UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Cancer Sciences

Volume 1 of 1

Evaluating the genomic landscape of B cell malignancies

by

Jade Forster

Thesis for the degree of of Doctor of Philosophy

September 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE

Cancer Sciences

Thesis for the degree of Doctor of Philosophy

EVALUATING THE GENOMIC LANDSCAPE OF B CELL MALIGNANCIES

Jade Forster

Splenic marginal zone lymphoma (SMZL) and chronic lymphocytic leukaemia (CLL) are B cell malignancies, predominately affecting the elderly. The disease course of both SMZL and CLL is highly variable, with some patients dying rapidly within a month whilst other remain stable and live a normal lifespan. Biomarkers are used to help to distinguish those patients who may progress. Genomic abnormalities such as copy number alterations and mutation of genes may have prognostic value in CLL and SMZL. Several technologies were used to assess the genomic landscape in B cell malignancies; low resolution technology such as multiplex ligation-dependant probe amplification (MLPA) and Sanger sequencing as well as higher resolution methods such as SNP 6.0 arrays and next generation sequencing technologies; whole exome sequencing, TruSeq and Nextera XT. MLPA was used to assess both the copy number alterations (CNA) and mutational abnormalities in the well-defined CLL4 clinical trial cohort. Though the technology is restrictive in terms of probe location and sensitivity, there was statistical concordance with Sanger sequencing and fluorescence *in situ* hybridisation (FISH). Novel CNA, independently associated with survival were uncovered; 1) a very indolent disease course in patients carrying a biallelic 13q deletion with *IGHV* mutated genes and 2) a very poor prognosis in patients with a 9p deletion. Exome sequencing of CLL (n=6) and SMZL (n=7) patients uncovered novel variants via a bioinformatical pipeline; this pipeline was validated using more traditional sequencing methods. The candidate genes; *U2AF1*, *BIRC3*, *POT1* and *MYD88* are implicated in CLL and were screened in a larger cohorts of patients. Analysis of the 11q loci in relation to *ATM* and *BIRC3* gene mutations, identified alterations of *ATM* that impact most significantly on survival. Nextera technology was optimised to screen ibrutinib treated CLL patients for *BTK* and *PLCG2* mutations. *DNAH9* and *SPEN* were also sequenced in a small cohort of patients. A breakpoint deletion was identified in the *DNAH9* gene, suggesting this gene may be relevant in CLL. Further studies are needed to examine

these newly defined high and low risk groups, and other genetic abnormalities in another well-defined CLL cohort, to ascertain their pathogenic and biological significance.

Table of Contents

List of Tables

List of Figures

DECLARATION OF AUTHORSHIP

I, Jade Forster, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

EVALUATING THE GENOMIC LANDSCAPE OF B CELL MALIGNANCIES

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. Parts of this work have been published as:

Longitudinal copy number, whole exome and targeted deep sequencing of 'good risk' IGHV-mutated CLL patients with progressive disease.
Rose-Zerilli MJ, Gibson J, Wang J, Tapper W, Davis Z, Parker H, Larrayoz M, McCarthy H, Walewska R, **Forster J**, Gardiner A, Steele AJ, Chelala C, Ennis S, Collins A, Oakes CC, Oscier DG, Strefford JC.
Leukemia. Epub 2016 Feb 5. PMID: 26847028

Telomere length predicts progression and overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial.
Strefford JC, Kadalayil L, **Forster J**, Rose-Zerilli MJ, Parker A, Lin TT, Heppel N, Norris K, Gardiner A, Davies Z, Gonzalez de Castro D, Else M, Steele AJ, Parker H, Stankovic T, Pepper C, Fegan C, Baird D, Collins A, Catovsky D, Oscier DG.
Leukemia. 2015 Dec;29(12):2411-4. PMID: 26256637

Low frequency mutations independently predict poor treatment-free survival in early stage chronic lymphocytic leukemia and monoclonal B-cell lymphocytosis.
Winkelmann N, Rose-Zerilli M, **Forster J**, Parry M, Parker A, Gardiner A, Davies Z, Steele AJ, Parker H, Cross NC, Oscier DG, Strefford JC.
Haematologica. 2015 Jun;100(6). PMID: 25710457

ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q deleted chronic lymphocytic leukemia: data from the UK LRF CLL4 trial.
Rose-Zerilli MJ, **Forster J**, Parker H, Parker A, Rodríguez AE, Chaplin T, Gardiner A, Steele AJ, Collins A, Young BD, Skowronska A, Catovsky D, Stankovic T, Oscier DG, Strefford JC.
Haematologica. 2014 Apr;99(4):736-42. PMID: 24584352

The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial.
Oscier DG, Rose-Zerilli MJ, Winkelmann N, Gonzalez de Castro D, Gomez B, **Forster J**, Parker H, Parker A, Gardiner A, Collins A, Else M, Cross NC, Catovsky D, Strefford JC.
Blood. 2013 Jan 17;121(3). PMID: 23086750

Whole exome sequencing identifies novel recurrently mutated genes in patients with splenic marginal zone lymphoma.
Parry M, Rose-Zerilli MJ, Gibson J, Ennis S, Walewska R, **Forster J**, Parker H, Davis Z, Gardiner A, Collins A, Oscier DG, Strefford JC.
PLoS One. 2013 Dec 13;8(12).
PMID: 24349473

Non-coding NOTCH1 mutations in chronic lymphocytic leukemia; their clinical impact in the UK CLL4 trial.
Larrayoz M, Rose-Zerilli MJJ, Kadalayil L, Parker H, Blakemore S, **Forster J**, Davis Z, Steele AJ, Collins A, Else M, Catovsky D, Oscier DG, Strefford JC.
Leukemia. 2016. October:1–10.
PMID: 27773930

Signed: .....

Date: .....

Acknowledgements

I would like to acknowledge my main supervisor Professor Jon Strefford for all of the support and guidance over the years and for giving me the opportunity to do this PhD project. I would also like to acknowledge my other supervisors Dr Mathew Rose-Zerilli and Professor Mark Cragg, for the additional support. Especially Mat whom was of great support in the lab as well as with data analysis and who provided and undertook a lot of the NGS Bioinformatical analysis in this study.

I would also like to say I am indebted to Dr Helen Parker for being another source of support and guidance throughout this PhD and especially in editing this thesis.

To all the other members of the Southampton Cancer genomics group, who have all helped me at one point during my research; Thank you all so much.

Also to Francesco Forconi and Ian Tracey for providing Southampton patient samples and clinical data. And the Southampton bioinformatical team for analysing the NGS data.

To the molecular pathology team at The Royal Bournemouth Hospital for providing samples, especially to Professor David Oscier, Dr Zadie Davis and Dr Hazel Robinson for running the MLPA experiments.

To my partner, Jason and my Daughter Jaya. Thank you for being my life, my support and my source of happiness.

All my friends and family for the support and encouragement over the years, especially my Mum and Dad. Mum, thank you for being an excellent Nana, looking after my daughter twice a week whilst also working full time. Dad, always there when I need you. My dear friend Beth, who as well as proof reading many of my chapters, is also a source of inspiration and one of the reasons I want to continue to work in cancer research.

And finally to my Grandad and all the other lives taken too early by cancer, I want to continue in this line of work for you.

Definitions and Abbreviations

Abbreviation	Definition
ALL	Acute Lymphoblastic Leukaemia
AID	Activated induced B cell deaminase
AML	Acute myeloid leukaemia
ATP	Adenosine Triphosphate
B2M	B2-microglobulin
BCR	B cell receptor
BLAT	Blast-Like Sequence Alignment Tool
BLAST	Basic Local Alignment Search Tool
BP	Basepairs
BTK	Bruton's Tyrosine Kinase
CDK	cyclin dependant kinase
cDNA	Complementary Deoxyribonucleic Acid
CHL	Chlorambucil
CLL	Chronic lymphocytic leukaemia
CML	Chronic Myeloid Leukaemia
CMML	Chronic Myelomonocytic Leukaemia
CNA	Copy number alteration
CSR	Class switch recombination
ddPCR	Droplet digital PCR
DLBCL	Diffuse large B cell lymphoma
DNA	Deoxyribonucleic acid
dsDNA	Double Stranded Deoxyribonucleic Acid
DTT	Dithiothreitol
ERIC	European research initiative on CLL
FACS	Fluorescence activated cell sorting
FC	fludarabine cyclophosphamide
FCS	Fetal Calf Serum
FDR	Fludarabine
FISH	Fluorescent *in situ* hybridisation
FL	Folicular lymphoma

GC	Germinal centre
GL	Germline
HBSS	Hank's Balanced Salt Solution
HDF	Hi-Di Formamide
HSC	Haematopoetic Stem Cell
HR	Hazard ratio
HRM	High resolution melting
Ig	Immunoglobin
IGHV	Immunoglobulin heavy chain variable
IGV	Integrated genome viewer
JMML	Juvenile Myelomonocytic Leukaemia
Kb	Kilobase
kDa	Kilodalton
LDH	Lactase dehydrogenase
LDT	Lymphocyte doubling time
LOH	Loss of heterozygosity
LPL	Lymphoplasmastic lymphoma
M CLL	Mutated IGHV CLL
M-CLL	Mutated Chronic Lymphocytic Leukaemia
MALT	gastric mucosa associated lymphoid tissue lymphoma
Mb	Megabase
MBL	Monoclonal B cell lymphocytosis
MCL	Mantle cell lymphoma
MDR	Minimally deleted region
MDS	Myeloidysplastic Syndrome
MDS	Myelodysplastic syndrome
miRNA	Micro RNA
MLPA	Multiplex ligation dependant probe amplification
MM	Multiple myeloma
MMQPCR	Monochrome multiplex quantitative PCR
MOPS	3-(n-Morpholino)Propanesulfonic Acid
mRNA	Messenger Ribonucleic Acid
MW	Molecular Weight
MZ	Marginal zone
MZL	Marginal zone lymphoma

NGS	Next generation sequencing
OS	Overall survival
PCR	Polymerase chain reaction
PFS	Progression free survival
PID	Patient identification number
F	Phenylmethanesulfonyl Fluoride
qPCR	Quantitative PCR
QRT-PCR	Quantitative Reverse Transcriptase Polymerase Chain Reaction
RCN	Relative Copy Number
REC	Research Ethics Committee
RNA	Ribonucleic acid
RPM	Revolutions Per Minute
RS	Ritchers syndrome
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SBS	Sequencing by synthesis
SHM	somatic hyper mutation
SIFT	Sorting Intolerant From Tolerant
SMZL	Splenic marginal zone lymphoma
SNP	Single nucleotide polymorphism
SNRNA	Small Nuclear Ribonucleic Acid
SnRNP	Small Nuclear Ribronucleoprotein
SNV	Single Nucleotide Variant
ssDNA	Singe Stranded Deoxyribonucleic Acid
STELA	Single Telomere Length Analysis
TBS	Tris Buffered Saline
TFS	Treatment free survival
TK	Thymidine kinase
TL	Telomere length
TTFT	Time to first treatment
UM CLL	Unmutated IGHV CLL
UMI	Unique molecular identifier
UTR	Untranslated Region
VAF	Variant allele frequency
WCL	Whole Cell Lysate

WES	Whole exome sequencing
WGA	Whole genome amplification
WGS	Whole genome sequencing
WHO	World Health Organisation
WM	Waldenströms macroglabulinemia
WT	wildtype

Chapter 1: Introduction

1.1 The hallmarks of cancer

For a normal cell to become malignant, requires several hallmark changes such as; continuous proliferative signalling, evade growth suppressors, activate invasion and metastasis, replicative immortality, induction of angiogenesis, resistance to apoptosis, the avoidance of immune destruction and deregulation of cellular energetics.[1] Under pinning all of these hallmarks is genomic instability that is driving and generating genetic diversity allowing the normal cell to obtain all these hallmarks of cancer.[1] Cancer has been shown to be comparable to the Darwinian, clonal evolutionary selection process, with competition between differing clones causing and driving malignancy.[2] Through natural selective pressures; intrinsic competition as well as extrinsic selection (i.e. chemotherapy), occur frequently to gain a survival advantage. Oncogenes are activated through protein altering mutations, chromosome/gene amplifications and translocations, whilst tumour suppressor genes are de activated through frameshift/stop codon mutations and chromosomal/gene deletions.[3]

1.2 B cell development and malignancies

All haematopoietic cells arise from a haematopoietic stem cell (HSC) through a process known as haematopoiesis. The lymphoid branch of this process involves the differentiation into a B cell (Figure 1).[4] B lymphocytes are a population of cells that express clonally diverse cell surface immunoglobulin (Ig) receptors that recognise specific antigens.[5] B cells develop in the bone marrow and the process involves the rearrangement of the V, D and J gene segments in the heavy chain and the V and J gene rearrangements in the light chain locus of the B cell receptor (BCR).[5] The B cell enters the periphery. Antigen induced B cell activation and differentiation in lymphoid tissue cause the germinal center (GC) reaction that is characterised by a clonal expansion and selection, the process involves class switch recombination (CSR) of the Ig heavy locus, somatic hyper mutation (SHM) of the V_H genes and the high affinity selection of the BCR.[5,6] CSR and SHM are implemented by activated induced cytidine deaminase (AID). Once immature B cells leave the bone marrow and enter the periphery they produce antibodies after encountering lipopolysaccharide antigens.[5] Lymphoid tissue (spleen/lymph node) B cells are presented with antigen from dendritic cells and differentiate to plasma cells or enter the GC reaction as previously mentioned.[5] Through studying the Ig gene rearrangements, as well as the cytogenetic, genetic and gene expression

profiling, have identified the B cell lineage of malignant cells from their normal counterpart.[5] Ig gene arrangement profiles have deduced that acute lymphoblastic leukaemia (ALL) is a pre B cell malignancy.[5,7] The Ig gene arrangement profiles in Hodgkin's lymphoma revealed a B cell origin.[5,8] Splenic marginal zone lymphoma (SMZL) originates from a marginal zone (MZ) of lymphoid follicles (spleen and mucosa associated tissues) memory B lymphocytes.[9] The cellular origin of chronic lymphocytic leukaemia (CLL) was derived from gene expression profiling that identify the IGHV unmutated subset of CLL as arising from a mature CD5 positive post GC B cell, whilst the other CLL subset with mutated IGHV genes arising from a CD5 and CD27 positive post GC B cell.[10]

Both of the B cell neoplasms, SMZL and CLL, will be explored in more detail in the next sections.

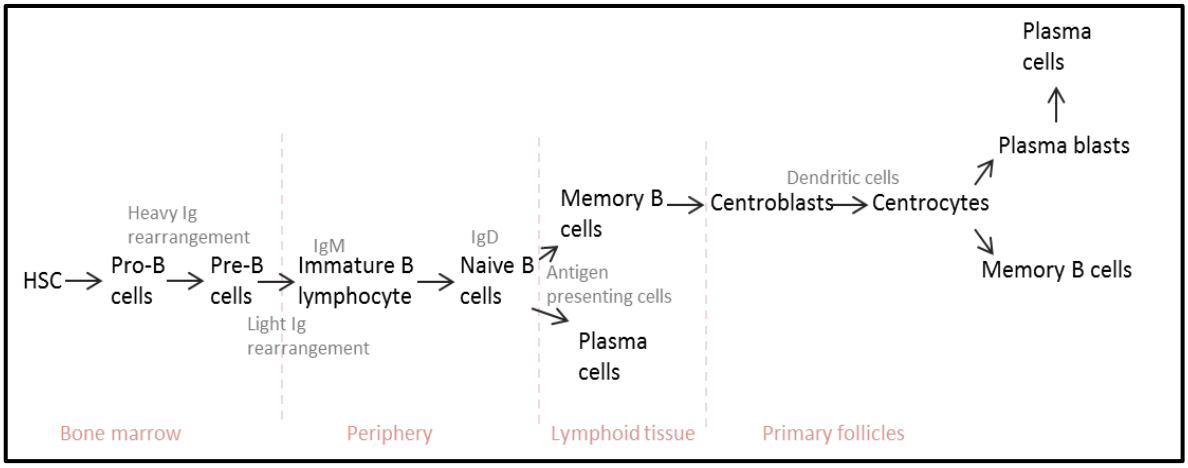


Figure 1 Simplified diagram of the process of B cell development from a haematopoietic stem cell (HSC) to differentiated mature B cell.
The earliest form of a B cell is a progentior B cell where the Ig heavy chain locus is rearranged to form a pre B cell. The light Ig rearrangement derives an IgM expressing immature B lymhocyte cell that then moves from the bone marrow to the periphery where it expresses IgD (naive B cell). G0 phase cells enter the lymphoid tissue where the cells are exposed to antigen presenting cells leading to differentiation to plasma or memory B cells. The memory B cells then go to the primary follicles where further exposure to antigen (dendritic cells) causes activation and differentiate into centroblasts; Ig isotype switching and somatic mutations in the variable region of the Ig generates high-affinity antibodies. Centroblasts differentiate to centrocytes and express Ig, some with high affinity antibodies differentiate into either memory B cells or plasmablasts, that migrate to the bone marrow and differentiate into plasma cells.[6]

1.2.1 **SMZL**

One of the major roles of marginal zone (MZ) B cells is the early response to microbial antigen in the blood and mucosal tissues, as well as removing senescent cells.[11] There are 3 distinct types of MZ lymphomas (MZL) including nodal MZL, extra nodal MZL of MALT type, and SMZL and other splenic B cell neoplasms (including splenic B cell lymphoma/leukaemia unclassified and hairy cell leukaemia and splenic diffuse red pulp lymphoma).[11]

SMZL is a rare chronic B cell lymphoproliferative disease, concerning the spleen, bone marrow and peripheral blood but it does not involve the lymph nodes.[11] It is so rare that only 2% of all B cell lymphomas are diagnosed as SMZL.[11] It is a disease predominant amongst the elderly, with a median survival around 10 years, 70% of patients require some form of intervening treatment and up to 10% of these patients can then progress to a more aggressive disease; large B cell lymphoma that has a median survival of around 4 years.[11,12] Although there is no standard treatment regime, there are good response rates to treatments, which include; splenectomy, rituximab as a single agent or in combination with a chemotherapy. The oldest and most commonly used treatment is splenectomy but this is associated with a high mortality and so current practice is starting to move away from using this radical form of treatment if possible.[11,12] The overall (OR) response rate for single use rituximab is more than 90% and a complete response (CR) in up to 50% of patients.[12]

Patterns of SHM suggest that this process is ongoing through disease progression, which suggests that the pathogenesis of SMZL is an antigen driven process whereby B cells are selected with a distinct BCR Ig rearrangements.[11] This has been corroborated with report of an association of hepatitis C infection being associated with the development of SMZL.[9]

SMZL is diagnosed using clinical, morphological, histological and immunophentypic information.[11] Prognostic markers in SMZL are scarce due to the rarity of this disease that has meant that SMZL has been understudied and so establishing markers of disease progression has been difficult. Haemoglobin, platelet count, serum albumin and other clinical markers have been used to help aid prognosis, but with the advent of next generation sequencing (NGS), recurrent genomic abnormalities have been uncovered in recent years and suggested to be of prognostic relevance.[11,13–15] Cytogenetic analysis using SNP 6.0 and aCGH studies have identified that more than 72% of SMZL cases harbour a chromosomal abnormality. Including that 40% of SMZL patients carry a 7q31-32 deletion and 3q duplications in around 30% of cases.[16] Other CNAs include gains of 8q, 9q, 12q and 18q, and deletions of 6q, 8p, 14q and 17p as well as translocations of the immunoglobulin genes.[11,16,17] The 7q deletion is a SMZL specific aberration as it is rarely seen in

other B cell malignancies. It can also be found as a sole abnormality in about 25% of 7q deleted cases with a MDR identified a 2.7Mb region on the 7q32, though no candidate genes have been identified.[11,18] Recent whole genome and exome analysis of SMZL have identified on average around 25 somatically acquired variants per patient and variants involved in Notch signalling, the NF-κB pathway, chromatin remodelling and cytoskeleton.[11,13,14,19–21] The most frequently mutated gene identified in up to 25% of SMZL patients, is a *NOTCH2* mutation.[13,15,19,20] This gene mutation again seems to be SMZL specific and rarely found in other B cell diseases, and it causes a constitutive activation of Notch signalling.[11] *NOTCH2* mutations and their prognostic relevance remains unclear though many studies have tried to establish their relevance. Rossi *et al*. found that patients harbouring a *NOTCH2* mutation had a significantly better survival than wildtype (WT) patients, while others have seen a reduced survival in mutated patients.[15,20] Other members of the Notch pathway, such as the *SPEN* and *DTX1* genes, have been identified as being frequently mutated in up to 11% of SMZL patients.[11] Another 34% of patients have an NF-κB pathway mutation including, *CARD11*, *MYD88*, *TRAF3*, *BIRC3*, *MAP3K14*, *TNFAIP3* and *IKBKB*.[19] *TP53* alterations have also been found implicated in up to 15% of SMZL patients and also impact negatively upon survival.[15] The *KLF* gene mutation has been shown to associate with other genomic abnormalities such as the 7q deletion as well as with the IGHV1-2*04 gene sub group.[15] More work is needed to investigate and characterise the genomic landscape of SMZL and relate this back to clinical associations.

Another B cell neoplasm, CLL has been utilised extensively to study cancer at the genetic level as a model of tracking, as it evolves, through out the disease stages.

1.2.2 **Chronic Lymphocytic Leukaemia**

CLL is a neoplasm of mature B cells, affecting the blood, bone marrow and lymph nodes, which is largely found in the elderly, with a median age at diagnosis around 72 years old [22], and has an incidence of around 4.2 cases per 100,000 people per year in developed countries [23]. The diagnosis of CLL is usually by immunophenotyping of surface antigen markers by fluorescent activated cell sorting (FACS). This includes positivity for the B cell antigens CD19 and CD23 and a negative result for FMC7, as well as Kappa or Lambda immunoglobulin light chain expression [24]. CLL is a progression of CD5 positive monoclonal B cell lymphocytosis (MBL), which is detectable in 3.5% of a normal healthy elderly population but of which only 1% then advance to CLL [25–27]. MBL patients have similar characteristics to CLL patients though the lymphocytosis is less than 50000/mL blood.[28] MBL has also been shown to have the same cytogenetic abnormalities, just at lower

frequencies, as CLL.[29,30] Why some MBL cases go on to develop CLL and others do not is currently unknown. Up to 90% of CLL patients are first diagnosed with an early stage-low risk disease, but disease course and outcome are highly variable between individuals with CLL, some having a relatively benign disease and normal life expectancy whilst others have a more aggressive disease, progress rapidly and may die within a few months[31]. An early stage patient could expect to have a median survival of more than 10 years, whilst a later stage patient can expect a median survival of around 6.5 years.[32] Overall the 5 year relative survival rate of a CLL patient is around 87.9% [33]. CLL can also progress onto diffuse large B cell lymphoma (DLBCL) or Richter's syndrome (RS). RS is a highly aggressive lymphoma of the same B cell origin as CLL, though it is notoriously difficult to treat. Around 10% of CLL cases go on to develop RS.[34] So this is why the standard approach to tackling CLL is to monitor the patient or 'watch and wait' for disease progression.

Over the years various biomarkers have been discovered that help to predict disease progression, some are widely used to aid in the prognostication of patients whilst others are less popular.

1.2.2.1 Clinical features of CLL

To estimate prognosis in CLL two staging systems were devised in the 1970's, by Rai [35]and Binet [36]. They are both based on the physical examination of a patient for lymphadenopathy and hepato-splenomegaly, as well as blood tests, which are used to detect lymphocytosis, anaemia and thrombocytopenia. After these tests the patient is placed into a risk group; 0-IV for Rai staging and A-C for Binet[35–37]. Both these staging systems have been widely used as standard risk stratification of a patient at diagnosis and are still used today.

Serum protein markers such as thymidine kinase (TK), lactase dehydrogenase (LDH) and B2-microglobulin (B2M) are also widely used and predict outcome in CLL. [37–39] TK can be a measure of proliferation of the tumour and in turn disease progression, whilst B2M correlates with disease stage [39]. For instance patients who were categorised to Binet stage A but that had high B2M levels, were shown to progress quicker. LDH is an enzymatic protein that is found in CLL at elevated levels and can be used to predict the development to RS[37,38].

Other markers of disease progression include lymphocyte doubling time (LDT); that is quite simply the length of time (months) it takes for the peripheral blood lymphocytes to double in number. An LDT of less than 12 months indicates a poor prognosis and a more aggressive disease, were as an LDT of more than 12 months relates to a benign disease and so good prognosis. The LDT

had been shown to be an independent prognostic marker, with some degree of correlation to the clinical staging.[40]

1.2.2.2 Treatment strategies

The standard regime for a CLL patient at diagnosis is a 'watch and wait' approach, but once a patient had progressed and needs treatment various options can be used. The current standard therapy that is suggested for physically fit patients is a standard chemotherapy such as fludarabine/cyclophosphamide (FC), bendamustine or chlorambucil, combined with an anti-CD20 antibody immunotherapy reagent such as rituximab, ofatumumab or obinutuzumab[31]. Commonly the FCR (fludarabine-cyclophosphamide-rituximab) treatment combination is offered to fit patients, but the immunotherapy obinutuzumab combined with chlorambucil has shown good results in less fit patients. Allogenic stem cell transplant remains the only option to completely 'cure' a patient of CLL, but this option is high risk and is only offered to those were all other treatments have failed and the patient meets the requirements for the treatment; i.e. has *TP53* mutation and/or deletion.[31,37,41]

New therapies, using small molecule inhibitors are starting to be utilised in the clinic as a last chance therapy, with some success, for instance kinase inhibitors (Ibrutinib) targeting the BCR pathway, these will be fully explored later.[42]

1.2.2.3 Mutational status of the immunoglobulin genes

In both normal and tumour B cells, signalling response to antigens is through the cell surface BCRs. The somatic hypermutation of the variable region of their immunoglobulin heavy (IGHV) of these receptors can be used to classify CLL into two almost equal subsets of the disease, with above 98% homology to a germline sequence being classified as unmutated CLL (U-CLL) and below 98% as mutated CLL (M-CLL).[5,43] Patients who lie on the borderline of these cut offs can be difficult to categorise and their predictive risk is less certain[37,43,44]. It has been suggested that the U-CLL cases arise from a naïve B cell population, while M-CLL cells were derived from post-germinal centre or memory B cells.[10] This suggests two distinct pathways of disease evolution, with M-CLL having a mainly benign disease course, with little treatment intervention unless progression occurs while U-CLL is associated with a more aggressive disease and poor prognosis [43,44]. CD38 and Zap-70, as well as TK protein expression has been shown to be highly correlated with IGHV mutational status, with U-CLL tending to have higher levels of all proteins[45–47]. CD38 is a cell surface marker that has been shown to bind to another cell surface CD31, this is important in cell to cell interactions that activate cell survival pathways in normal and CLL B cell lymphocytes.[45] Zap-70 is a protein tyrosine

kinase that is normally expressed by T cells and is involved in intracellular signalling.[48] Both CD38 and Zap-70 have been proposed as an alternative, cheaper, less labour intensive biomarkers for distinguishing IGHV status, as they are determined via flow cytometry, with a more than 30% expression of either protein relating to a poor prognosis or an unmutated IGHV status.[44,46]

1.2.2.4 Genomics of CLL

1.2.2.4.1 Chromosomal aberrations

Unlike other haematological malignancies, CLL is characterised by partial gains but mainly deletions of chromosomal material, rather than translocations which tend to be rare events in CLL.[41] Recurrent copy number alterations (CNA) were first discovered in CLL in the 1980s, with the discovery of an additional copy of chromosome 12.[49] This pivotal discovery was then followed up with the discovery of other recurrent CNAs including the partial loss of 13q, 11q and 17p .[50–52] In 2000, Döhner *et al*. used a panel of specific fluorescent *in situ* hybridisation (FISH) probes revealing chromosomal aberrations in 82% of CLL patients. [53] This pivotal publication, for the first time provided a prognostic, hierarchical-model of survival, based on the cytogenetics of a CLL patient. The most common FISH abnormality were deletions of 13q14 (55%), 11q23 (18%) and 17p11 (7%), and trisomy 12 (16%). Less prevalent abnormalities were also observed, including; deletion of 6q (6%), trisomy 8q (5%), trisomy 3q (3%) and a translocation on 14q (4%). The abnormalities were then related to survival, with a deletion of 13q, as a sole aberration showing the most favourable prognosis (133 months), followed by normal/no cytogenetic abnormality (114 months), trisomy 12 (111 months), 11q deletions (79 months) and finally 17p deletions (32 months) with the worst prognosis (Figure 2).[53] This cytogenetic, hierarchical classification of CLL has been validated across multiple studies.[24,54–57]

The deletion of 13q is the most frequent CNA observed in CLL cells associated with favourable outcome for patients who carry this as a sole abnormality. Deletion of this locus is an early event in CLL pathogeneis and can also include the additional loss of the other copy of 13q (biallelic loss) that has been associated with a reduced survival compared to single loss.[58–60] The 13q14 region has also been implicated in various other malignancies such as mantle cell lymphoma, multiple myeloma and prostate cancer[61].The deletion can vary in size between patients, from many megabases to a smaller more defined region. A 9kb minimally deleted region (MDR) that is common to most 13q deletions, though some cases do not include this MDR. The MDR is located at 13q14 and contains multiple genes including *DLEU1*, *DLEU2*, and *miR15a/16-1* have been suggested to be involved in the pathogenesis of CLL. [37,41] Two mouse models have been created to study the mechanistic role of

13q in CLL pathogenesis; with loss of all MDR genes in one model, and focal loss of the miR15a/16-1 cluster in another. Both these regions, when deleted, caused the mice to develop B-cell like diseases such as MBL and CLL. [62,63] Two further interesting observations have been made in 13q-deleted CLL. Firstly, larger deletion size have been noted and seem to impact on disease progression and survival, with the involvement of the *RB1* gene seen as a potential driver of progression.[64,65] Secondly, the number of tumour cells harbouring a 13q deletion, has been implicated is disease progression with those carrying a larger clone having a reduced time to first treatment.[66]

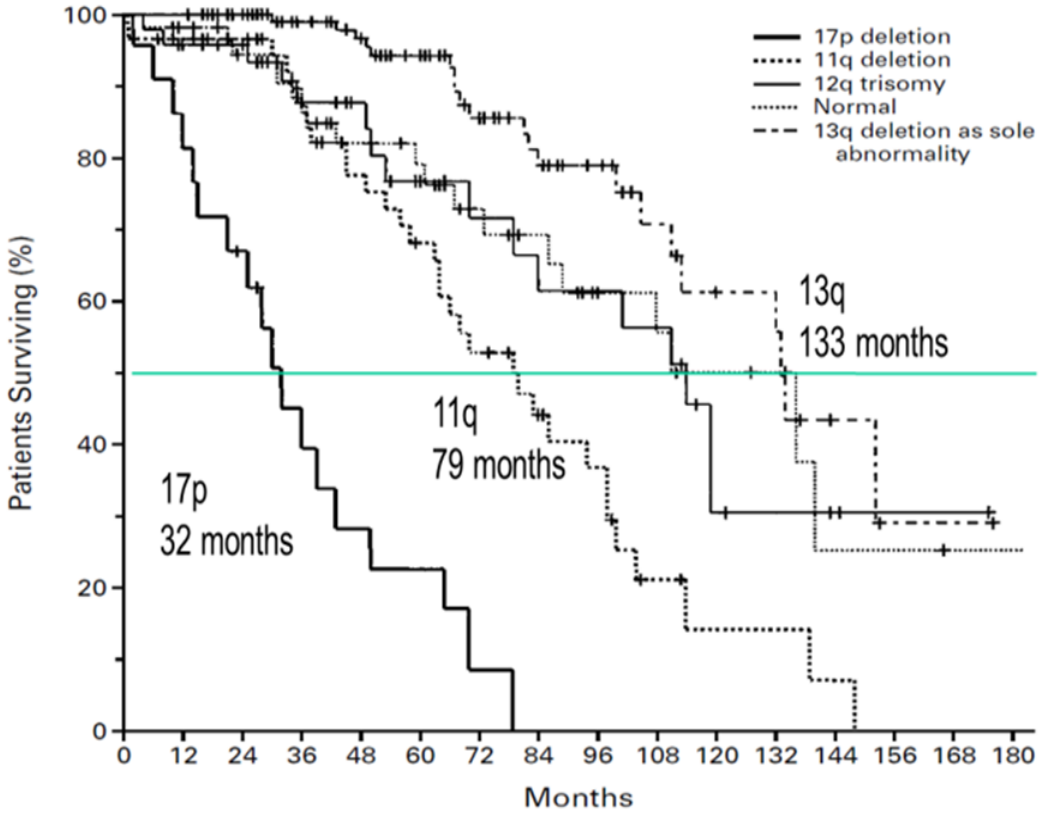


Figure 2. Hierarchical model of survival in CLL based on the recurrent FISH lesions. Figure taken from Dohner *et al*. NEJM 26 (2000)[53]**. The patients with a 13q deletion as a sole abnormality have the best outcome in terms of survival, whilst those patients harbouring a 17p deletion do the worst.**

Trisomy 12 as previously stated is associated with an intermediate prognosis, atypical immunophenotype and lymphocyte morphology, as well as over expression of integrins and adhesion molecules. Trisomy 12 is also associated with having concurrent trisomies of chromosomes 18 and 19, the impact of which is less known[67,68]. Although the additional copy of chromosome 12 was one of the first abnormalities to be confirmed in CLL, it is yet to be fully characterised in terms of CLL pathogenesis and biology; though the *MDM2* gene, which is a regulator of P53, has

been suggested as a likely candidate due to the over expression seen in trisomy 12 patients.[69] Patients carrying this abnormality tend to respond well to standard treatments and have been shown to improve drastically with the addition of an antiCD20 immunotherapy[37,41].

Deletions of the 11q22-23 region are associated with poor outcome and this abnormality is seen enriched in progressive patients and in contrast to 13q deletions, only monoallelic deletion tends to occur[41]. A 3Mb MDR has been identified in 11q deleted patients, this region contain an important tumour suppressor gene, *ATM,* which loss of has been associated with genomic instability and lymphoid malignancies.[70] ATM has important roles in cell cycle regulation, through DNA damage response and activating TP53 pathway[37]. The *ATM* gene is also found to be have somatic and germline mutations in around a third of CLL patients; 10% found in the 11q deleted patients and up to 25% of patients without an 11q deletion.[70] The biallelic inactivation of both copies of the *ATM* gene is associated with a worse survival than 11q or *ATM* gene mutation alone.[70] The confirmation of the MDR to the *ATM* gene is only seen in 5% of 11q deleted patients, therefore implicating that the many hundreds of genes also deleted along with *ATM* could also play a pathogenic role in 11q deletions.[41,71] Other genes to be suggested as interesting candidates include the genes encoding for the microRNAs miR34B and miR34C that a form a part of the TP53 network.[41]

17p deletions are highly predictive in respect to therapy as they tend to be progressive cases, often get poor responses to standard chemotherapy treatment options and so tend to have a very poor prognosis. The deletion tend to be a rare CNA (<5%) at diagnosis but are present in over 50% of chemo refractory patients[41,53]. Deletion often result from loss of the whole p arm of chromosome 17 but rarer focal deletions of 17p13.1 including the *TP53* gene have also been resported. TP53 is an important tumour suppressor protein that is involved in DNA damage repair, genome integrity, and apoptosis as well as having many other important roles in cell regulation[37]. Along with deletions, *TP53* has also found to be mutated in CLL patients with and without an 17p deletion. *TP53* disruption (deletion and/or mutation) has been implicated in a third of RS patients, as a pathway to disease transformation[37,72,73].

