 4424-31.
537. Delgado, P., et al., *Essential function for the GTPase TC21 in homeostatic antigen receptor signaling.* Nat Immunol, 2009. **10**(8): p. 880-8.
538. Krysov, S., et al., *Surface IgM of CLL cells displays unusual glycans indicative of engagement of antigen in vivo.* Blood, 2010. **115**(21): p. 4198-205.
539. Kaucka, M., et al., *Post-translational modifications regulate signalling by Ror1.* Acta Physiol (Oxf), 2011. **203**(3): p. 351-62.
540. Cho, S.H., et al., *PARP-14, a member of the B aggressive lymphoma family, transduces survival signals in primary B cells.* Blood, 2009. **113**(11): p. 2416-25.
541. Tanigaki, K. and T. Honjo, *Regulation of lymphocyte development by Notch signaling.* Nat Immunol, 2007. **8**(5): p. 451-6.
542. Wang, L., et al., *SF3B1 and other novel cancer genes in chronic lymphocytic leukemia.* N Engl J Med, 2011. **365**(26): p. 2497-506.
543. Oscier, D.G., et al., *The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial.* Blood, 2013. **121**(3): p. 468-75.
544. Landau, D.A., et al., *Evolution and impact of subclonal mutations in chronic lymphocytic leukemia.* Cell, 2013. **152**(4): p. 714-26.
545. Berry, W.L. and R. Janknecht, *KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells.* Cancer Res, 2013. **73**(10): p. 2936-42.
546. Strahl, B.D., et al., *Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1.* Curr Biol, 2001. **11**(12): p. 996-1000.
547. Khoury-Haddad, H., et al., *PARP1-dependent recruitment of KDM4D histone demethylase to DNA damage sites promotes double-strand break repair.* Proc Natl Acad Sci U S A, 2014. **111**(7): p. E728-37.
548. Fu, L., et al., *HIF-1alpha-induced histone demethylase JMJD2B contributes to the malignant phenotype of colorectal cancer cells via an epigenetic mechanism.* Carcinogenesis, 2012. **33**(9): p. 1664-73.
549. Kim, J.G., et al., *Histone demethylase JMJD2B-mediated cell proliferation regulated by hypoxia and radiation in gastric cancer cell.* Biochim Biophys Acta, 2012. **1819**(11-12): p. 1200-7.
550. Zhao, L., et al., *JMJD2B promotes epithelial-mesenchymal transition by cooperating with beta-catenin and enhances gastric cancer metastasis.* Clin Cancer Res, 2013. **19**(23): p. 6419-29.
551. Ye, Q., et al., *Genetic alterations of KDM4 subfamily and therapeutic effect of novel demethylase inhibitor in breast cancer.* Am J Cancer Res, 2015. **5**(4): p. 1519-30.
552. Cho, J.H., et al., *Arginine methylation-dependent regulation of ASK1 signaling by PRMT1.* Cell Death Differ, 2012. **19**(5): p. 859-70.
553. Li, B., et al., *miR-503 suppresses metastasis of hepatocellular carcinoma cell by targeting PRMT1.* Biochem Biophys Res Commun, 2015. **464**(4): p. 982-7.
554. Zhang, T., et al., *Inhibition of Nonsmall Cell Lung Cancer Cell Migration by Protein Arginine Methyltransferase 1-small Hairpin RNA Through Inhibiting Epithelial-mesenchymal Transition, Extracellular Matrix Degradation, and Src Phosphorylation In Vitro.* Chin Med J (Engl), 2015. **128**(9): p. 1202-8.
555. Avasarala, S., et al., *PRMT1 Is a Novel Regulator of Epithelial-Mesenchymal-Transition in Non-small Cell Lung Cancer.* J Biol Chem, 2015. **290**(21): p. 13479-89.

556. Infantino, S., et al., *Arginine methylation of the B cell antigen receptor promotes differentiation.* J Exp Med, 2010. **207**(4): p. 711-9.
557. Xie, Y., et al., *Virtual screening and biological evaluation of novel small molecular inhibitors against protein arginine methyltransferase 1 (PRMT1).* Org Biomol Chem, 2014. **12**(47): p. 9665-73.
558. Wan, Y. and C.J. Wu, *SF3B1 mutations in chronic lymphocytic leukemia.* Blood, 2013. **121**(23): p. 4627-34.
559. Su'ut, L., et al., *Trisomy 12 is seen within a specific subtype of B-cell chronic lymphoproliferative disease affecting the peripheral blood/bone marrow and co-segregates with elevated expression of CD11a.* Br J Haematol, 1998. **101**(1): p. 165-70.
560. Hu, Z., et al., *GEP100/Arf6 is required for epidermal growth factor-induced ERK/Rac1 signaling and cell migration in human hepatoma HepG2 cells.* PLoS One, 2012. **7**(6): p. e38777.
561. Xie, C.G., et al., *Down-regulation of GEP100 causes increase in E-cadherin levels and inhibits pancreatic cancer cell invasion.* PLoS One, 2012. **7**(5): p. e37854.
562. Menju, T., et al., *Engagement of overexpressed Her2 with GEP100 induces autonomous invasive activities and provides a biomarker for metastases of lung adenocarcinoma.* PLoS One, 2011. **6**(9): p. e25301.
563. Carey, G.B., et al., *IL-4 protects the B-cell lymphoma cell line CH31 from anti-IgM-induced growth arrest and apoptosis: contribution of the PI-3 kinase/AKT pathway.* Cell Res, 2007. **17**(11): p. 942-55.
564. Planelles, L., et al., *APRIL promotes B-1 cell-associated neoplasm.* Cancer Cell, 2004. **6**(4): p. 399-408.

10.0 APPENDIX

Appendix A1. **Confirmation of the multi-feature weighting approach using machine learning.** Provided and described courtesy of Yawwani Gunawardana:

Linear regression approach was employed to observe whether the features weights obtained by our ad-hod method have achieved the optimum solution. In fact, this is a machine learning approach which obtains optimum weight values with respect to linear prediction. Suppose we have a set of m samples $(\{\boldsymbol{x_i}, y_i)\}_{i=1,\dots,m}$ where $\boldsymbol{x_i} \in \mathbb{R}_n$ and $y_i \in \mathbb{R}$ are input and targets respectively. Here the weights are obtained by minimizing the squared loss between the target and the regression prediction $f(x) = \langle \boldsymbol{w}, \boldsymbol{x} \rangle + b$ as given below;

$$\min_{w,b} \{y - (\langle \boldsymbol{w}, \boldsymbol{x} \rangle + b)\}^2,$$

where $\boldsymbol{x}$ represents the input matrix of covariates, y is the response vector and $\boldsymbol{w}$ is the unknown weight vector which minimizes the loss function. Five input features variables were normalized to zero mean and one standard deviation before performing the linear regression. The table below shows the correlation between the weights described in **Chapter 4** and the new weights obtained by the linear regression model. We observe that these two weight vectors are highly correlated (R = 0.96). Therefore, weights obtained by our ad-hoc technique have essentially reached to the optimal solution.

Linear Regression Ad-hoc Weights	Isolation Interference [%] -10	Ion Inject Time [ms] -9	Intensity 3	PEP -5	Charge -5	Old Weights to New Weight Correlation
50.0:1	-0.3431	-0.2569	0.1361	-0.1035	-0.2319	0.9622
5.6:1	-0.3396	-0.2147	0.0886	-0.1053	-0.2484	0.9229
1.5:1	-0.3251	-0.1837	0.0563	-0.1061	-0.253	0.8815

Appendix A2. **The SPIQuE user guide**

SPIQuE – user guide

Statistical Processing for Isobaric Quantitation Evaluation

SPIQuE is designed to assist in the determination of differentially expressed proteins from isobaric-tagged proteomics data by the calculation of quality-weighted statistics and ratios.

Generally, iTRAQ and TMT quantitations are assigned to proteins by averaging the ratios from corresponding peptide spectrum matches (PSMs). Giving equal consideration to all quantitations results in compressed ratios due the effects of precursor co-isolation. Approaches where data are filtered are detrimental to any statistical analyses.

SPIQuE allows users to upload their own data and uses a carefully defined multi-feature weighing model to maximise the emphasis of those quantitations with the highest probability of low co-isolation. As described in 'Quality-Weighted Statistics Promotes Differentially Expressed Protein Determination by Isobaric Tag Quantitation', this weighting improves the ratios and statistics for isobarically tagged experiments.

Through SPIQuE this weighing can be easily implemented to generate proteome-wide ratios and statistics to assist in defining differentially expressed proteins. Options are available for normalisation, statistical testing, group testing and PSM quality weighting all available through the online SPIQuE interface.

SPIQuE is designed primarily for use with iTRAQ and TMT data, but has applicability to other forms of multiplex labelling. It has so far been tested with 4-plex, 6-plex, 8-plex and 10-plex data.

All that is required is a list of PSMs with raw reporter intensity values, features and a protein accession number - easily exportable from many proteomics software.

Quick Guide:

Recommended input preparation:	Output usage:
Settings: Disable any co-isolation filtering	All rows will align identically when columns from each ratio file are combined
Show reporter ion intensities (raw values)	To align SPIQuE data with protein descriptions, coverage, peptides etc.;
Enable quantitation value corrections	Copy columns and sort by accession no. (Alphabetically)
Disable peptide grouping	Use 'match' function in excel to ensure alignment
Enable intensity values	=match('**individual** description accession', 'SPIQuE accession **column**', 0)
Export: PSMs with 'Used' quantitations	Sort by match column and remove any unquantitated proteins/ unrequired columns
Remove all but accession, reporter intensities, and features	
Save file as .CSV	
Input to SPIQuE with settings below	

The recommended setting are as default and the following weightings are recommended:

PEP (5 -), isolation interference (10 -), ion injection time (9 -), intensity (3 +) and charge (5 -)

Data input

SPIQuE reads data from a comma delimited .csv file containing all the PSMs and their associated reporter ion intensities from any given iTRAQ-, TMT- or similarly-labelled bottom-up experiment. Once exported from the users search software eg. MaxQuant, Proteome Discoverer, Protein Pilot etc. the PSMs can be formatted using Excel, or similar spreadsheet software, to match the description below, see also example data. The sheet can then be saved as a .csv file and uploaded to www.spiquetool.com.

When selecting the PSMs for export it is recommended to turn off filters such as those that disregard quantitations where co-isolation is >30%, as these still have statistical value and will be compensated for by the weighting.

PSM Features

SPIQuE's secondary feature is to apply PSM-specific weightings to the ratio and statistical calculations to give the greatest emphasis to PSMs with the least co-isolation and highest quality. While the weighting is optional, it provided a superior output, as demonstrated in 'Quality-Weighted Statistics Promotes Differentially Expressed Protein Determination by Isobaric Tag Quantitation'. Minimally recommended, is the use of a measure of co-isolation eg. 'isolation interference'. The use of a further correlating feature such the precursor intensity or ion injection time improves the output further. The use of other features such as pep score or equivalent measure of peptide match probability can also improve the weighting efficiency. Favouring peptides with a charge state of +2 also improves the weighting. For recommended weighting factor see 'user defined settings'.