Chromothripsis is a massive global chromosomal rearrangement resulting from a single catastrophic mitotic cycle followed by inaccurate repair of the remaining fragments.[74] Chromothripsis as an event was first identified in CLL and has subsequently using NGS has been identified in 2-3% of CLL patients, predominately in patients with IGHV unmutated status and with a *TP53* aberration.[54,75]

1.2.2.5 Next generation sequencing

Next generation sequencing (NGS) differs dramatically to Sanger sequencing. Sanger sequencing uses the dye-terminator chemistry and capillary electrophoresis technology to sequence small segments of genes, whereas NGS uses massively parallel sequencing methods to capture the entire genome or exome [76]. Sanger sequencing established the first fully sequenced human genomes in 2001, at estimated costs of up to 3 billion US dollars, with NGS now being able to sequence a whole genome around $1000 in a matter of days, it is seen as a much more attractive method to use [77].

Massively parallel sequencing involves 3 main sequencing platforms (454/Roche, SOLiD/Applied biosystems and the Solexa/illumina platforms. All of these methods involve an initial form of fragmenting/shearing the DNA into smaller fragments followed by amplification and then a final detection phase [78].

The first NGS technology; the Roche 454 sequencing involves a bead-emulsion PCR step followed by pyrosequencing technology, which relies on the detection of pyrophosphate released during nucleotide incorporation [76,79]. The Applied Biosystems Sequencing by Oligo Ligation Detection (SOLiD) utilises the two-base sequencing based on ligation sequencing technology [76]. It is similar in that it also has a bead emulsion PCR step but then the bead is attached to a glass slide, probes are then ligated to the DNA to distinguish the sequence [78]. The (Solexa) Illumina platform on the other hand uses the sequencing by synthesis (SBS) sequencing technology [76], by which a flow cell is used to perform bridge amplification (PCR) and then utilises the reversible dye terminator chemistry to identify the sequence [78]. These three NGS technologies all have slightly differing capabilities when it comes to read length, accuracy, consumables and costs but none stand as outperforming the others though Illumina bench top sequencers are the most widely utilised.

Other NGS technologies include Ion Personal Genome Machine (PGM). PGM uses semi-conductor sequencing technology, which detects a change in pH if a proton is released during nucleotide incorporation by the polymerase. [76] Ion torrent and Illumina are targeted towards a clinical setting, hence why these technologies are widely used.[76] Newer, third generation sequencing technologies such as single molecule real time (SMRT) from Pacific bioscience, which is able to sequence read lengths up to 20kb and Oxford nanopore sequencing that is able to sequence read lengths up to 10kb on a platform (minION) able to fit in the palm of ones hand and vastly cheaper than the bench top sequencers available at present.[80] These 'third' generation technologies hope to improve upon the previous NGS platforms by emitting the PCR steps and capturing single molecules in real time.[76,80] While both these technologies have yet to stand up to

the sensitivity of the established technologies and platforms widely used, their promise is tantalising and could change the whole field of NGS whilst making the technology more affordable and accessible.

The massively parallel sequencing technologies have been used widely in recent years to capture the whole genome or exome of individuals with neoplasms or disease to improve our understanding of the genetic burden within disease. This has included CLL that has seen a large number of studies using these NGS technologies in the last few years, this will be further discussed in the following chapters.

CLL has been used to study cancer at the genetic level, due to the natural prolonged disease course of CLL and the tumour material is easily accessible, unlike most solid tumours. This has allowed a novel insight into the evolution of CLL and the dynamics of clonal selection.[81,82] The insights of knowledge and understanding gained into the CLL genome and disease pathogenesis since the advent of NGS will again be further explored in the coming chapters.

Chapter 2: Aims of the research

The pathogenesis of B cell malignancies, CLL and SMZL remains uncertain. CLL patents with good risk markers can still progress to a higher grade lymphoma, RS and/or require therapy. Whilst in SMZL a lack of clinically useful biomarkers makes it hard to predict the disease course. This study hopes to utilise current technology to help dissect and understand the genomes of these diseases in a quest to find novel prognostic biomarkers.

The first section of this study hopes to examine both common and rarely studied copy number changes and gene mutations of 3 frequently mutated genes within the context of CLL, using a relative PCR based method; MLPA.

- One of the aims is to assess the clinical applicability of MLPA as a technology by comparing to the gold standard of evaluating copy number alterations, FISH and Sanger sequencing for the genetic changes
- Another aim is to utilise the CLL4 clinical trial as a cohort to see if any novel biomarkers are identified

The second section aims to assess the genetic changes of both CLL and SMZL, using Whole exome sequencing technology.

- Matched tumour and germline CLL and SMZL samples will be whole exome sequenced
- A bioinformatical pipeline will be developed (with Southampton Bioinformatical group) to identify novel gene mutations in patients
- Novel variants will be Sanger sequenced to confirm bioinformatical pipeline

The third section aims to further assess novel genes identified from the literature that may play a role in CLL.

- Four understudied genes; *BIRC3*, *POT1*, *MYD88* and *U2AF1* are to be screened in CLL cohorts to find the frequency they present and at different disease stages
- A highthroughput screening method will be developed to screen all genes
- This section also aims to perform functional validation of the *BIRC3* gene to understand how mutations impact at the mRNA and protein level

As next generation sequencing technology progresses, the next sections of this study aims to incorporate NGS into the project and assess this technologies validity for use as a mutational screening method.

- Develop an amplicon based method, TruSeq, to target up to 25 genes
- Apply TruSeq to a local CLL cohort to identify gene mutations
- Associate any mutations to clinical data and identify any novel biomarkers
- Assess the utility of TruSeq as a clinically relevant tool
- Evaluate the functional impact of gene mutations, identified by TruSeq, upon mRNA expression
- Develop assays for use with Nextera XT to deep sequence on a NGS platform
- Screen a small cohort of CLL patients on a small molecule inhibitor therapy for *BTK* and *PLCG2* mutations

Chapter 3: Material and methods

3.1 Patient samples

Patient cohorts varied depending on the particular study, and are detailed in each results chapter. In summary, patient samples were obtained from;

1) The UK CLL4 Leukaemia Research Fund clinical trial. The trial was conducted between 1999 and 2004, and randomly assigning patients to first-line treatment with chlorambucil (CHL), fludarabine (FDR) or fludarabine plus cyclophosphamide (FC). Samples were taken prior to the initiation of therapy. All patients were diagnosed using standard morphologic and immunophenotypic criteria. Informed consent was obtained from all patients in accordance with the Helsinki declaration, and this study was approved by national or regional research ethics committees. Although the CLL4 clinical trial assessed traditional chemotherapy, a plethora of clinical data has been collected, over a decade of follow up, and allowing important associations to genomic features to be made.

2) Previously untreated CLL patients recruited in the lymphoproliferative disorder study at time of initial evaluation at the Department of Hematology of the Southampton University Hospital Trust from January 2001 to May 2015. Diagnosis of CLL was according to the 2008 International Workshop on Chronic Lymphocytic Leukemia (IWCLL2008)/National Cancer Institute (NCI) criteria and confirmed by a flow cytometry "Matutes score" in all cases. The lymphoproliferative disorder study was approved by the institutional review boards at the University of Southampton (228/02/t). All patients provided informed consent prior to inclusion in the study;

3) Previously untreated patients diagnosed and treated at the Royal Bournemouth Hospital.

Diagnostic Fluorescence *In-Situ* hybridization (FISH) was performed (by RBH cytogeneticists) on all cases using commercially available probes (LSI TP53 (17p13.1); D12Z3 (centromere12), D13S25 (13q14.3), LSI ATM (11q22.3)) according to the manufacturer's instructions (Vysis from Abbott Laboratories, Berkshire, UK). 6q21 was analysed separately (courtesy of Dr S Stilgenbauer). 200 cell nuclei were examined and scored for each probe by two independent investigators. The cut-off points for defining loss were > 5 for 11q, 13q14, 17p13.1 and 6q21, and > 3 for trisomy 12. ZAP70 and CD38 expression was determined as previously described, where 10 and 30 positive cells were classed as positive, respectively. *IGHV* genes were sequenced as previously described

and a cut-off of >98 germ-line homology was taken to define the unmutated sub-set.

3.2 DNA extraction from tumour cells

Genomic DNA was needed for applications such as WGA, HRM, PCR, Sanger sequencing, MLPA, ddPCR, SNP 6.0 and NGS.

DNA was extracted using the DNeasy mini kit (Qiagen). The general protocol is as follows:

1. AW1 and AW2 were made up with 25µl and 30µl 100% ethanol, respectively.
2. Up to $5x10^6$ cells (per column) were centrifuged for 5 minutes at 300g and resuspended in 200µl PBS.
3. 20µl proteinase K and 4µl of RNase A (100mg/ml) was added and incubated at room temperature for 2 minutes; then 200µl buffer AL was added, mixed by vortexing and incubated at 56ºC for 10 minutes.
4. 200µl 100% ethanol was added and mixed by vortexing
5. The reaction mixture was pipetted into a DNeasy mini spin column placed in a 2ml collection tube and centrifuged for 1 minute at 6000g. The flow through and collection tube were discarded
6. The column was placed in a new collection tube. 500µl buffer AW1 was added to the column and centrifuged as above. The flow through and collection tube were discarded
7. The column was placed in a new collection tube. 500µl buffer AW2 was added to the column and centrifuged for 3 minutes at 20,000g to dry the membrane. The column was carefully removed from the collection tube to prevent any ethanol carryover. The flow through and collection tube were discarded
8. The column was placed in a new collection tube and 200µl buffer AE was added directly to the membrane. It was incubated at room temperature for 1 minute then centrifuged for 1 minute at 6000g. This step was repeated using the same elutant. The extracted DNA was stored at 4ºC or -20ºC

3.3 Extraction of DNA from saliva

To obtain large amounts of pure germline DNA from CLL patients, to use as a reference sample in sequencing studies, DNA was extracted from patients' saliva using the Oragene kit (DNAgenotek).

1. Patients rinsed their mouths out with 5mls of sterile water for 5 seconds, a total of three times. A 500ml sample of saliva was then collected in the Oragene tubes (performed by research nurses).

2. Samples were mixed by inversion and incubated at 50°C for a minimum of 1 hour in a water bath.
3. 500µl sample was transferred to a tube and 20µl of Oragene purifier was added, vortex mixed and incubated on ice for 10 minutes.
4. Tubes were centrifuged at 13000 rpm for 5 mins.
5. The clear supernatant was transferred to a fresh tube and an equal volume of room temperature 95-100% ethanol was added and mixed 10x by inversion.
6. The DNA precipitated over 10 minutes, then the tube was centrifuged at 13000rpm for 2 minutes. The supernatant was discarded and the DNA pellet retained.
7. 250µl of 70% ethanol was used to wash the DNA, at room temperature for 1 minute, then removed. 100µl of TE buffer was used to dissolve the DNA pellet. DNA was stored at -20°C until needed.

3.4 RNA extraction

RNA extracted from target cells may be used to investigate the effect of disease on gene expression levels.

RNA was extracted using the RNeasy mini kit (Qiagen). The protocol is as follows:

1. The DNase stock solution was prepared by suspending the lyophilized DNase 1 in 550µl RNase free water and gently mixing. 10µl aliquots were stored at -20°C
2. Buffer RLT was supplemented with 10µl β-mercaptoethanol (BDH) per 1ml, in the fume hood. This solution was stable for 1 month.
3. 4 volumes 100 ethanol were added to buffer RPE to obtain a working solution.
4. To lyse the cells, 350µl buffer RLT was added to $<5x10^6$ pelleted cells. 600µl buffer RLT was added to a maximum of 10^7 cells.
5. To completely homogenise the cells, they were passed through a 20-gauge needle fitted to an RNase free syringe, five times.
6. One volume 70% ethanol was added and mixed by pipetting. 700µl was applied to a column and centrifuged at 10,000g for 15 seconds at room temperature. The flow through was discarded. If the sample volume exceeded 700µl, successive volumes were applied to the column and centrifuged as above
7. 350µl buffer RW1 was applied to the column and centrifuged as above. The flow through was discarded

8. 70µl buffer RDD was added to a 10µl aliquot of DNase 1 solution, and mixed by inverting the tube. This solution was applied directly onto the RNeasy silica-gel membrane and incubated for 15 minutes at room temperature
9. 350µl buffer RW1 was added to the column and centrifuged as above. The column was then transferred to a new 2ml collection tube
10. 500µl buffer RPE was applied to the column and centrifuged as above. The flow through was discarded. Another 500µl buffer RPE was added and centrifuged at 10,000g for 2 minutes to dry the membrane
11. The column was transferred to a new 1.5ml collection tube and centrifuged at 10,000g for 1 minute
12. 50µl RNase free water was added to the membrane and incubated at room temperature for 10 minutes. The column was spun at 10,000g for 1 minute. This step was repeated once. The RNA was stored at -80°C

3.5 Nucleic acid assessment

After DNA or RNA extraction, the nucleic acids were assessed to check the quality and quantity before being applied for downstream applications.

3.5.1 **Gel electrophoresis**

Gel electrophoresis was used to check DNA quality and PCR amplification. As the phosphate backbone of DNA is negatively charged, it can be loaded onto a semi permeable gel and an electrical current applied that pushes the DNA towards a positive anode. As it migrates through the gel, fragments of DNA separate by size with large high molecular weight fragments moving more slowly than those with a smaller molecular weight. An intercalculating dye is used to bind the DNA and fluoresce under UV light, allowing an assessment of DNA quality to be made. DNA comprised of different sized fragments will pass through the gel at different times creating a ladder like image. Genomic DNA is comprised of very large molecules that will not migrate far, creating a discrete band. A smear indicates DNA degradation or digestion (Figure 3).

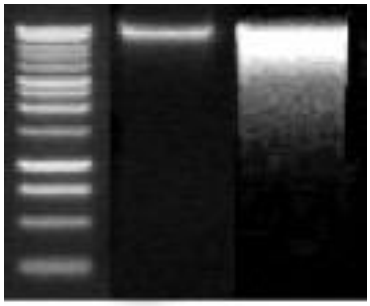


Figure 3. Agarose gel image.
DNA ladder (lane 1), genomic DNA (lane 2) and degraded genomic DNA (lane 3)

make a 1 agarose gel:

1. 1g agarose powder (Promega) and 100ml 1xTAE were combined and heated in the microwave for 2 minutes. After cooling slightly, 5µl SafeView was added and the gel poured into a casting tray.
2. 2µl 6x DNA loading buffer (Promega) was added to 1-8µl DNA, depending on the expected concentration, and made up to 10µl with nuclease free water. The sample was loaded into one of the wells. Loading buffer contains glycerol which pulls the DNA to the bottom of the well, and dyes to allow the sample to be tracked down the gel. A 1kb DNA ladder was also loaded for reference (Promega).
3. The gel was run for 40-60 minutes at 120V
4. A UV transilluminator was used to visualise the gel and a photo was taken of the gel

3.5.2 DNA quantification by nanodrop

Nano drop is a spectrophotometer that assesses the amount of DNA/RNA and contaminants present in a sample by measuring the amount of light that it absorbs. The sample is pipetted onto a fiber optic cable (the receiving cable) and a second fiber optic cable (the source cable) is bought into contact with the sample, causing the liquid to bridge the gap between the fiber optic ends. A pulse of light originating in the source cable is passed through the sample. When a photon encounters a DNA molecule it is absorbed and the intensity of light reaching the receiving cable is reduced and measured. DNA absorbs UV light at a wavelength of 260nm, proteins absorb light at 280nm and 230nm. Other contaminates such as carbohydrates also absorb at 230nm. Absorbance is measured at these three wavelengths, allowing the concentration and purity of the DNA sample to be determined (Figure 4)

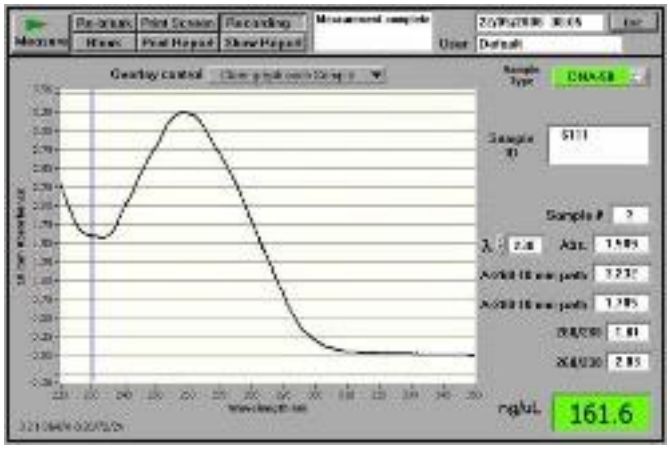


Figure 4. Nanodrop DNA profile showing the concentration and absorbance ratios

1. With the sampling arm in the down position the NanoDrop-1000 software was started and the nucleic acid application selected.
2. NanoDrop-1000 was calibrated by applying 1.5µl nuclease free water to the lower measurement pedestal, lowering the sampling arm, and selecting blank.
3. The water was wiped from both pedestals. 1.5µl nuclease free water was measured and a flat base-line was returned. Both pedestals were wiped.
4. 1.5µl of each DNA sample was measured. Each profile was saved after measurement and the pedestals were wiped.
5. The software provided a concentration in ng/µl. This is calculated in the following way:

   Optical density (absorbance reading at 260nm) x 50 (1 absorbance unit at 260nm= 50µg/µl DNA)
6. DNA purity is calculated by dividing the absorbance at 260nm by the absorbance at 280nm and at 230nm. This ratio should fall between 1.5 and 2.0. A ratio <1.5 indicates high contamination.

3.5.3 DNA quantification by Qubit (Qubit™ 3.0)

Qubit fluorometric quantitation incorporates fluorescent dyes selectively into DNA, so increasing the sensitivity and specificity compared to spectrophotometer methods, that can be compounded by contaminants and nucleotides. DNA concentration is measured by comparing the fluorescent signal to the signals generated by a range of known concentration standards. Qubit was used to assess DNA quantity prior to and during any post amplification quantification steps in the NGS methods. The dsDNA broad range assay kit was used for genomic DNA while the dsDNA high sensitivity assay kit was used for amplified DNA.

1. 2µl DNA sample was transferred to a Qubit tube.
2. 10µl of standard 1 and standard 2 were put into separate tubes.

3. The qubit working solution was prepared by diluting the dsDNA reagent 1:200 in dsDNA buffer. DNA samples and standards were made up to 200µl with the working solution.
4. Samples were vortexed and briefly centrifuged then incubated in the dark for 2 mins to incorporate the fluorescence dye
5. Samples were analysed on the Qubit V3.0 for DNA concentration

3.5.4 **RNA and High Sensitivity DNA assays (Agilent Bioanalyser 2100)**

The Agilent RNA and DNA chips use the principles of gel electrophoresis to assess nucleic acid quantity and quality. RNA chips (Figure 5) were used to assess RNA quality prior to cDNA conversion whilst DNA chips (Figure 6) were used to assess DNA during NGS methods.

RNA;

1. The electrodes were decontaminated by filling the wells of an electrode cleaner with 350µl RNAseZAP (Invitrogen) and placing it inside the bioanalyser for 1 minute. A second electrode cleaner filled with 350µl water was used to wash excess RNAseZAP from the electrodes. The lid was left open for 10 seconds to allow excess water to evaporate.
2. The reagents were equilibrated to room temperature for 30 minutes then 550µl gel matrix was applied to the top receptacle of a spin filter. The filter was centrifuged at 1520g for 10 minutes. The eluate was aliquotted in 65µl volumes.
3. The dye concentrate was vortexed and spun briefly to sediment any particles. 1µl was added to 65µl gel matrix and spun at 16000g for 10 minutes.
4. A RNA nano chip was placed in the chip priming station and 9µl gel-dye mix was pipetted to the bottom of the first gel well. The station was closed and the plunger depressed from 1ml until it was held by the syringe clip. This was held for exactly 30 seconds to force the gel into the microfluidics channels. The clip was then released.
5. After 5 seconds the station was opened and another 9µl gel-dye mix was pipetted into the remaining two gel wells.
6. 5µl Nano marker was pipetted into all the remaining wells. 6µl marker was added to any unused sample wells. The marker contains a 50bp DNA fragment and provides an internal standard which is used to align the ladder data with the sample data
7. 1µl sample and 1µl ladder were denatured at 70ºC for 2 minutes to minimize secondary structure and then pipetted into the appropriate wells.
8. The chip was vortexed at 2400rpm for 1 minute and placed inside the bioanalyser.

9. In the 2100 Expert software the appropriate assay (nano or pico) was selected and the run started. When completed, the chip was removed and the electrodes decontaminated as described in step 1.

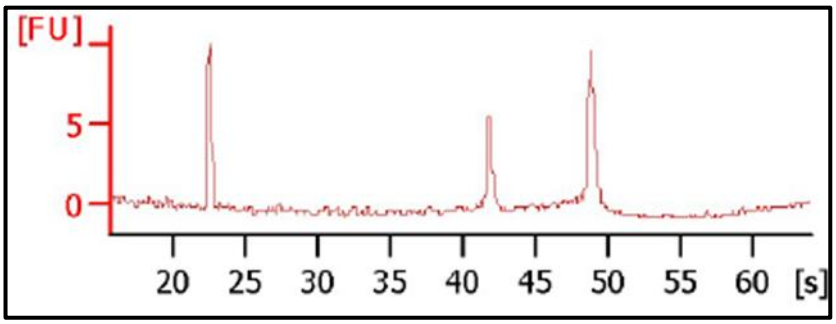


Figure 5 An RNA trace generated by the Nanochip software

DNA;

The protocol described above was used, with the following changes: electrodes were decontaminated with water alone. 15µl dye concentrate was added to the high sensitivity DNA gel matrix prior to centrifugation in the spin filter. The plunger of the chip priming station was depressed for 60 secs. The ladder and samples were not denatured prior to use.

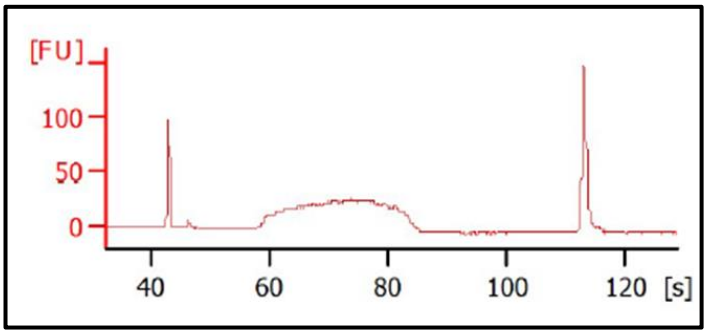


Figure 6 A DNA trace generated by the Nanochip software.

3.6 Multiplex Ligation-dependant Probe Amplification

MLPA was performed using the SALSA MLPA P037 and P038 CLL probemix (MRC Holland, Amsterdam, Netherlands). Both kits combined have a total of 105 probes, including 44 probes targeted to regions with prognostic relevance in CLL (13q14, 11q22.3, 17p13.1 and trisomy 12) and 32 probes for regions frequently affected by copy number changes in CLL (2p24, 6q25-26, 8q24, 9p21, 10q23, and chromosome 19). Probes for the detection of hotspot mutations in *NOTCH1* (P2515*fs~), *SF3B1* (K700E) and *MYD88* (L265P) are included in kit P038, and 26 reference/control probes targeting chromosomal locations typically unaffected in CLL are distributed between the two kits. The 3p control probes were included in downstream analysis as this region has been implicated in CLL (See Appendix A).

3.6.1.1 DNA denaturation step

1. 5µl (~20-100ng) of DNA was denatured in a PCR tube (0.2ml) for 15 minutes at 98°C then cooled to 25°C before removing from thermocycler.

3.6.1.2 Hybridisation step

2. 1.5µl of buffer A and 1.5µl of buffer B were vortex mixed.
3. 3µl of the prepared master mix was added to the denatured DNA and pipette mixed.
4. Samples were transferred to the thermocycler and incubated for 1 minute at 95°C then overnight (16-20 hours) at 60°C.

3.6.1.3 Ligation step

1. 3µl Ligase Buffer A and Ligase Buffer B, and 1µl Ligase-65 enzyme was added to the master mix and pipette mixed.
2. 32µl of the ligation master mix was added to each sample and pipette mixed at 54°C (on the thermocycler).
3. The samples were left to ligate for 15 minutes at 54°C, with a final temperature increase to 98°C to deactivate the ligase enzyme, before reducing to 20°C.

3.6.1.4 Amplification step

1. 7.5µl dH_20, 2µl SALSA PCR primer mix and 0.5µl SALSA Polymerase, were prepared per reaction, and pipette mixed.
2. 10 µl of the PCR master mix was added to each sample and mixed using a pipette.
3. Samples were put onto the thermocycler, using the following PCR programme: 95°C for 30 seconds, 35 cycles of 60°C for 30 seconds and 72°C for 30 seconds, 72°C for 20 minutes, and a 15°C hold.
4. After amplification the samples were moved from the pre PCR room to the post PCR lab and stored in the fridge (in the dark) for one hour, then prepared for capillary electrophoresis and fragment analysis. At this point samples could be stored in the freezer, wrapped in tin foil to avoid light degradation for future analysis.

3.6.1.5 Fragment separation by capillary electrophoresis

1. 1µl dH_2O, 0.5µl size standard and 13.5µl Hi-Di Formamide (HDF) was prepared per sample.

2. 15 µl of the master mix was added to a new sequencing tube with 0.5 µl of the PCR product and pipette mixed.
3. The PCR products were denatured in the HDF at 92°C for 2 minutes on the thermocycler, then put on ice.
4. The samples were loaded onto the ABI-Prism 310 genetic analyser. The polymer used was POP4 and the primer dye was FAM labelled, therefore the ABI 500 TAMRA size standard was used in the fragment analysis mastermix. The following run parameters were used:
    - Module-GS STRPOP4 (1ml)
    - Matrix-Bogus
    - Injection time- 25 (inj.secs)
    - Injection voltage- 1.6 (inj.KV)
    - Run voltage- 15 (KV)
    - Run temperature - 60 (°C)
    - Run time- 27 (minutes)

3.6.1.6 MLPA analysis

A copy number change in the MLPA probe target sequence will cause an increase or decrease in the relative amount of probe product, meaning it is possible to distinguish between deletions and amplifications by comparing the relative signal to a normal reference sample. The MLPA data is normalized by dividing each probes peak area by the average peak area of the sample. The normalized peak pattern is then divided by the average normalized peak of the reference samples included in the same experiment. When cells have a normal DNA copy number a relative copy number of 1.0 is the expected result, but when a deletion or amplification has occurred, the relative copy number will deviate to 0.5 or 1.5, accordingly [83]. The fragment data was analysed in GeneMarker (version 2.4.2) ID software (Softgenetics, Bio-gene). Using the mean ratio of 36 aged matched controls, ±2SD, the normal reference range for each probe was established, as described by Abdool *et al*. Based on the mean of all the upper and lower ranges, a universal cut-off was then generated for both kits and all probes as follows; <0.86, <0.3 and >1.13, for deletions, monoallelic deletions and gains, respectively. These data were further analysed in Microsoft Excel (2010); the data was sorted by probe genomic location and conditional formatting was applied. The data was visually inspected and probes/genomic loci could finally be called as deleted, gained or normal. Typically, three or more consecutive aberrant probes were required for a copy number aberration (CNA) call, although this was subjective and dependent upon the region being analysed; for example, all probes on chromosome 12 must be gained for a Tri12 call and in regions covered by only 2 probes (e.g. 9p) both must be aberrant and not on the threshold

of the cut-off values to be called aberrant. Probe values on the threshold of the cut-offs are more likely to be technical artefacts, although they may reflect a small sub-clonal population size. The 10q (PTEN) probes were excluded from further analysis due to the high variation observed across the samples, including the controls.

3.7 Whole genome amplification (WGA)

Genomic DNA was amplified using the illustra GenomiPhi V2 DNA amplification Kit. This allows highly uniform amplification over the entire genome, so that locus representation remains extremely close to the original template. WGA DNA can be used in downstream applications such as sequencing, and prevents the depletion of precious archived material.

1. 1µl genomic DNA (10ng/µl) was transferred to new plate/tube
2. 9µl sample buffer was added to the DNA and heated to 95°C for 3 minutes
3. Samples were incubated on ice whilst master mix was prepared (for multiple samples): 9µl reaction buffer and 1µl enzyme mix per sample
4. Thermocycling conditions were: 30°C for 1.5 hours, then 65°C for 10 minutes (enzyme inactivation)
5. WGA DNA was diluted to a working concentration (~10ng/µl; 4µl WGA DNA added to 96µl H_2O) and stored in the freezer

3.8 Primer design

Sequences for the genomic regions of interest were extracted from Ensembl (http://www.ensembl.org/index.html) (GRCh37/hg19 genome assembly). The sequence was uploaded to the on-line Primer3 (http://frodo.wi.mit.edu/) software, which designs primers based on specific parameters e.g product size and primer Tm. Sanger sequencing primers had an optimum annealing temperature of 55°C, and the product length was between 450-500bp long. PCR Blat (BLAST-like alignment tool) (http://genome.ucsc.edu/cgi-bin/hgPcr?org=Human&db=hg18&hgsid=248670511) was used to confirm primers align with the correct genomic location and the online oligo calculator; http://www.basic.northwestern.edu/biotools/oligocalc.html was used to ensure primers would not self anneal in primer-dimers or hairpin structures. Primers were purchased from Eurogentec.

3.9 High Resolution Melt analysis (HRM)

HRM analysis was used to detect mutations in genomic DNA sequence. PCR amplification of target sequences is performed in the presence of a saturating dye that fluoresces brightly when intercalated with dsDNA. The dsDNA is then exposed to increasing temperatures. As it dissociates into ssDNA, the dye molecules are released and the fluorescent signal diminishes. The rate of decline is detected by the machine optics. The observed Tm is characteristic of a particular DNA sample and is influenced by sequence length, GC content and DNA sequence complementarity. Melting curves are generated (Figure 7), which display the decrease in fluorescent signal with increasing temperature. HRM analysis can be used to identify single base changes, insertions, deletions and base pair substitutions. It can quantitatively detect a small proportion of variant DNA in a background of wild-type sequence, at sensitivities approaching 5%. However, as the technique can not specify the nature of the variant i.e. type of mutation/SNP or where the mutation occurs, any variant peak was then Sanger sequenced to confirm the mutation.

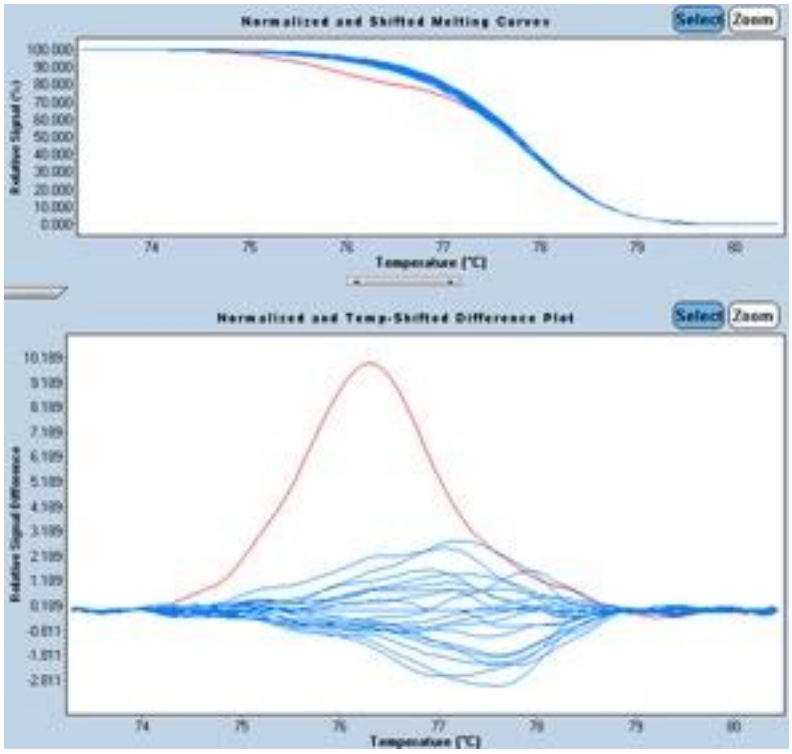


Figure 7. HRM melting curve
Blue lines indicate a normal melting profile, whilst the red line represents an abnormal melting curve.

1. Test and normal reference DNA was diluted to a 5ng/µl concentration. 5mM SYTO 9 dye (Invitrogen) was diluted 1:100 to give a stock solution.

2. The following master mix (Table 1) was prepared for each sample and dispensed into a LightCycler 480 Instrument 96 well plate (Roche);

Table 1.HRM master mix

Reagents	Volume/well (µl)	Concentration
Platinum Taq 10x buffer (Invitrogen)	1	**1x**
5mM dNTPs	0.4	**0.2mM**
Platinum Taq DNA polymerase (5U/µl) (Invitrogen)	0.05	**0.025U**
50mM $MgCl_2$ (Invitrogen)	0.3	**1.5mM**
10µM F and R primers	0.5	**0.5µM**
DNA	2	
SYTO 9	1	
Water	**4.75**	

3. The plate was sealed and loaded into the LightCycler 480 System (Roche). PCR was performed with Taq activation at 95°C for 10 minutes, followed by 40 cycles of 15 seconds at 95°C, 30 seconds at 60°C and 30 seconds at 72°C.
4. The samples were heated to 95°C for 1 minute to denature the dsDNA, and cooled to 40°C for 1 minute to allow the complimentary single strands to reanneal.
5. The temperature was gradually increased from 65°C to 95°C, and the fluorescent signal was measured 25 times for every 1°C increase. The plate was cooled to 40°C for 1 minute.
6. The data was analyzed with the LightCycler 480 software.

3.10 Sanger Sequencing

Sanger sequencing is used to detect the presence of mutations in genomic DNA, relative to germline DNA or a reference sequence. Primers were designed as described in section 3.8.

3.10.1 PCR amplification

The following Master mix was prepared: 10µl GoTaq master mix (Promega), 0.35µl each of 10µM forward and reverse primers, 4µl DNA (10ng/µl), 5.3µl H_2O. Samples were placed in the thermocycler under the following touchdown conditions: 95°C for 10 minutes, followed by 10 cycles of 30 seconds at 95°C, 30 seconds at 60°C (decreasing by 1°C per cycle) and 30 seconds at 72°C, followed by 30 cycles of 30 seconds at 95°C, 60 seconds at 56-60°C (depending on primer Tm) and 30 seconds at 72°C,ending with 72°C for 10 minutes and a 4°C hold.

3.10.2 PCR purification

PCR products were purified using the MinElute 96 UF PCR Purification Kit (Qiagen)

1. PCR product was transferred to a clean well on the purification plate.
2. The plate was attached to a vacuum pump for 10 minutes at room temperature.
3. 50µl of dH_2O was added to the sample wells and vacuumed for 10 minutes.
4. 20µl of dH_2O was added to the sample wells, pipetted up and down 20 times to dissolve the DNA from the plate membrane and then transferred to a new plate

3.10.3 Sequencing

10µl of PCR product (10ng/µl) and 10µl of primers (3.2pg/µl) were sent to Source Bioscience (Oxford) for Sanger sequencing.

3.10.4 Sequencing Analysis

The raw sequence data was then visualised, aligned and interpreted using SeqMan (DNASTAR lasergene) software to confirm if any mutations were present within the sequence, by a comparison to the wild-type (WT) or reference genomic sequence (Figure 8).

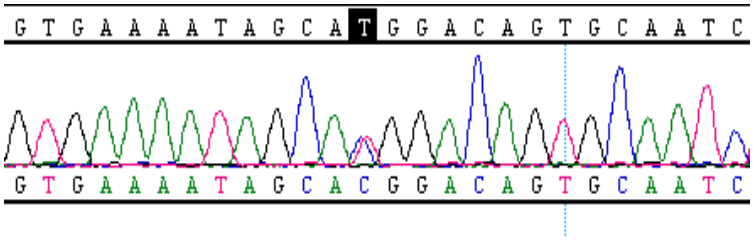


Figure 8. Sanger Sequencing trace.

Each base is represented by a peak (A, T, C and G are displayed as green, red, blue and black, respectively). The reference sequence is displayed above the trace and the sequence of the sample is displayed below. In the centre of the trace, two overlapping peaks can be seen; a red 'T' (highlighted in black in the reference sequence) and a blue 'C'. This sample contains a T>C mutation at this locus, in approximately 50% cells (deteremined by the equal size of the two peaks)

Mutation surveyor (Softgenetics) was also used once mutations were confirmed by SeqMan, to determine the clonality of the mutation present in the sample (peak height ratio) as well as making it easier to visualise and determine the exact location of any indels.

In silico predictions of the mutational effect on protein function was performed using the sorting tolerant from intolerant (SIFT) software. SIFT is a multi-step algorithm that uses a sequence homology based approach to classify amino acid substitutions [84]. The software gives a SIFT score scaled on the probability of the amino acid substitution being tolerated. Highly conserved

positions tend to be intolerant to substitution and are given a score below 0.05, which are predicted to have a damaging affect to protein function [84].

3.11 Monochrome Multiplex quantitative PCR (MMqPCR)

MMQPCR is a quantitative PCR method of measuring the telomere length. It was developed by Cawthon and the method involves multiplexing both the telomere (T) and single copy gene (S) in one reaction with a single fluorescent DNA intercalating dye. The S primers (Table 2) are designed with a number of GC clamps so that the T signal is analysed in the early cycles, before S signals rise above baseline, while the single copy gene signal can be collected at a temperature that fully melts the telomere product. An average T/S value is then produced that is a relative measurement of the telomere length [85].

Table 2. MMqPCR primers.

Telomere primers are designed to generate fixed product length whilst the albumin primers have GC clamps (highlighted in bold) to allow for a much higher melting temperature than the telomere product.

Primer	Gene	Sequence	tm	Product size
telg	Telomere	ACACTAAGGTTTGGGTTTGGGTTTGGGTTTGGGTTAGTGT	74°C	79bp
telc	Telomere	TGTTAGGTATCCCTATCCCTATCCCTATCCCTATCCCTAACA	74°C	79bp
albu	Albumin	**CGGCGGCGGGCGGCGCGGGCTGGGCGG**AAATGCTGCACAGAATCCTTG	88°C	98bp
albd	Albumin	**GCCCGGCCCGCCGCGCCCGTCCCGCCGG**AAAAGCATGGTCGCCTGTT	88°C	98bp

A six-point standard curve was prepared by diluting 30µl of stock (50ng/µl) standard DNA in 23.28µl H_2O and performing a serial dilution.

Samples were plated out in a 96 well plate in triplicate. The following mastermix was prepared (

Table 3)

Table 3. MMqPCR mastermix

Reagent	Stock concentration	Volume per reaction	Final concentration
Titanium Taq buffer	x10	2.5µl	x1
dNTPs	2.5µM	2µl	2.5 µM
Betaine	x5	5µl	x1
Titanium Taq	x5	0.125µl	0.025
SYBR Green	x100	0.188µl	x0.75
Forward TEL (telg) primer	20µM	1.125µl	0.9µM
Reverse TEL (telc) Primer	20µM	1.125µl	0.9µM
Forward ALB (albu) primer	20µM	1.125µl	0.9µM
Reverse ALB (albd) primer	20µM	1.125µl	0.9µM
PCR grade water		5.687µl	
DNA template	10ng/µl	5µl	50ng/µl

Samples were transferred to the LightCycler480 (Roche), using the following programme:

- 95°C 10 minutes (ramp rate 3.3)

- 94°C 15 seconds (ramp rate 2) 2 cycles
- 49°C 15 seconds (ramp rate 2)

- 94°C 15 seconds (ramp rate 3.3) 45 cycles
- 62°C 10 seconds (ramp rate 2)
- 74°C 15 seconds (ramp rate 3.3) Single acquisition
- 84°C 10 seconds (ramp rate 3.3)
- 88°C 15 seconds (ramp rate 3.3) Single acquisition

- 72°C 30 seconds (ramp rate 1)
- 98°C (ramp rate 0.5) Continuous acquisition (12 acquisitions per °C)
- 40°C 1 seconds (ramp rate 2)

The output data was exported into the excel MMqPCR template file, reformatted and uploaded to an online PCR mining programme (http://ftp.ewindup.info/miner/Version2/data_submit.htm). These results were put back into the template file and the instructions followed; Standard curves for the telomere signal and the single copy gene were generated. For each sample the average Telomere/Single copy (T/S) ratio was calculated, which is proportional to the average telomere length per cell. Samples with a T/S >1.0 or <1.0 have an average telomere length longer, or shorter than the standard DNA, respectively.

3.12 Droplet digital PCR

Droplet digital PCR (ddPCR) can detect mutant DNA present in 0.1%, in a WT background; some studies have reported an improved sensitivity of just 0.0005%. The test and reference assays contain a single set of primers and two competitive probes to detect the WT and mutant alleles (labelled with HEX and FAM, respectively). A modified 5' fluorescent moiety oligonucleotide probe is quenched by a 3' quencher moiety that when bound to a specific target DNA, is cleaved by the 5' to 3' nuclease activity of the *Taq* polymerase, releasing the 5' fluorescent nucleotide and so generating a signal[86].

1. The following master mix (dependant on the gene of interest) was prepared:
    a. NOTCH1:10µl 2x ddPCR supermix was mixed with 1µl 20x NOTCH1 p.P2514fs*4 (FAM) mutant assay, 1µl 20x NOTCH1 WT Human (HEX) and 3µl H_2O
    b. SF3B1: 11µl 2x ddPCR supermix was mixed with 0.55µl 40x SF3B1 custom assay (AHGS8IT) and 4.95µl H_2O
2. 5.5µl DNA sample (10ng/µl) was added to the mastermix
3. 20ul of sample was transferred to the middle row of the droplet generator (DG8™) cartridge and 70µl droplet generator oil was pipetted into the first row.
4. A gasket was hooked over the cartridge then loaded onto the droplet generator
5. Oil and sample were drawn up through the microfluidic channels and droplets were generated and placed into the droplet well
6. 40µl of the emulsion was transferred carefully to a PCR plate, which was heat sealed and put on the thermocycler under the following conditions; 95°C for 10 minutes (enzyme activation); 40 cycles (with a ramp rate of 2°C/sec) of 30 seconds at 95°C (denaturation) and 60 seconds at 55°C (annealing/extension); followed by 10 minutes at 98°C (denaturation)
7. The plate was then transferred to the QX200 droplet reader; this aspirates the sample into singular droplets and detects and reads each droplet to determine if it is carrying a WT or mutant (or both) allele.
8. Data analysis was performed using the Quantasoft™ software (Bio-Rad). This generates four distinct clusters of droplets; 1) Negative droplets containing no DNA template, 2) WT droplets, 3) mutant droplets, 4) WT and mutant containing droplets. The concentration of the WT and mutant are calculated using Poisson statistics based on the number of droplets in each cluster (Figure 9).

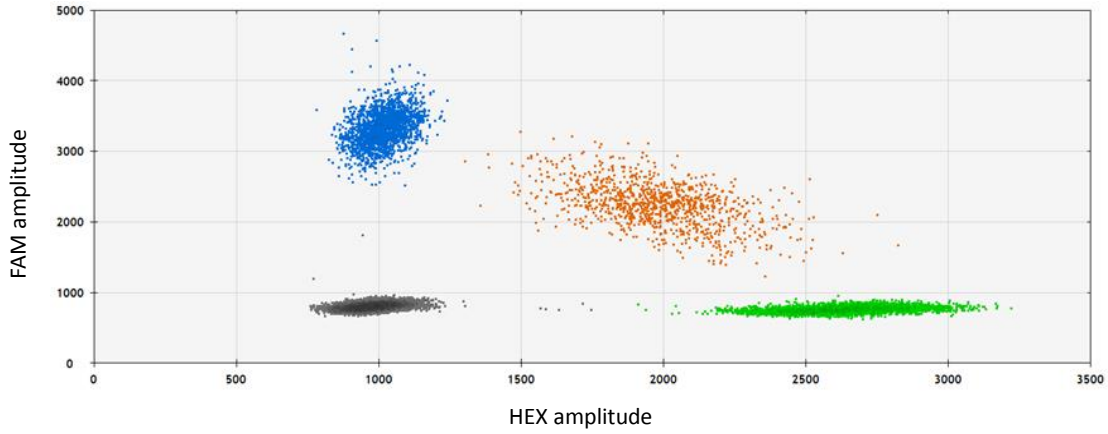


Figure 9. Droplet Digital PCR analysis.
Four distinct clusters of groups are generated; negative droplets (grey), wildtype (green), mutant (blue) and heterozygous (orange).

3.13 Reverse Transcription PCR (RT-PCR)

RT-PCR is used to determine if a particular sequence of DNA is being transcribed into mRNA. RNA samples are reversed transcribed into cDNA, which is subsequently amplified by a PCR reaction.

RT-PCR was performed on 1 µg of total RNA isolate using the Improm™II RT-PCR kit (Promega, UK). All RT-PCR reaction set-ups contained positive (Kanamycin RNA), Reverse Transcription negative (RT-ve) and no-template controls (NTC) to control for reagent failure, internal and external gDNA contamination respectively.

1. 2µl RNA, 1µl Random primers and 2µl dH_2O were mixed and denatured at 70°C for 5 minutes on a thermocycler to remove any secondary RNA structures. The samples were immediately put onto ice.
2. The following master mix (Table 4) was prepared and added to the primer/template;

Table 4. RT-PCR master mix

Reagents	Kanamycin Reaction	RT-ve Reaction	Experimental/ NTC Reaction
Nuclease-Free water	4.2 µl	8.8 µl	7.3 µl
5 x ImPromII™ Reaction buffer	4.0 µl	4.0 µl	4.0 µl
$MgCl_2$ (Final conc 1.5 mM)	4.8 µl	1.2 µl	1.2 µl
dNTP mix	1.0 µl	1.0 µl	1.0 µl
RNasin Ribonuclease Inhibitor (20 U)	0.0 µl	0.0 µl	0.5 µl
ImPromII™ RTase	1.0 µl	0.0 µl	1.0 µl
Final Volume	15 µl	15 µl	15 µl

3. Tubes were transferred to the thermocycler programme and heated to 25°C for 5 mins, 42°C for 1 hr, 70°C for 15 mins and then finally cooled and held at 20°C. The cDNA was

diluted to a working concentration (sample >20ng/µl 1:7 dilution and samples <20ng/ µl 1:5 dilution) and then stored at -20°C until needed.

3.14 Quantitative Real Time-Polymerase Chain Reaction (qRT-PCR)

This technique can be used for RNA, DNA and cDNA quantitation. The Taqman system uses a primer and probe combination specific for the gene of interest. The probe contains a fluorescent reporter dye on the 5' end and a quencher molecule on the 3' end. While the probe is intact, the proximity of the quencher molecule greatly reduces the fluorescence emitted by the reporter dye. The probe anneals to a target sequence downstream of the primer site and as the primer is extended by Taq DNA polymerase its 5' nuclease activity cleaves the reporter dye from the quencher. This allows the fluorescence of the reporter to increase. Each cycle of denaturation, primer annealing and primer extension cleaves another probe. In this study FAM was used as the fluorescent reporter dye. Data collection occurs throughout the process and reactions are characterized by the point in time during cycling when the fluorescence exceeds the threshold. The cycle number needed to generate a defined amount of fluorescence when the PCR is in its linear phase is called the cycle threshold (Ct). The higher the Ct value the less nucleic acids present in the sample. Abolute quantitation can be performed using standard curves and relative quantitation measures the amount of DNA or RNA relative to an endogenous control gene, such as housekeeping gene, 18s (Figure 10).

1. A standard curve was created using B cell control DNA/cDNA at the following concentrations: 20ng/µl, 15ng/µl, 1.25ng/µl, 0.31ng/µl, 0.1ng/µl, 0.019ng/µl, 0ng/µl.
2. Samples were assayed in triplicate, using an appropriate calibrator on the Roche LC480. The following master mix was prepared (per reaction): 10µl 2x Taqman universal PCR mastermix, 1µl 20x Taqman assay mix. 9µl DNA/cDNA was added and mixed onto a plate. For each gene a well containing no nucleic acids was included as a negative control.
3. The plate was sealed with an adhesive film (Applied Biosystems) and briefly centrifuged to collect the samples. It was placed in the Roche LC480 under the following cycling conditions; 10 minutes at 95°C (denaturation), followed by 40 cycles of 15 seconds 95°C and 1 minute at 60°C (Primer annealing and extension) and 72°C for 5 mins.

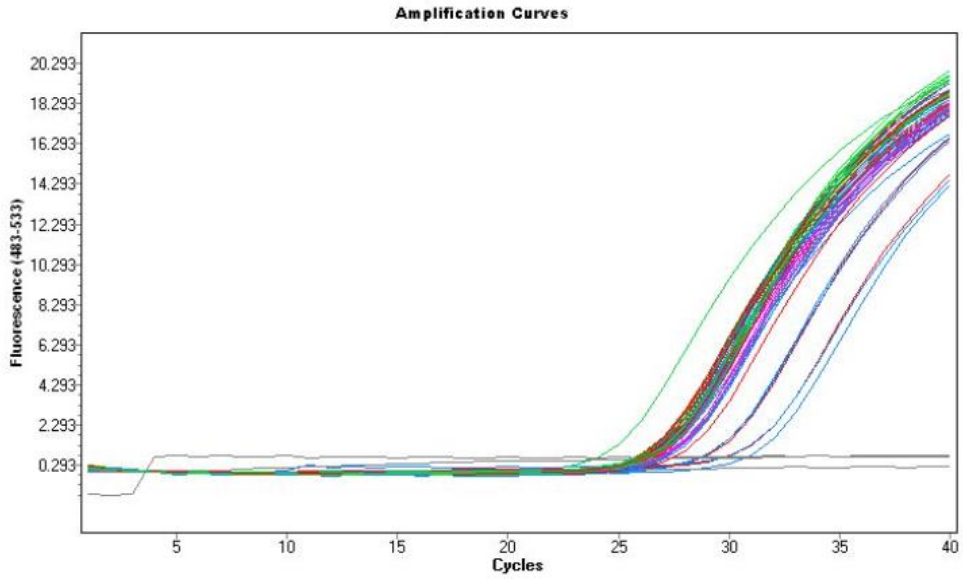


Figure 10. QRT-PCR amplification curves.

As the amount of nucleic acid present increases with amplification, the fluorescence generated by FAM also increases. In this example the number of cycles required for fluorescence detection varies from approximately 24-30 cycles, representing samples with decreasing amounts of starting material.

3.15 Protein extraction

Extracted proteins are used to determine if genomic aberrations are impacting on protein translation and function.

Cells stored in DMSO were removed from the -80°C freezer and thawed on ice.

1. 500µl of the thawed cells were transferred to a 1.5ml Eppendorf tube and 500µl of cold HBSS was added.
2. Samples were centrifuged at 4°C for 5 minutes at 2,500rpm. The supernatant was removed and discarded.
3. 500µl of HBSS was added and centrifuged as before. The supernatant was removed.
4. 70µl of lysis buffer (100µl whole cell lysate (WCL), 1µl PMFS (17.4mg/ml DMSO) and 1µl of protease inhibitor was added and pipet mixed.
5. The sample was centrifuged at 4°C at 13,000 rpm for 10 minutes to pellet the nucleic acids.
6. The lysate was transferred to a new tube and stored at -20°C until needed.

3.16 Protein quantification assay

After extraction the protein was quantified using the DC protein assay (Biorad) kit as follows;

1. Control samples were used to create a standard curve at the following concentrations: 1.52mg/µl, 0.76mg/µl, 0.38mg/µl, 0.19mg/µl, 0.0095mg/µl and finally a WCL blank.
2. 5µl protein sample was diluted into 20µl WCL, briefly centrifuged then put on ice. 5µl of the diluted sample was loaded onto the plate in duplicate. Reagent A* was made by mixing 20µl reagent S with 980µl of reagent A. 25µl reagent A* and 200µl reagent B was added to each of the sample wells. The plate was incubated for 15 minutes at room temperature then loaded onto the Varioskan using the Skanit RE programme; Wavelength-750nM, Bandwidth-5nM, Measurement time: 100s
3. The results were exported into MS Excel, to create the standard curves and calculate the protein concentration of the samples. From this the amount of protein needed for western blot analysis was calculated per sample (diluted in WCL).

3.17 Western blotting

1. 20µl protein sample and 10µl 4x Loading Dye (7.5µl Loading dye (Invitrogen) and 2.5µl DTT (1 M) (Sigma)) were mixed.
2. Samples were heated to 95°C for 5 minutes
3. The white strip and the comb were removed from the gel.
4. The gel tanks were assembled and the gel running buffer was made: 50 ml 20X stock MOPS (Invitrogen) + 950ml dH_2O
5. The central reservoir was filled with running buffer and 500µl antioxidant (Invitrogen) was added and mixed using a Pasteur pipette. Any large bubbles were removed from the buffer surface.
6. The outer reservoirs were filled with running buffer and the gel wells were washed 3 times using a syringe.
7. 30µl sample was loaded into the wells. One lane was loaded with Novex® Sharp Pre-stained Protein Standard (LC5800) and 6µl rainbow marker. The gel was run for 1 hour (150v 75mA 12.5W).
8. Transfer buffer was made (20ml 20X trans Buffer at 4°C, 80ml Methanol, 300ml dH_2O and 400µl Antioxidant) and poured into 2 containers with 5 sponges and 1 Hybond (Amersham) membrane.
9. The gel was removed from the tank and the front casing was opened.
10. A wet piece of 8.5cm x 6cm 3MM filter paper (Whatman) was placed over the gel and it was turned over.

11. The gel spatula was used to remove the back casing. A small amount of the transfer buffer was put onto the gel and the wet Hybond was placed directly onto the gel. A second 3MM filter paper soaked in transfer buffer was put on top of the Hybond.
12. A plastic pipette was rolled over the top of the gel, hybond and 3MM paper to remove any bubbles which can reduce protein transfer to the membrane.
13. The wet sponges were arranged in the blotting cassette and the front cover was placed on the blotting rig and placed into the tank.
14. The blotting cassette was clamped into place and filled with transfer buffer until the blotting pads were covered. The outer reservoirs were filled with water to cool the blotting cassette.
15. The gel was run for 1h 30mins at 25V, 125mA, 15.0W.
16. The hybond membrane was removed and put in dH_2O in a tray, stained for 30 seconds with Ponceau S (Sigma) and rinsed in H_20.
17. The membrane was then washed in TBS tween for 5 minutes to remove the Ponceau S.
18. The membrane was wrapped around the inside of a falcon tube with 5ml VETO and 250µl FCS for 30mins on the roller to block non-specific Ab binding.
19. The solution was removed and replaced with fresh 5ml VETO and 250ul FCS along with the appropriate antibody. This was left to incubate overnight (>16 hrs) at room temperature.
20. The following day the membrane was removed and washed in TBS tween for 40 minutes (with 4 buffer changes) to remove non-specific AB.
21. The secondary antibody was added into 10ml VETO and 0.5ml FCS
22. The membrane was incubated for 2 hours, then washed in TBS tween for 45mins, changing the wash buffer at least 4 times
23. 350µl of Reagent 1 (peroxidase solution) was mixed with 350ul reagent 2 (luminol enhancer solution).
24. The membrane was blotted and ECL was added. It was incubated for 1 min and wrapped in cling film. The membrane was imaged.

3.18 Illumina TruSeq Next generation sequencing

3.18.1.1 Hybridisation of the Oligo Pool

1. DNA quantity and quality was assessed using an agarose gel and Qubit. DNA was diluted to 250ng (10µl).

2. 5µl CAT was added to the gDNA. 35µl OHS2 was added and pipet mixed.
3. The plate was sealed with an adhesive aluminium foil, centrifuged (1000 x g) for 1 minute and placed on the Hybex heating block for 1 minute at 95°C.
4. The temperature was reduced to 40°C and the plate left to incubate for 80 minutes

3.18.1.2 Removal of Unbound Oligos

5. The filter plate assembly unit (FPU) was assembled. 45µl SW1 was added to each sample well and centrifuged at 2400 x g for 10 minutes.
6. The sample plate was removed from the heating block and centrifuged for 1 minute at 1000x g.
7. The entire sample was transferred to the FPU plate and centrifuged at 2400 x g for 2 minutes
8. 45µl SW1 was added to each sample well and centrifuged for 2 minutes (2400 x g). this step was repeated.
9. The flow through was removed, 45µl of UB1 was added to each sample and centrifuged for 4 minutes at 2400 x g.

3.18.1.3 Extension-Ligation of Bound Oligos

A DNA polymerase extends from the upstream oligo through the targeted region, followed by ligation to the 5' end of downstream oligo using DNA ligase, resulting in the formation of products including the regions of interest and amplification sequences.

10. 45µl ELM4 was added to each sample well on the FPU plate. The plate was sealed with an adhesive aluminium foil seal, then incubated at 37°C for 45 minutes. The plate was centrifuged for 4 minutes at 2400 x g.
11. 25µl of 50mM NaOH was added to each sample well, pipetted mixed 5 times then incubated at room temperature for 5 minutes. Pipette mixing was repeated and 20µl of sample was transferred to a new plate.

3.18.1.4 PCR amplification

Indexing PCR primers are used to amplify and multiplex samples as described in Appendix B

12. 4µl of appropriate i5 primer was added to each sample followed by 4µl of appropriate i7 primer.

13. The following master mix was prepared (per sample): 0.58µl TDP1 and 29.2µl PMM2, pipette mixed 20 times. 22µl PCR master mix was transferred to each well, the plate was sealed and centrifuged for 1 minute at 1000 x g.
14. The plate was placed on thermocycler under the following conditions; 95°C for 3 minutes, 24 cycles of 95°C for 30 seconds, 66°C for 30 seconds and 72°C- 60 seconds, 72°C for 5 minutes and a 10°C hold.

3.18.1.5 PCR clean up

15. Following amplification, the plate was centrifuged for 1 minute at 1000 x g.
16. 5µl PCR product was loaded into a 3 agarose gel (125V for 40 minutes) to confirm the library successfully amplified.
17. 45µl (room temperature) AMPure XP beads were added to a MIDI plate. The remaining PCR product was transferred to the MIDI plate and the plate was sealed.
18. The plate was shaken for 2 minutes, then removed and incubated for 10 minutes at room temperature followed by incubation on a magnetic stand for 2 minutes.
19. The supernatant was removed whilst the plate was on the magnetic stand. 200µl 80% ethanol was added, incubated for 30 seconds and then removed. This wash was removed.
20. The samples were air dried for 10 minutes, then 30µl EBT was added to each sample well. The plate was sealed and shaken for 2 minutes, then removed from the shaker and incubated for 2 minutes at room temperature.
21. The plate was placed on the magnetic stand for 2 minutes and 23µl of the EBT was then transferred onto a new plate.
22. Qubit HS DNA quantification was performed and the libraries were diluted to 4nM accordingly

3.18.1.6 Library Pooling and MiSeq Sample Loading

23. 5µl of each 4nM normalised sample was combined into a single tube (PAL), the lid was secured with Parafilm and then incubated at 95°C for 5 minutes.
24. The PAL was removed and put on a shaker for 5 minutes (1800rpm)
25. 5µl PAL was transferred to a new tube (20pM PAL) with 5µl 0.2M NaOH, vortex mixed and centrifuged then incubated at room temperature for 5 minutes.
26. The library was diluted to 20pM with 990µl HT1.
27. 375µl 20pM PAL was transferred to a new tube (DAL). 219µl HT1 and 6µl of 12.5pM PhiX was added to give a concentration of 12.5pM. The DAL was vortex mixed and centrifuged then incubated at 95°C for 5 minutes.

28. The DAL was immediately put in an ice bath for ~10 minutes then loaded onto a defrosted MiSeq V2 500 cycle reagent cartridge. The cartridge was loaded into the Illumina MiSeq along with a prepared flow cell and the library was sequenced.

3.19 Nextera XT Sequencing

The Nextera XT kit (Illumina) was used to simultaneously fragment and tag (Tagment) DNA with sequencing adapters, for sequencing on the Illumina MiSeq. The Nextera protocol allows multiplexed, sequencing ready libraries to be generated very rapidly, using as little as 50ng input DNA

3.19.1 **Tagmentation**

1. 10µl Tagment DNA (TD) buffer was added to the wells of a 96 well plate.
2. 5µl of sample DNA (0.2ng/µl) was added to the wells and then gently pipette mixed 5 times.
3. 5µl Amplicon Tagment Mix (ATM) was added to the wells and pipette mixed 5 times
4. The plate was sealed and centrifuged, then put onto the thermocycler using the tagmentation programme: 55°C for 5mins, 10° hold
5. 5µl Neutralise Tagment (NT) buffer was added to the samples and pipette mixed 5 times, then incubated at room temperature for 5 minutes

3.19.2 **Amplification**

6. 15µl NPM was added to each sample.
7. Referring to the TruSeq indices combination layout (Appendix B), 5µl index 2 (i5) primers were added to the samples followed by 5µl index 1 (i7) primers. This was pipette mixed 5 times.
8. The plate was put on the thermocycler using the Nextera PCR programme: 72°C for 3 minutes, 95°C for 30 seconds, 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds and 72°C for 30 seconds, 72°C for 5 minutes and a 10°C hold.

3.19.3 **PCR clean up**

9. 75µl AMPure XP beads were added to each sample and pipette mixed 10 times, then incubated at room temperature for 5 minutes.

10. The plate was put on a magnetic stand for 2 minutes, then the supernatant was removed.
11. 200µl freshly prepared 80 ethanol was added to the samples, incubated on the magnetic stand for 30 seconds then removed. This wash step was repeated.
12. Samples were air dried for 15 minutes. The plate was removed from the magnet and 52.5µl Resuspension Buffer (RSB) was added, pipette mixed 10 times, then incubated at room temperature for 2 minutes.
13. The plate was put back on the magnetic stand for a further 2 minutes. 50µl of the supernatant was transferred to a new plate.

3.19.4 **Preparing libraries for MiSeq**

14. The DNA was quantified using Qubit (HS Assay) and Agilent high sensitivity DNA chips were used to check Tagmentation had occurred successfully.
15. The purified Nextera product was diluted 1:20 and pooled at 4nM per sample. 122µl H_2O was added to complete the dilution.
16. 5µl of the pooled library was put into a new Eppendorf. 5µl 0.2M NaOH was added and incubated at room temperature for 5 minutes.
17. 990µl cold HT1 added and mixed well.
18. 390µl was transferred to a new Eppendorf tube with 180µl HT1 to make the final library concentration of 13pM, with 30µl 13pM PhiX (5% PhiX spike in to compensate for the low complexity library).
19. This was incubated at 95°C for 5 minutes and immediately put into an ice bath for 10 minutes.
20. All of the 13pM library was loaded onto a MiSeq cartridge and run on the MiSeq.

3.20 Visualising next generation sequencing data

Integrative genomics viewer (IGV) was designed by the Broad Institute to visualise next generation sequencing data (http://www.broadinstitute.org/igv/). IGV was used to confirm the precence of variants identified by whole exome sequencing, TruSeq and Nextera data. BAM files of respective patients were uploaded to the IGV software and the genomic location of interest selected (Figure 11).

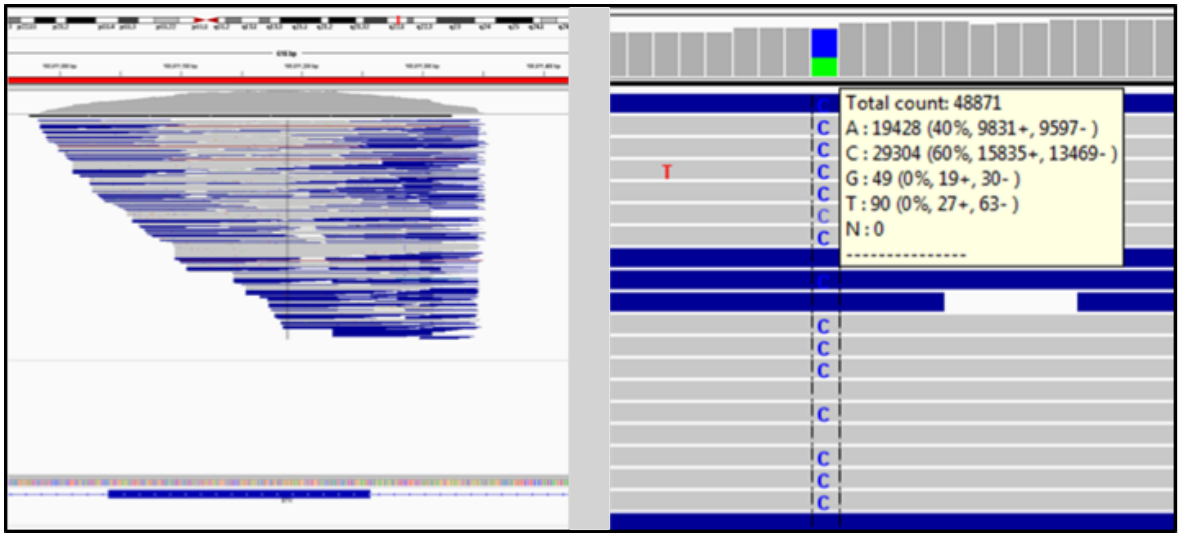


Figure 11. Visualisation of NGS data in IGV.

In the left hand window, the sequencing reads for a region of the BTK gene is shown. Each vertical blue/grey line represents a single sequencing read. The genomic location is indicated by a Chromosome ideogram and genomic ladder (top panel). The genomic sequence is shown in the bottom panel. Zooming in allows each base to visualised in detail. (right hand window). The blue/green bar in the top panel represents a variant base, and the associated statistics are shown, including total number of reads for this base and the relative proportion of reads that call an A', 'T', 'C' or 'G'.

Chapter 4: The application of Multiplex ligation dependant probe amplification to the UK LRF CLL4 trial

4.1 Introduction

The genome of approximately 80% of CLL patients is characterised by the presence of copy number abnormalities (CNA); the most well-studied of which are recurrent deletions of 13q, 11q and 17p and trisomy 12. Traditionally fluorescence *in situ* hybridisation (FISH) has been used to detect and characterise these CNA.[53] More recently higher resolution methods such as microarrays and next generation sequencing methods have paved the way to analyse whole genomes for copy number changes and/or gene mutations. These techniques are costly, and technically and computationally challenging, so targeted approaches have been developed. Multiplex ligation-dependant probe amplification (MLPA) is a PCR based method developed by MRC Holland that can be used to detect the copy number state of up to 50 DNA sequences in a single reaction, as described in Figure 12. MLPA can also be used to detect mutations, due to the very sensitive enzyme used in the probe ligation reaction, ligase-65, which detects mismatches at the 3' end of the left probe oligonucleotide ligation site. Any mismatches mean the probe will not ligate and no signal is generated. MLPA can be used to reliably detect point mutations in samples with a heterozygous mutation in around 5-10% of the whole cell population, which is comparable to pyrosequencing sensitivity.[87] The main advantages of employing MLPA as a technique for CNA analysis are; a) it allows the simultaneous processing of many samples in one experiment; b) it is relatively easy to perform; c) it can target a large number of genomic loci in a single reaction; d) data processing can be automated; e) the technique is cost effective; f) MLPA can detect not only CNA, but single nucleotide polymorphisms (SNPs) and mutations, analyse methylation patterns and quantify mRNA expression.[87] As such, several studies have employed MLPA in both a research and clinical setting across a host of diseases. However, MLPA cannot detect most inversions or translocations, and cannot identify any copy number changes that lie outside of the sequence detected by the MLPA probes.[87] Another major limitation of MLPA is the possibility of false positive or negative results, many of which can be caused by incorrect set up of the experimental protocol and mis-interpretation of the results.

The current gold standard diagnostic technique, FISH, requires highly trained cytogeneticists and is labour intensive. Cost comparisons have shown that MLPA would be between 80-86% cheaper, as well as being quicker and easier to adapt to high throughput settings, making MLPA a very

attractive proposition for future use in the clinic. In CLL, many studies have compared the sensitivity of MLPA to FISH, showing between 80 to 100% concordance with variation due to technical differences [88–93] . Fabris and collegues demonstrated a 95% concordance to FISH if the clone size was above 30% and purified/pre-selected CD19 positive cells were used[93]. Other studies have shown that MLPA can detect clone sizes ranging from 10-35%. In dilution experiments to detect 13q deletions, Al Zaabi *et al*. found that only deletions present in >36% cells could reliably be identified [92], whilst adjusting the cut off values per probe to allow for the normal variation seen in the reference samples, allowed Abdool *et al* to reduce the level of detection to around 20% aberrant cells, without the need to pre enrich the cell population [90]. Adjusting each individual probe cut off has also been shown to allow for a more accurate and consistent interpretation of the results and can be related to the clone size of an aberration [90]. For example, in one CLL MLPA study, cut off values <0.95 and >1.05 for loss and gain respectively were used, and they were able to detect a clone size down to just 10% [83] These studies show that MLPA is comparable to FISH and although it is less sensitive it is still accurate and specific. MLPA is also able to identify rare micro deletions in CLL that are not detectable by FISH and to detect genomic abnormalities in regions not covered by the standard CLL FISH panel, which have been validated by microarray technologies [83]. MLPA has been used successfully for confirmation of aCGH results, with 100% concordance between the two methods.[87]

MLPA can allow researchers to assess numerous novel genomic aberrations and to investigate how these interact at a genetic level, as well as helping to understand the genomic complexity of CLL, which FISH severely underestimates, and how these aberrations may impact on patient survival and treatment strategies. However, the utility of MLPA as a diagnostic tool in the CLL clinic has not yet been thoroughly tested in a well characterised trial cohort,

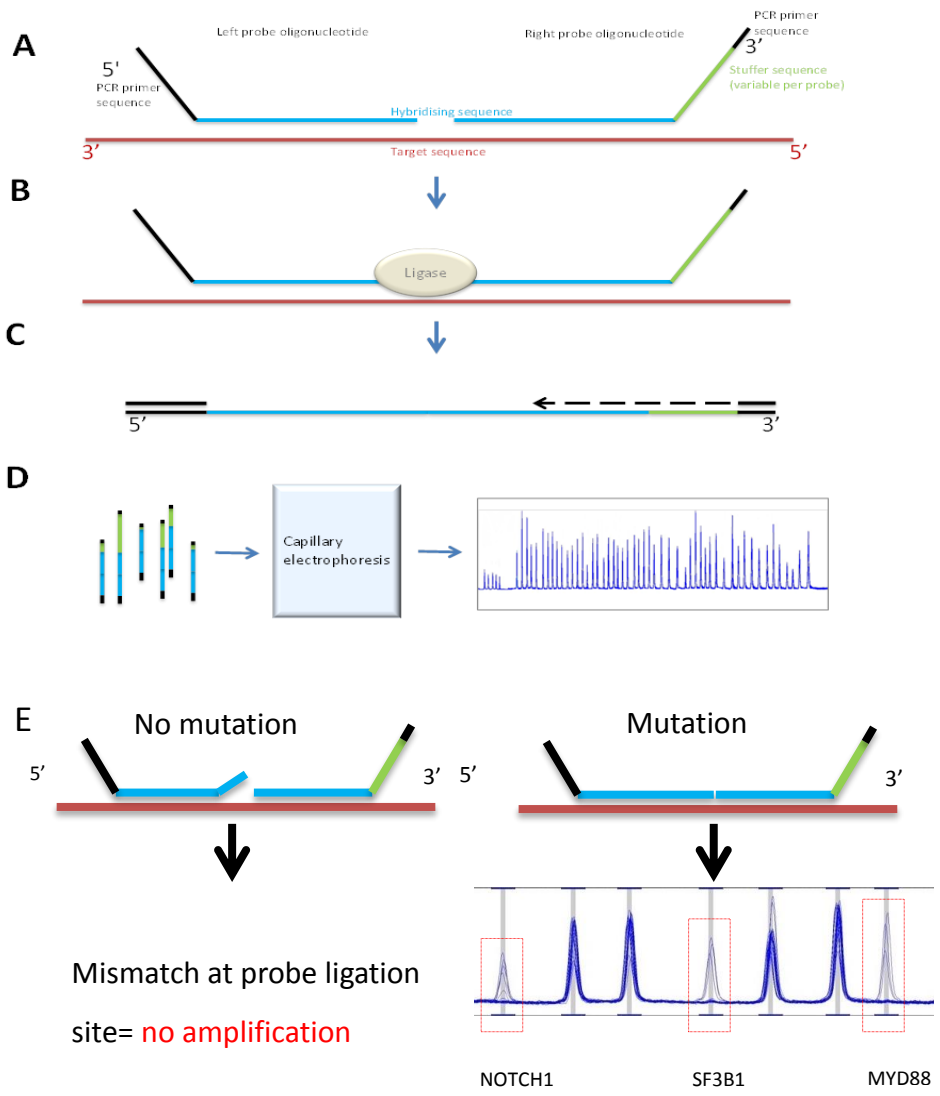


Figure 12 MLPA probe structure.A) An MLPA probe consists of two oligonucleotides that hybridise adjacent to each other on a target sequence. They consist of a hybridising sequence of ~55-80 nucleotides, complementary to the target sequence. On the 5' end is one of the universal PCR primers. The 3' probe has a stuffer sequence adjoined to the hybridising sequence, the length of which is unique to each probe, allowing each MLPA product to differ in size (range; 130 to 480 nucleotides) and be identifiable. B) After hybridisation, the sequences are ligated by a thermostable ligase. C) All probes (not the genomic DNA) are amplified in a multiplex PCR using universal primers. D) Amplified products are separated by capillary electrophoresis, producing an electropherogram of the samples. E) Mutation detection using MLPA; The probe will only bind to the target sequence if the mutation is present, the relative peak size is a reflection of clonality. In this example MYD88 has the largest peak, compared to *NOTCH1* and *SF3B1*, which is consistent with our understanding of *MYD88* mutations as early initiating events and therefore present in a high proportion of cells. [87]

4.2 Aims

- This study will employ MLPA to detect a number of chromosomal aberrations both commonly and rarely studied in CLL, in the context of a clinical trial setting.
- In addition to CNA, MLPA will be used to detect mutations in three genes, *SF3B1, NOTCH1* and *MYD88*.
- The utility of MLPA will be considered by drawing a comparison to other methods of CNA/mutational detection such as FISH, SNP 6.0 micro arrays and Sanger sequencing.