Required .csv file formatting

- 1 PSM per row (ungrouped peptides with used quantitations)
- Column 1; protein or protein group accession number
- Column 2+; raw reporter ion intensities
- Then; features used for weightings (up to 8 features)

Other columns/rows may cause the algorithm to fail.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Protein Group Accessions	126	127_N	127_C	128_N	128_C	129_N	129_C	130_N	130_C	131	PEP	Isolation Interference [%]	Ion Inject Time [ms]	Intensity	Charge
2	Hs_Q07065	5314	2944	1316	266	191	204	832	1509	3361		7.08E-08	0	200	167331	2
3	Hs_P49840	4058	1819	900	97			440	897	2165		5.90E-08	0	200	91953	2
4	Hs_P62937	37120	18280	9553		2758		4186	5491	15910		5.28E-08	0	17	1077591	3
5	Hs_Q14247	538	395	329		178	91	920	929	1108		2.52E-07	0	200	0	3
6	Hs_Q14671	15250	9781	6773	3399	8195	4676	3644	4722	8155	361	2.34E-08	0	168	140874	3
7	Hs_P62937	24020	11580	5440	1622	3997	1841	3812	6665	13420		7.71E-08	7	33	440988	3
8	Hs_E9PRY8	17370	10700	6130	2672	5057	2653	5065	8867	13260	338	1.51E-08	0	50	332450	3
9	Hs_O43707	7769	3346	2236	1038	2100	1507	2745	5114	9476	186	7.91E-08	19	187	100955	3
10	Hs_B8ZZ47	7257	2977	1795	995	872		1170	1822	6069		5.08E-08	0	67	237631	3

User defined settings

After successful upload, SPIQuE will read the headers of the file and the data in the first line of each column. Based on the data, it will try to automatically assign these as either a feature or a reporter. The unique protein identifier from the first column should be set as ‘ignore’.

Feature weighting and correlation

Feature weightings are proportional to one another eg. 1 and 5 would give the same result as 10 and 50. The correlation defines whether the feature correlated positively or negatively with quality and isolation purity eg. higher precursor intensity correlates positively with the probability of a reliable quantitation. Recommended values are as follows:

PEP	7.081E-08	Feature	5	-
Isolation Interference [%]	0	Feature	10	-
Ion Inject Time [ms]	200	Feature	9	-
Intensity	167331.0781	Feature	3	+
Charge	2	Feature	5	-

Other options

Normalisation

Median is typically recommended for most data, but options for normalising by the mean and scaled mean absolute deviation are included

Statistical test

T-test provides a quick (1-15 minutes per ratio), statistical test to determine whether the values for any given protein and any given ratio demonstrates a significant difference in expression levels. It is improved by the use of weightings. P-values are provided before and after multiple test correction.

Permutation testing is a far slower (3-12 hours per ratio) test of significance but typically gives more proteins significantly regulated at $p<0.05$. Permutation testing takes all the associated ratios for a protein and compares these with random samples of the data as a whole. The number of permutations designates the number of samplings performed. The p-value is then calculated as the number of times the sampling returns the random numbers as more regulated than the protein's quantitation over the number of times they are not. E.g. when only 2 of 1000 permutations return a greater regulation than the protein tested, the p-value of this observation would be 0.002.

While more permutations returns more accurate p-values, for the majority of purposes, 1000 permutations should be sufficient when a p-value of 0.05 is considered significant. Time taken is proportional to the number of permutations. Permutations take approximately 100+ times longer than T-testing and are not improved by weightings, but offers a good alternative to t-testing.

Replace missing values

In some circumstances missing values are of biological interest and proteins with low intensities or substantial differential expression can be identified as statistically significant by inserting the minimum detected reporter intensity.

Final exponential weighting

The emphasis of the final weighting factor (between 0 and 1) can be manipulated by raising it to a power. E.g. $13 = 1$, $0.13 = 0.001$. An increase in the exponential weighing factor increases the number of significantly regulated proteins. However this also increases the number of false significant regulations. A value of 1 is recommended.

Output files

Once the analysis is complete, the following files will be available for download as a compressed archive file. Unzipping this file will provide:

Protein ratios and significance values – for each user defined ratio. Again as .csv files, best opened as a spreadsheet

Accession numbers – derived from the input file

Log2_'reporterA:reporterB' – the log 2 (ratio) calculated from the weighted, individual PSM ratios.

Use =2^[num] to convert to ratios

neglog10_p 'reporterA:reporterB' – the negative log 10 of the uncorrected p-values

neglog10_padj 'reporterA:reporterB' – the negative log 10 of the multiple-test corrected p-values – these are the most statistically sound values

use =10^[num] to convert to p-values, or filter at $-\log_{10}(0.05) = 1.301$ to get significantly regulated proteins ($p<0.05$)

Peptide counts – the number of individual quantitations used to derive the ratio and p-value for each protein

Mean and SD of weights – the mean and standard deviation of all the weighting factors of the PSMs used to derive the ratios and statistics.

	A	B	C	D	E	F	G	H
1		Protein.Group.Accessions	log2_X126:X130_C	neglog10_p X126:X130_C	neglog10_padj X126:X130_C	Peptide Counts	Mean Weights	SD Weights
2	Hs_A0AV96	Hs_A0AV96	-0.047781745	0.009494006	0.001903414	5	0.455418624	0.302170259
3	Hs_A0AVT1	Hs_A0AVT1	-0.523245782	0.391470336	0.106633076	20	0.336649641	0.211285922
4	Hs_A0JLT2	Hs_A0JLT2	-0.141859895	0.181765547	0.031287443	6	0.223258304	0.196454552
5	Hs_A0JNW5	Hs_A0JNW5	1.534598932	NA	NA	1	0.634843306	NA
6	Hs_A0JP02	Hs_A0JP02	-0.383841907	0.207449434	0.038015776	16	0.373099847	0.247219775
7	Hs_A0MZ66	Hs_A0MZ66	-0.541411928	0.986997913	0.405052875	67	0.49564643	0.285544485
8	Hs_A1BQX2	Hs_A1BQX2	-0.629748732	0.527784412	0.161120518	5	0.223157318	0.194937599
9	Hs_A1L188	Hs_A1L188	0.132487024	0.132392627	0.018459228	2	0.338682053	0.045266977
10	Hs_A1L390	Hs_A1L390	0.23599169	0.23959705	0.048683618	5	0.164678658	0.142343712
11	Hs_A1X283	Hs_A1X283	1.417417788	2.676577145	1.572886886	14	0.392071744	0.25106064
12	Hs_A2A2F0	Hs_A2A2F0	-0.4360362	0.11611606	0.014215467	4	0.400440356	0.170907818
13	Hs_A2A2N6	Hs_A2A2N6	-0.687304917	NA	NA	1	0.183579329	NA
14	Hs_A2A2V1	Hs_A2A2V1	0.087609166	0.084814353	0.012204448	2	0.248372929	0.194233547
15	Hs_A2A2V2	Hs_A2A2V2	0.820018611	1.484197355	0.729595343	32	0.484524397	0.291426644
16	Hs_A2AAT0	Hs_A2AAT0	0.657995533	NA	NA	1	0.881077254	NA
17	Hs_A2ABF8	Hs_A2ABF8	-0.002842543	0.005441494	0.001263943	16	0.349350537	0.269309996
18	Hs_A2IDC6	Hs_A2IDC6	0.693623808	0.468452029	0.136580507	7	0.409277107	0.225058719
19	Hs_A2TJX0	Hs_A2TJX0	-0.122227329	NA	NA	1	0.136156382	NA
20	Hs_A2VDF0	Hs_A2VDF0	2.81440009	NA	NA	1	0.123489964	NA
21	Hs_A3KFL1	Hs_A3KFL1	0.536453859	1.050842486	0.446995606	3	0.387236655	0.35794866
22	Hs_A3KMH1	Hs_A3KMH1	0.543585852	0.533039034	0.162835674	9	0.19333626	0.088928734
23	Hs_A3KN83	Hs_A3KN83	0.577153994	0.505348943	0.150360974	20	0.49344452	0.275647993

Other files

Settings.csv – The experimental design file containing the user defined inputs required by R and are defined on the first page of the interface. This provides a reference of these initial settings. W_Name is the name assigned for within R, W_Value is the weighting factor specified for that feature and Sub_Vector specifies if the correlation of a feature is positive (1) or negative (-1) (see Feature weighting and correlation)

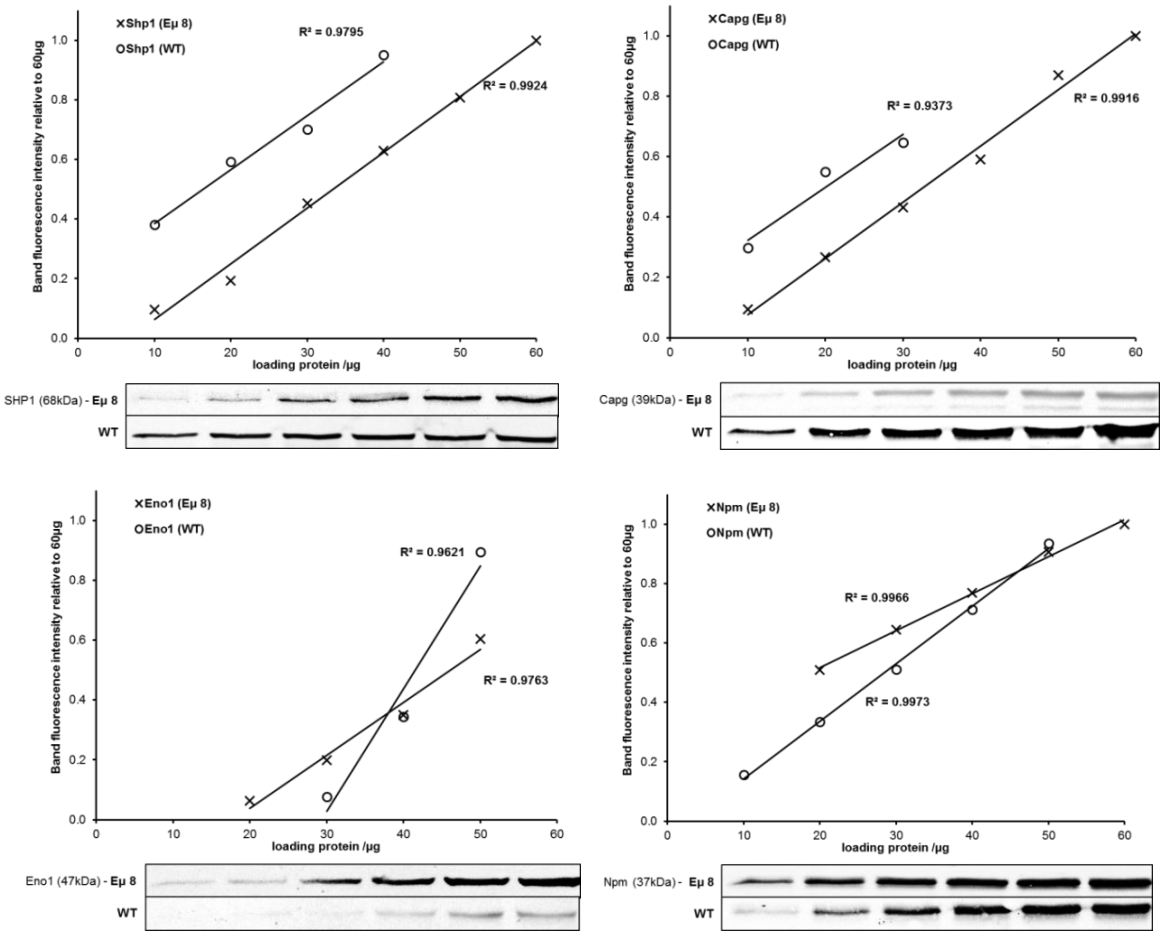
	A	B	C	D	E	F	G
1	Name	Condition	Replicate	Weights	W_Name	W_Value	Sub_Vector
2	126	Condition_1	1	PEP	Feature_1	5	-1
3	127_N	Condition_2	2	Isolation Interference [%]	Feature_2	10	-1
4	127_C	Condition_3	3	Ion Inject Time [ms]	Feature_3	9	-1
5	128_N	Condition_4	4	Intensity	Feature_4	3	1
6	128_C	Condition_5	5	Charge	Feature_5	5	-1
7	129_N	Condition_6	6				
8	129_C	Condition_7	7				
9	130_N	Condition_8	8				
10	130_C	Condition_9	9				
11	131	Condition_10	10				