4.3 Methodology

4.3.1 Cohort description

DNA was extracted from B-cell samples (all with >80% tumour purity) from 385 of the 777 patients enrolled on the UK CLL4 treatment trial (the chi square test was used to show that the MLPA cohort is representative of the full cohort) (Table 6. Clinical characteristics of the MLPA cohort compared to the complete CLL4 cohortTable 6). The CLL4 trial is described in Chapter 3.1. A further 36 aged matched samples (range 54-78 years) were obtained from the bone marrow of healthy individuals who had undergone hip replacement surgery, for use as control samples for the MLPA analysis.

4.3.2 MLPA

MLPA was performed and analysed as described in Chapter 3.6.

4.3.3 Mutation screening

NOTCH1, *SF3B1* and *MYD88* mutations (P2515*fs~, K700E and L265P, respectively), were screened using High Resolution Melt (HRM) analysis (Chapter 3.9) and confirmed by Sanger sequencing (Chapter 3.10). See Table 5 for primer sequences.

Table 5. Mutation screening primer sequences

Gene Name	Sequencing Techique	Forward primer	Reverse Primer
NOTCH1	HRM	AGCAAACATCCAGCAGCAG	GCTCTCCTGGGGCAGAATA
	Sanger	GAGCTTCCTGAGTGGAGAG	CGAGGAGTAGCTGTGCTG
SF3B1	HRM	TTGGGGCATAGTTAAAACCTG	AAATCAAAAGGTAATTGGTGGA
	Sanger	TTGGCTGAAGCAGCAACTC	CCATAAGGAGTTGCTGCTTCA
MYD88	HRM	CAGGTGCCCATCAGAAGC	GGTTGGTGTAGTCGCAGACA
	Sanger	GCAGAAGTACATGGACAGGCAGACAGATAC	GTTGTTAACCCTGGGGTTGAAG

4.3.4 **Droplet digital PCR (ddPCR)**

Digital droplet PCR was performed as described in Chapter 3.12 to confirm *NOTCH1* and *SF3B1* mutations.

4.3.5 **Quantitative PCR**

QPCR was used to validate the 3p (n=3) and 9p (n=3) aberrations detected by MLPA as described in Chapter 3.14. The two 9p assays used covered the *CDKN2A* (TaqMan Hs01354804) and *CDK2NB* (TaqMan Hs02237898) genes, whilst for the 3p region an assay within the *SETD2* gene (TaqMan Hs06623092) was use. *RNaseP* was used as the reference gene.

4.3.6 **Monochrome Multiplex Quantitative PCR (MMqPCR)**

Telomere length was assessed in 287 samples using MMqPCR as described in Chapter 3.11.

4.3.7 **Affymetrix SNP 6.0 array comparative genomic hybridisation (aCGH)**

Archival SNP 6.0 data (Affymetrix, Santa Clara, CA, USA) was available for 71/385 cases, aligned onto the human genome sequence (GRCh37) and analysed in Partek Genomics Suite (Partek Inc, Missouri, USA) as reported previously.

4.3.8 **Statistical analysis**

Statistical analysis was performed using IBM SPSS Statistics (V.21). Pearsons Chi square statistical test was used to establish that the subset of CLL4 cases studied was representative of the entire cohort (Table 1). Cohens kappa statistical analysis was used to establish the concordance between MLPA, FISH and Sanger sequencing data. Pearsons Chi square ($\chi 2$) or Fishers exact test (sample size < 5) were used to look for significant associations ($p \leq 0.05$) between CNA and clinically relevant biomarkers. Univariate survival analysis was performed by Kaplan-Meier, log rank and Cox regression analysis, whilst multivariate analysis, accounting for confounding factors, was performed using Cox proportional hazard analysis. Overall survival (OS) was defined as the time from randomisation to death, or date of last follow up (August 2012) for survivors. Progression free survival (PFS) was defined as time from randomisation to relapse needing further therapy, progression or death; or date of last follow up (Oct 2010; final LRF CLL4 PFS update) for those with no progression or death.

Table 6. Clinical characteristics of the MLPA cohort compared to the complete CLL4 cohort

		MLPA cohort		Full CLL4 cohort		
Characteristic	**Subgroup**	**Cases (%)**	**Total no of cases with data**	**Cases (%)**	**Total no of cases with data**	**p-value**
Sex	Male	287 (75)	382	573 (74)	777	NS
	Female	95 (25)		204 (26)		
Age	Mean	64	382	64	777	NS
	Range	38-85		35-86		
Binet stage	A	88 (23)	382	191 (25)	777	NS
	B	171 (44)		352 (45)		
	C	123 (32)		234 (30)		
Treatment	Chl	200 (52)	382	387 (50)	777	NS
	FC	79 (21)		196 (25)		
	FDR	103 (27)		194 (25)		
IGHV	Unmutated	212 (62)	344	327 (61)	533	NS
	Mutated	130 (38)		206 (39)		
CD38 positivity	positive	140 (45)	312	236 (44)	535	NS
	negative	172 (55)		299 (56)		
Zap-70 positivity	Positive	103 (33)	314	146 (30.5)	478	NS
	Negative	211 (67)		332 (69.5)		
FISH 17p deletion	Normal	285 (82)	347	496 (85)	581	NS
	Deleted	62 (18)		85 (15)		
FISH 11q deletion	Normal	273 (79)	347	462 (80)	578	NS
	Deleted	74 (21)		116 (20)		
FISH trisomy 12	Normal	287 (83)	347	487 (84)	578	NS
	Deleted	60 (17)		91 (16)		
FISH 13q deletion	Normal	149 (43)	347	232 (40)	578	NS
	Deleted	198 (57)		346 (60)		
SF3B1 (K700E) Sanger	Unmutated	232 (91)	256	359 (91)	396	NS
	Mutated	24 (9)		37 (9)		
NOTCH1 (P2515*fs~) Sanger	Unmutated	290 (91)	318	450 (92)	491	NS
	Mutated	28 (9)		41 (8)		
MYD88 (L265P) Sanger	Unmutated	139 (98)	142	205 (98)	209	NS
	Mutated	3 (2)		4 (2)		
All *SF3B1* mutations (genotyping)	unmutated	233 (83.5)	279	364 (83)	437	NS
	mutated	46 (16.5)		73 (17)		
All *NOTCH1* mutation (genotyping)	unmutated	280 (90)	310	420 (90)	466	NS
	mutated	30 (10)		46 (10)		

4.4 Genomic abnormalities detected by MLPA

Chromosomal abnrmalities were identified in 77.6% of patients, with between 1 and 5 CNA per patient (1 =54%, 2 =17%,.4 ≥3 6%). Deletions of 13q (total n=200 (51.9%); monoallelic n=140 (60%) biallelic), 11q (n=63, 16.4%,), 17p (n=26, 6.8%) and trisomy 12 (n=53, 13.8%,) were detected at a frequency expected for the studied cohort, aligning with published data and demonstrating the validity of MLPA screening. Other frequent CNAs included 6q deletion (n=21, 5.5%), 2p gain (n=14, 3.7%), 8q gain (n=8, 2.1%), 14q deletions (n=6, 1.6%), 8p deletions (n=5, 1.3%) and trisomy 19 (n= 5, 1.3%). Less frequent abnormalities included; deletions of 19p and 9p, and gain of 19p (n=4, 1%), 3p deletions (n=2, 0.5%), and finally deletions of 2p, 8q and 19q, and gains of 3p and 12q (n=1, 0.3%) (Figure 13). In total, 27% (n=104) of cases had a mutation in either *NOTCH1* (n=52, 13.5%), *SF3B1* (n=51, 13.4%) or *MYD88* (n=6, 1.6%), including 49 patients (12.8%) with no concurrent CNA. Whilst mutations in *MYD88* were mutually exclusive, both *NOTCH1* and *SF3B1* were mutated in 1.6% of cases. There were 44 patients (11.5%) had no detectable chromosomal abnormalities or mutations.

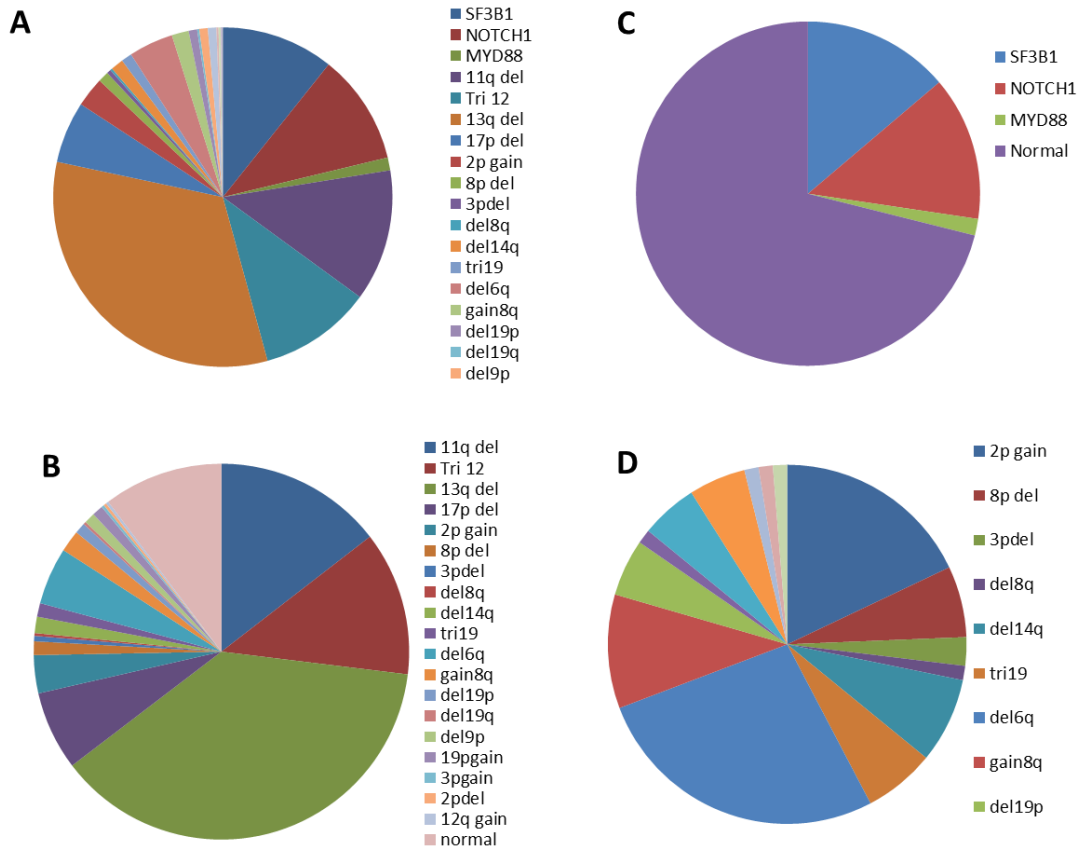


Figure 13 Frequency and distribution of genomic abnormalities detected by a combination of P037 and P038 MLPA kits (MRC Holland) A) Complete genomic abnormalities, including copy number alterations and gene mutations as detected by MLPA B) Chromosomal aberrations detected by MLPA, the less frequent one have yet to be confirmed by an alternative method C) Gene mutations detected by MLPA D) The less frequent copy number alterations identified using MLPA

4.4.1 Structural architecture of 13q deletions

The 13q14 region had the most comprehensive probe coverage (15 probes across a 3587.2Kb region) spanning countless genes implicated in the pathogenesis of CLL. Copy number profiles were created for all 13q deleted patients to look at the copy number architecture (Figure 14).

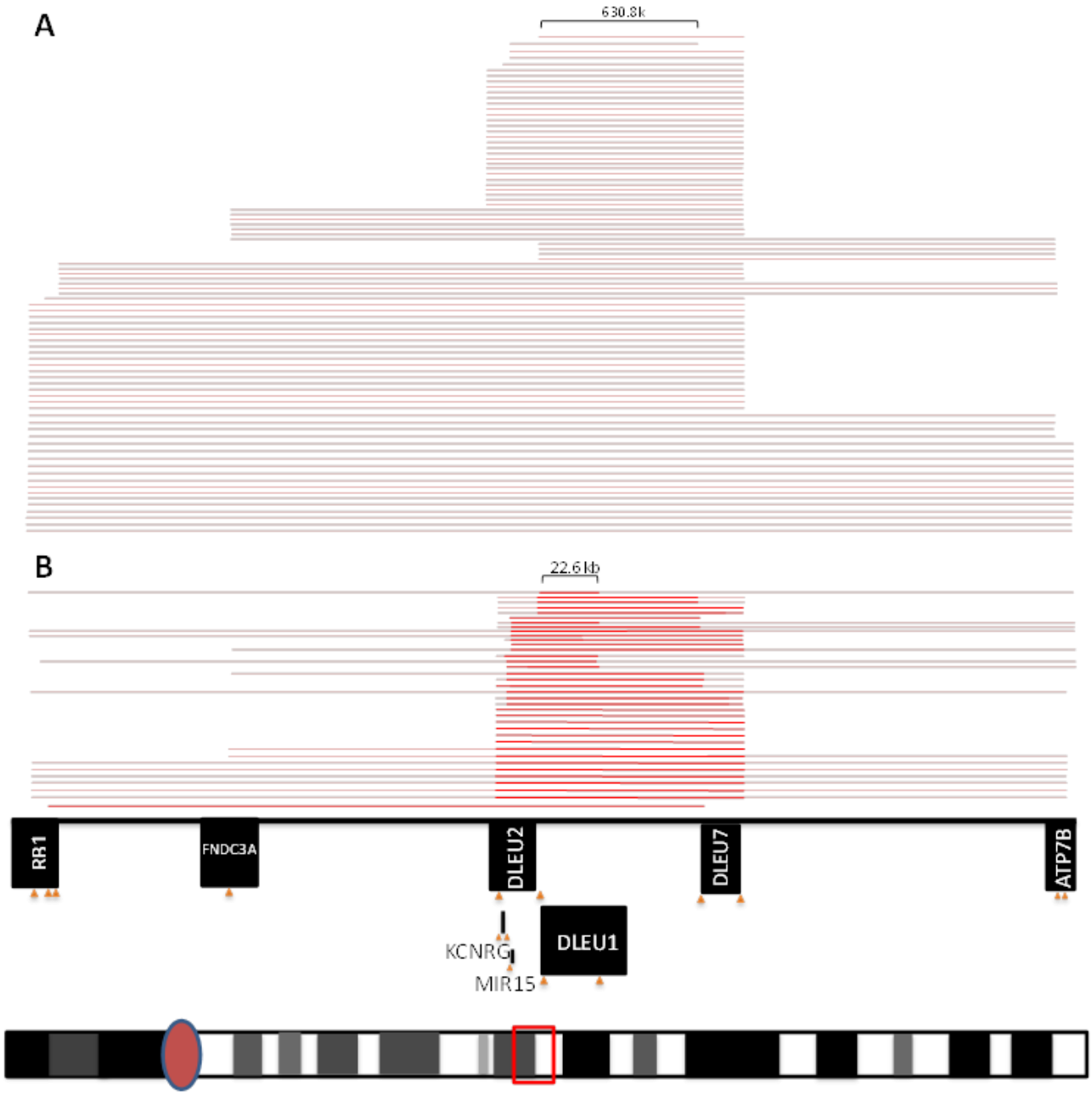


Figure 14 13q deletion profiles. Each bar represents the deletion identified in a single patient. A) monoallelic deletion profiles (pink bars); The MDR identified between DLEU2 and DLEU7, encompassing the DLEU1 gene is shown. B) Biallelic deletion profiles (red bars), including concomitant monoallelic (pink) deletions; the MDR targeting the DLEU1 gene is shown. 14 monoallelic and 3 biallelic cases were excluded from the deletion profiling as the RCN values at the extremities of the deletions were borderline to the cut-off values and therefore the exact extent of the deletion was unclear.

Deletion size was very heterogeneous ranging from ~664kb to ~3.6Mb (loss of 5 to 15 probes, respectively). A 630.8kb minimally deleted region (MDR), defined by two patients, encompassing the *DLEU1* gene, was identified in all cases. Biallelic deletions were generally larger than the MDR (~860.5kb) and included *DLEU2* and *DLEU7*. However, a more refined MDR of 23.6kb, resulting in partial loss of *DLEU1,* was delineated by a single case with biallelic del13q. Deletions were classified on the basis of size; large deletions (n=27, 22.6) included the loss of some or all of the *RB1* probes.

4.4.2 **SNP 6.0 validation of MLPA copy number data**

Archived SNP 6.0 data was available for 10 cases with biallelic 13q deletions. This data confirmed the size and location of the MDR defined by MLPA (Figure 15), validating the accuracy of MLPA as a tool for building 13q deletion profiles. Archival SNP 6.0 data was also used to confirm the presence of 12 rare CNAs identified by MLPA (Table 7). A single aberration was not confirmed and was retrospectively discounted; a trisomy 19 (ID 4260) that had 3/5 MLPA relative copy number (RCN) values borderline to the cut-offs, and 2/5 below the cut-off. In addition, SNP 6.0 data facilitated the retrospective identification of two CNA (ID 4478 and 4107) that were overlooked because there were less then 3 consecutive aberrant probes, one of which (PID 4478) was just a single MLPA probe deletion.

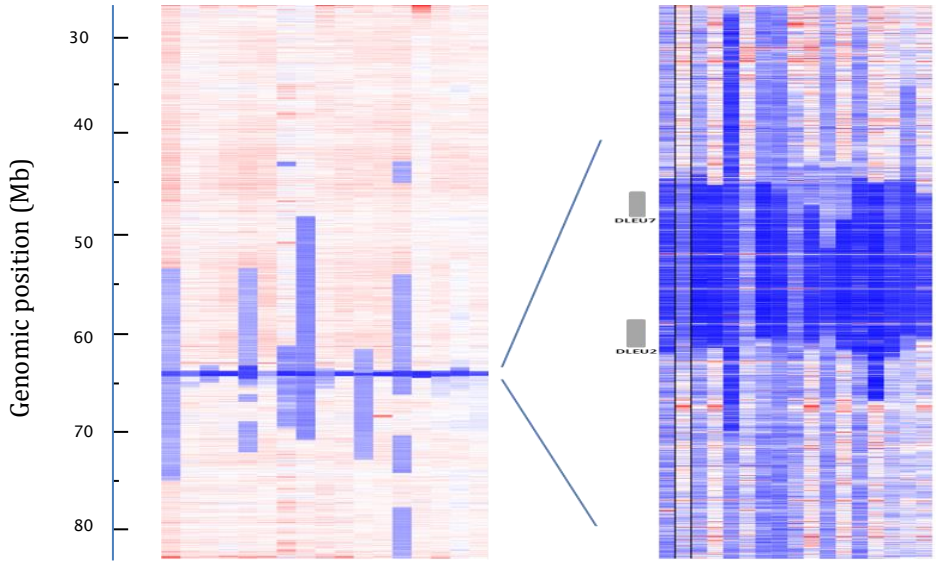


Figure 15. SNP 6.0 heatmap profiles for CLL4 patients with biallelic 13q deletion.
The genomic position of CNA is indicated by the ladder on the left hand side. Red, blue, dark blue and white bars represent gains, deletions, biallelic deletions and normal CN, respectively. The MDR is shown in more detail in the right hand panel, located between *DLEU2* and *DLEU7* (grey bars), as defined by MLPA

Table 7. Confirmation, by SNP 6.0 array data, of rare CNA detected by MLPA

ID	MLPA	SNP 6.0
4154	14q deletion	Confirmed
4168	8p deletion	Confirmed
4433	2p gain	Confirmed
4479	2p gain	Confirmed
4534	2p gain	Confirmed
4536	8p deletion	Confirmed
4662	2p gain	Confirmed
4698	8p deletion and 8q gain	Confirmed
4727	8p, 9p and 19p deletions	Confirmed
4107	Confirmed	2p gain*
4478	Confirmed	19p deletion*
4260	Trisomy 19	Not confirmed

*Abnormalities identified by SNP 6.0 and retrospectively identified by MLPA

4.4.3 Quantitative PCR validation of 9p and 3p CNA

As the 9p deletion is extremely rare and only covered by two probes in the MLPA kits, quantitative PCR assays (covering *CDK2NA* and *CDK2NB* genes) were used to confirm the deletions detected by MLPA. Also as the 3p region in the MLPA kits was part of the control probes, this was also validated using a qPCR assay (*SETD2* gene). Quantitative PCR was able to confirm the 3p CNA identified in patients 4434 and 4370 (Table 8), whilst the deletion in patient 4436 was not confirmed. However, the RCN values were well below borderline (<0.86) indicating this is a true deletion. DNA was available for 3 of the 4 patients with del9p. Of these three, two deletions were validated by qPCR (ID 4049 and 4727). The RCN values for patient 4432 were borderline to the cut-off values, indicating that this is either a false positive result, or a sub clonal deletion that is below the level of sensitivity for qPCR detection.

Table 8.Rare 9p and 3p MLPA variants were assessed using qPCR analysis. RCN values from MLPA are used to determine copy number status (<0.86 for deletion and >1.12 for a gain) similarily qPCR values are also used in this way.

PID	Chromosome	Gene	CNA	MLPA (RCN)	qPCR	Status
4434	Chr3	SETD2	deletion	0.56	0.65	Confirmed
4436	Chr3	SETD2	deletion	0.59	1.21	Not confirmed
4370	Chr3	SETD2	gain	1.28	1.99	Confirmed
4510	Chr9	CDK2NA	deletion	0.46	NA	Confirmed
4510	Chr9	CDK2NB	deletion	0.53	NA	Confirmed
4049	Chr9	CDK2NA	deletion	0.64	0.77	Confirmed
4049	Chr9	CDK2NB	deletion	0.70	0.79	Confirmed
4432	Chr9	CDK2NA	deletion	0.84	0.94	Not confirmed
4432	Chr9	CDK2NB	deletion	0.83	1.03	Not confirmed
4727	Chr9	CDK2NA	deletion	0.58	0.74	Confirmed
4727	Chr9	CDK2NB	deletion	0.60	0.76	Confirmed

4.4.5 **Further validation of *NOTCH1* mutations**

For validation of the mutations detected by MLPA, data was available from archived Sanger sequencing (Southampton Cancer genomics group) and TruSeq (Author kindly acknowledges Mr Stuart Blakemore for access to TruSeq data). By Sanger sequencing, 27 out of 42 cases analysed were confirmed to be *NOTCH1* mutated. By TruSeq, 10 out of 32 cases were confirmed to be *NOTCH1* mutated (5/9 of these also confirmed by Sanger). To account for low frequency mutations ddPCR was employed to confirm mutations not identified by Sanger sequencing. 19/22 cases analysed by ddPCR, were confirmed to be *NOTCH1* mutated (Figure 16). Of the 51 cases that were mutated by MLPA, 8 were not confirmed to be mutated by either Truseq, Sanger or ddPCR. 5 of these cases (PID 4069, 4731, 4762, 4779, 4799) did not have any DNA or material left to analyse and the remaining 3 cases (PID 4481, 4717, 4750) did not have any mutations present by any of the other methods.

4.4.6 **Further validation of *SF3B1* mutations**

The *SF3B1* mutation was also validated through the analysis of archived Sanger sequencing and TruSeq data with further Sanger sequencing and ddPCR. 20 of the 32 cases with Sanger sequencing were identified as mutated. 27 of the 37 analysed by TruSeq harboured a mutation (11/15 of these also Sanger confirmed). Further Sanger sequencing in 19 cases identified 18 as having a mutant Sanger trace (Figure 17), whilst ddPCR in 8 cases identified 6 of these to be mutated (Figure 18). Overall, 45 of the 53 *SF3B1* mutations that were identified by MLPA, were confirmed to be real mutations by TruSeq, Sanger sequencing and/or ddPCR. 7 of those not confirmed (PID 4647, 4660, 4740, 4691, 4727, 4207 and 4273) did not have any DNA material left to analyse. The remaining 2 cases (4011 and 4048) were confirmed to be negative by ddPCR.

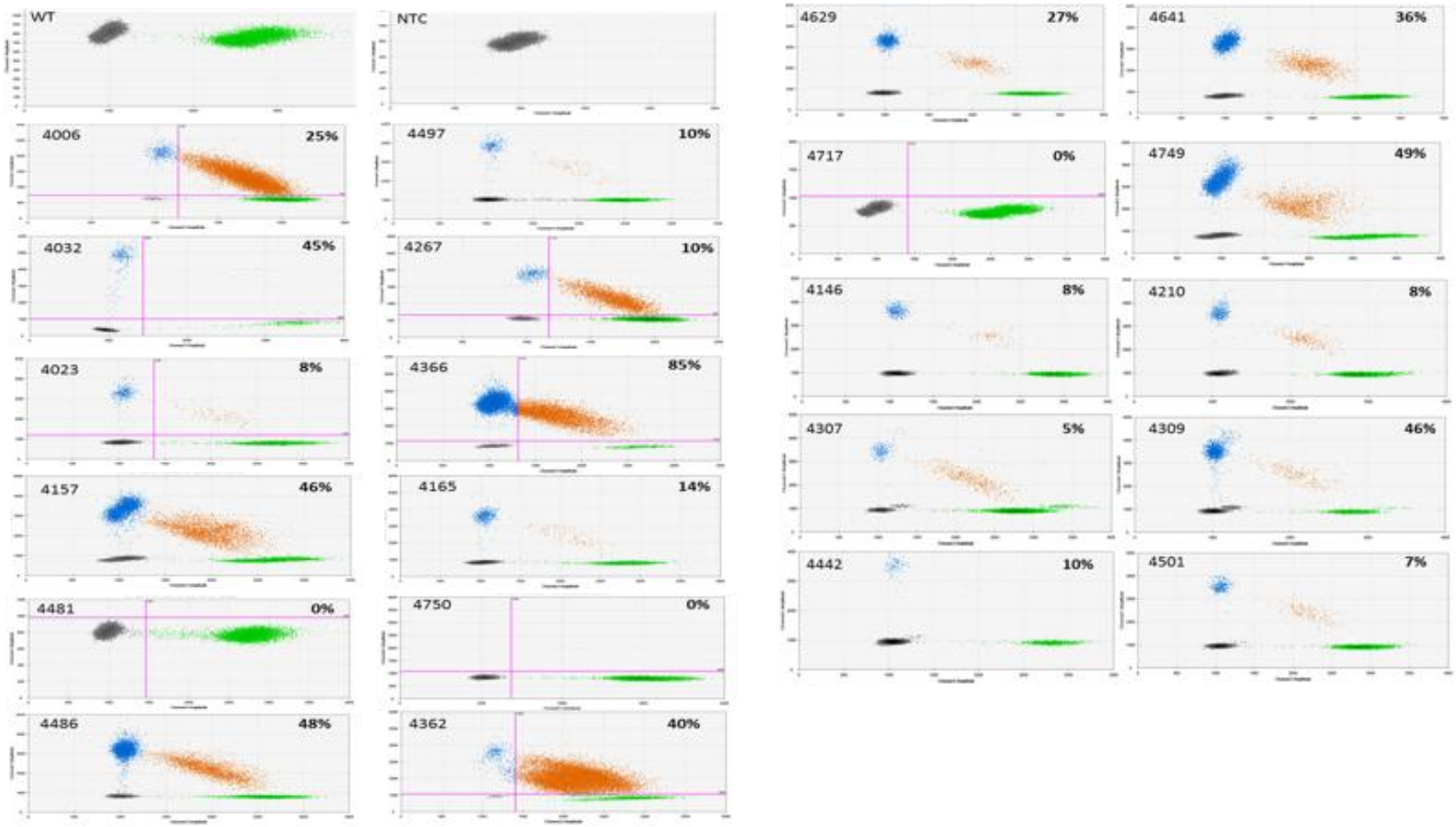


Figure 16 Validation of NOTCH1 mutation identified by MLPA, by ddPCR

Wildtype (WT) samples had no mutant populations, only negative (black) and WT (green). Mutatnt populations are represented by orange populations. The percentage of mutated cells is shown in the top right corner of each plot.

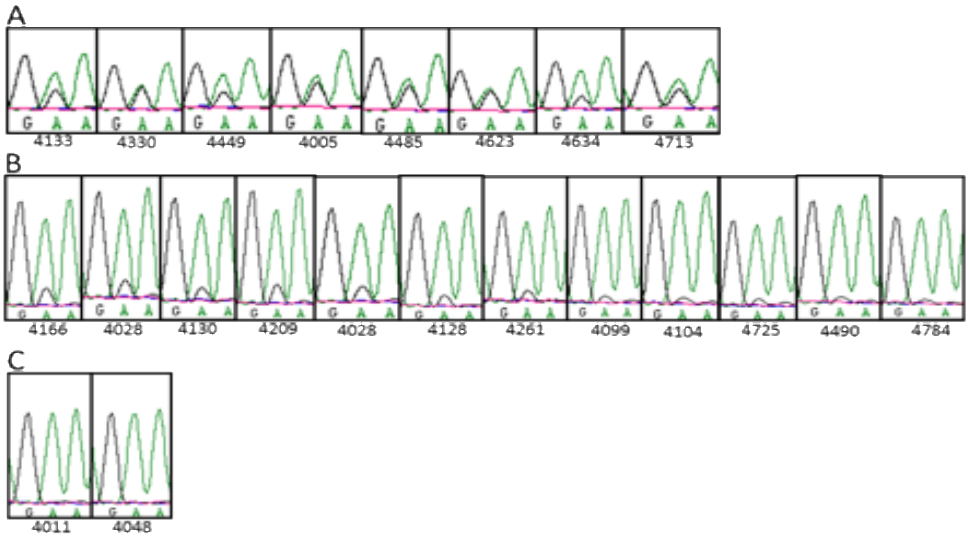


Figure 17. Sanger sequencing of *SF3B1* variants detected by MLPA

A) Clonal mutations. B) Sub-clonal mutations. C) Mutations in a sub-clonal fraction, below the detection limit of Sanger sequencing. These were validated using ddPCR

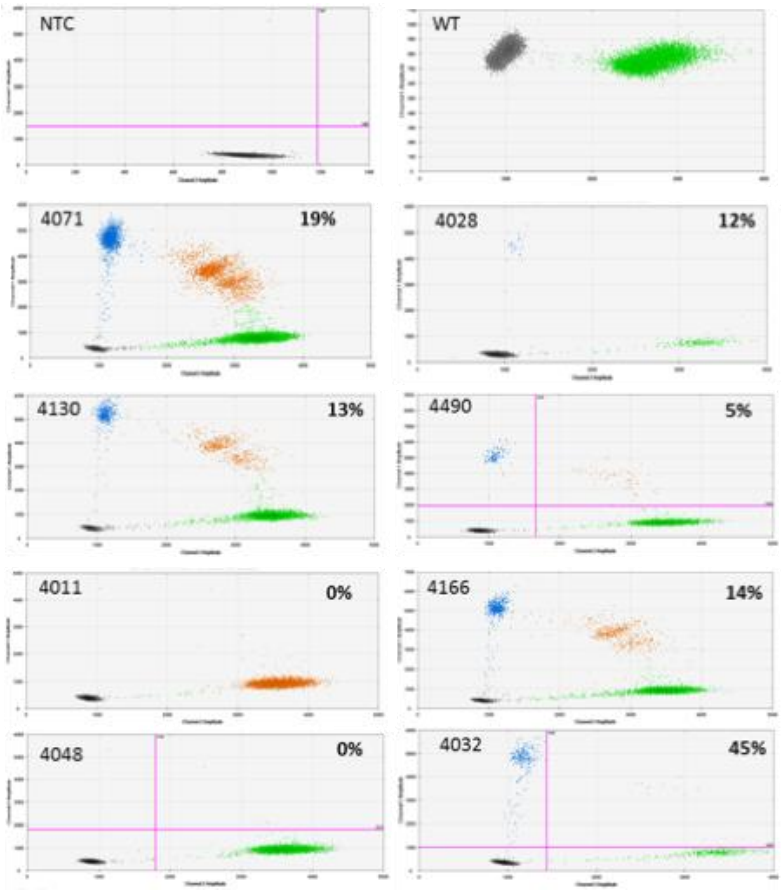


Figure 18. Validation of the *SF3B1* mutations that were identified by MLPA, by ddPCR.

Sensitivity and specificity of MLPA

Matched FISH data was available for 6q deletions (n=243), 11q deletions (n=347), trisomy 12 (n=347), 13q deletions (n=347) and 17p deletions (n=347), and Sanger sequencing data was available for *NOTCH1* (n=314), *SF3B1* (n=252) and *MYD88* (n=140) mutations. Cohens Kappa statistical analysis was conducted to test the agreement between MLPA and these 'gold standards' methods of aberration detection, as well as the sensitivity and specificity of MLPA. (Table 9-16). Overall there was substantial agreement between FISH and MLPA, for all CNA (κ=0.726, p<0.0001), and between MLPA and Sanger sequencing for all mutations (κ=0.742, p<0.0001). The level of concordance, sensitivity and specificity varied according to the abnormality, which is likely a reflection of the clonal population size.

Table 9-16. MLPA vs the 'Gold standards' for CNA and mutation detection; FISH and Sanger sequencing, respectively. The concordance, sensitivity and specificity of MLPA was determined using the Kappa statistical test.