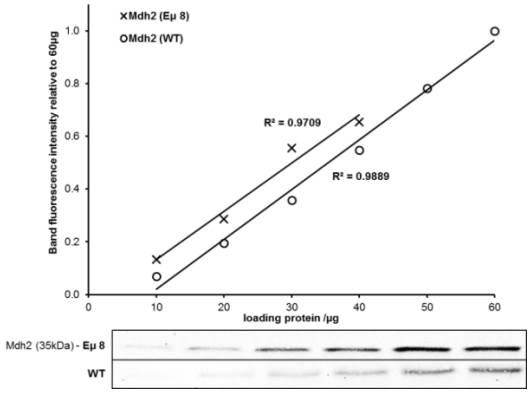
Weighting_Results.csv – This details the weighting breakdown for each PSM. Each feature is scored by percentile ranking and the weighting factor for each feature is used to generate a final cumulative weighting score 'Weight_Formula_Result'. This score is further percentile ranked to give a score between 0 and 1 'Final_Weighting'.

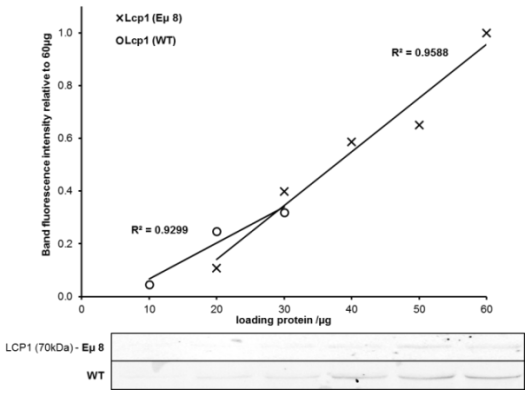
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Protein. Group. Accessions	PEP	PEP_w	Isolation Interference [%]	Isolation Interference [%]_w	Ion Inject Time [ms]	Ion Inject Time [ms] _w	Intensity	Intensity _w	Charge	Charge _w	Weight_ Formula_ Result	Final_ Weighting
2	Hs_P10809	1.61E-06	0.965	0.000	0.9999848	1	0.999	49585780	0.996346	2	1	10.951	1.00000000
3	Hs_Q09666	1.34E-06	0.970	0.000	0.9999848	2	0.993	29567526	0.992213	2	1	10.933	0.99999239
4	Hs_P08727	3.00E-06	0.935	0.000	0.9999848	1	0.999	86617368	0.998653	2	1	10.929	0.99998478
5	Hs_P14618	2.08E-06	0.954	0.000	0.9999848	2	0.993	32123972	0.993149	2	1	10.919	0.99997716
6	Hs_P47914	3.54E-06	0.925	0.000	0.9999848	1	0.999	80787736	0.998462	2	1	10.918	0.99996955
7	Hs_P04792	3.81E-06	0.920	0.000	0.9999848	1	0.999	74043632	0.998135	2	1	10.912	0.99996194
8	Hs_P29401	3.40E-06	0.928	0.000	0.9999848	1	0.999	41636116	0.995379	2	1	10.911	0.99995433
9	Hs_P04792	3.91E-06	0.918	0.000	0.9999848	1	0.999	75145168	0.998211	2	1	10.910	0.99994672
10	Hs_P08727	3.81E-06	0.920	0.000	0.9999848	1	0.999	62250948	0.997458	2	1	10.910	0.99993910
11	Hs_P51991	1.13E-07	0.998	0.000	0.9999848	4	0.978	15342814	0.982515	2	1	10.902	0.99993149
12	Hs_P37802	4.19E-06	0.914	0.000	0.9999848	1	0.999	56623916	0.997008	2	1	10.902	0.99992388
13	Hs_F5H1A8	1.39E-06	0.969	0.000	0.9999848	3	0.985	18922598	0.986390	2	1	10.899	0.99991627
14	Hs_P14618	1.96E-06	0.957	0.000	0.9999848	3	0.985	23654354	0.989594	2	1	10.896	0.99990866
15	Hs_P11142	4.67E-06	0.906	0.000	0.9999848	1	0.999	61337108	0.997351	2	1	10.895	0.99990104
16	Hs_P38646	1.76E-06	0.961	0.000	0.9999848	3	0.985	20213798	0.987455	2	1	10.894	0.99989343
17	Hs_Q15149	1.88E-06	0.959	0.000	0.9999848	3	0.985	20220082	0.987463	2	1	10.892	0.99988582
18	Hs_Q16543	9.66E-07	0.979	0.000	0.9999848	4	0.978	17312316	0.984829	2	1	10.890	0.99987821

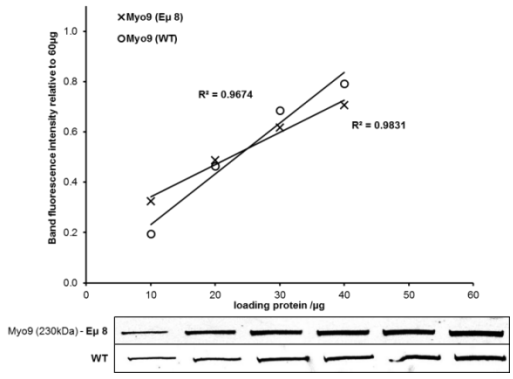
Input.csv – the input file used for the analysis

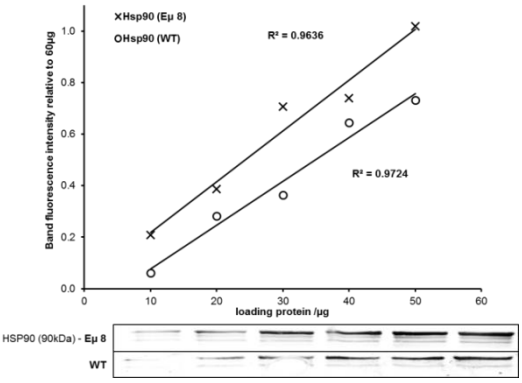
Appendix A3. **Determination of the linear phase of quantitative WB protein detection.** WB detection was conducted for lysates of WT B cells and lymphoma (Eμ8 - Eμ-*myc*-derived lymphoma cell line) evaluating detection of 10-60 μg of protein loaded. The lower portion of each graph shows the linear portion of the upper graph with a trend line fitted and correlation coefficient described. This was conducted for all 10 candidates and GAPDH, as described in **Chapter 3**.

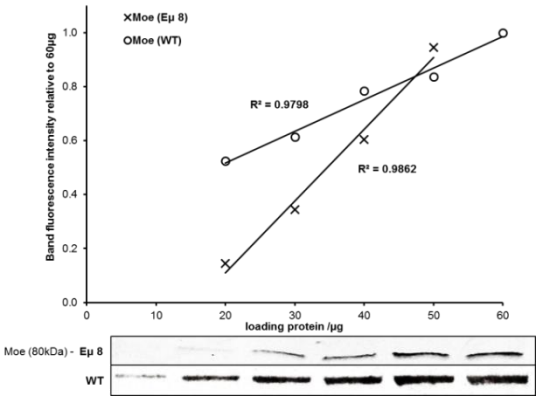


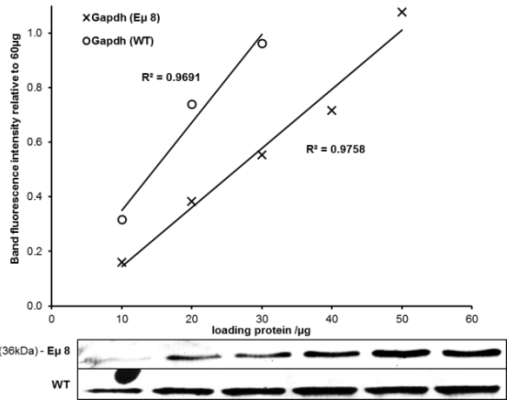

×Mdh2 (Eμ 8)
○Mdh2 (WT)
R² = 0.9709
R² = 0.9889
Band fluorescence intensity relative to 60μg
loading protein /μg
Mdh2 (35kDa) - Eμ 8
WT


×Lcp1 (Eμ 8)
○Lcp1 (WT)
R² = 0.9588
R² = 0.9299
Band fluorescence intensity relative to 60μg
loading protein /μg
LCP1 (70kDa) - Eμ 8
WT

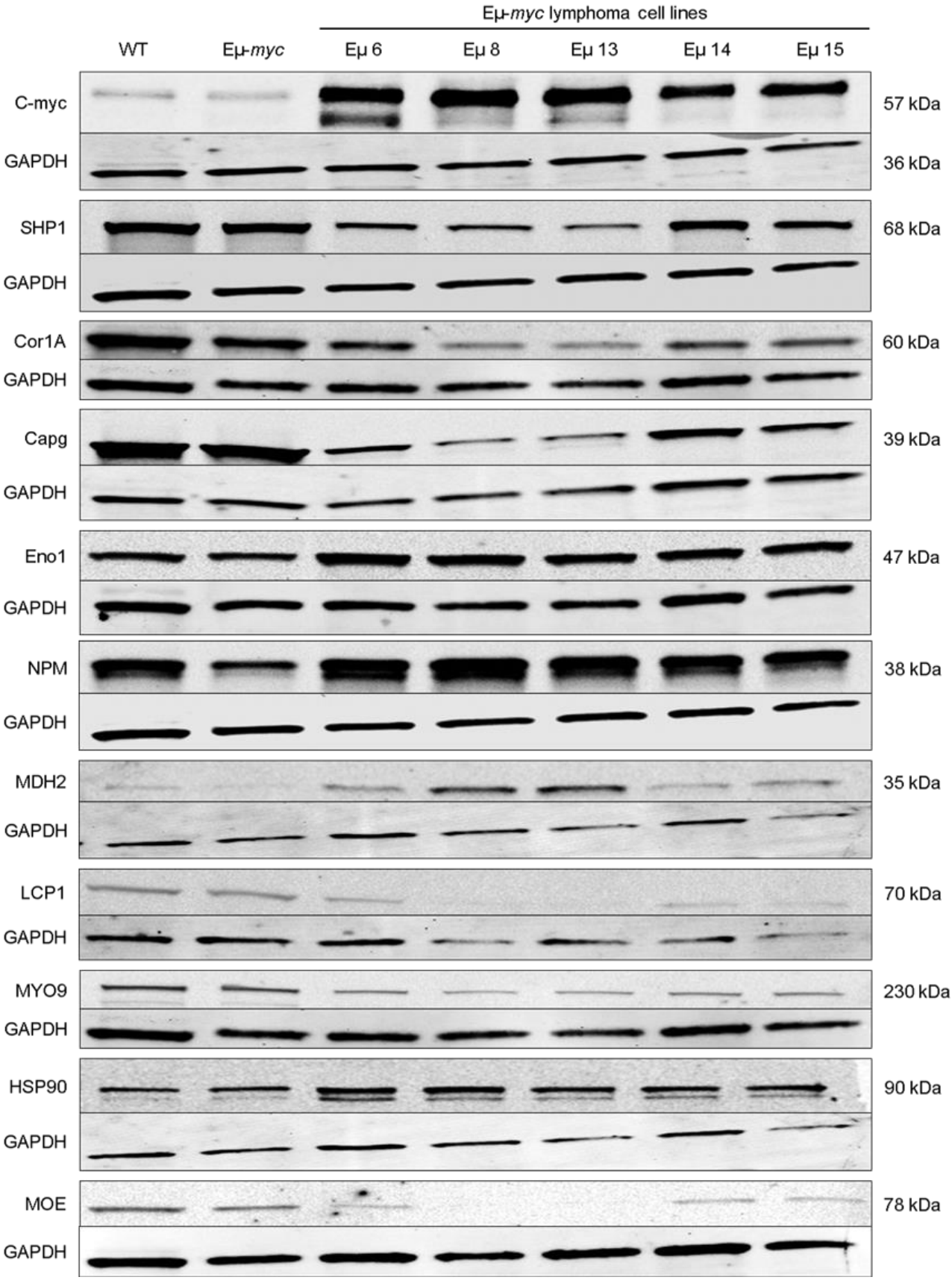

×Myo9 (Eμ 8)
○Myo9 (WT)
R² = 0.9674
R² = 0.9831
Band fluorescence intensity relative to 60μg
loading protein /μg
Myo9 (230kDa) - Eμ 8
WT


×Hsp90 (Eμ 8)
○Hsp90 (WT)
R² = 0.9636
R² = 0.9724
Band fluorescence intensity relative to 60μg
loading protein /μg
HSP90 (90kDa) - Eμ 8
WT

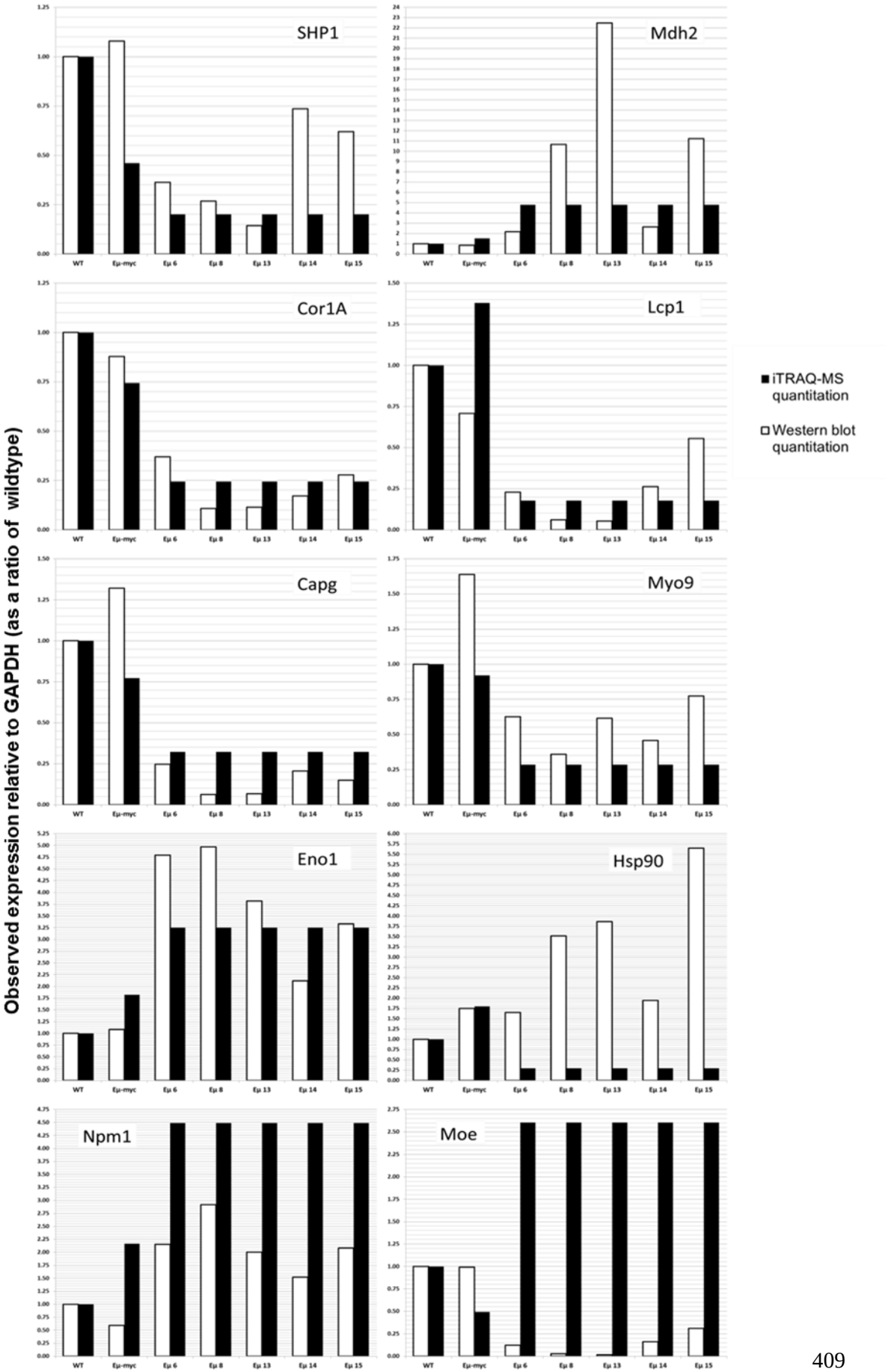

×Moe (Eμ 8)
○Moe (WT)
R² = 0.9798
R² = 0.9862
Band fluorescence intensity relative to 60μg
loading protein /μg
Moe (80kDa) - Eμ 8
WT


×Gapdh (Eμ 8)
○Gapdh (WT)
R² = 0.9691
R² = 0.9758
Band fluorescence intensity relative to 60μg
loading protein /μg
Gapdh (36kDa) - Eμ 8
WT

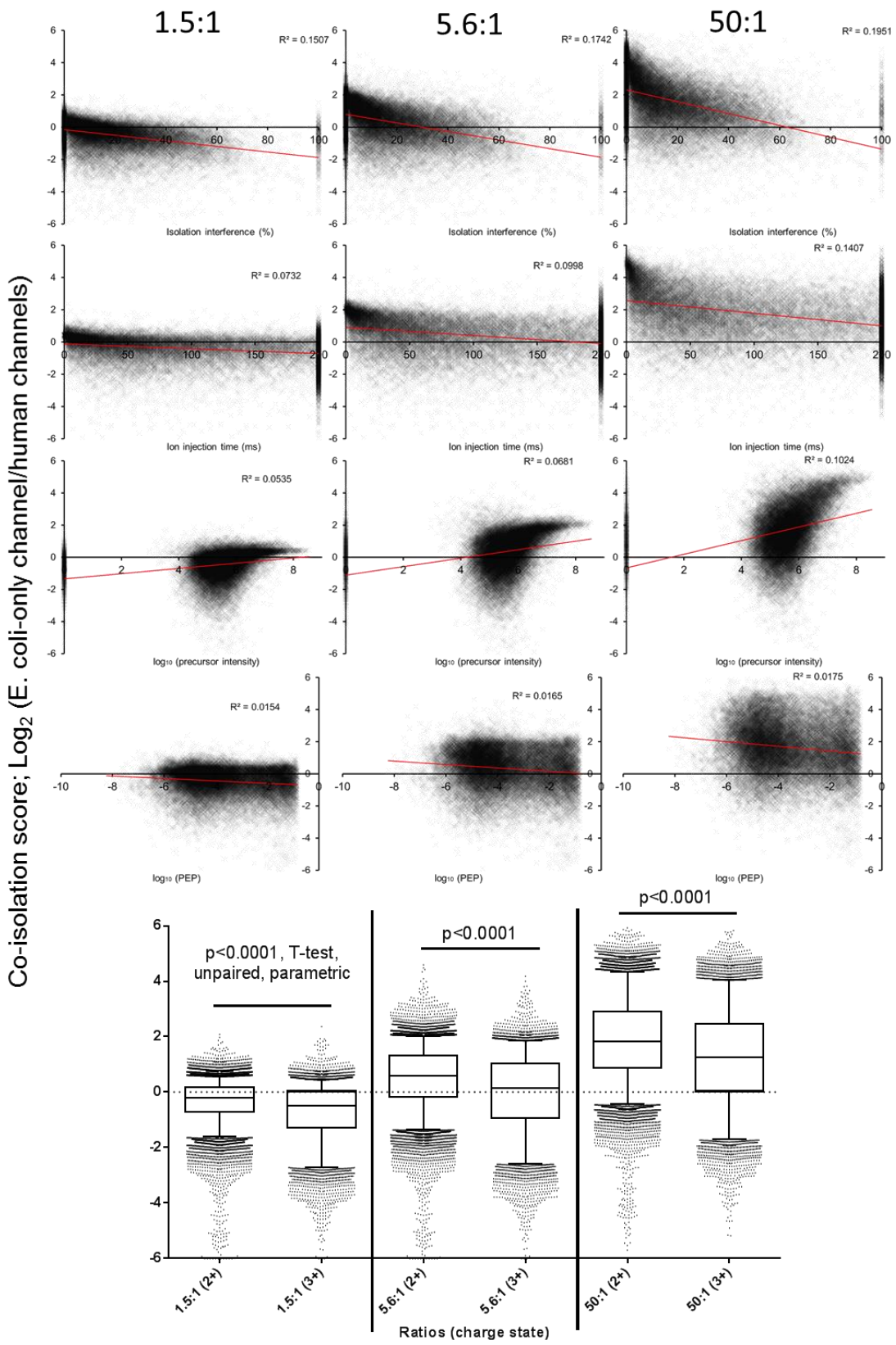
Appendix A4. **The expression of the 10 validation candidates and myc in Eμ-*myc* lymphoma and pre-lymphoma samples, relative to WT B cells.** WB of lysates derived from a pool of splenic B cells from three WT mice (WT), a pool of splenic B cells from three Eμ-myc mice prior to any tumour development (Eμ-*myc*) and five individual Eμ-*myc*-derived lymphoma cell line lysates (Eμ6, Eμ8, Eμ13, Eμ14 and Eμ15). GAPDH serves as a loading control for each WB.