Table 9 MLPA vs FISH 11q deletion

	MLPA								
FISH	**Positive**	**Negative**	**Total**	**Concordance**	**Sensitivity**	**Specificity**	**Kappa**	**Agreement**	**p value**
Positive	56	18	74	94.2%	75.7%	99.3%	0.814	Substantial	p<0.0001
Negative	2	271	273						
Total	57	290	347						

Table 10 MLPA vs FISH Trisomy 12

	MLPA								
FISH	**Positive**	**Negative**	**Total**	**Concordance**	**Sensitivity**	**Specificity**	**Kappa**	**Agreement**	**p value**
Positive	48	12	60	96.5%	80.0%	100.0%	0.869	Almost perfect	p<0.0001
Negative	0	287	287						
Total	48	299	347						

Table 11 MLPA vs FISH 13q deletion

	MLPA								
FISH	**Positive**	**Negative**	**Total**	**Concordance**	**Sensitivity**	**Specificity**	**Kappa**	**Agreement**	**p value**
Positive	170	28	198	89.9%	85.9%	95.3%	0.798	Substantial	p<0.0001
Negative	7	142	149						
Total	177	170	347						

Table 12 MLPA vs FISH 17p deletion

	MLPA								
FISH	**Positive**	**Negative**	**Total**	**Concordance**	**Sensitivity**	**Specificity**	**Kappa**	**Agreement**	**p value**
Positive	23	39	62	88.8%	37.1%	100.0%	0.492	Moderate	p<0.0001
Negative	0	285	285						
Total	23	324	347						

Table 13 MLPA vs FISH 6q deletion

	MLPA								
FISH	**Positive**	**Negative**	**Total**	**Concordance**	**Sensitivity**	**Specificity**	**Kappa**	**Agreement**	**p value**
Positive	15	7	22	94.7%	68.2%	97.3%	0.668	Substantial	p<0.0001
Negative	6	215	221						
Total	21	222	243						

Table 14 MLPA vs Sanger *SF3B1* mutation

	MLPA								
Sanger	**Positive**	**Negative**	**Total**	**Concordance**	**Sensitivity**	**Specificity**	**Kappa**	**Agreement**	**p value**
Positive	20	2	22	94.8%	90.9%	95.2%	0.727	Substantial	p<0.0001
Negative	11	219	230						
Total	31	221	252						

Table 15 MLPA vs Sanger *NOTCH1* mutation

	MLPA								
Sanger	**Positive**	**Negative**	**Total**	**Concordance**	**Sensitivity**	**Specificity**	**Kappa**	**Agreement**	**p value**
Positive	27	1	28	95.2%	96.4%	95.1%	0.757	Substantial	p<0.0001
Negative	14	272	286						
Total	41	273	314						

Table 16 MLPA vs Sanger *MYD88* mutation

	MLPA								
Sanger	**Positive**	**Negative**	**Total**	**Concordance**	**Sensitivity**	**Specificity**	**Kappa**	**Agreement**	**p value**
Positive	3	0	3	98.6%	100.0%	98.5%	0.743	Substantial	p<0.0001
Negative	2	135	137						
Total	5	135	140						

The smallest clone size in which MLPA could accurately detect the aberration varied from 6-31% and is likely dependant on the probe coverage of the genomic region, as smaller subclones were detected for 13q and 11q CNA, the two regions with the greatest coverage (Figure 17.) The smallest detected clone with 11q deletions was in 6% cells. All MLPA probes were deleted and the RCN values were well below borderline. Clones at 11%, 13% and 20% were also detected. However, a del11q in 21% of cells was missed; RCN values were within the normal copy number range (0.915-1.026). Similarly, 13q deletions were detected in samples with 8%, 12%, 14% and 19% aberrant cells but were missed in a case with 24% aberrant cells (ID 4377). An 11q deletion in 17% of cells was also unidentified in this case. The smallest clone size in which Trisomy 12 and 6q were detected was 23% and 31%, respectively. The sensitivity of 17p detection by MLPA was low (37.1%) because a large proportion of the deletions (64.5%) had a clone size below 30%, and 26% aberrant cells was the lower limit of detection (Figure 19).

Only three mutations identified by Sanger sequencing were missed by MLPA (*NOTCH1* n=1, *SF3B1* n=2).

4.4.7 **Genomic aberrations associate with clinically relevant biomarkers**

NOTCH1 mutations were associated with Trisomy 12 ($p<0.0001$) and trisomy 19 ($p<0.0001$). 17p deletions were associated with 8p deletions ($p=0.45$), 6q deletions ($p=0.013$), 8q gains ($p= 0.016$), 19p and 9p deletions ($p=0.029$). 8q deletions associated with 19q deletions ($p=0.003$). When the 13q deletion size was taken into account a large deletion (inclusion of *RB1*) was significantly associated with 11q deletions ($p=0.019$).

NOTCH1 mutations were associated with unmutated *IGHV* status ($p=0.003$), as were 11q deletions ($p<0.0001$) and trisomy 12 ($p=0.001$), whereas *MYD88* mutations were associated with a mutated IGHV status ($p=0.031$)

NOTCH1 mutations ($p<0.0001$), 11q deletions ($p=0.008$) and trisomy 12 ($p=0.002$) were associated with CD38 positivity whereas 13q deletions, specifically biallelic deletions, were observed in CD38 negative cases ($p=0.001$). In fact, in cases with del13q as the sole abnormality, 95% of cases are CD38 negative. Four of the five *MYD88* mutated patients are also CD38 negative but this was not significant, due to the small sample size.

Whilst 14q deleted patients were associated with Zap70 positivity ($p=0.043$), 13q deletions were associated with Zap70 negativity (77% cases negative) ($p<0.0001$), with a bias toward biallelic

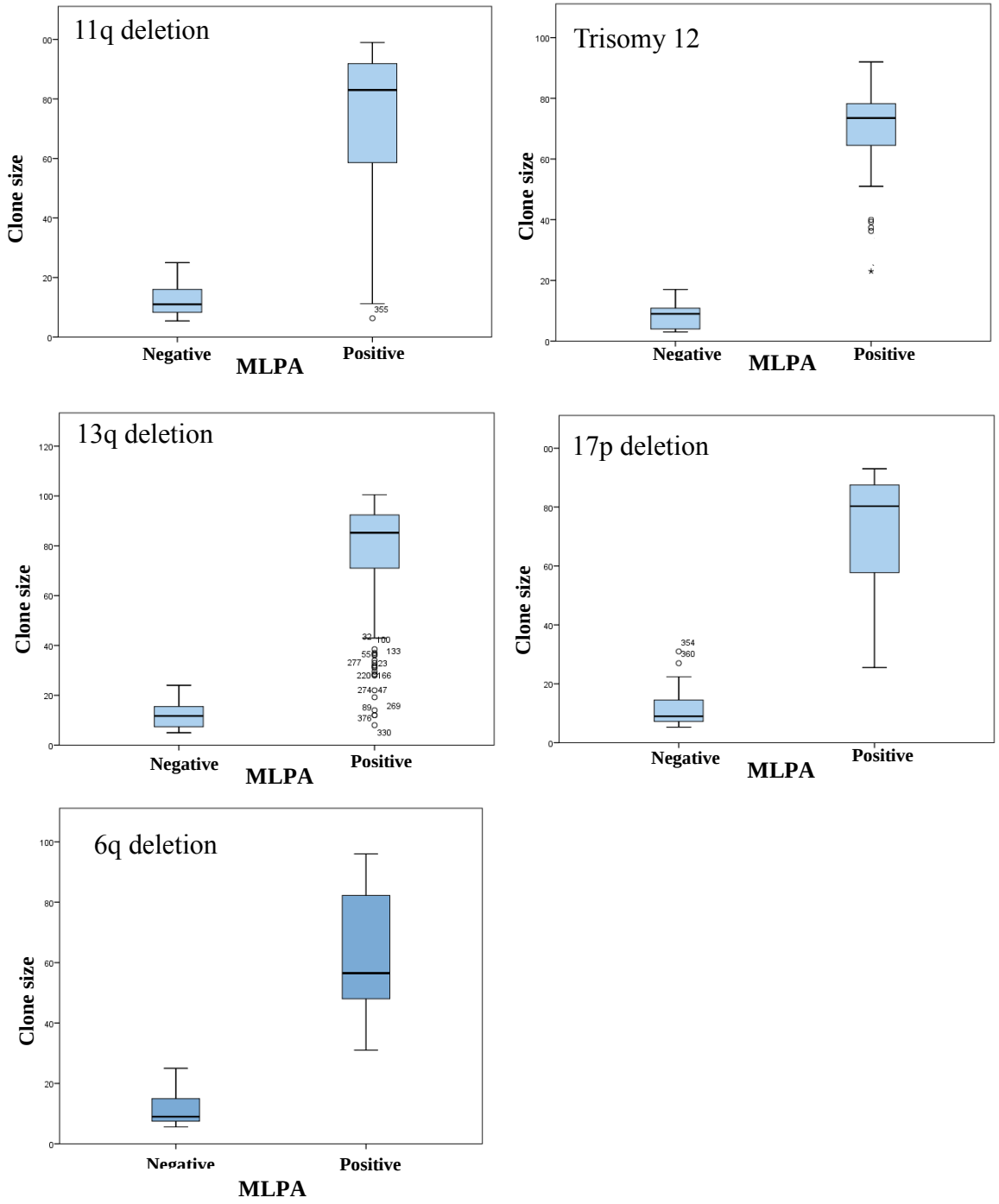


Figure 19. MLPA sensitivity is determined by clone size.
Small sub-clones (<30%), as determined by FISH, tend to be below the detection threshold of MLPA

deletions (89.5% cases negative) (p=0.37). All *MYD88* mutated patients were Zap70 negative, though this was not significant. Binet Stage A was associated with 13q deletion (p=0.016), whilst stage B/C was associated with *NOTCH1* mutation (p=0.043).

The use of Circos plots (Figure 20 and Figure 21) allows the associations between different genomic abnormalities to be visualised; a large proportion of 11q deleted patients also have a 13q deletion, trisomy 12 significantly associates with NOTCH1 and Trisomy 19 (p<0.0001), 13q deleted patients can have a complex genome, 17p deletions associate with most other genomic abnormailities, consistent with a more complex and unstable genome, *NOTCH1* and *SF3B1* mutated patients tend to have a much more complex genome than *MYD88* mutated patients, *NOTCH1* mutated patients may have trisomy 12, but interestingly it is not seen in the trisomy 12/19 patients, a large proportion of *SF3B1* mutated patients also have a 13q deletion, *MYD88* mutations do not associate with any of the other poor risk CNA such as 17p or 11q, although some have a 13q deletion.

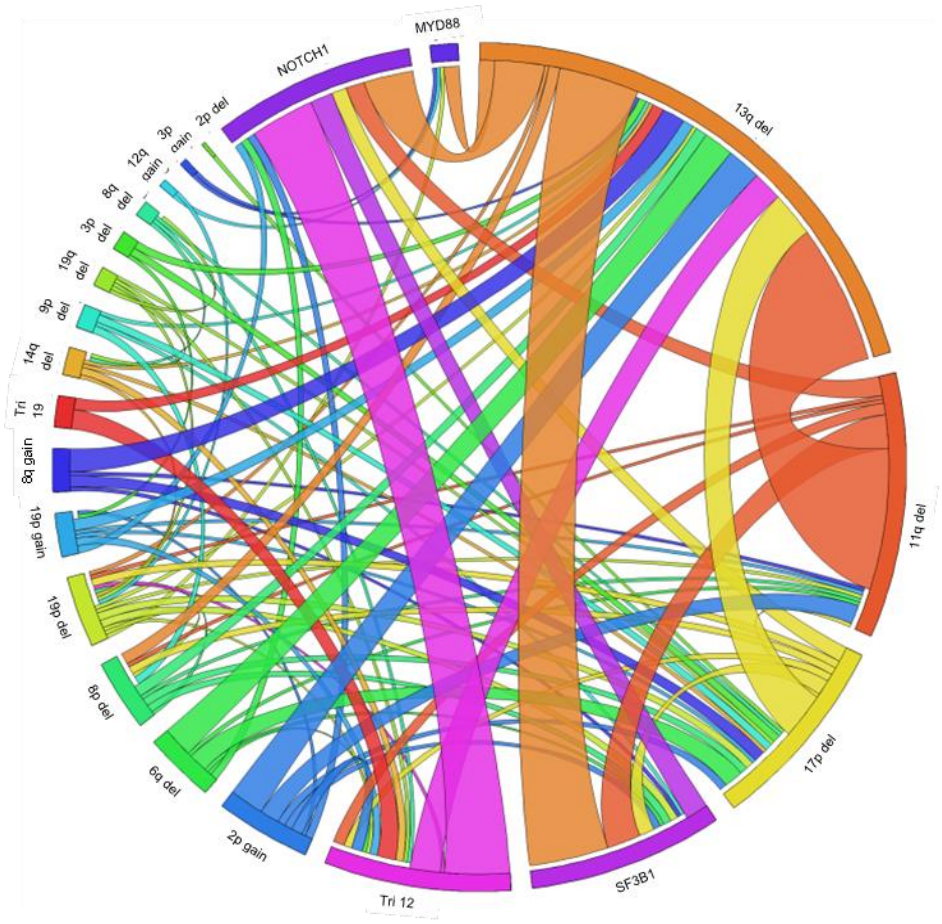


Figure 20. Circos diagram illustrating the genomic complexity and the interplay between all genomic aberrations detected by MLPA. Visually inspecting circus plots can provide an insight into genetic associations, for instance; a large proportion of *NOTCH1* mutated patients also have a trisomy 12, whilst 11q and 13q deletions can co occur.

!

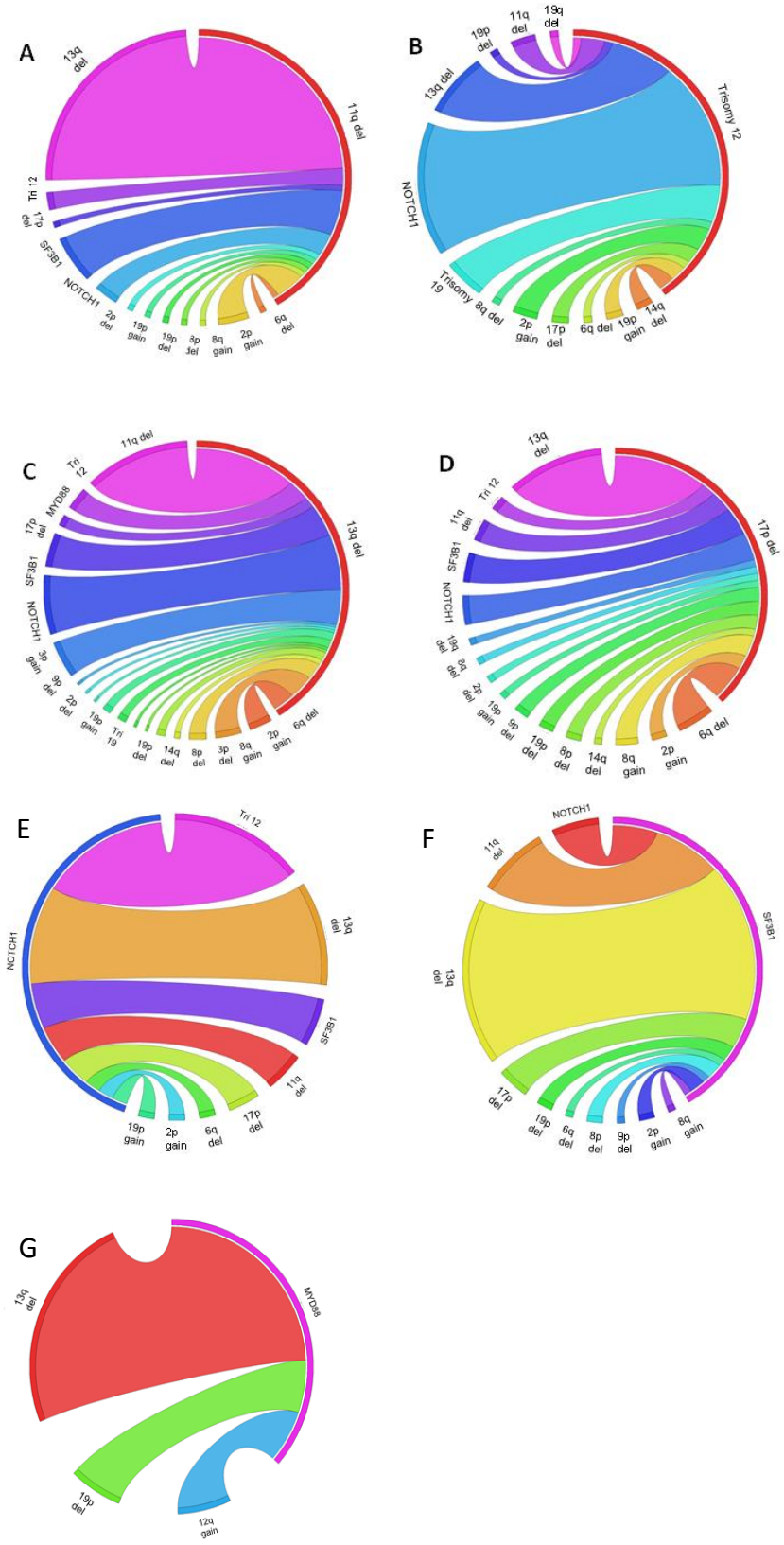


Figure 21. 22 Circos plots displaying the genomic complexity associated with established CNA and mutations. *MYD88* mutated patients are wholly less complex than other genomic abnormalities such as 13q or 17p deleted patients.
A) 11q deletion. B) Trisomy 12. C) 13q deletions. D) 17p deletions. E) *NOTCH1* mutations. F) *SF3B1* mutations. G) *MYD88* mutations

4.4.8 The impact of genomic lesions upon survival

The established recurrent CNA had the following impact on survival (13q deletions are discussed in the next section): 17p deleted patients had a reduced median OS and PFS compared to WT cases (3.8 vs 21.5 months and 30.5 vs 99 months, respectively). Both the median OS and PFS were also reduced in del11q cases compared to WT (17 vs 24 months and 71.7 vs 99 months, respectively). Trisomy 12 was associated with a reduction in OS compared to WT (70.9 vs 97.1 months), but PFS was not significantly affected (Table 17 and Figure 22).

The less frequent CNA had the following impact: 8p deletions, 6q deletions, 8p gains, 19p deletions and 9p deletions were all associated with a significant reduction in both median OS and PFS compared to the wild type cases (median OS= 94.9 months and PFS=23 months for WT cases). 2p gains were associated with a significantly reduced OS (46.4 months) but PFS was not affected. Refer to Table 17 and Figure 23 for survival time in months and p-values.

Whilst *NOTCH1* mutations significantly reduced OS from 99 to 68 months, *SF3B1* and *MYD88* mutations did not significantly affect OS or PFS, although there was a trend towards a decreased and increased OS, respectively compared to WT (Table 17 and Figure 24).

Table 17.Univariate cox regression anlysis for progression free survival (PFS) and overall survival (OS) for the statistically significant copy number alterations and gene mutations detected by MLPA.

	Progression-free survival (months)						Overall survival (months)					
Genomic abnormality	Events/Total	median	95% CI	HR	95% CI	p value	events/total	median	95% CI	HR	95% CI	p value
13q deletion	182/200	23	20-26	0.87	0.71-1.07	0.189	137/200	100	88-113	0.67	0.53-0.85	0.001
11q deletion	62/62	17	13-21	1.73	1.31-2.28	<0.0001	52/62	72	52-91	1.50	1.10-2.02	0.009
Trisomy 12	48/52	22	15-29	0.96	0.71-1.30	0.792	43/52	71	40-102	1.45	1.04-2.00	0.026
NOTCH1	48/51	19	15-23	1.32	0.97-1.79	0.068	45/49	68	47-90	2.03	1.46-2.80	<0.0001
6q deletion	20/21	18	9-27	1.62	1.03-2.56	0.031	20/21	66	12-120	1.74	1.10-2.75	0.016
2p gain	14/14	10	1-19	1.52	0.89-2.60	0.119	12/14	46	3-90	2.25	1.29-3.94	0.004
8q gain	8/8	7	0-22	2.77	1.37-5.63	0.003	8/8	42	0-120	2.28	1.12-4.62	0.019
8p deletion	5/5	6	2-10	2.44	1.00-5.90	0.039	5/5	42	0-87	5.55	2.25-13.66	<0.0001
17p deletion	27/27	4	3-5	3.48	2.34-5.20	<0.0001	26/27	31	25-36	3.91	2.59-5.91	<0.0001
19p deletion	4/4	9	1-17	3.13	1.16-8.47	0.016	4/4	20	0-66	3.03	1.13-8.16	0.021
9p deletion	4/4	2		10.24	3.70-28.33	<0.0001	4/4	10	0-27	22.75	7.91-65.38	<0.0001

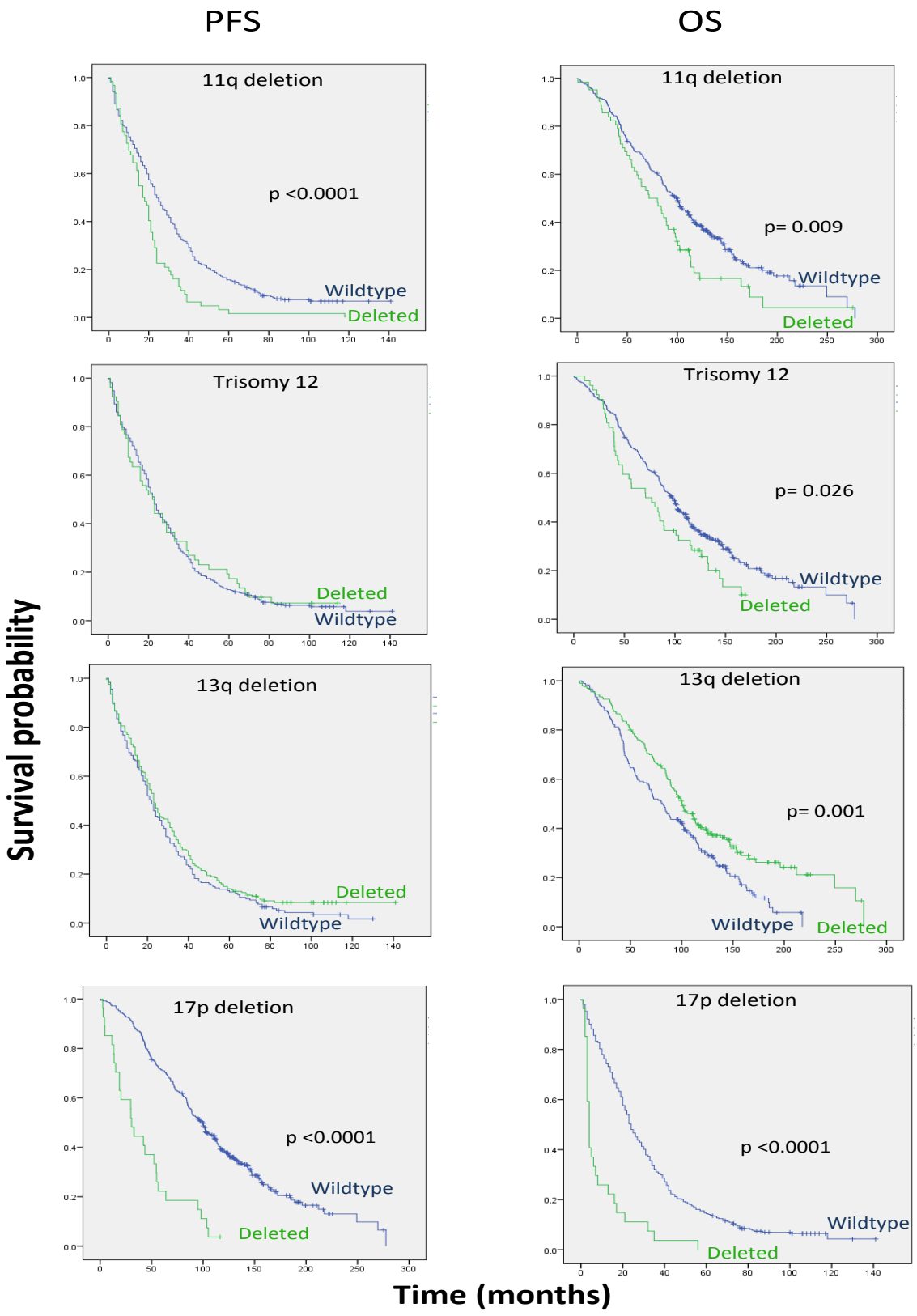


Figure 22. Kaplan Meier survival analysis for the established recurrent CNAs in CLL, as detected by MLPA. Deletions of 11q and 17p impact significantly on both PFS and OS, while 13q deletion and trisomy 12 only reduce OS significantly.

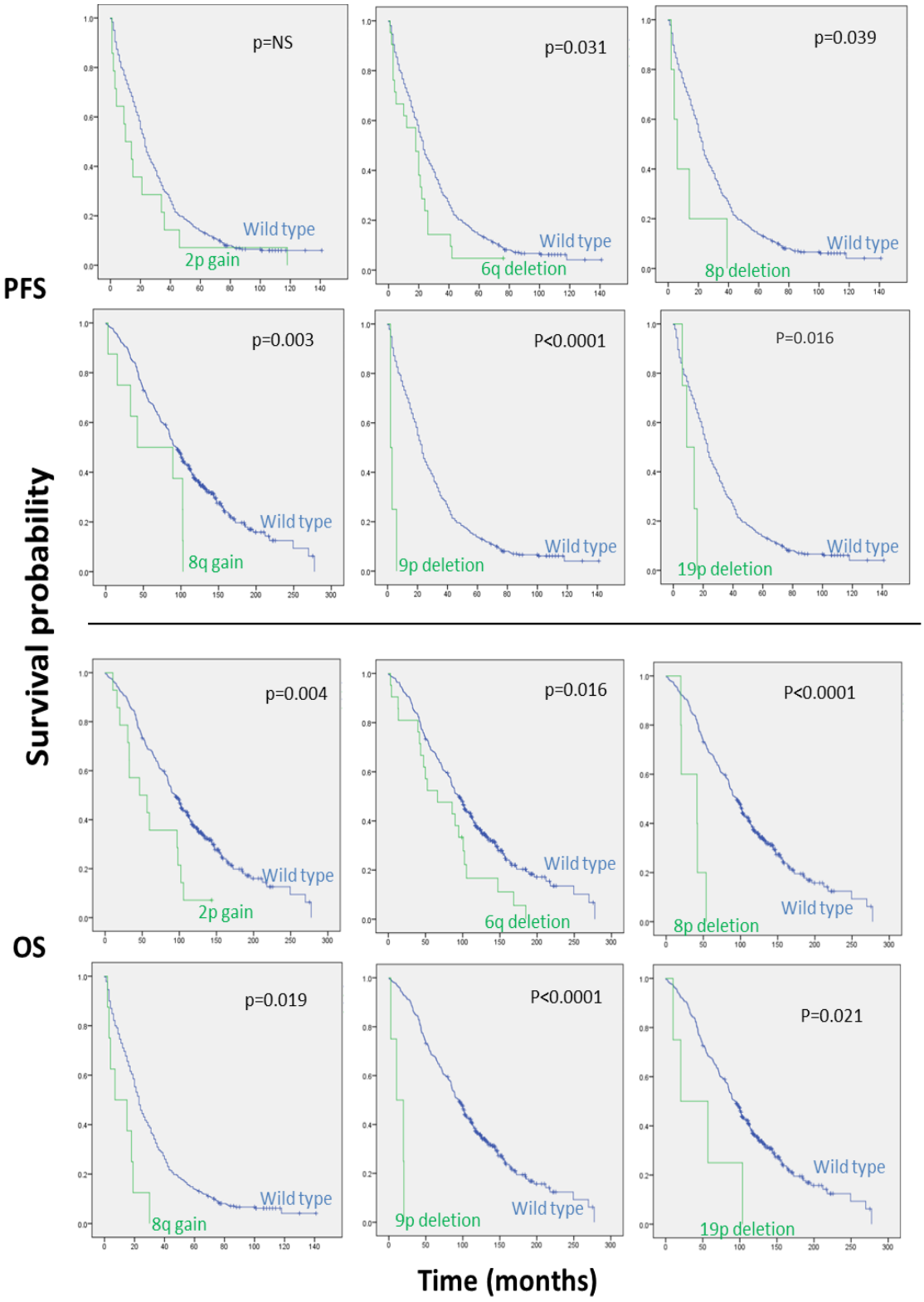


Figure 23. Kaplan Meier survival analysis for the less frequent CNAs in CLL, as detected by MLPA, impact significantly on survival. 2p and 8q gains, 6q, 8p, 9p and 19p deletions all demonstrate a significant reduced survival (PFS and/or OS) in patients carrying these CNAs.

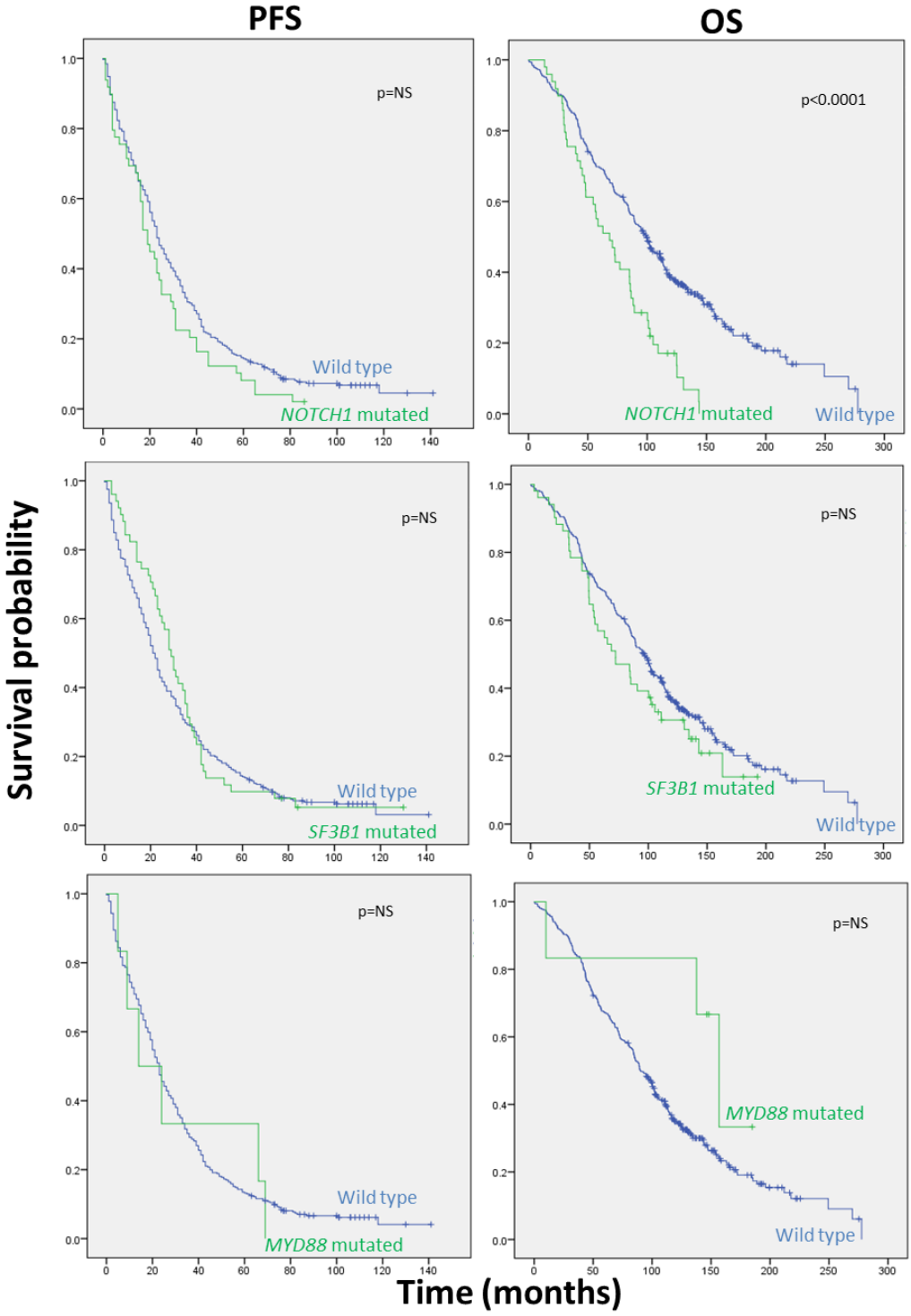


Figure 24. Kaplan Meier analysis for the mutations detected by MLPA. None of the mutations had a significant impact upon PFS, though *NOTCH1* mutated patients did significantly reduce OS.

4.4.9 **Examining the impact of 13q deletion upon survival**

13q deletions had a positive impact on median OS compared to WT (100.4 vs 81 months) (Table 17 and Figure 22). Stratifying 13q deletions by mono/biallelic status and size showed that biallelic deletions have an even more favourable effect on OS; 95 months (monoallelic loss) vs 212 months (biallelic loss), whereas large deletions (inclusion of *RB1*) reduced OS compared to small deletions (89.3 vs 111 months). Patients were then stratified by the size of the 13q clone; < 80% and ≥ 80%. The results were not statistically significant for either OS or PFS, but there was a trend towards an increased OS in patients with a clone size ≥ 80% (Figure 25).

13q as a sole abnormality was then considered (n=130; monoallelic n=100, biallelic n=30). The OS for monoallelic cases was similar WT patients (111 vs 102 months, respectively), whilst patients with biallelic deletions had a significantly increased OS (212 months). To check the accuracy of these results the statistical analyses, using the full CLL4 trial cohort FISH data was performed. As seen in the MLPA cohort, a biallelic deletion was associated with an increased OS, compared to monoallelic loss and WT, respectively (158 vs 95 vs 74 months). *IGHV* status in these cases with biallelic loss was then investigated. An unmutated *IGHV* (n=4) significantly reduced both PFS and OS (9 and 76.5 months) compared to mutated *IGHV* cases (n=24) (31 and 212 months). (Figure 26 and Figure 27).

With one exception (a *TP53* mutation), the *IGHV* unmutated cases have no other poor risk markers (e.g. Zap-70 or CD38 positivity, *SF3B1*/*NOTCH1* mutations, other CNAs, Binet stage C). Therefore, the presence of either a mono or biallelic deletion, as a sole aberration, in patients with a mutated *IGHV* status, confers the increased survival seen in the 13q deletion.