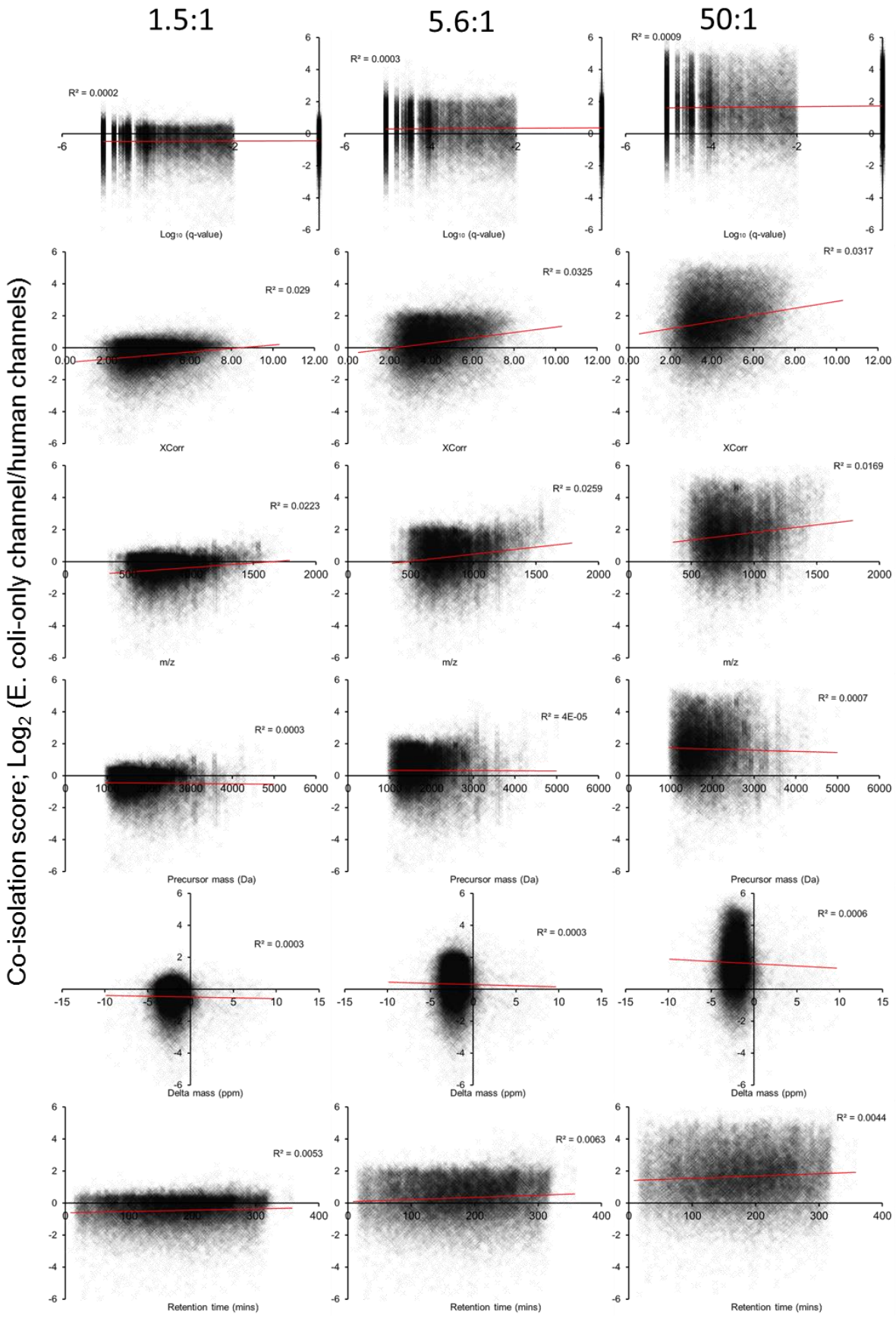

Eμ-myc lymphoma cell lines
WT
Eμ-myc
Eμ 6
Eμ 8
Eμ 13
Eμ 14
Eμ 15
C-myc
57 kDa
GAPDH
36 kDa
SHP1
68 kDa
GAPDH
Cor1A
60 kDa
GAPDH
Capg
39 kDa
GAPDH
Eno1
47 kDa
GAPDH
NPM
38 kDa
GAPDH
MDH2
35 kDa
GAPDH
LCP1
70 kDa
GAPDH
MYO9
230 kDa
GAPDH
HSP90
90 kDa
GAPDH
MOE
78 kDa
GAPDH

Appendix A5. **Comparison of the MS-determined and WB-determined ratios of protein expression.** Ratios of protein expression in splenic isolated B cells from pre-neoplastic Eμ-myc mice (Eμ-*myc*) and Eμ-myc-derived cell lines from spontaneous splenic tumours (Eμ 6-15) were determined relative to wildtype B cells (WT) for WB- determined (open bars) and MS-determined (closed bars) quantitation methods.


SHP1
Mdh2
Cor1A
Lcp1
Capg
Myo9
Eno1
Hsp90
Npm1
Moe
WT
Eμ-myc
Eμ 6
Eμ 8
Eμ 13
Eμ 14
Eμ 15
Observed expression relative to GAPDH (as a ratio of wildtype)
iTRAQ-MS quantitation
Western blot quantitation

Appendix A6. **The correlation of PSM features with the co-isolation score.** The co-isolation scores described in **Section 2.20.3** and **Section 4.4** for each *E. coli* PSM were plotted against each respective PSM feature.

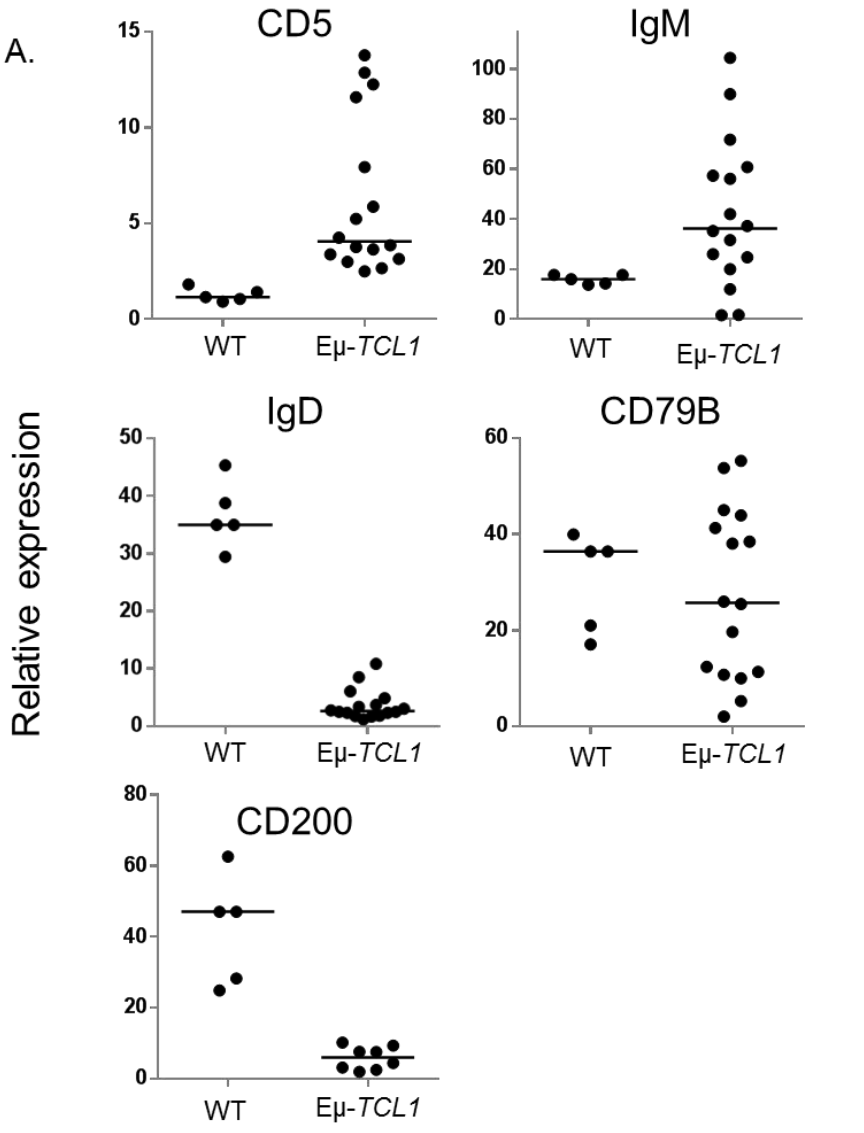



1.5:1
5.6:1
50:1
Co-isolation score; Log2 (E. coli-only channel/human channels)
R² = 0.0002
R² = 0.0003
R² = 0.0009
Log10 (q-value)
R² = 0.029
R² = 0.0325
R² = 0.0317
XCorr
R² = 0.0223
R² = 0.0259
R² = 0.0169
m/z
R² = 0.0003
R² = 4E-05
R² = 0.0007
Precursor mass (Da)
R² = 0.0003
R² = 0.0003
R² = 0.0006
Delta mass (ppm)
R² = 0.0053
R² = 0.0063
R² = 0.0044
Retention time (mins)

Appendix A7. **B-cell and B-cell tumour samples collected for MS analysis.** The characteristics of the B cells isolated from of WT, Eμ-*myc* and Eμ-*TCL1* splenocytes. Cellularity was determined by coulter counter and $CD19^+$ $CD3^-$ cell percentages after B-cell isolation by flow cytometry as described in **Section 2.7** and **2.14**. Cell death was determined by PI exclusion described in **Section 2.14.2**. ND – not determined.