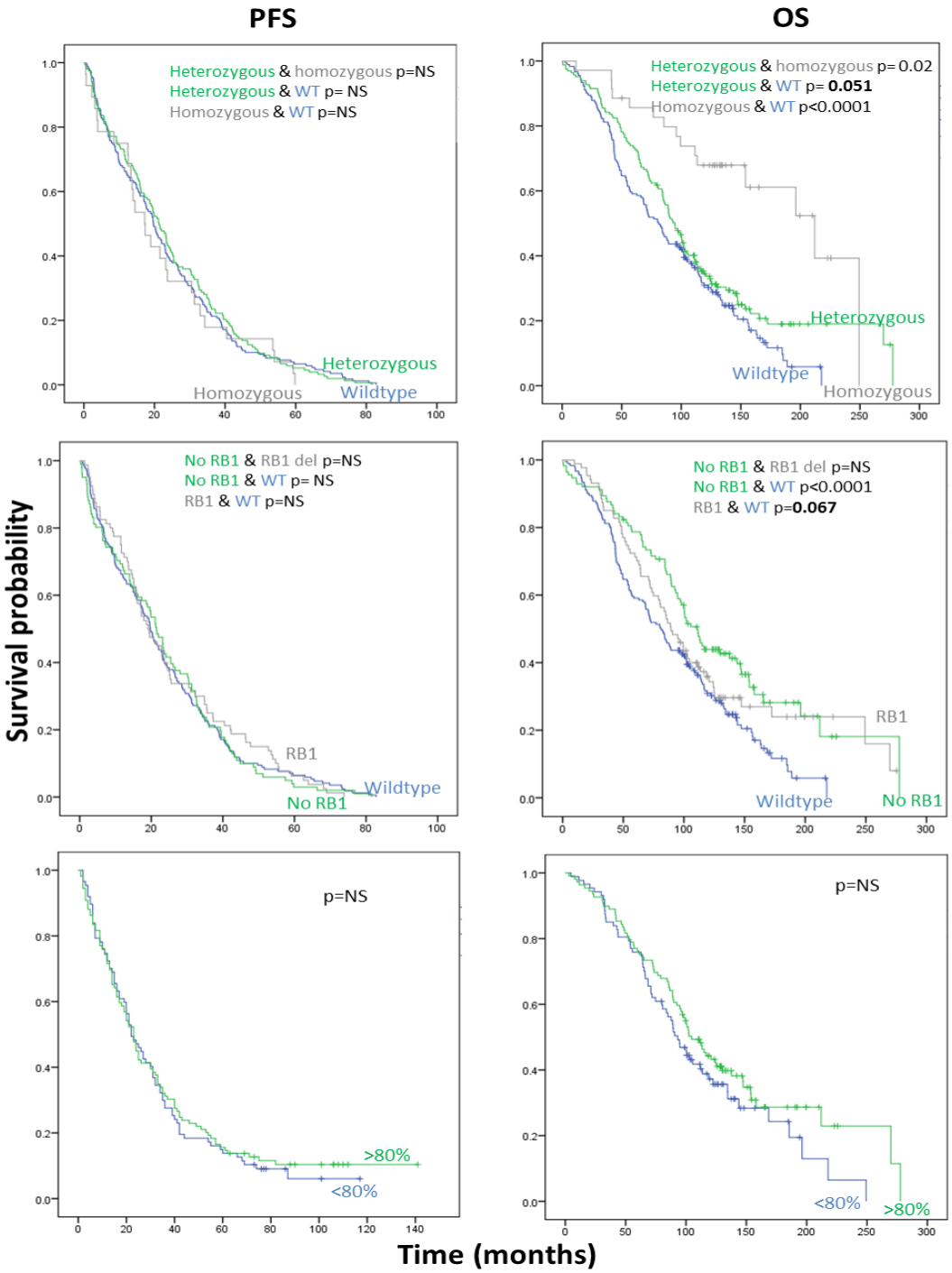
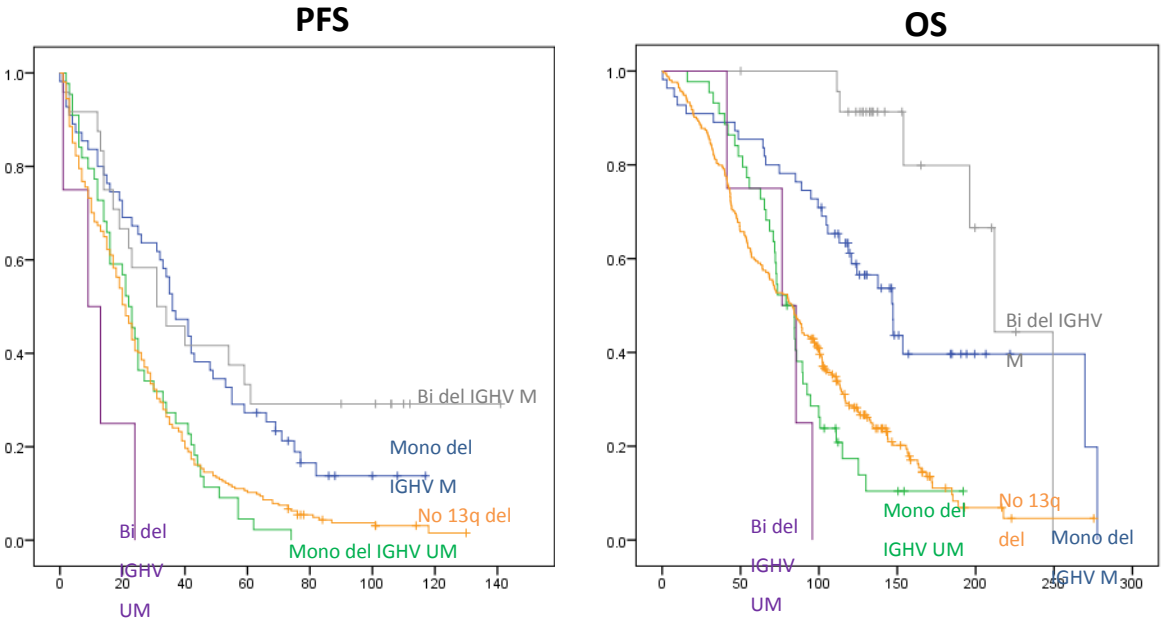


Figure 25. Kaplan Meier analysis of the impact of 13q deletion.

Patients were stratified by mono- and bi-allelic deletions, deletion size and clone size. Deletion size (inclusion of *RB1* gene deletion) and clone size (</> 80%) did not impact upon survival, whilst a homozygous deletion had a significantly increased survival compared to heterozygous deletion.

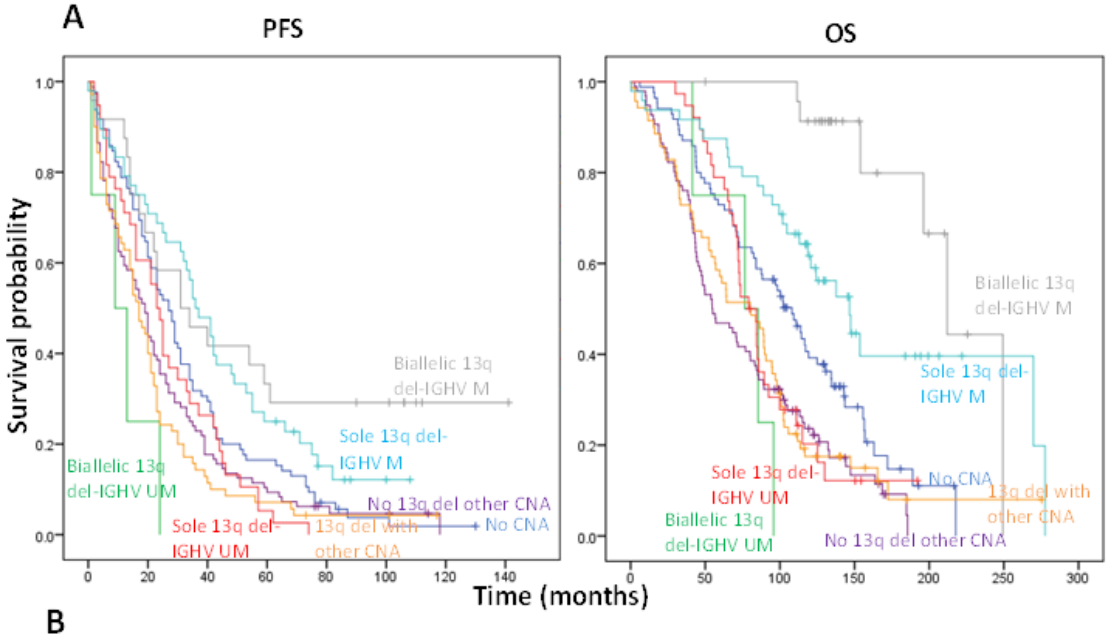


	PFS				OS			
	Total number	Total events	Median	95% CI	Total number	Total events	Median	95% CI
mono del-IGHV mutated	55	46	36	28.7-43.3	55	30	147.253	120.3-174.2
mono del-IGHV UM	44	44	22	17.7-26.3	44	37	78.916	66.5-91.4
bi del IGHV M	24	17	31	10.6-51.4	24	6	212.008	33.2-119.7
bi del IGHV UM	4	4	9	0.0-20.8	4	4	76.452	94.4-147.0
No 13q deletion	254	245	21	18.7-23.3	254	205	82.5	71.3-93.7

	mono del-IGHV mutated	mono del-IGHV UM	bi del IGHV M	bi del IGHV UM	No 13q deletion
mono del-IGHV mutated	NA	<0.0001	0.451	0.001	<0.0001
mono del-IGHV UM	<0.0001	NA	0.002	0.038	0.715
bi del IGHV M	0.451	0.002	NA	0.005	0.001
bi del IGHV UM	0.001	0.038	0.005	NA	0.060
No 13q deletion	<0.0001	0.715	0.001	0.060	NA

Figure 26. Kaplan Meier survival analysis of 13q deletion stratified according to mono-/bi-allelic status intergrated with IGHV status (un/mutated). Log rank statistical test was used to determine significance. Biallelic with mutated IGHV genes patients had a significantly increased survival than mono deleted-IGHV mutated patients. Whilst carrying IGHV unmutated genes regardless of deletion status did not have the benign disease course and actually gave a significantly reduced survival.

A



B

	PFS				OS			
Genomic abnormality	total number	Total events	median	95% CI	total number	Total events	median	95% CI
bi-IGHV mut	24	17	31	10.6-51.4	24	6	212.008	182.8-241.3
sole 13q del-mono	102	94	32	25.6-38.4	102	67	111.113	92.4-129.9
no CNA	86	83	27	21.8-32.2	86	64	102.341	78.8-125.9
13q with other CNA	70	67	17	12.9-21.1	70	60	80.131	51.6-108.6
bi-IGHV UM	4	4	9	0.0-20.8	4	4	76.452	33.2-119.7
no 13q del other CNA	95	91	19	15.0-23.0	95	81	54.899	35.4-74.4

	no CNA	bi-IGHV UM	bi-IGHV mut	no 13q del other CNA	13q with other CNA	sole 13q del-mono
no CNA	NA	0.01	0.01	0.088	0.008	0.364
bi-IGHV UM	0.01	NA	0.005	0.139	0.242	0.002
bi-IGHV mut	0.01	0.005	NA	0.002	0.001	0.047
no 13q del other CNA	0.088	0.139	0.002	NA	0.325	0.009
13q with other CNA	0.008	0.242	0.001	0.325	NA	<0.0001
sole 13q del-mono	0.364	0.002	0.047	0.009	<0.0001	NA

	no CNA	bi-IGHV UM	bi-IGHV mut	no 13q del other CNA	13q with other CNA	sole 13q del-mono
no CNA		0.097	<0.0001	0.002	0.013	0.135
bi-IGHV UM	0.097		<0.0001	0.734	0.474	0.047
bi-IGHV mut	<0.0001	<0.0001		<0.0001	<0.0001	<0.0001
no 13q del other CNA	0.002	0.734	<0.0001		0.722	<0.0001
13q with other CNA	0.013	0.474	<0.0001	0.722		<0.0001
sole 13q del-mono	0.135	0.047	<0.0001	<0.0001	<0.0001	

Figure 27. Survival analysis of 13q deleted patients. Patients whom carry a concurrent abnormality along with a 13q deletion do not show a survival advantage as those where 13q deletion is the only genomic lesion. A) Stratified by 13q deletion type, IGHV status and presence of concurrent CNA. B) & C) Univariate survival analysis and log rank statistical test for significance

4.4.10 Univariate cox regression analysis: re- stratifying genomic abnormalities

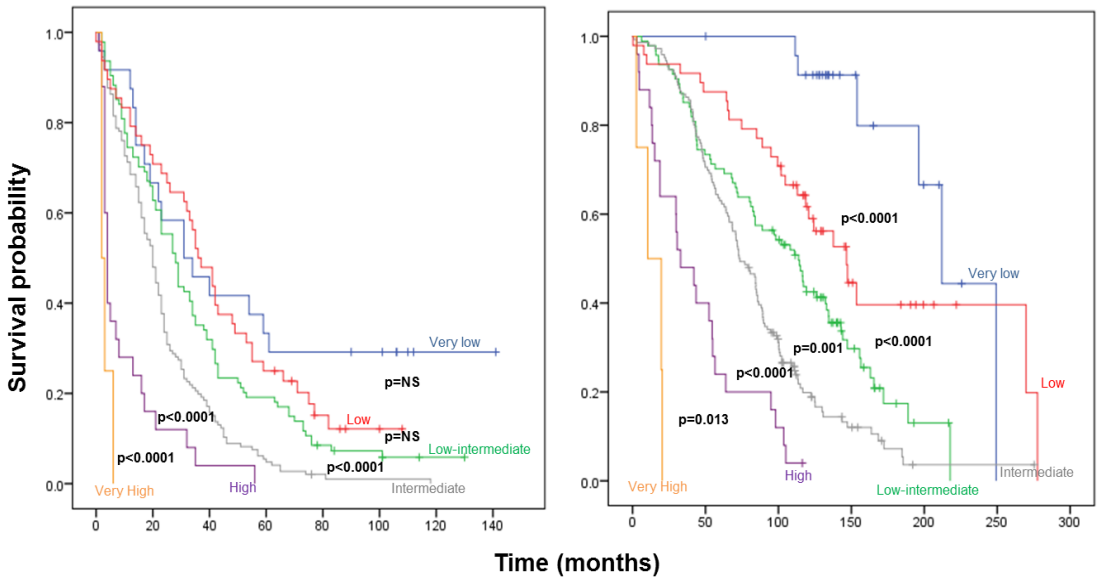
To further assess the impact of the CNAs and mutations, univariate Cox regression analysis was conducted to predict the risk of death, by assigning a hazard ratio (HR). Table 18 summarises the statistically significant univariate data, with the inclusion of the hazard ratios; 13q deletion was deemed the most benign CNA, with a low HR for OS (0.67), whilst 9p deletion presented the highest HR for OS of 22.75. 19p deletions had a similar HR to 17p deletions; 3.13 vs 3.48 (PFS) and 3.03 vs 3.91 (OS), respectively. However, this is not representative of the survival data already discussed; integration of the aberrations that co-occur allows for better characterisation and stratification of patients' risk. Table 17 summarises the integrated approach to univariate analysis, including the survival data, hazard ratios obtained by Cox regression analysis and log rank statistical test for significance. In this case the most benign cases become the biallelic 13q deleted-IGHV mutated patients (HR 0.17), followed by trisomy 19, *MYD88* mutated patients and then monoallelic 13q-*IGHV* mutated (HR 0.59) cases, although it is only the former and latter which are statistically significant.

Table 18. Intergrated approach to stratifying patient survival. Univariate survival analysis (Cox regression) PFS and OS of the statistically significant copy number alterations and gene mutations detected by MLPA. Aberrations were compared to WT populations. Biallelic 13q deletion with IGHV mutated status gave the most benign/low risk disease course, whilst 9p deleted patients predicted the poorest/high risk disease for patients with this abnormality.

	Progression-free survival (months)						Overall survival (months)					
Genomic abnormality	Events/Total	median	95% CI	HR	95% CI	p value	events/total	median	95% CI	HR	95% CI	p value
Bi-13q-IGHVM	17/24	31	11-51	0.47	0.29-0.77	0.003	6/24	212	183-241	0.17	0.07-0.37	<0.0001
Trisomy 19	4/5	10	1-19	0.73	0.27-1.96	0.525	3/5	166		0.73	0.24-2.29	0.592
MYD88	6/6	14	0-32	1.02	0.46-2.30	0.961	3/6	157	128-185	0.43	0.14-1.34	0.134
mono-13q-IGHVM	46/55	36	29-43	0.47	0.32-0.70	<0.0001	30/55	147	120-174	0.59	0.43-0.81	0.001
No CNA	43/44	27	18-36	1.12	0.81-1.54	0.493	32/44	114	98-130	1.16	0.80-1.68	0.426
13q deletion	182/200	23	20-26	0.87	0.71-1.07	0.189	137/200	100	88-113	0.67	0.53-0.85	0.001
sole SF3B1	18/20	29	22-36	0.78	0.49-1.26	0.385	13/20	84	0-213	0.83	0.48-1.45	0.517
14q deletion	4/6	22	0-50	0.466	0.17-1.25	0.115	4/6	80	23-137	1.043	0.39-2.80	0.934
mono-13q-IGHVUM	44/44	22	18-26	1.30	0.92-1.85	0.143	37/44	79	67-100	1.13	0.82-1.55	0.461
Bi-13q-IGHVUM	4/4	9	0-21	2.73	1.02-7.36	0.047	4/4	77	33-120	1.81	0.67-4.87	0.241
SF3B1	48/49	29	24-34	0.90	0.67-1.23	0.503	39/51	72	45-100	1.28	0.91-1.79	0.158
11q deletion	62/62	17	13-21	1.73	1.31-2.28	<0.0001	52/62	72	52-91	1.50	1.10-2.02	0.009
Trisomy 12	48/52	22	15-29	0.96	0.71-1.30	0.792	43/52	71	40-102	1.45	1.04-2.00	0.026
NOTCH1	48/51	19	15-23	1.32	0.97-1.79	0.068	45/49	68	47-90	2.03	1.46-2.80	<0.0001
6q deletion	20/21	18	9-27	1.62	1.03-2.56	0.031	20/21	66	12-120	1.74	1.10-2.75	0.016
NOTCH1-trisomy12	16/16	16	7-25	1.24	0.75-2.05	0.273	15/16	48	36-61	2.77	1.63-4.72	<0.0001
2p gain	14/14	10	1-19	1.52	0.89-2.60	0.119	12/14	46	3-90	2.25	1.29-3.94	0.004
8q gain	8/8	7	0-22	2.77	1.37-5.63	0.003	8/8	42	0-120	2.28	1.12-4.62	0.019
8p deletion	5/5	6	2-10	2.44	1.00-5.90	0.039	5/5	42	0-87	5.55	2.25-13.66	<0.0001
17p deletion	27/27	4	3-5	3.48	2.34-5.20	<0.0001	26/27	31	25-36	3.91	2.59-5.91	<0.0001
19p deletion	4/4	9	1-17	3.13	1.16-8.47	0.016	4/4	20	0-66	3.03	1.13-8.16	0.021
9p deletion	4/4	2		10.24	3.70-28.33	<0.0001	4/4	10	0-27	22.75	7.91-65.38	<0.0001

4.4.11 Risk stratifying patients using an integrated genomic hierarchical model

Patients were then classified into 6 hierarchical groups according to their univariate survival data, to risk stratify patients from low to high risk; 1) Very low risk; the *IGHV* mutated cases with biallelic del13q; 2) Low risk; the patients with monoallelic del 13q and *IGHV* mutated; 3) Low-intermediate risk; patients with no genomic abnormalities (by MLPA), *SF3B1* mutations as a sole abnormality and trisomy 12; 4) Intermediate risk; cases with del11q, del6q, del13q (unmutated *IGHV*) and *NOTCH1* mutated; 5) High risk; del17p; 6) Very high risk; patients with a 9p deletion. Figure 28 shows the Kaplan Meier and survival data for this model of risk prediction. PFS for groups 1 and 2 was not significantly different (31 vs 36 months, respectively), whereas all remaining groups were statistically different to each other. For OS, all groupings were significantly ($p<0.0001$) different from the subsequent stratification.



		PFS				OS			
Risk group	Genomic abnormality	total number	Total events	median	95% CI	total number	Total events	median	95% CI
Very low	biallelic-IGHV mutated 13q del	24	17	31	10.6-51.4	24	6	212.008	182.8-241.3
Low	Mono allelic 13q deletion –IGHV mutated	48	41	36	28.1-43.9	48	26	146.7	114.7-178.7
Low-intermediate	no CNA or mutations /sole*SF3B1* mutation/trisomy 12	94	88	28	22.8-33.2	94	68	113.8	94.7-132.9
Intermediate	11q 6q/13q deletion-IGHV unmutated /*NOTCH1* mutation	146	145	20	17.0-23.0	146	125	73.0	62.9-83.1
High	17p deletion	25	25	4	3.0-5.0	25	24	32.953	13.1-52.9
Very high	9p deletion	4	4	2		4	4	10.382	0.0-27.1

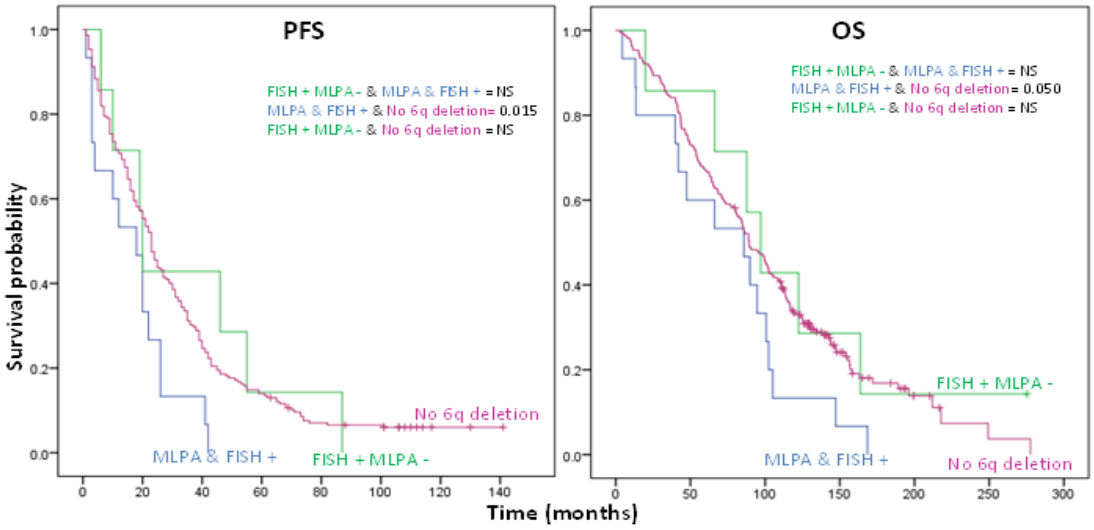
Figure 28. Integrated mutational and CNA abnormality model using a hierarchical approach of risk stratification, better stratifies patients into very low to very high risk groups. This model includes 345/384 patients. The 95% CI is missing for the very high risk group as two patients progressed to treatment as soon as they were enrolled onto trial.

4.4.12 **Low clone size still impacts upon survival**

To assess the utility of MLPA in a clinical setting the impact of genomic abnormalities present in a small sub-clonal population on survival was investigated, as these CNA are likely to be missed by MLPA. Patients were stratified according to MLPA status; MLPA positive/FISH positive (large clone size) vs MLPA negative/FISH positive (small clone size), and survival analysis was conducted. The clone size for deletions of 6q, 11q and 13q and trisomy 12 did not affect PFS or OS, whereas large clone was associated with a reduced survival in del17p cases (Figure 29, Figure 30, Figure 31, Figure 32, Figure 34)

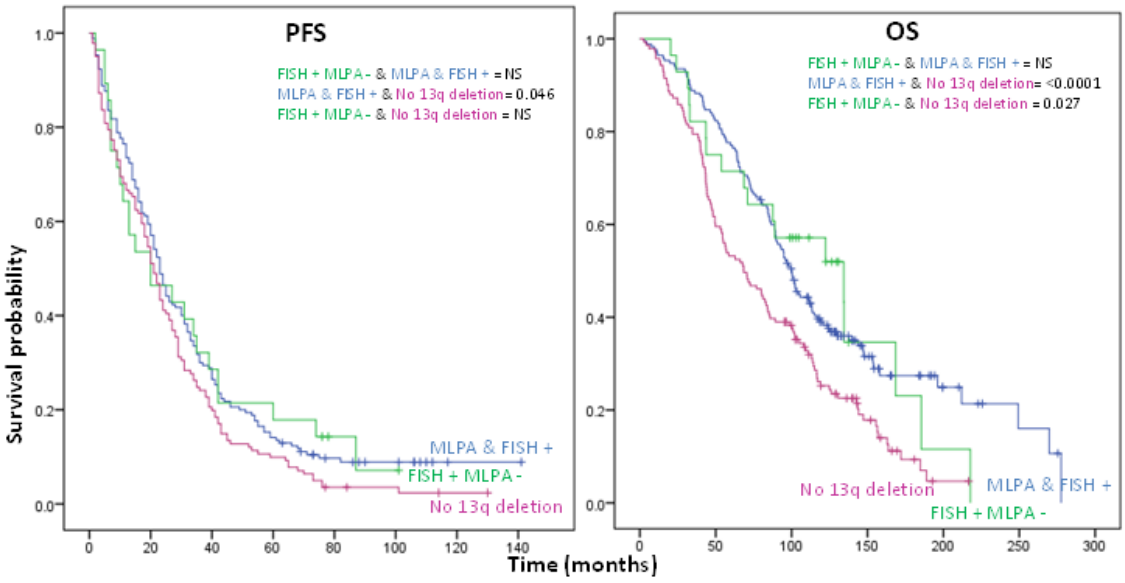
It is hard to determine the true nature of a low clone size upon survival as they are not significantly different to the respective large clone size (MLPA positive) but they are also not significantly different from the wild type population (except for 13q deletion), suggesting an intermediate effect. Conversely it was only 17p deletions that had a clone size below 30%, which MLPA cannot accurately detect that does not impact upon OS significantly and has a similar OS to the none 17p deleted patients.

The 17p deletion clone size, based on FISH data, was sub-divided into three groups; <10%, >10% but <20%, and >20%. Only the deletions in >20% of cells carried a significant risk for both OS and PFS (Figure 34). To account for cofounders such as *IGHV* status or *TP53* mutations, a chi square statistical test was conducted between the groups. The majority of *TP53* mutations were found in the >20% group (n=18), with only 2 patients found in the <20% group (p=0.003). In the <10% group, the *IGHV* mutational status was equally represented (52.4% unmutated, 47.6% mutated), whereas the proportion of unmutated *IGHV* cases rises to 73.9% in the >20% group. Although this was not significant, it demonstrates that patients with larger clone sizes exhibit poor risk features, and that the low clones might be noise, given the strong link between *TP53* and I*GHV* status.



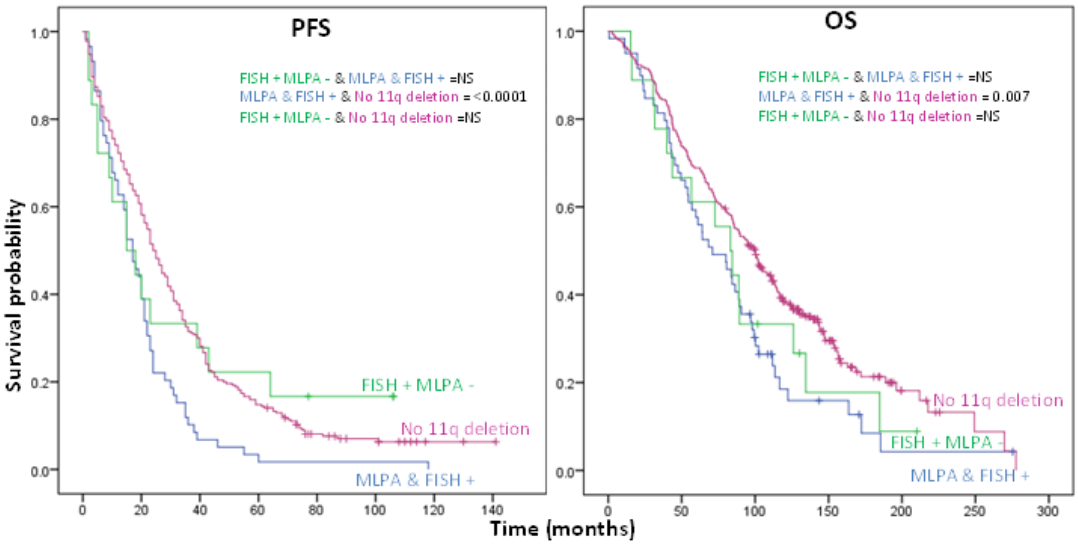
Genomic abnormality	PFS total number	Total events	median	95% CI	OS total number	Total events	median	95% CI
MLPA and FISH positive	15	15	18	8.5-27.5	15	15	85.9	32.5-139.2
FISH pos, MLPA neg	7	7	20	17.4-22.6	7	6	96.9	73.2-120.6
No 6q deletion	215	201	23	20.1-25.9	215	169	89.3	78.6-100.0

Figure 29. Survival analysis for 6q deletion as determined by clone size.



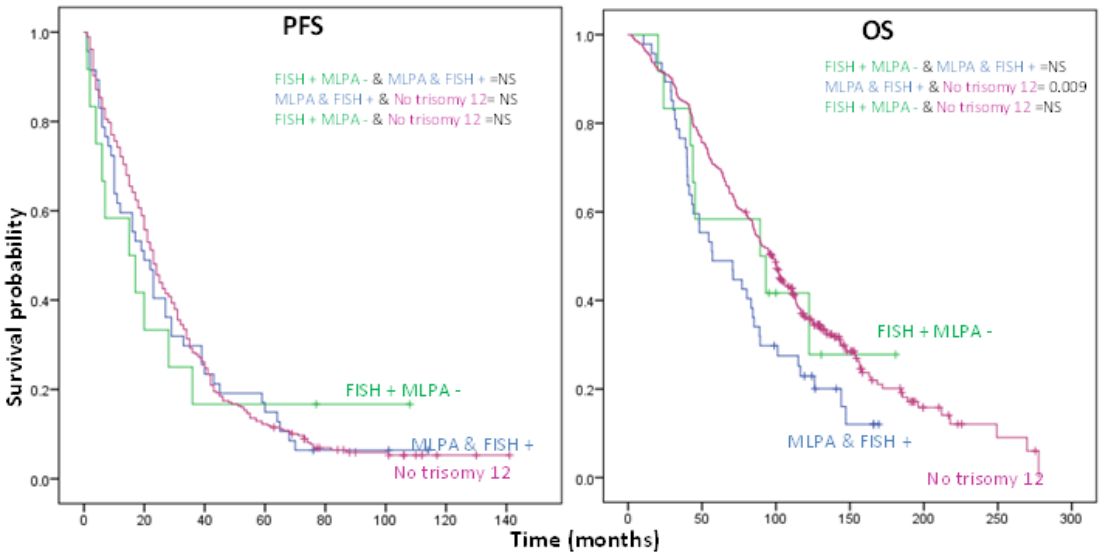
Genomic abnormality	PFS total number	Total events	median	95% CI	OS total number	Total events	median	95% CI
MLPA and FISH positive	170	154	23	20.1-25.9	170	118	99.9	90.9-108.9
FISH pos, MLPA neg	28	25	20	1.9-38.1	28	18	134.4	61.1-207.7
No 13q deletion	141	137	21	17.9-24.1	141	118	68.3	49.8-86.9

Figure 30. Survival analysis for 13q deletions as determined by clone size



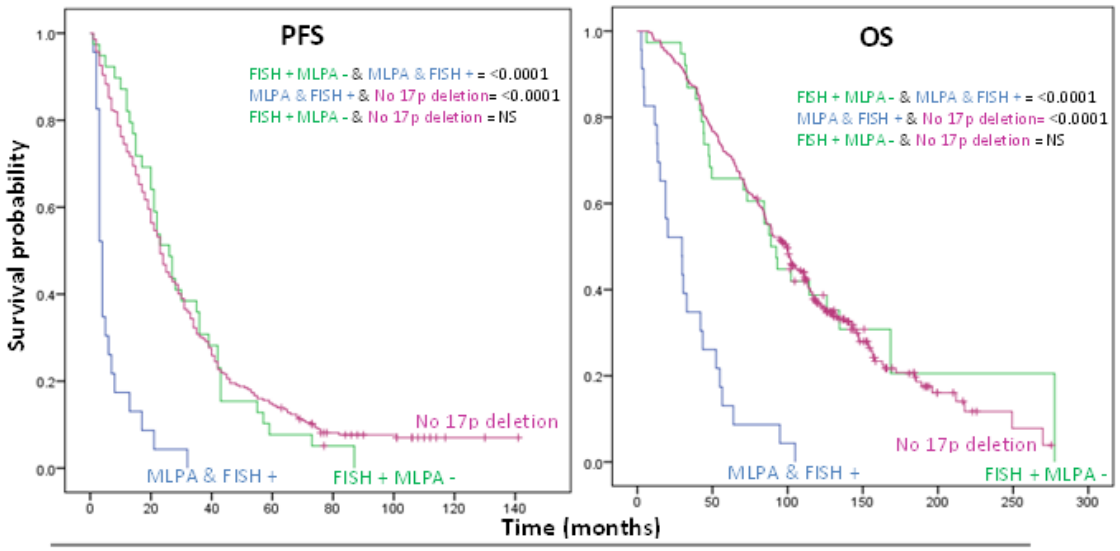
	PFS				OS			
Genomic abnormality	total number	Total events	median	95% CI	total number	Total events	median	95% CI
MLPA and FISH positive	59	59	17	12.8-21.2	59	50	70.9	46.6-95.3
FISH pos, MLPA neg	19	16	15	3.6-26.4	19	16	83.1	63.8-102.5
No 11q deletion	270	250	24	20.9-27.1	270	194	99.9	87.1-97.8

Figure 31. Survival analysis for 11q deletions as determined by clone size



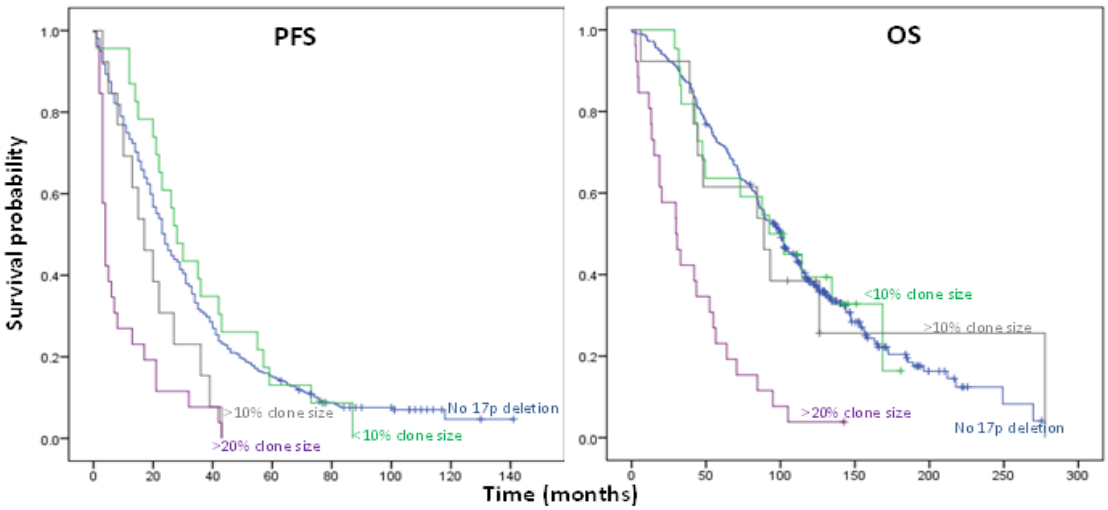
Genomic abnormality	PFS total number	PFS Total events	PFS median	PFS 95% CI	OS total number	OS Total events	OS median	OS 95% CI
MLPA and FISH positive	47	44	20	13.3-26.7	47	39	57	26.5-87.5
FISH pos, MLPA neg	12	10	15	0.0-32.0	12	8	41.4	8.1-170.5
No trisomy 12	287	269	23	20.6-25.4	287	259	98.1	88.6-107.6

Figure 32. Survival analysis for Trisomy 12 as determined by clone size



Genomic abnormality	PFS total number	PFS Total events	PFS median	PFS 95% CI	OS total number	OS Total events	OS median	OS 95% CI
MLPA and FISH positive	23	23	4	3.2-4.8	23	23	29.8	12.5-47.0
FISH pos, MLPA neg	39	38	26	19.9-32.1	38	37	88.9	75.8-102.0
No 17p deletion	283	260	23	20.2-25.8	283	207	99.5	88.6-110.4

Figure 33 17p deletion survival. MLPA positive (large clone size) plays an important role in reduced survival compared to MLPA negative (low clone size)



Genomic abnormality	PFS total number	Total events	median	95% CI	OS total number	Total events	median	95% CI
no 17p del	319	295	24	21-27	320	232	100	90-111
<10% clone	23	22	28	22-34	22	15	92	49-136
>10% clone	13	13	17	9-25	13	10	89	36-142
>20%	26	26	4	3-5	26	25	30	14-46

	no 17p del	<10% clone	>10% clone	>20%
no 17p del	NA	0.586	0.052	<0.0001
<10% clone	0.586	NA	0.02	<0.0001
>10% clone	0.052	0.02	NA	0.058
>20%	<0.0001	<0.0001	0.058	NA

	no 17p del	<10% clone	>10% clone	>20%
no 17p del	NA	0.986	0.993	<0.0001
<10% clone	0.986	NA	0.802	<0.0001
>10% clone	0.993	0.802	NA	0.006
>20%	<0.0001	<0.0001	0.006	NA

Figure 34. Deletion of 17p status catergorised by clone size (as determined by FISH analysis), groups included; <10% clone size, >10% (but <20%) clone size and >20% clone size. <10% clone size had a similar survival to no 17p deleted patients.