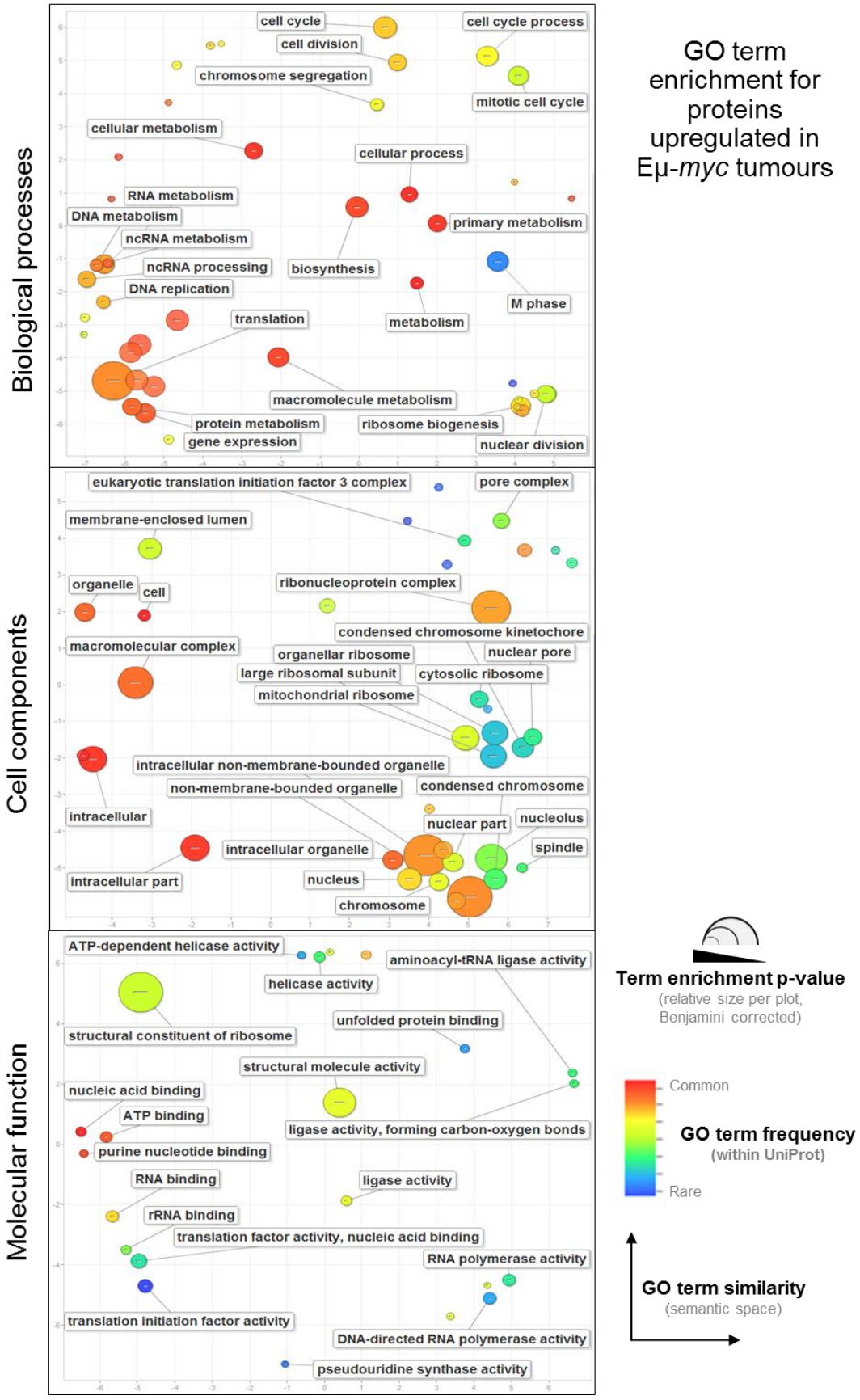
B-cell sample	Genotype	Age (days)	Spleen Cellularity ($x10^6$ cells)	B cells ($x10^6$)	$CD19^+$ $CD3^-$ cells	Cell death
6 week WT Pool **A**	WT	42	460	90	92 %	ND
6 week WT Pool **B**			410	90	90 %	ND
200 day WT Pool **A**		200	340	100	92 %	29 %
200 day WT Pool **B**			310	100	94 %	16 %
6 week Eμ-*myc* Pool **A**	Eμ-*myc*	42	650	125	81 %	ND
6 week Eμ-*myc* Pool **B**			300	95	87 %	16 %
6 week Eμ-*TCL1* Pool **A**	Eμ-*TCL1*		370	105	85 %	ND
6 week Eμ-*TCL1* Pool **B**			315	75	93 %	22 %
UN1	Eμ-*myc*	68	380	50	91 %	32 %
JY2		82	400	120	90 %	35 %
KA3		110	150	90	75 %	22 %
UN5		41	135	50	75 %	25%
BG3	Eμ-*TCL1*	318	750	200	92 %	31 %
BG8		318	880	320	93 %	33 %
AM4		374	1400	180	90 %	8 %
GB2		255	900	600	70 %	23%

Appendix A8. **WB and immunophenotyping of Eμ-*TCL1* tumours.** Relative splenic B-cell protein expression of WT and terminal Eμ-*TCL1* determined by; A. flow cytometry. B. Western blotting. These characterisations were performed by Matthew Carter.



Appendix A9. **Gene ontology term enrichment analysis of the Eμ-*myc* and Eμ-*TCL1* tumours**. GO term enrichment p-values (Benjamini-corrected) were determined for those proteins with significant differential regulation (RS>0.5 or < -0.5, $p<0.05$) in all 4 tumour pools using DAVID with all fully quantitated proteins as background. Significantly enriched GO terms ($p<0.05$) were visualised with REVIGO. These are separated into biological processes, cell components and molecular function. The circle size is proportional to the $-\log_{10}$ (term enrichment p-value), the colour indicates the GO term frequency amongst all mouse genes, and the axis of the plot are based on semantic space, clustering terms that are more frequently co-annotated together.

Appendix A9



GO term enrichment for proteins upregulated in Eμ-*myc* tumours

Term enrichment p-value
(relative size per plot, Benjamini corrected)

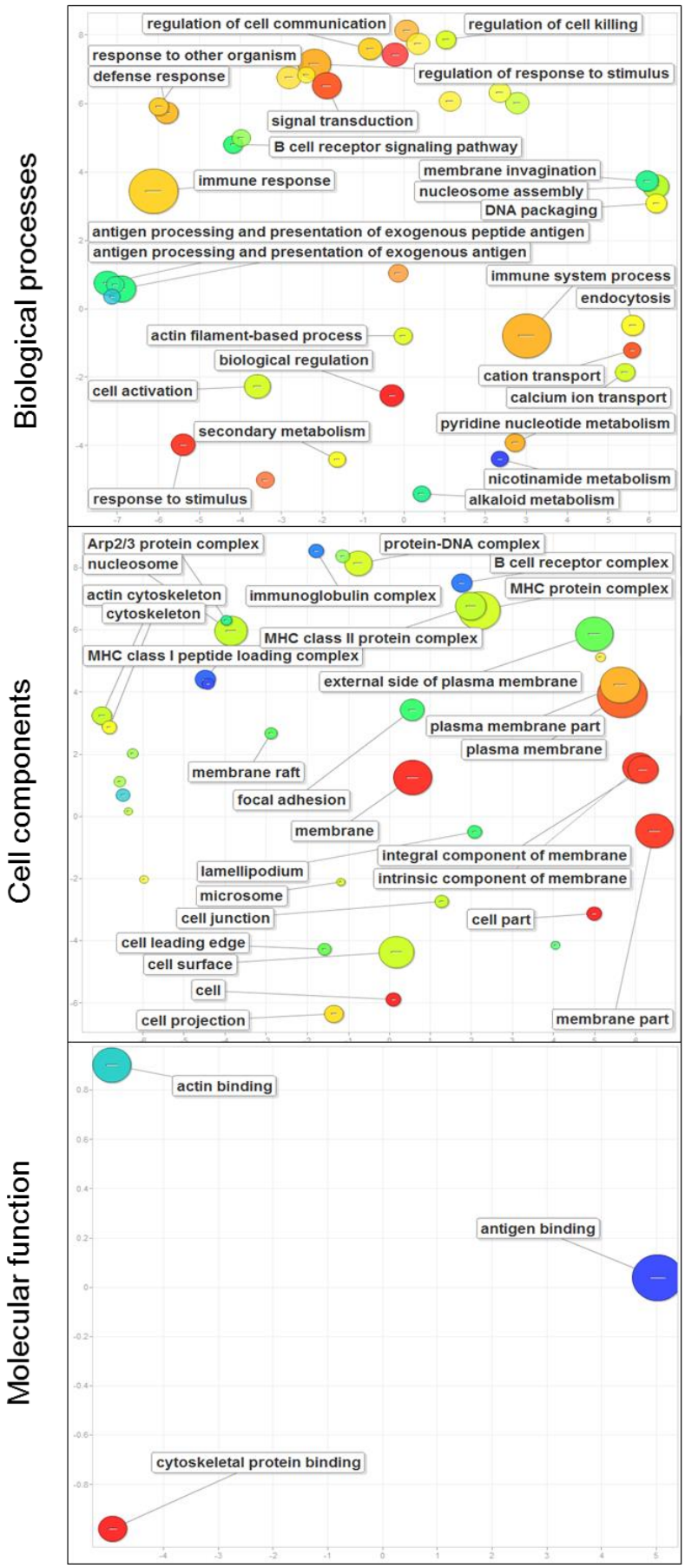
Common

GO term frequency
(within UniProt)

Rare

GO term similarity
(semantic space)

Appendix A9



GO term enrichment for proteins downregulated in Eμ-*myc* tumours

Term enrichment p-value
(relative size per plot, Benjamini corrected)

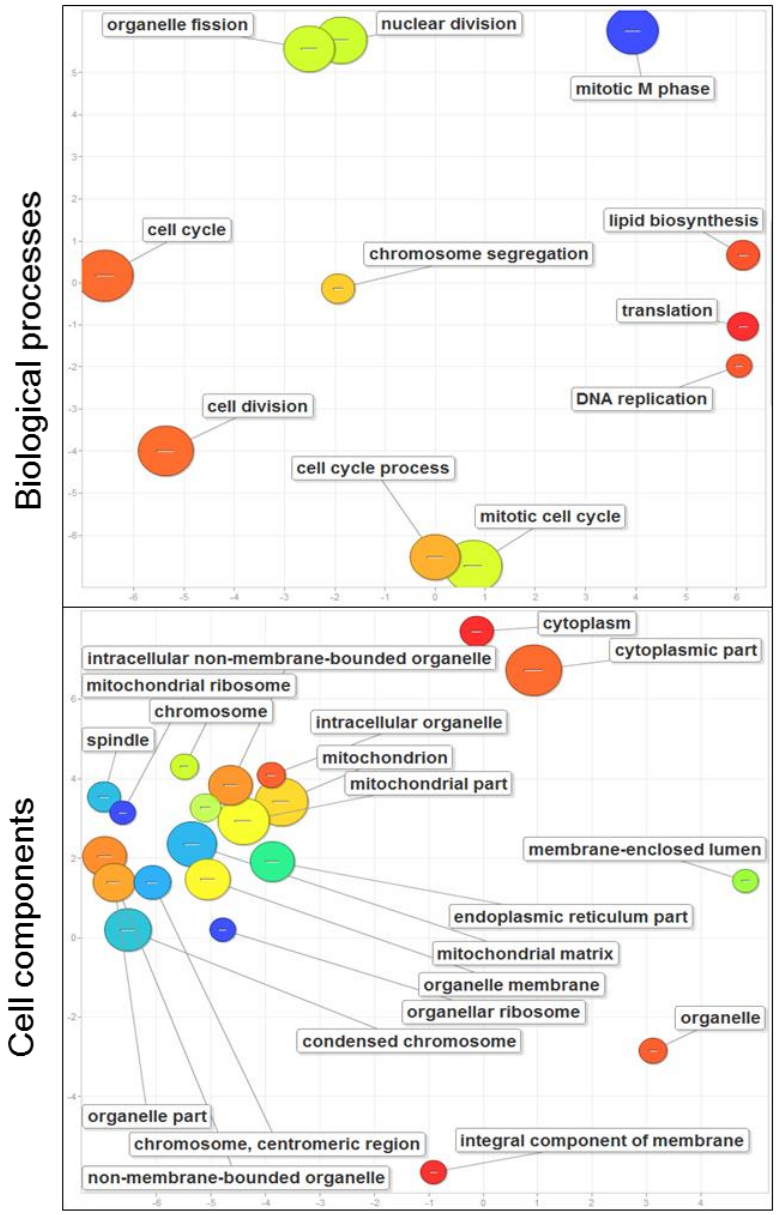
Common

GO term frequency
(within UniProt)

Rare

GO term similarity
(semantic space)

Appendix A9



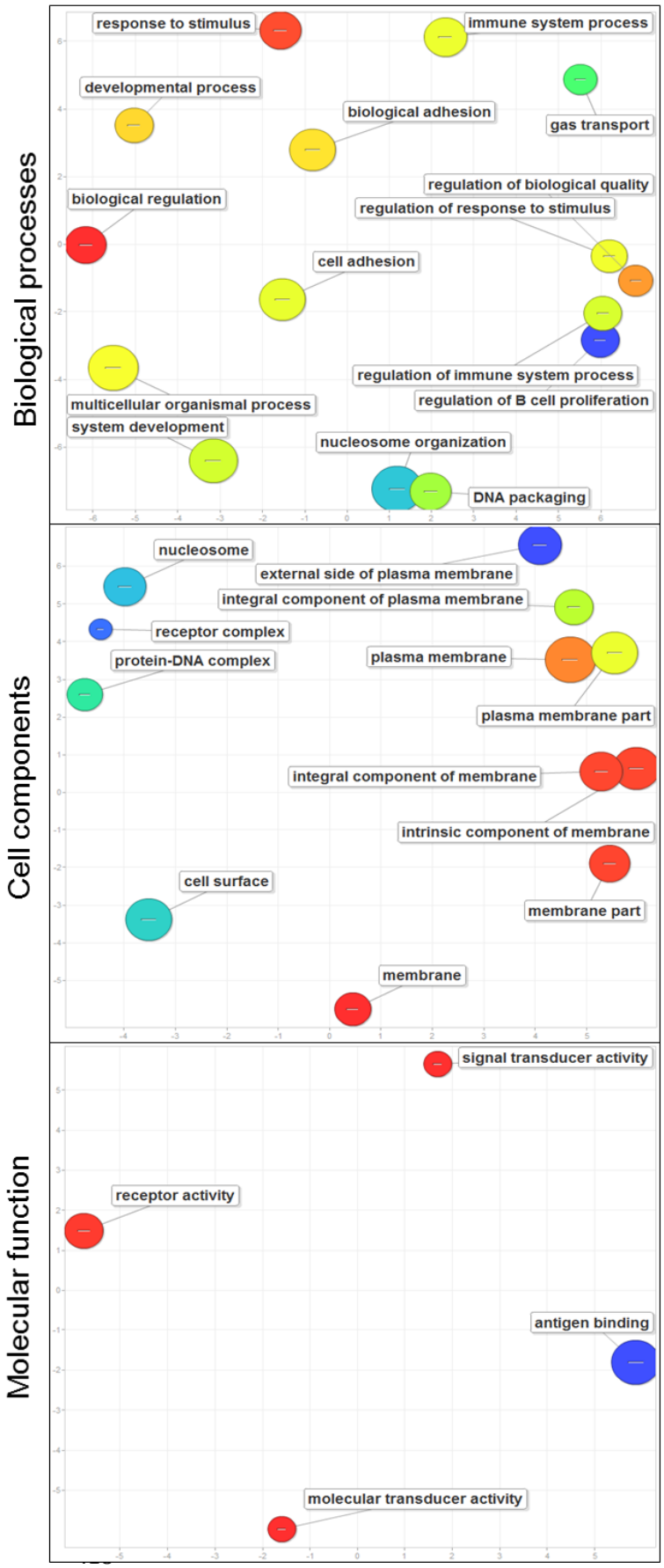
GO term enrichment for proteins upregulated in Eμ-*TCL1* tumours

Term enrichment p-value
(relative size per plot, Benjamini corrected)

Common
GO term frequency
(within UniProt)
Rare

GO term similarity
(semantic space)

Appendix A9



GO term enrichment for proteins downregulated in Eμ-*TCL1* tumours

Term enrichment p-value
(relative size per plot, Benjamini corrected)

Common

GO term frequency
(within UniProt)

Rare

GO term similarity
(semantic space)

Appendix A10. **Pathways identified as significantly enriched from the proteins regulated in Eμ-*myc*, Eμ-*TCL1* and both tumours.** Significantly ($p<0.05$), differentially expressed ($Rs>0.5$ or < -0.5) proteins in Eμ-*myc*, Eμ-*TCL1* and both tumours were analysed using DAVID to identify significantly enriched pathways. Red headers highlight those pathways enriched from overexpressed proteins and green headers for underexpressed proteins. Shaded values indicate significance after multiple test correction

Appendix A10

Pathways Defined by Upregulated Proteins	Both Tumour Models			Eμ-*myc* tumours			Eμ-*TCL1* tumours		
BIOCARTA	Fold Enrichment	p-value	Benjamini	Fold Enrichment	p-value	Benjamini	Fold Enrichment	p-value	Benjamini
RB Tumor Suppressor/Checkpoint Signaling in response to DNA damage	8.98	3.4E-05	4.5E-03	5.19	3.1E-03	7.7E-02	9.09	3.2E-05	4.5E-03
Regulation of p27 Phosphorylation during Cell Cycle Progression	7.05	8.0E-04	5.1E-02	3.97	2.9E-02	3.1E-01	5.95	6.9E-03	1.3E-01
Cell Cycle: G1/S Check Point	4.51	1.1E-03	4.7E-02	4.19	8.9E-05	4.5E-03	4.00	5.6E-03	1.5E-01
Role of Ran in mitotic spindle regulation	7.84	2.2E-03	6.9E-02	7.40	7.4E-05	5.6E-03	6.35	2.0E-02	2.5E-01
p53 Signaling Pathway	5.29	3.5E-03	8.9E-02	3.57	1.9E-02	2.6E-01	5.36	3.3E-03	2.1E-01
Cell Cycle: G2/M Checkpoint	4.29	3.8E-03	8.0E-02	3.73	1.4E-03	4.2E-02	3.73	1.8E-02	2.4E-01
Cyclins and Cell Cycle Regulation	4.29	3.8E-03	8.0E-02	4.55	3.7E-05	5.7E-03	4.35	3.5E-03	1.5E-01
Estrogen-responsive protein Efp controls cell cycle and breast tumors growth	9.41	5.7E-03	1.0E-01	6.35	1.7E-02	2.6E-01	9.52	5.5E-03	1.8E-01
CDK Regulation of DNA Replication	8.06	9.4E-03	1.4E-01	5.44	2.8E-02	3.3E-01	8.16	9.1E-03	1.5E-01
mTOR Signaling Pathway	3.85	1.5E-02	2.0E-01	3.46	5.0E-03	1.0E-01			
Cyclin E Destruction Pathway	6.27	2.0E-02	2.4E-01	4.23	5.8E-02	4.2E-01	6.35	2.0E-02	2.5E-01
AKAP95 role in mitosis and chromosome dynamics	6.27	2.0E-02	2.4E-01	5.29	9.3E-03	1.6E-01	6.35	2.0E-02	2.5E-01
E2F1 Destruction Pathway	5.64	2.8E-02	2.9E-01	3.81	7.7E-02	4.9E-01	5.71	2.7E-02	3.0E-01
ATM Signaling Pathway	3.92	3.2E-02	3.0E-01	3.17	3.2E-02	3.0E-01	3.97	3.1E-02	3.1E-01
Role of BRCA1, BRCA2 and ATR in Cancer Susceptibility	3.71	3.9E-02	3.3E-01	3.01	4.0E-02	3.4E-01			
Tumor Suppressor Arf Inhibits Ribosomal Biogenesis	4.70	4.6E-02	3.6E-01						
Eukaryotic protein translation				8.16	0.00020	0.00766			
KEGG_PATHWAY									
Ribosome	8.75	1.6E-28	2.4E-26	10.36	1.5E-70	2.6E-68	4.34	3.4E-08	5.8E-06
Pyrimidine metabolism	5.74	2.2E-14	1.7E-12	4.87	1.8E-17	1.0E-15	3.26	4.8E-05	1.4E-03
Cell cycle	4.75	2.3E-13	1.2E-11	4.30	4.1E-18	3.5E-16	3.45	2.6E-07	1.5E-05
DNA replication	8.69	5.6E-11	2.2E-09	7.19	1.8E-13	7.4E-12	5.78	1.0E-05	4.4E-04
Aminoacyl-tRNA biosynthesis	6.79	1.2E-08	3.8E-07	5.99	1.6E-11	4.5E-10	6.13	1.7E-07	1.4E-05
Purine metabolism	3.39	2.9E-08	7.6E-07	3.36	6.7E-13	2.2E-11	2.11	4.7E-03	7.7E-02
RNA polymerase	7.04	5.4E-06	1.2E-04	7.54	2.3E-11	5.6E-10	5.45	4.4E-04	9.5E-03
One carbon pool by folate	9.50	7.7E-06	1.5E-04	7.49	1.2E-06	2.3E-05	5.75	9.2E-03	1.0E-01
p53 signaling pathway	3.86	4.8E-05	8.2E-04	2.78	4.3E-04	6.0E-03	3.73	6.7E-05	1.6E-03
Ubiquitin mediated proteolysis	2.24	4.8E-03	7.2E-02	2.47	1.5E-05	2.5E-04			
Oocyte meiosis	2.31	6.9E-03	9.2E-02	1.98	6.4E-03	7.4E-02	2.24	8.9E-03	1.0E-01
Small cell lung cancer	2.46	1.3E-02	1.5E-01	2.11	9.9E-03	1.0E-01	2.60	5.8E-03	8.7E-02
Base excision repair	3.32	1.7E-02	1.8E-01	2.70	1.6E-02	1.4E-01			
Valine, leucine and isoleucine biosynthesis	6.91	1.7E-02	1.7E-01	4.36	5.7E-02	3.3E-01	4.80	2.4E-05	8.2E-04
Nucleotide excision repair	3.09	2.3E-02	2.2E-01	4.46	8.8E-07	1.9E-05			
Glycine, serine and threonine metabolism	3.56	2.4E-02	2.1E-01	2.62	4.6E-02	2.9E-01			
Progesterone-mediated oocyte maturation	2.24	3.3E-02	2.6E-01	2.40	1.5E-03	1.9E-02			
Steroid biosynthesis	4.47	5.6E-02	3.9E-01						
Arginine and proline metabolism	2.51	5.7E-02	3.8E-01				3.12	7.1E-03	9.7E-02
Cytosolic DNA-sensing pathway	2.42	6.7E-02	4.1E-01	2.18	3.6E-02	2.7E-01			
Mismatch repair				5.45	3.3E-05	5.1E-04			
N-Glycan biosynthesis							4.00	6.7E-04	1.3E-02
PANTHER_PATHWAY									
p53 pathway feedback loops 2	4.38	5.6E-05	5.2E-03				3.78	4.6E-04	2.5E-02
De novo pyrimidine deoxyribonucleotide biosynthesis	8.34	1.1E-04	5.0E-03				6.73	1.4E-03	4.9E-02
p53 pathway	2.69	1.6E-03	4.9E-02				2.89	2.8E-04	3.0E-02
De novo purine biosynthesis	4.53	3.6E-03	8.0E-02				3.05	7.7E-02	8.9E-01
Cell cycle	5.43	3.9E-03	7.0E-02						
Formyltetrahydroformate biosynthesis	11.32	3.9E-03	5.9E-02						
De novo pyrimidine ribonucleotides biosythesis	5.96	8.2E-03	1.0E-01						
DNA replication	5.96	8.2E-03	1.0E-01						
Serine glycine biosynthesis	13.59	1.7E-02	1.9E-01						
Folate biosynthesis	13.59	1.7E-02	1.9E-01						
Ubiquitin proteasome pathway	2.24	4.4E-02	3.7E-01						