Multivariate survival analysis

The impact of each deletion/mutation found by MLPA on OS and PFS was further assessed using multivariate Cox proportional hazard analysis, to control for the effects of any cofounding variables. Two multivariate models were created for the analysis; model 1 (based on 371 patients) included all aberrations that were statistically significant for PFS and/or OS (Table 19), to see which had the biggest independent impact. Deletions of 8p, 13q, 6q and 19p and trisomy 12 did not have a significant impact in model 1 for PFS or OS, and were therefore excluded from model 2, whilst all other abnormalities were included. In addition, several important clinical biomarkers were included in model 2; age, sex, Binet stage, *IGHV* status, *TP53* mutation status (from the CLL4 trial data) and it was based upon 279 patients (Table 20). From model 2, the following factors have a significant impact on PFS; 11q deletion, 9p deletion, biallelic 13q deletion with mutated and unmutated *IGHV* as a sole abnormality, treatment, 17p deletion, *TP53* mutation and 2p gain. The factors significantly impacting on OS were; IGHV status, 9p deletion, biallelic 13q deletion with mutated and unmutated *IGHV* as a sole abnormality, *NOTCH1* mutations, age, stage, 17p deletion and 2p gain.

Table 19. Multivariante model 1.
All the CNAs and mutations detected by MLPA and found to have a significant impact upon PFS and OS in a univariate analysis were included in this model. All those that stayed (as significant) in model 1 were then taken into model 2.

PFS					OS			
Variable	p value	HR	95% CI		P value	HR	95% CI	
			Lower	Upper			Lower	Upper
11q deletion	**0.001**	1.728	1.241	2.406	**0.027**	1.519	1.048	2.2
MLPAdel8p	0.931	0.949	0.294	3.068	0.856	1.137	0.284	4.55
9p deletion	**0.009**	6.517	1.584	26.821	**<0.0001**	22.06	4.374	111.25
13q-IGHV-M	**0.047**	0.64	0.413	0.994	**0.032**	0.565	0.335	0.952
13q-IGHV-UM	0.289	1.27	0.817	1.974	**0.021**	1.769	1.091	2.87
bi13q-IGHV-UM	**0.019**	3.447	1.221	9.737	**0.051**	2.841	0.994	8.122
bi13q-IGHV-M	**0.02**	0.494	0.273	0.895	**<0.0001**	0.184	0.071	0.476
NOTCH1 mutation	**0.036**	1.408	1.023	1.938	**0.001**	1.788	1.27	2.516
Trisomy 12	0.496	0.889	0.632	1.248	**0.094**	1.373	0.948	1.989
13q deletion	0.845	1.032	0.752	1.417	0.52	0.894	0.635	1.258
8q gain	**0.035**	2.224	1.059	4.672	0.059	2.102	0.973	4.541
17p deletion	**<0.0001**	3.27	1.991	5.37	**<0.0001**	3.72	2.192	6.313
MLPAdel19p	0.836	0.868	0.226	3.338	0.241	0.457	0.124	1.69
2pgain	0.414	1.265	0.719	2.225	**0.018**	2.011	1.125	3.596
6q deletion	0.203	1.366	0.846	2.207	0.539	1.171	0.707	1.938

Table 20. Multivariate model 2. Statistically significant MLPA genomic abnormalities from model 1 and important clinical CLL biomarkers were included in this model. Interestingly, even in the presence of such strong predictors of survival (such as IGHV status and *TP53* abnormalities) biallelic 13q-IGHV M and 9p deletion along with 2p gains stayed in this model as significant independent predictors of survival (for both PFS and OS).

PFS					**OS**			
Variable	P value	HR	95% CI		P value	HR	95% CI	
			Lower	Upper			Lower	Upper
11q deletion	**0.001**	1.874	1.303	2.693	0.072	1.435	0.968	2.126
IGHV status	0.128	0.75	0.518	1.086	**<0.0001**	0.427	0.272	0.668
9p deletion	**<0.0001**	9.117	2.633	31.574	**<0.0001**	29.069	8	105.626
13q-IGHV-M	0.247	0.749	0.459	1.222	0.754	0.905	0.486	1.687
13q-IGHV-UM	0.519	1.139	0.767	1.691	0.333	1.234	0.806	1.89
bi13q-IGHV-UM	**0.033**	3.18	1.099	9.201	**0.035**	3.094	1.084	8.828
bi13q-IGHV-M	**0.025**	0.455	0.229	0.905	**0.011**	0.245	0.083	0.725
NOTCH1 mutation	0.444	1.156	0.797	1.678	**0.011**	1.671	1.124	2.484
Age	0.701	0.997	0.982	1.012	**<0.0001**	1.042	1.024	1.06
Treatment	**<0.0001**	0.606	0.468	0.786	0.603	0.926	0.692	1.238
Sex	0.248	1.195	0.884	1.616	0.386	1.172	0.818	1.678
Stage	0.713	1.062	0.77	1.465	**0.014**	1.597	1.098	2.323
8q gain	0.095	2.28	0.867	5.998	0.216	1.794	0.71	4.532
17p deletion	**<0.0001**	3.449	1.819	6.537	**<0.0001**	6.021	2.772	13.076
TP53 mutation	**0.024**	2.038	1.097	3.787	0.7	1.157	0.551	2.43
2pgain	**<0.0001**	3.554	1.831	6.897	**0.023**	2.173	1.115	4.235

4.4.13 **Telomere length (TL) is affected by genomic aberrations**

A proportion of the MLPA patients (n=287) had telomere length data, determined by MMQPCR (Figure 35). This data was compared to the genomic MLPA data to see if any correlations could be made.

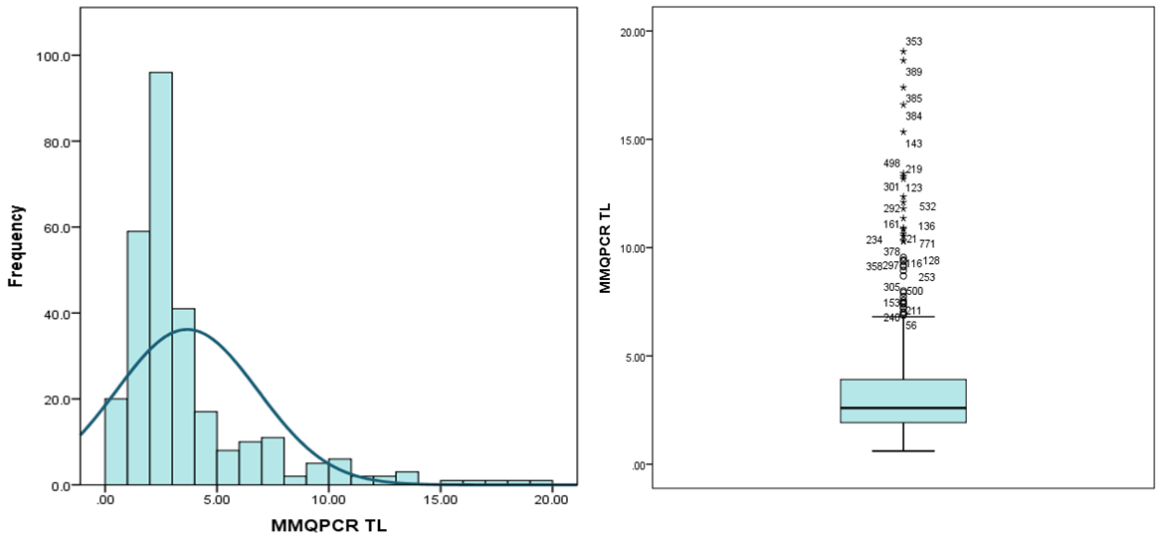


Figure 35. Data distribution for telomere length assessed by MMQPCR within the MLPA cohort (n=287). The mean, median and range of telomeric length was 3.67, 2.60 and 0.61-19.05 respectively and the 25th, 50th and 75th percentiles are 1.92, 2.60, 3.92, respectively

Patients with mutated *IGHV* genes (n=100) had significantly longer telomeres than those with unmutated IGHV genes (n=164) (p<0.001). Patients positive for Zap-70 and CD38 expression, clinical markers of *IGHV* status, also had significantly shorter telomeres (p<0.001 and p=0.001 respectively) (Figure 36).

Both 11q (n=48) and 17p (n=17) deleted patients had significantly shorter TL than WT patients (p<0.001), whilst they were significantly longer in patients with 13q deletions (n=155) (p=0.02). Biallelic and monoallelic 13q deletions did not differ significantly. Trisomy 12 (n=34) did not impact upon TL, however, in the presence of a concurrent trisomy 19 (n=3), telomeres were significantly longer (p=0.010) (Figure 37). Patients with 2p gains (n=11), 8q gains (n=7) and 6q deletions (n=16) have significantly shorter telomeres than those WT for these CNA (p=0.009, <0.001, and 0.02, respectively). Although there is a trend towards shorter telomeres in patients with a 9p deletion (n=3), this is not significant. Patients without any CNAs or mutations as detected by MLPA (n=33), have a trend towards an increased TL but this is not significant. *SF3B1* mutated patients (n=36) had a significantly reduced TL when compared to WT cases (p=0.018), whereas telomeres were significantly longer in *MYD88* mutated patients (n=6) (p=0.027). *NOTCH1* mutations (n=40) did not affect TL (Figure 37)

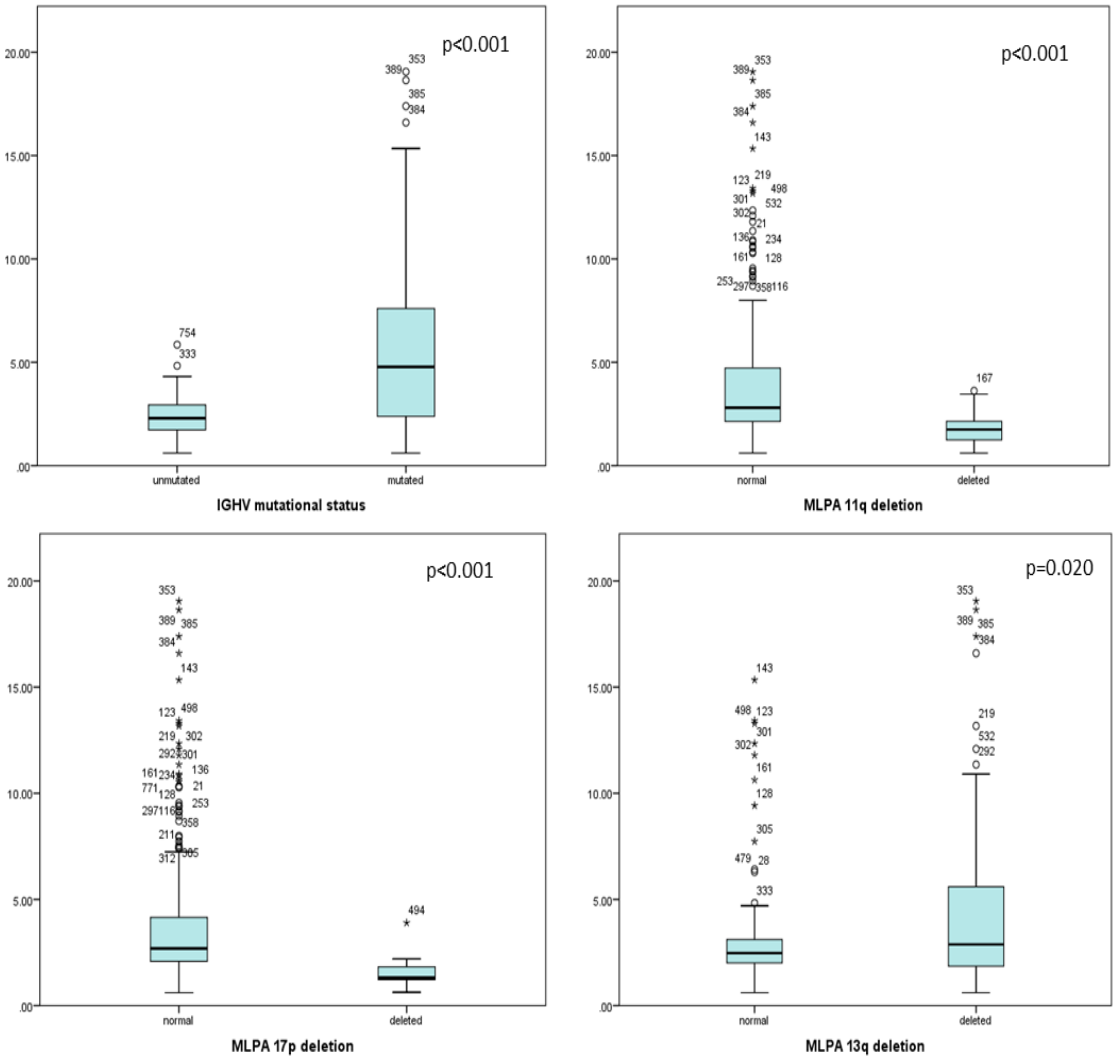


Figure 36. The clinically important genomic abnormalities; IGHV gene status, 17p deletion, 11q deletion and 13q deletion all impact significantly on telomere lengths

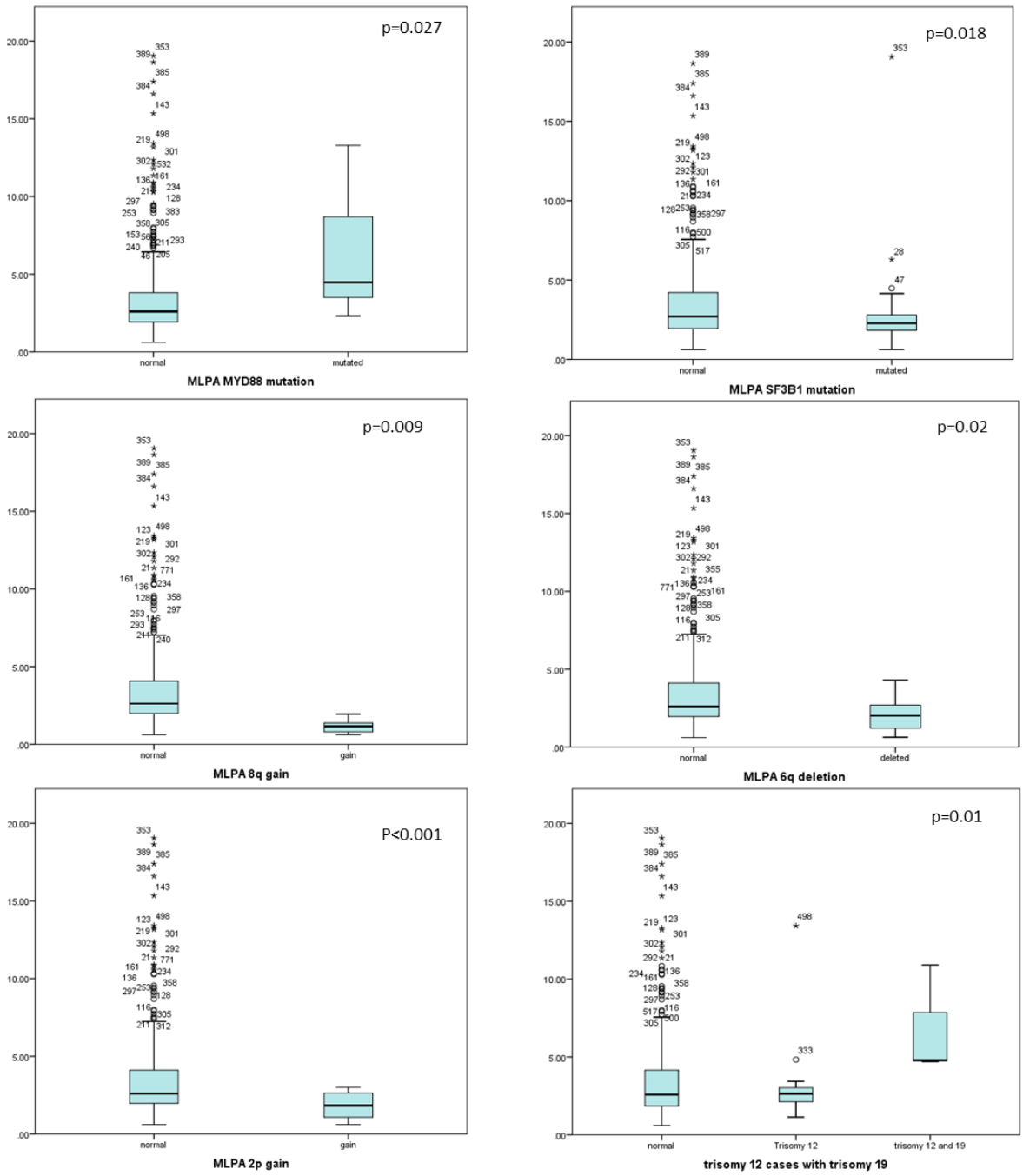


Figure 37. Mutations in *SF3B1* and *MYD88*, and the rare chromosomal defects; 2p and 8q gain, 6q deletion and trisomy 12/trisomy 19, all were shown to impact upon TL in CLL patients.

4.5 Discussion

In this study multiplex ligation dependant probe amplification (MLPA) was used to detect copy number alterations and hotspot mutations in a cohort of 385 CLL4 trial patients that is representative of the full CLL4 trial cohort. This is the first example of the use of MLPA in a clinical trial cohort and will help the community better understand the utility of MLPA in both a clinical and research setting. Matching FISH data for 11q, 13q, 17p and 6q deletions and trisomy 12 was available in 348 of the cases, and Sanger sequencing data was available for *NOTCH1*, *SF3B1* and *MYD88* hotspot mutations in 318, 256 and 142 patients, respectively. As FISH is the 'gold standard' methodology for detecting CNA, the MLPA data was compared to the FISH data using statistical analysis. At first glance the analysis showed good concordance but a large proportion of the aberrations were missed by MLPA if the aberrant cell population was below 30%, the level of sensitivity proposed by the manufacturers of the MLPA technology (MRC Holland). This was especially apparent in the 17p deleted cases, were the majority of the deleted cases had a clone size well below 30%. Published studies in CLL have also compared MLPA to FISH and most found a similar cut off to the manufacturers, although some stated they could detect aberrant populations down to 10 [83,88–93]. However, these studies have included relatively few cases with *TP53* and/or *ATM*/11q loss as detected by FISH. MLPA identified CNA in 6% of cells for 11q deletions and down to 26% for del17p. For 13q deletions five patients were below this level of detection (8-19%). Therefore, using the suggested 30% cut off limit for reliable MLPA detection gives 100% concordance to FISH for 6q, 11q and 13q deletions and trisomy 12, whilst for 17p deletions there is still 96% concordance, which is similar to the literature. It has been suggested that *TP53* mutations can disrupt the binding of the MLPA probes and so may give a discordant result for deletions [88]. Analysis of the extensive CLL4 clinical trial data, which includes *TP53* mutation status showed this was not the case for this study. The detection of clone sizes well below the expected cut-off could be attributable to the pre-enrichment of some samples, or alternatively, the true clone sizes may have been larger as FISH is a highly subjective method of estimation. The high probe coverage of the 13q and 11q deleted regions could also result in more sensitive detection of low clone populations. Pre-enrichment or incorrectly called clone sizes could also explain the cases where larger clone sizes were missed. In addition, samples may not have been optimally prepared for MLPA, e.g. diluting DNA in EB buffer and removing salts/ethanol/phenol/chloroform contaminants, as these can cause false positive results. The sensitivity of MLPA at detecting low level clones is the major limitation to using this technology in the clinic, although it has been suggested that a pre enrichment for CD19 positive cells and altering each individual probes cut off values can increase the sensitivity of MLPA down to a 10% clone size[88]. Recent NGS studies have proved the relevance of low level *TP53* mutation clones that still have an impact upon progression and survival of a CLL patient [94]. This is especially

important in respect to therapy choice as chemo toxic regimes may be detrimental to *TP53* abnormal patients. Therefore, MLPA may not be a suitable method for detecting 17p deletions. Rather than a universal cut off, individual probe cut-offs could improve the accuracy and increase the number of CNA identified from our data, such as micro deletions and areas excluded from analysis such as 10q (*PTEN*). This improvement would eliminate false positives and make single probes easier to interpret and assign a relative copy number to, as currently all single probe copy number changes are discounted. For example, a micro deletion on 19p was identified in the archival SNP6.0 data but was only retrospectively called by MLPA because it is a single probe deletion (*CCND1* probe).

The comparison of MLPA to FISH data allowed for further analysis in regards to patient survival, which gave a novel insight into CLL stratification. To assess the impact of a CNA present in a small sub-clonal population, patients were stratified according to their MLPA status; MLPA negative but FISH positive (small sub-clone) or MLPA and FISH positive (large clone). The hypothesis being that a small sub-clone would not impact upon survival, and therefore MLPA could be implemented as a clinical diagnostic technique. There was no significant difference to OS or PFS between large and small clones for 6q, 11q and 13q deletions or Trisomy 12. The exception to this was 17p deletions; the patients which were negative by MLPA had a better OS and PFS, comparable to 17p WT patients. When investigated further, it was discovered that the poorer survival associated with 17p deletion was only captured in patients with a clone size > 20%, and that this data was confounded by the poor prognostic marker, *IGHV*-UM. This is in contrast to current published data, which show subclonal *TP53* aberrations impact upon survival [94].

The mutations we detected using MLPA were compared to Sanger sequencing data, with very high concordance. This may be due to the different probe chemistry for detecting mutations compared to CNA. Whilst CNA probes bind and give a relative signal reduction for loss or increase for gain, probes for the detection of mutations can only bind if the exact mutation is present. From the mutational data, the frequency that we identify *SF3B1* and *NOTCH1* mutants does seem to be a lot higher than would be expected from the literature. For example, the *SF3B1* codon 700 mutation that the MLPA kit detects normally accounts for approximately 50% of all *SF3B1* mutations, where as we have found a frequency of 13.2% just of this variant. A similar frequency was found for the *NOTCH1* delta CT mutation, although this variant does account for over 90% of the *NOTCH1* mutations it still suggests that we are picking up a larger than expected number of mutants. This may mean that some false positive results were detected, via contamination from other samples, or that MLPA may be more sensitive than Sanger sequencing at detecting at mutations in smaller clones.

MLPA has been assessed for its clinical utility and has been shown to be effective at detecting rare and established CNA in CLL, as well as clinically relevant mutations. Despite the fact that a small proportion of sub-clonal CNA with clinical relevance may be missed, the low cost and ease at which MLPA can be performed, and the large amount of combined copy number and mutational data generated in a single reaction still makes MLPA an attractive alternative to standard working practices. MLPA could be used as a diagnostic test to screen for genomic abnormalities along with the standard 'watch and wait' approach to patient care, until disease progression, in which case a more sensitive method of CNA can be implemented, such as FISH prior to treatment therapy. This approach could work in the clinic to give the maximum amount of information in a short amount of time, as well as being a cheaper and easier option to exploring genomic complexity and mutations in CLL at all stages of disease.

The 13q deletion as a sole abnormality has long been used as a clinical marker for good prognosis, although some of these patients do ultimately progress and/or die. The data generated in this study shows 13q deletions had a significant impact upon survival, with an increased OS compared to the 13q WT patients, in line with the literature[53]. Further stratification of the data by monoallelic or biallelic loss, and interestingly observed increased survival in the biallelically deleted cases. The literature on this subject is currently inconclusive, as some studies suggest that deletion size or more importantly, the inclusion of the *RB1* gene may impact survival, whilst others suggest that the clone size is crucial [64,65]. The data presented shows that inclusion of the *RB1* gene does not impact on the patient survival significantly, although the trend does suggest a slightly worse OS compared to a small 13q deletion. However, large deletions were significantly associated with 11q deletions, which are linked to poorer survival. In addition, a significant association between del13q clone size and survival was not found. Conversely there is an inclination towards a better survival for those with a clone size above 80%.

Once the deletion size and clone size were excluded, 13q biallelic deletions were looked at in more detail. Only patients with 13q deletion as their sole abnormality had a survival advantage and were in fact similar to patients with no chromosomal changes, which is reflected in the literature[66]. Patients with biallelic del13q as a sole abnormality were then stratified by mutated and unmutated *IGHV* status. This allowed further insight into survival, showing that only those with mutated *IGHV* genes had better survival; the unmutated patients had a similar survival to patients without a 13q deletion. This data is novel and suggests that patients with mutated *IGHV* genes, harbouring a biallelic del13q and no other chromosomal abnormalities, do in fact have a much longer survival outcome. Most studies looking at 13q deletions in CLL have shown or speculated that the homozygous loss of 13q would lead to a worse survival, although most of these were small scale, non-clinical trial studies. However, in a study of 627 CLL patients with a 13q deletion, Puiggros *et al*, did not associate a poor outcome biallelic loss, although *IGHV* status

was not accounted for [58]. Another study recently associated an *IGHV* mutated, del13q genotype with increased survival, although they did not account for loss of the other allele [95]. A study of 203 newly diagnosed CLL patients, using SNP 6.0 arrays, showed a trend towards a better survival for biallelic 13q deletion compared to monoallelic, although this did not reach significance [96].

Due to the high number of probes (n=15) covering a 3587.2kb region, including probes for genes implicated in CLL (*RB1*, *DLEU1*, *DLEU2*, *DLEU7* and *MIR15)*, chromosomal profiles were created for the 13q14 loci. Deletion size varies from patient to patient, but a minimally deleted region (MDR) seen in all patients was identified. For the monoallelic 13q deletions a 630.8kb MDR was defined between *DLEU2* and *DLEU7*, encompassing the *DLEU1* gene. The biallelic MDR was much smaller (22.6Kb), located between two probes encompassing part of the *DLEU1* gene. These MDRs are in line with those identified in the CLL8 clinical trial, using SNP 6.0 arrays; the monoallelic MDR included *DLEU1* and the first two exons of *DLEU2* and the biallelic MDR included a partial deletion of *DLEU1*[54]. This study confirms the validity of MLPA as a reliable and comparable technology to more sensitive platforms such as SNP 6.0 arrays. Archived SNP 6.0 data was available to compare the biallelic 13q deletion in the CLL4 trial; this confirmed that the biallelic MDR targeted the *DLEU1* gene. The MDR from our study did not contain *MIR15a* and *MIR16-1,* microRNAs that have been implicated as important in CLL pathogenesis, although the majority of the patients with del 13q did include these genes [62,97].

In a study looking at the interplay between microRNAs and the *TP53* pathway, Fabbri *et al* identified *TP53* binding sites upstream of the *miR15a/16-1* loci, thought to control the regulation of downstream genes, and binding sites for *miR15a* and *miR16-1* were also discovered in the 3' UTR of *TP53*, suggesting these microRNAs directly regulate *TP53* expression [98]. The biallelic 13q deleted patients in their study (n=22) showed higher levels of *TP53* expression than heterozygous patients and both sets of patients had a significantly reduced miR15/16-1 expression compared to cytogenetically normal patients, at both the mRNA and protein level. In cell line experiments the overexpression of miR15a and miR16-1 reduced TP53 protein expression by 82 as well as TP53 downstream effectors such as CDKN1A, BBC3 and BCL2. The re-expression of miRs 15a/16b in biallelically 13q deleted patient B cells reduced TP53, miR34a/b/c and increased ZAP70 protein expression (Figure 38). It is interesting to note that one of the biallelically 13q deleted/*IGHV* mutated patients also has a *TP53* mutation (PID 4). Despite having this poor risk feature the patient has done remarkably well in terms of survival. This provides evidence that having the biallelic deletion with mutated *IGHV* genes may increase TP53 production, therefore if one allele is affected by mutation, the remaining allele is compensating through the upregulation of TP53 via removal of the miRs.

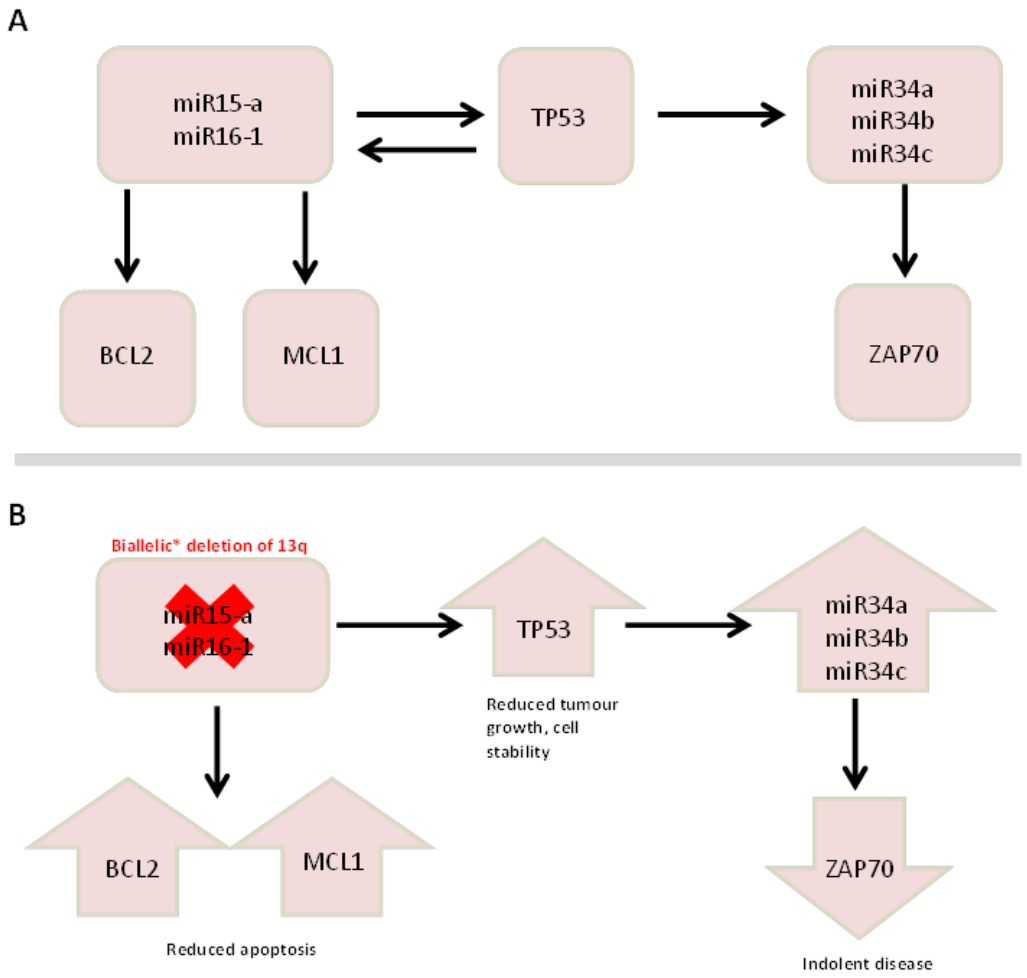


Figure 38. Proposed involvement of TP53 in 13q deleted patients with an indolent disease course.
A) normal feedback mechanisms of the microRNAs with TP53. The miR15-1/16-1 inhibit TP53 mRNA directly and so TP53 expression is reduced. TP53 also binds to these miRs which in turn reduces their effects on TP53 levels and so an equilibrium is reached via a feedback loop. TP53 controls the miRNAs 34a/b/c levels via a transactivating binding site which in turn can inhibit ZAP70 mRNA. B) The biological effects of biallelic 13q deletion using microRNA interactions with the TP53 network. The increased cell survival effects (reduced apoptosis) as seen from an increase in BCL2 and MCL1 levels are thought to be counteracted by the increased TP53 levels, which also lead to an increase of miR34a/b/c/ that in turn reduce ZAP70 levels. The end results in a benign disease course. This may also be seen in the mono allelic deletions to an extent and so the effect may be limited and so may not affect survival as seen in the biallelic cases * biallelic deletion includes the loss of miR15-a/16-1 loci. Only 1 case in our cohort did not include the miRs within the biallelic deletion. (Figure adapted from Fabbri et al, 2011)

The data from this chapter has provided a novel insight into the 13q deletion, different to the current literature, which may be because no other study has looked at all the conflicting variables in such close detail in a large, well defined clinical trial such as CLL4. It is this well annotated, long term follow up that has provided a platform for the 13q deletion story to be fully appreciated.

Deletion of 11q, namely *ATM*, are associated with a poor prognosis and response to DNA damaging agents and this poor survival was confirmed in patients with 11q deletions detected by MLPA. However, whilst the MLPA kit contained probes that covered the *ATM* and *RDX* genes, the architecture of 11q deletions could not be further dissected as probes did not extend beyond these loci. Importantly no probes covered the *BIRC3* gene, is a key component in the negative regulation of non canonical *NF-κB* signalling[19]. Deleted and/or mutation of *BIRC3* have also been implicated in aggressive disease, with patients having a similar survival to patients with *TP53* abnormalities[57,99]. No additional survival analysis was undertaken due to the limitations of the MLPA design.

The gain of an additional chromosome 12 is found in 10-20% of CLL patients and is associated with an intermediate response, although data linking this aberration to progression to Richters syndrome has been published [53,68,100]. Trisomy 12 is significantly associated with *NOTCH1* mutations and there is a trend towards poorer survival compared to patients with *NOTCH1* mutation and normal Chr12 CN. These patients have a more 'aggressive' phenotype including features such as; unmutated *IGHV* genes, and Zap-70 and CD38 positivity but not del11q or del17p. This ties in with what has been previously reported, namely patients with both trisomy 12 and *NOTCH1* mutations have a more aggressive disease course and may precede the transformation to Richters syndrome [100,101]. In addition, the trisomy 19 aberration in this data always co-occurs with a trisomy 12, a phenomenon reported in the literature[67,68]. Interestingly, none of the trisomy 12/19 patients had a *NOTCH1* mutation, suggesting that this subset of trisomy 12 patients may have a different disease biology. These observations are limited however, because of the low frequency of these events.

MLPA is able to detect less frequent CNA that may have an as yet undetermined role in CLL pathogenesis. Univariate analysis showed that 8p, 9p, 19p and 6q deletions and 2p and 8q gains can have a significant impact on PFS and OS. Both losses and gains of chromosome 8 have been documented at a low frequency (~2.5%) in CLL, although they are enriched in 17p deleted cases (80% (8p del)) and 44% (8q)), and have been shown to be an independent risk factor for poor survival[69,102].

9p deletions in CLL have been documented but are fairly rare events and are most common after transformation to RS. The gene/s thought to be involved in CLL pathogenesis on the 9p21.3

region are the *CDKN2A*/*B* genes. The cyclin- dependent kinase (CDK) inhibitors, CDKN2A (p16) and CDKN2B (p15), are well known tumour suppressor proteins that function by negatively regulating CDKs thus controlling cell cycle progression[103,104]. P16 can be alternatively splice to form p14, which is an important cell cycle inhibitor and part of the p53 pathway[104]. Richters syndrome can occur in up to 15 of CLL patients and is defined as the transformation to an aggressive diffuse large B-cell lymphoma[100]. The transformation event is thought to occur through two main pathways, one is a trisomy 12/*NOTCH1* mutation lead transformation, whilst the other is via a *TP53*/*CDKN2A* inactivating mechanism and MYC enriched pathway[100]. Fabbri *et al,* also suggest that the two most common events in the transformation to RS is *CDKN2A* gene loss, and *NOTCH1* activating mutations[34]. Both of these studies only found the loss of the *CDKN2A* gene in the RS patients and it was notably absent from the CLL phase of disease[34,100]. However, Haidar *et al*, used Southern blot analysis on 81 CLL patients to identify 8 CLL patients with *CDKN2A* loss and 7 with a *CDKN2B* deletion, including 1 homozygous deletion for both genes. Additional sequencing analysis of the *CDKN2A* gene identified 4 of the patients had concomitant small deletions on the remaining allele[105]. Other published studies, also using MLPA, have identified the loss of 9p21 (*CDKNA/B* genes) in CLL patients; Fabbris *et al* and Buijs *et al* detected by MLPA and confirmed with FISH, the loss of 9p21 in CLL patients (n=1 and n=2, respectively)[93],[106].