Pathways Defined by Downregulated Proteins	Both Tumour Models			Eμ-*myc* tumours			Eμ-*TCL1* tumours		
BIOCARTA									
Caspase Cascade in Apoptosis	5.58	3.0E-02	7.7E-01	4.55	9.6E-04	1.6E-01			
FAS signaling pathway (CD95)	4.04	6.9E-02	8.2E-01	2.47	8.5E-02	8.3E-01			
B Cell Receptor Complex	29.28	6.6E-02	8.9E-01						
B Lymphocyte Cell Surface Molecules	10.98	2.7E-02	9.3E-01	5.97	2.3E-02	6.5E-01	11.56	2.4E-02	8.3E-01
Apoptotic DNA fragmentation and tissue homeostasis	10.98	2.7E-02	9.3E-01						
KEGG_PATHWAY									
Hematopoietic cell lineage	8.15	3.9E-08	4.4E-06	3.79	2.9E-05	7.5E-04	9.76	9.5E-10	3.9E-01
Systemic lupus erythematosus	6.64	3.9E-07	2.2E-05	2.89	9.6E-04	7.7E-03	3.98	7.6E-03	5.0E-01
Asthma	12.76	2.0E-06	7.3E-05	5.15	6.9E-04	7.0E-03			
Cell adhesion molecules (CAMs)	4.44	2.7E-05	7.7E-04	2.48	8.0E-04	7.2E-03	2.66	4.5E-02	1.6E-01
Viral myocarditis	5.60	6.0E-05	1.1E-03	3.62	2.6E-05	8.0E-04	3.11	7.4E-02	5.5E-01
Intestinal immune network for IgA production	7.80	5.9E-05	1.3E-03	3.15	1.2E-02	7.7E-02			
Allograft rejection	7.26	9.4E-05	1.5E-03	4.40	6.2E-05	1.4E-03			
B cell receptor signaling pathway	5.92	1.1E-04	1.6E-03	6.11	5.8E-12	8.9E-10			
Autoimmune thyroid disease	5.85	3.7E-04	4.6E-03	3.54	4.6E-04	5.9E-03			
Graft-versus-host disease	6.35	7.1E-04	7.9E-03	4.03	3.1E-04	4.7E-03			
Type I diabetes mellitus	5.85	1.1E-03	1.1E-02	3.71	6.1E-04	6.7E-03			
Antigen processing and presentation	4.63	1.5E-03	1.4E-02	4.20	7.7E-07	4.0E-05			
MAPK signaling pathway	2.19	2.7E-02	2.1E-01	2.09	5.9E-04	6.9E-03			
Tryptophan metabolism	5.26	3.9E-02	2.7E-01	3.19	3.8E-02	2.0E-01	7.32	4.4E-03	3.9E-03
Amino sugar and nucleotide sugar metabolism				4.35	8.5E-04	7.3E-03			
Chemokine signaling pathway				2.34	8.0E-04	7.7E-03			
Endocytosis				2.21	1.1E-03	8.8E-03			
Apoptosis				3.42	1.8E-04	3.1E-03			
Natural killer cell mediated cytotoxicity				3.31	1.2E-05	4.8E-04			
T cell receptor signaling pathway				2.88	3.6E-04	5.1E-03			
Fc gamma R-mediated phagocytosis				4.34	9.1E-08	7.0E-06			
Leukocyte transendothelial migration				3.04	1.2E-04	2.3E-03			
Regulation of actin cytoskeleton				2.15	1.2E-03	8.5E-03			
Arginine and proline metabolism							4.42	6.0E-02	1.1E-07
PANTHER_PATHWAY									
B cell activation	4.44	6.9E-04	2.6E-02	5.04	5.7E-11	5.4E-09			
T cell activation	3.63	6.1E-04	4.5E-02	3.51	5.4E-08	2.6E-06			
Ras Pathway	3.91	3.5E-03	8.6E-02	3.11	2.1E-04	5.1E-03			
PDGF signaling pathway	2.64	6.8E-03	1.2E-01	2.14	1.3E-03	2.5E-02			
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	2.67	1.6E-02	2.2E-01	2.01	1.1E-02	1.2E-01			
FAS signaling pathway	4.19	6.6E-02	5.8E-01	3.12	2.2E-02	1.9E-01			
Inflammation mediated by chemokine and cytokine signaling pathway	1.66	9.4E-02	6.6E-01	2.11	1.3E-05	4.2E-04			
Cytoskeletal regulation by Rho GTPase				2.51	2.0E-03	3.1E-02			
EGF receptor signaling pathway				2.20	3.5E-03	4.6E-02			
Blood coagulation							6.41	1.7E-04	1.2E-02

Appendix A11. **Pathways identified as significantly enriched by IPA from the proteins determined as significantly regulated in CLL.** IPA canonical pathway analysis of upregulated, downregulated and all regulated proteins and the subsequent enrichment values and p-values.

	Upregulated		Downregulated		All regulated	
Ingenuity Canonical Pathways	$-\log_{10}$ (p-value)	%	$-\log_{10}$ (p-value)	%	$-\log_{10}$ (p-value)	%
Granzyme A Signaling	**6.78**	**47.1%**	**2.44**	**23.5%**	**9.49**	**70.6%**
Cleavage and Polyadenylation of Pre-mRNA	**12.00**	**83.3%**			**9.17**	**83.3%**
Fatty Acid β-oxidation I	**6.63**	**33.3%**	**1.55**	13.3%	**7.68**	**46.7%**
B Cell Receptor Signaling	0.53	5.3%	**9.96**	16.4%	**7.56**	**21.6%**
PI3K Signaling in B Lymphocytes	0.58	5.7%	**8.71**	17.9%	**6.91**	**23.6%**
Hereditary Breast Cancer Signaling	**8.59**	18.3%	0.26	4.0%	**6.13**	**22.2%**
iCOS-iCOSL Signaling in T Helper Cells	0.57	5.8%	**6.35**	16.5%	**5.19**	**22.3%**
Primary Immunodeficiency Signaling	0.23	4.4%	**6.90**	**26.1%**	**5.00**	**30.4%**
Agrin Interactions at Neuromuscular Junction	0.51	6.0%	**5.03**	17.9%	**4.17**	**23.9%**
Macropinocytosis Signaling			**6.55**	**20.6%**	**4.09**	**23.5%**
IL-4 Signaling	0.43	5.5%	**4.64**	16.4%	**3.70**	**21.9%**
Apoptosis Signaling	**1.93**	10.2%	**2.14**	10.2%	**3.68**	**20.5%**
T Cell Receptor Signaling	0.96	7.3%	**3.46**	12.5%	**3.66**	**19.8%**
Leucine Degradation I	**4.84**	**55.6%**			**3.48**	**55.6%**
Ethanol Degradation II	**2.11**	16.7%	**1.55**	13.3%	**3.36**	**30.0%**
Assembly of RNA Polymerase II Complex	**5.24**	**22.0%**			**3.29**	**24.0%**
NF-κB Activation by Viruses			**5.38**	17.8%	**3.19**	**20.5%**
Ephrin B Signaling			**4.64**	16.4%	**3.19**	**20.5%**
IL-15 Signaling	0.52	6.1%	**3.66**	15.2%	**3.16**	**21.2%**
Induction of Apoptosis by HIV1	**1.43**	10.2%	**2.11**	11.9%	**3.13**	**22.0%**
IL-2 Signaling	0.41	5.7%	**3.73**	17.0%	**3.05**	**22.6%**
FcγRIIB Signaling in B Lymphocytes	0.28	4.9%	**3.81**	**19.5%**	**2.89**	**24.4%**
B Cell Development			**5.10**	**28.6%**	**2.88**	**28.6%**
γ-linolenate Biosynthesis II (Animals)	0.29	5.9%	**3.43**	**29.4%**	**2.80**	**35.3%**
Mitochondrial L-carnitine Shuttle Pathway	0.29	5.9%	**3.43**	**29.4%**	**2.80**	**35.3%**
Valine Degradation I	**4.18**	**33.3%**			**2.65**	**33.3%**
Fatty Acid Activation	0.37	7.7%	**2.91**	**30.8%**	**2.59**	**38.5%**
DNA Methylation and Transcriptional Repression Signaling	**2.92**	**25.0%**	0.26	5.0%	**2.39**	**30.0%**
Nucleotide Excision Repair Pathway	**4.09**	**22.9%**			**2.23**	**22.9%**
Spliceosomal Cycle	**2.75**	**100.0%**			**2.18**	**100.0%**
Glutaryl-CoA Degradation	**2.93**	**33.3%**			**1.90**	**33.3%**
IL-9 Signaling			**3.55**	**20.6%**	**1.76**	**20.6%**
Ascorbate Recycling (Cytosolic)			**2.35**	**66.7%**	**1.73**	**66.7%**