2p/*MYCN* gains have also been implicated in an aggressive disease course in CLL and in the transformation to Richter's syndrome. Chapiro *et al*, found 2p gains were associated with Binet stages B/C, an unmutated *IGHV* status and a reduced OS, but the aberration was never seen as a sole abnormality and was associated with 6q deletions and 1p gains[107]. The MLPA probes cover 3 oncogenes on the p arm of chromosome two; *REL*, *ALK* and *MYCN* genes. Expression analysis in cells with and without 2p gains found no difference in REL expression but an upregulation of MYCN expression providing evidence that this gene is the target of 2p gains and may exert a pathogenic role in CLL; it is a member of the MYC family of transcription factors[107]. However, NGS studies in CLL have identified frequent mutations (~5) in the *XPO1* gene also located in the 2p region, centromeric to *REL,* making it an potential target of the del2p[75]. Unfortunately the *XPO1* gene is not covered by the MLPA probes in the kits used in this study. The XPO1 protein is one of the most well characterised nuclear exporter proteins that transports over 200 proteins and RNA from the nucleus to the cytoplasm, including the cell cycle inhibitors p53 and IκB. *XPO1* expression has been shown to be increased in many different cancers and is also correlated to poor prognosis and therapy resistance. A study using a small molecule inhibitor to bind to XPO1 demonstrated, in a mouse model of CLL, reduced disease progression, increased apoptosis and improved OS.[108] Another study investigated the clinical and molecular characteristics of CLL patients harbouring an *XPO1* mutation.[109] Using a NGS amplicon panel *XPO1* mutations were identified in 7.8% of their cohort (n=486), which were associated with UM *IGHV* status, 13q deletions, Zap70 and CD38

positivity, and showed a trend towards increased survival. There was no significant association, in this study, between 2p gain and 11q deletion, as seen by others[96]. Only a small number of 2p gained patients carried a concomitant 11q deletion (5/12). Future CLL trials may wish to consider these less frequent chromosomal aberrations to assess their impact on disease progression, especially in regards to transformation to RS and response to newer therapies.

Splicing factor 3b, subunit 1 (*SF3B1*), which forms an essential component for RNA splicing, was found to be one of the most frequently mutated genes (15) in CLL[110]. Approximately 50% of the mutations are found at codon 700 (K700E), followed by G742 in around 19% of cases, K666 in 12% and H662 in 4% of cases[110–113]. MLPA can only detect K700E mutations, which were identified in 13.6% of the patients. They did not have a significant impact on PFS or OS, which is supported by published data; analysis of the CLL8 trial found no association between reduced OS and *SF3B1* mutated patients [114], and Guieze *et al*, and Schnaiter *et al*, did not find reduced survival in *SF3B1* mutated patients [115,116]. Other studies suggest *SF3B1* mutations do impact upon survival. This discrepancy may be caused by the inability of MLPA to detect all *SF3B1* mutations; perhaps K666 or G742 mutations have a different impact on protein function and therefore survival outcome. Stratifying the *SF3B1* mutations in future large clinical studies may provide an insight into the type of mutation on patient survival. From the MLPA data there was no significant association of *SF3B1* mutations and 11q deletion, and in fact only 15% (n=8) of the *SF3B1* mutated patients had a concurrent 11q deletion. This lower than previous reports, which suggest 11q deletions can occur in around 50% of *SF3B1* patients [110,111,116]. This again may reflect the type of *SF3B1* mutation the MLPA kit can pick up. Raa *et al*, suggest that specific amino acid substitutions in *SF3B1* may explain the variation in expression differences of *TP53*/*ATM* target genes, in patients with *SF3B1* mutations as a sole abnormality[117]. It has also been suggested that one of the other hotspot mutations, G742D is specific to CLL, as it is rarely found in other cancers, so this may play a more pathogenic role in CLL disease progression[118]. Along with the lack of 11q association, there was also no association with any other poor risk biomarkers, such as Zap70, CD38 and *IGHV* unmutated status [110,119]. Patients with *SF3B1* mutation as a sole abnormality (n=20) do not have a significantly reduced survival compared to patients with a mutation in conjunction with other CNA. It would seem that in the latter sub-group additional CNA may be confounding the results, although this would need further investigation to rule out low level CNAs and/or other gene mutations. Another possible explanation for the lack of association between *SF3B1* and poor risk biomarkers could be that MLPA has detected 11 patients with a low clone population, that Sanger sequencing missed, which may have less of an impact upon survival.

The myeloid differentiation primary response 88 (MYD88) protein is involved in the in interleukin-1 receptor/Toll like receptor (IL-1/TLR) activation of canonical NF-κB signalling[120]. The point mutation, L265P, detected by MLPA is also used in the diagnosis for Waldenström's macroglobulinemia and has been implicated in other lymphomas[121]. *MYD88* mutations detected by MLPA were infrequent (n=6), so the data for survival analysis was limited. However, the majority of patients exhibited good risk features such as mutated *IGHV* status, which was in keeping with the literature [120,122].

NOTCH1 mutations were found in 13.4% of the MLPA cohort, which is a slightly higher frequency than reported in the literature[111,123–125]. The exon 34 delCT *NOTCH1* mutation accounts for over 90% of the *NOTCH1* mutational spectra and results in the removal of the PEST domain and a truncated isoform of the NOTCH1 protein, which constitutively activates the NOTCH1 pathway[124]. *NOTCH1* mutations have been suggested to be an important clinical marker with respect to therapy choice, as data from the German CLL8 clinical trial showed that patients harbouring a *NOTCH1* mutation did not have an improved survival when treated with rituximab along with the standard fludarabine-cyclophosphamide regime, but there is a benefit from treated with alemtuzumab[114,125]. This data suggests that patients should be screened for *NOTCH1* mutations prior to treatment implementation.

With the development of higher resolution technologies the CLL research community, in recent years, has uncovered a plethora of novel genomic aberrations, including somatic gene mutations that have been shown to drive disease progression and evolution. This has enabled CLL biology and pathology to be better examined and understood, and improved patient stratification. For example, Rossi *et al*, built upon the original hierarchical 'Dohner' model, published in 2000, by integrating mutational data for *TP53*, *NOTCH1*, *SF3B1* and *BIRC3* genes. As a result, a large proportion of the low risk CLL cases were re-stratified to a higher risk CLL. However, the inclusion of *BIRC3* mutations in this model is contentious; as it is the involvement of the *ATM* gene that impacts on risk, and *BIRC3* deletions almost always co-occur with *ATM* deletion.[53,57] In a similar study, Jeromin *et al.* in a large cohort of CLL patients (n=1160), assessed the impact of novel gene mutations that were found via NGS; *SF3B1*, *NOTCH1*, *FBXW7*, *MYD88*, *XPO1* and *TP53.* A significant impact upon survival was only seen for *NOTCH1*, *SF3B1* and *TP53* mutations, though only the *SF3B1* and *TP53* disruption were deemed to have an independent prognostic value.[126] But the authors confirmed the findings of the previous study in that an integrated cytogenetic and mutational model better stratifies patients into subgroups, which predicts survival more accurately.[57,126] The European Research Initiative on Chronic lymphocytic leukaemia (ERIC) more recently published a multicentre collaborative study, which screened 3490 diagnostic/pre-treatment patients for mutations in *NOTCH1*, *SF3B1*, *TP53*, *MYD88* and *BIRC3* genes.[95] They aimed to assess the prevalence and clinical relevance of each mutation in one of the largest

cohorts of patients ever studied in this manner. Mutations were detected in *NOTCH1* (8%), which also associated with advanced stage, IGHV unmutated CLL and trisomy 12. *SF3B1* (11.2%) had associations again with advanced disease and UCLL as well as 11q deletions. *TP53* (10.4%) were enriched in UCLL and 17p deleted cases. *BIRC3* mutations were less frequent (2.5%) but again were mainly UCLL and advanced stage patients. *MYD88* (2.2%) mutations were exclusively found in mutated CLL and associated with 13q deletion[95]. *NOTCH1*, *SF3B1* and *TP53* mutations and/or deletions were all significantly associated with a poor TTFT, whilst only *SF3B1*, *TP53* aberrations, UCLL and 11q deletions were independently significant in a multivariate analysis. Additionally mutated IGHV patients with only 13q deletion or no genomic aberrations had a superior outcome compared to unmutated IGHV patients of the same profile. The authors argue that any future prognostic models must include the IGHV mutational status to further improve and validate the robustness of prognosis[95].

Drawing from the literature and incorporating the MLPA data; an alternate model of predicting progression and overall survival was created. Tthe initial step was to stratify patients according to the most common CNAs as detected by MLPA. This model stratified the majority of the patients into 6 risk groups, ranking from very low to very high risk as previously described. All groupings were significantly different from each other in terms of OS. This model provides a more detailed stratification for predictors such as the 'low risk' 13q deletion and uncovered a novel very high risk group with a 9p deletion. Despite there being no statistical difference between the first three low risk groups for PFS there is a highly significant difference for OS between all the risk groups justifying this heirarchy. However, as most genomic abnormalities do not appear as sole aberrations, and CLL has been shown to be genomically complex in many cases, the effects of individual variables seen in univariate analysis may be confounded by other variables, so skewing survival analysis. To validate each of the CNAs and mutations that were found to be significant in univariate analysis, multivariate analysis was conducted using cox proportional hazard regression models to account for the effects of these other variables and to see which have the strongest impact upon survival. The first multivariate model included all the significant CNAs that MLPA detected, and indicated that 9p, 17p deletions and 2p gains may be independent prognostic factors for poor survival, with an especially poor survival in 9p deleted cases. However, only 4 patients had a 9p deletion and although statistically significant, this finding will need to be confirmed in a larger cohort of 9p deleted patients. The majority of 9p deleted patients (3/4) also had *TP53* dysfunction, in the form of either a mutation or deletion of the gene (Figure 39). This was accounted for in the multivariate analysis and del9p was still an independent risk factor.

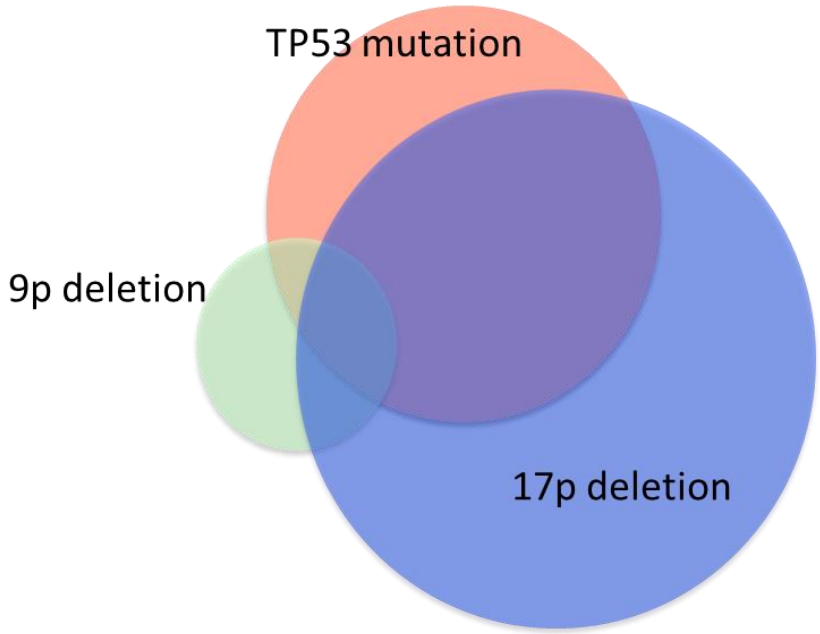


Figure 39 Interaction between 9p deletion and 17p abnormalities. The majority of 9p deletions also have additional TP53 disruption in the form of a deletion and/or a mutation (n=3/4). 1 case did not have any form of *TP53* disruption, but may have a sub clonal mutation that was not detected by Sanger sequencing. This suggests that 9p deletions may be a particularily progressive subgroup of patients harbouring *TP53* aberrations.

Telomere length in CLL has been shown to be a useful biomarker and predictor of survival. A large proportion of the CLL4 trial analysed by MLPA also had available DNA for telomere analysis using a quantitative method, MMQPCR. This technique was developed by Cawthon and allows the simultaneous analysis of a single copy gene and the telomeres in one reaction, using an intercalculating dye, to give a relative telomere measurement.[85] Previous studies in CLL have identified severly shortened telomeres in poor prognostic subgroups such as 17p and 11q deleted patients and patients with UM *IGHV* genes, a finding replicated in this cohort.[127,128] *SF3B1* mutations, *IGHV* UM status, and 17p and 11q deletions all associated with significantly shortened telomeres. The TL of 13q deleted patients was significantly longer, as were *MYD88* mutated patients, consistent with the link to a less progressive disease course. Novel associations were identified in the rarer genomic sub groups; 2p and 8q gains and 6q deletions all had a significantly shortened TL, whilst concurrent trisomy 12 and 19 patients had significantly longer telomeres than trisomy 12 alone. This suggests that this subset of trisomy 12 may in fact have a more benign disease course, although this will need further investigation.

As new genomic regions are characterised, their clinical impact on patients and usefulness as a biomarker will need to be determined. Are they drivers or passengers? Do they impact upon disease progression, clonal evolution or response to therapy? A recent review by Sutton & Rosenquist, highlights the importance of the interplay between tumour genetics and microenvironment interaction. Is it genomics driving disease or the microenvironment supporting and so influencing genetic changes? These questions have yet to be answered but in reality it is probably a combination of both these factors; considering both in future studies may help elucidate a better understanding of CLL biology and in doing so, we can create more efficient therapies that can target and destroy the driving force behind CLL. [129]

Chapter 5: Exome sequencing of B cell malignancies

5.1 Introduction

With the advent of NGS technology there has been a vast leap forward in the quest to find frequently mutated genes and biological pathways targeted in CLL. This technology has now been applied to thousands of patients at different stages of the disease and with different clinical features, to try to elucidate the mechanisms behind CLL pathology.[75,110,122,130,131] Initial observations show that the mutational burden is much lower than seen in solid tumours, with approx. 0.6-0.87 mutations per megabase (Mb) of genomic DNA.[75] Using WES of 180 CLL patients identified 2,444 nonsynonymous and 837 synonymous mutations with a mean somatic mutation rate of 0.6 ± 0.28 per mb and an average of 15.3 nonsynonymous mutations per patient [132].

The most common mutational signature is a G>A/C>T transition, suggesting that CLL mutations are likely introduced by error-prone polymerase η. A second signature; A>C/T>G transversions, specific to *IGHV* mutated patients is thought to reflect the different underlying molecular mechanisms between IGHV mutated and unmutated cases [110,122,131].

The early discovery NGS studies used whole genome sequencing technology on relatively few patients, to uncover novel recurrently mutated genes in CLL. In 2011, the earliest study included five newly diagnosed/untreated CLL patients and found somatic mutations in 39 genes, including a recurrent *NOTCH1* mutation (2/5 cases) [130]. This study extended their analysis to additional 48 patients, with matched germline material, equally representing both *IGHV* sub-groups (mutated (M) and unmutated (UM)). When the authors screened for mutations in the genes identified by the discovery panel, the identified recurrent mutations in *TGM7* (3.8%), *BIRC3* (3.8%), *PLEKHGS* (3.8%) and in *NOTCH1 (*15.1%). Further screening identified *NOTCH1* mutations in an additional 10/120 patients (8.3%) and it was established that they were an independent predictor of poor outcome, and associated with an aggressive phenotype, as enrichment was seen in chemorefractory (20.8%) patients and those with transformation to Richters Syndrome (31%)[130].

Another of the early landmark studies, published in 2011 by Puente and collegues from the Spanish CLL Consortium, combined whole genome and exome sequencing of two unmutated and two mutated *IGHV* patients.[122] The authors identified 45 mutated genes, which they validated in a further 169 patients, focusing on 26 genes that were expressed at the RNA level. They identified four recurrently mutated candidate genes; *NOTCH1*, *MYD88*, *XPO1* and *KLHL6*. Further analysis in 363 CLL

cases found that *NOTCH1* (12% of cases) and *XPO1* mutations (2.4%) were associated with UM-CLL, whilst *MYD88* (2.9%) and *KLHL6* mutations (1.8% cases) were associated with M-CLL. This study showed that these genes were possibly pathogenic and/or had implications as potential biomarkers [122]. In a second paper published in 2012, a larger WES study of 105 CLL patients (60 M-CLL and 45 UM-CLL), the authors identified somatic mutations in 78 genes, including recurrent mutations in *SF3B1*, *POT1* (exclusive to UM-CLL) and *CHD2* and *LRP1B* that were both M-CLL exclusive. Additional screening of *SF3B1* in 279 individuals found an incidence of 9.7%, associated with aggressive forms of CLL[131]. In the same year, the BROAD institute sequenced 88 exomes and 3 genomes. This study also identified frequent mutations in *SF3B1* (15%), *MYD88* (10%) and *NOTCH1* (4%) and novel, rarer mutations in *ZMYM3*, *MAPK1*, *FBXW7* and *DDX3X* [110]. Davide Rossi and colleagues, also found recurrent *SF3B1* mutations in 3 of 11 fludarabine refractory (*TP53* wildtype) patients, prompting further analysis in 301 patients . A large proportion (17%) of the mutated cases were found to be refractory to therapy [112]. Together, these studies identified a number of novel genes and pathways implicated in CLL pathogenesis, and provided further evidence of the distinction between UM/M-IGHV CLL.

Interestingly, whilst there is significant overlap in the recurrently mutated genes identified by these pivotal NGS studies, there is also a level of discordance, particularly in the less frequently mutated genes. This variation was probably caused by the relatively small patient numbers, different cohort composition and different sequencing platforms and bioinformatics pipelines used. Further studies went on to validate some of these genes in larger cohorts and clinical trials and documented their prognostic value. One of the most recent and extensive studies, again from the Spanish CLL Consortium, performed whole genome sequencing of 452 CLL (62 M-CLL, 35 UM) and 54 MBL patients, sampled pre-treatment using both WGS (n=150) and WES (n-440) (Figure 40).[75] A critical finding was the evidence of the functional and clinical importance of mutations within non coding sequences; one 330KB from the *PAX5* locus chromosome within a telomeric enhancer element, and the second in the 3' UTR of *NOTCH1*. By introducing a mutation to the enhancer region, using CRISPR/Cas-9, the authors showed a 40 fold reduction in *PAX5* expression (15 other genes in a 1Mb region were unaffected). The *PAX5* gene encodes a B-cell differentiation transcription factor. Sanger sequencing also identified these mutations in DLBCL (29), FL (23) and MCL (5). A recurrent 3' UTR *NOTCH1* mutation was found in 11/150 cases, with two additional cases having mutations several bases downstream. RNA-Seq was undertaken on six of these patients, along with RT-PCR, immunohistochemistry and western blot analysis, to assess the effect of the mutation on *NOTCH1* expression and function. These data showed that the mutations resulted in a novel splicing event that removesd the PEST domain of the Notch1 protein, thereby increasing stability and constitutively

activating NOTCH1. These patients had a reduced TTFT and OS compared to patients with a coding *NOTCH1* mutation. [75]

Additionally, the study identified 12 novel recurrently mutated genes in CLL including; *ZNF292, ARID1A, ZMYM3, SP110, SP140, SMARCC1, SETD2, NFKB2, IRF4, ASXL1* and *PTPN11*, and confirmed mutations frequently target eight key pathways; BCR signalling, cell cycle regulation, apoptosis, DNA damage response, chromatin remodelling, NF-κB signalling, NOTCH1 signalling and RNA metabolism. Several of these mutations were found to be independent prognostic indicators of TTFT (*BRAF*, *ZMYM3*, *IRF4*, *NFKB2*) and OS (*ASXL1* and *POT1*). The driver mutational burden was significantly lower in MBL compared to CLL, providing further evidence that CLL is an evolution of MBL that has accumulated mutations to further drive progression. [75]

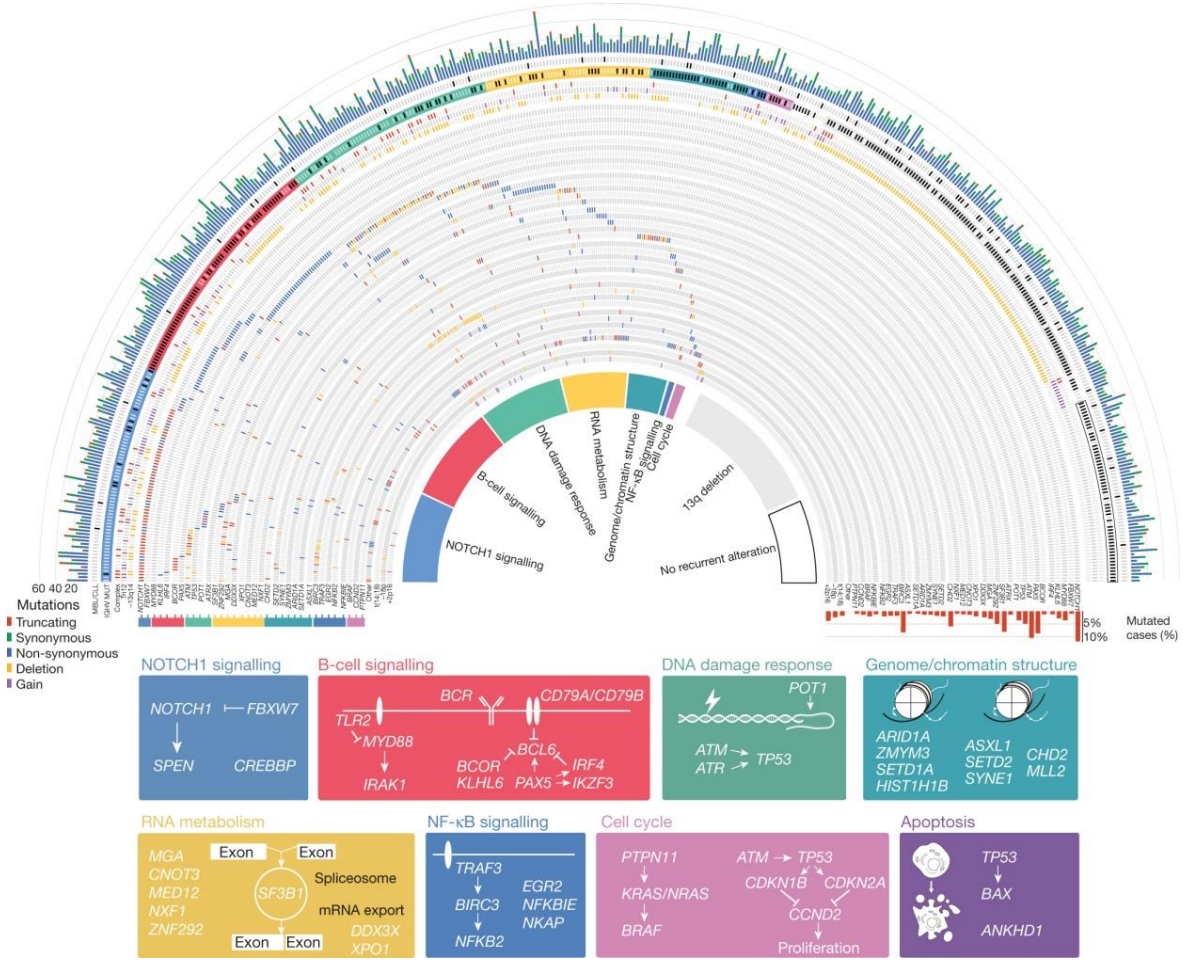


Figure 40. Figure from Puente et al.showing the mutational and genomic spectrum in CLL and comparing IGHV mutated and unmutated cases. Novel *NOTCH1* UTR mutations were identified along with an UTR on chromosome 9p which subsequent analysis found to be a *PAX5* gene enhancer. 8 distinct pathways were identified that drive disease.[75]

5.2 Aims

- To use exome sequencing to identify new candidate genes, that may have clinical or biological importance, with potential utility as biomarkers or therapeutic targets in SMZL and CLL

5.3 Methodology

5.3.1 Patient cohorts

DNA was extracted from the peripheral blood (n=2) or spleen (n=5) of seven CD19 positive SMZL patients, as described in Chapter 3.2. All patients met diagnostic criteria, had a histology that is typical of SMZL with no evidence of progression to a higher grade lymphoma. Each patient harboured chromosomal aberrations targeting 7q and *IGHV1-2*04,* ensuring the exclusion of other types of splenic lymphoma and maximizing the likelihood of identifying pathogenic mutations within related biochemical pathways. All cases had matched germline material from saliva, extracted as described in Chapter 3.3.

The CLL patients (n=6) were all initially diagnosed as MBL or Binet Stage A/ Rai stage 0 CLL according to the 2008 IWCLL/NCI guidelines. The selected patients had mutated IGHV genes, low CD38 expression and no 11q or 17p deletion, but subsequently had progressive disease requiring treatment. This cohort was chosen in the hope of identifying new genomic mechanisms of progression. DNA was extracted from CD19 positive B cells that were taken at a median of 1 year (0-7.3 years) from diagnosis. All cases had matched germline material from saliva. Both studies were approved by the local REC and informed consent was obtained according to the Declaration of Helsinki.

5.3.2 Validation of whole exome sequencing pipeline

High throughput exome sequencing was outsourced and performed using the Agilent SureSelect Human All Exon 51mb V4 and 50mb V3 platforms (Illumina, HighSeq) on paired tumour & germ-line samples from patients with CLL (n=6) and SMZL (n=7). The exome sequencing data was aligned to the human genome reference sequence (hg19/GRch37) using Novoalign software. Duplicate reads, resulting from PCR clonality or optical duplicates, and reads mapping to multiple locations were excluded from downstream analysis. The depth and breadth of sequence coverage was calculated with custom scripts and the BedTools package (v2.13.2). The matched tumour and germline were paired and then analysed using the Varscan (2.3.3) bioinformatical pipeline. This pipeline took into

account normal variation, by cross referencing with SNP databases (dbSNP135), as well as somatic variants from other databases such as 1000 genomes project, and the 4300 European American samples from The National Heart Lung and Blood Institute Exome Sequencing Project Exome Variant Server (http:// evs.gs.washington.edu/EVS/), (ESP6500 release) this allowed the reduction of false positive results and kept somatic variants. The variant allele frequency threshold was set at 10 with a read depth of 4. This analysis was performed by collegues in the University of Southampton Faculty of Genomic Medicine group. To confirm the presence of the somatically acquired mutations found by the pipeline, variants were visually inspected using integrated genomics viewer (IGV) (The broad institute).

5.3.3 **Sanger Sequencing validation**

Sanger sequencing primers were designed, as described in Chapter 3.7, to sequence variants in both the tumour and germline sample, to confirm that the mutation was somatically acquired (present in the tumour and absence in the germ-line) (See Appendix E and F for primers). Sequencing was performed as described in Chapter 3.10. The raw sequence data was aligned to the germline material, visualised, and interpreted using SeqMan (DNASTAR lasergene) software.

5.4 Results

5.4.1 WES validation in CLL

WES was performed on 6 CLL patients, with a mean read depth coverage from the exome sequencing data was 77x (min-max: 43-127), with more than 86% of all bases covered at >20x. The bioinformatical pipeline identified 343 somatically acquired non synonymous/stopgain gene variants across all patients, this was on average 49 genes mutated per patient (range 9-207). Further filtering resulted in 32 variants (Table 21), 15 of which could not be confirmed by the integrated genome viewer (IGV) software. The remaining variants (n=18) were Sanger sequenced. Eight variants were confirmed in this manner, targeting; *DNAH9*, *SPEN*, *NOMO1*, *NID1*, *C9orf66*, *TNFRSF10C*, *LOC646508* and *NAV2* (Figure 41).

Table 21. CLL variants identified by the Varscan bioinformatical pipeline. Variants were assessed by IGV and confirmed by and Sanger sequencing

PID	Genomic location	Gene Name	Variant type	Genomic variant	Coding Variant	VAF (%)	Confirmed by IGV	Confirmed by Sanger sequencing
2	chr11 6524473	*DNHD1*	Non frameshift insertion	c.5728_5729ins TCCTACTGCATG	p.A1910ins	NA	No	No
2	chr1 201453573	*CHIT1*	stopgain	c.1073_1074ins AGGGACTGGGC GGGGCCATGGT CC	p.W358ins*	NA	No	No
12	chr11 19858064	*NAV2*	Non-synonymous	c.G393T	p.Q131H	NA	Yes	Yes
28	chr1 201453581	*CHIT1*	stopgain	c.1065_1066ins CATGGTCTAGG GACTGGGCGGG GA	p.A355ins	NA	No	No
2	chr11 1607783	*KRTAP5-5*	Non frameshift insertion	c.137_138insCT CCGGCTGTGG	p.G46ins	19	No	No
2	chr16 16135841	*ABCC1*	Non-synonymous	c.G3923T	p.R1308L	20	Yes	No
2	chr19 58479881	*LOC646508*	Non-synonymous	c.G1249A	p.E417K	26	Yes	Yes
12	chr8 23025194	*TNFRSF10C*	Non-synonymous	c.C77T	p.A26V	33	Yes	Yes
28	chr9 205383	*C9orf66*	Non-synonymous	c.T14C	p.V5A	36	Yes	Yes
24	chr11 130290380	*SNX19*	Non-synonymous	c.G665T	p.G222V	41	Yes	No
10	chr1 234275572	*NID1*	Non-synonymous	c.A560G	p.D187G	41	Yes	Yes
24	chr16 14854889	*NOMO1*	Non-synonymous	c.A800G	p.Y267C	48	Yes	Yes
24	chr19	*CACNA1A*	stopgain	c.C2344T	p.R782X	53	Yes	No

	13271103							
2	chr4 8664762	*CPZ*	Non-synonymous	c.G1336T	p.G446W	59	Yes	No
24	chr1 16075511	*SPEN*	Non-synonymous	c.T632G	p.L211R	67	Yes	Yes
24	chr18 59136282	*BCL2*	Non-synonymous	c.G598A	p.G200S	NA	Yes	No
10	chr17 11508865	*DNAH9*	splicing			NA	Yes	Yes
2	chr16 28749775	*ATXN2L*	splicing			NA	No	NA
2	chr4 141030560	*MAML3*	Non frameshift deletion	c.1469_1480del	p.490_494del	NA	No	NA
2	chr8 8272280	*SGK223*	Non frameshift insertion	c.1049_1050insGCAGCT	p.G350delinsGSC	NA	No	NA
8	chr2 172290370	*DYNC1I2*	splicing			NA	No	NA
8	chr1 116923808	*IGSF3*	Non frameshift insertion	c.3123_3124insGGC	p.D1041ins	NA	No	NA
8	chr17 78348919	*TBCD*	splicing			NA	No	NA
10	chr7 65236693	*CRCP*	splicing			NA	No	NA
10	chr2 172290370	*DYNC1I2*	splicing			NA	No	NA
12	chr10 50205013	*C10orf71*	frameshift deletion	c.2106_2107del	p.702_703del	NA	No	NA
12	chr12 131128064	*EP400*	Non-synonymous	c.C9262G	p.P3088A	NA	No	NA
12	chr4 104046206	*NHEDC1*	Non-synonymous	c.A1246G	p.T416A	NA	No	NA
24	chr4 104046206	*NHEDC1*	Non-synonymous	c.A1246G	p.T416A	NA	No	NA
28	chr18 19378914	*NPC1*	splicing			NA	No	NA
28	chr19 43069248	*WDR87*	Non frameshift deletion	c.6769_6786del	p.2257_2262del	NA	No	NA
8	chr6 42719928	*UBR2*	splicing			NA	No	NA

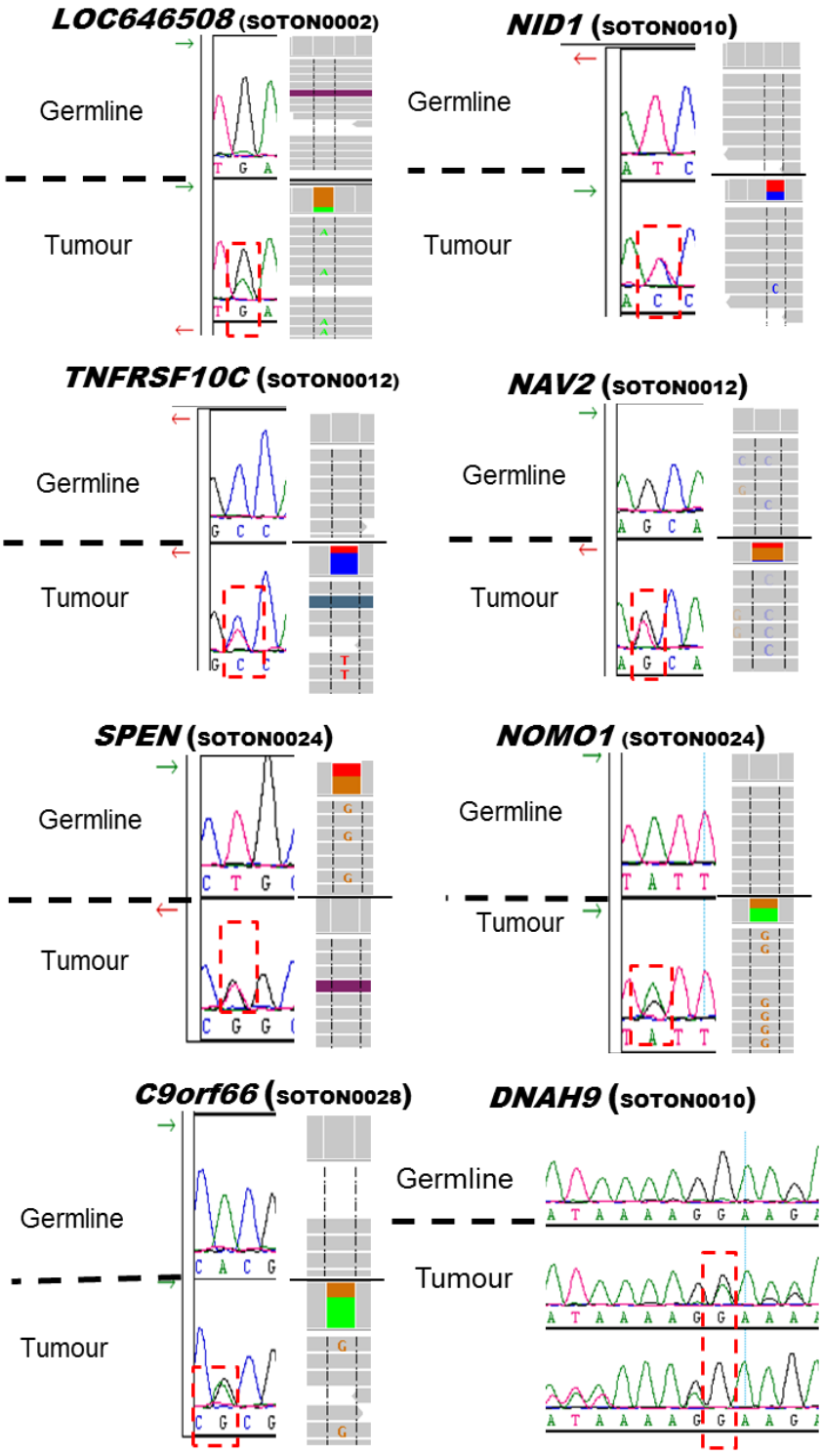


Figure 41. The somatic variants identified by the Varscan bioinformatical pipeline in CLL samples, validated by both IGV and Sanger sequencing.
For each variant, the germline and tumour sanger sequencing traces are shown, with the peak representing the variant highlighted by a red dashed box.

5.4.2 WES bioinformatical pipeline validation-SMZL

The exome sequencing data from seven SMZL patients was analysed with the bioinformatical pipeline outlined in section 4.3.2. Approximately 41.9 million reads per sample were captured with a mean sequencing depth of 69x (range, 43-109x). The mean total of target sequences captured at 20x was 82.2 (range, 70-95). The pipeline also identified 176 somatic non-silent variants in 165 genes, the variants were base-pair transitions (n=34), transversions (n=28), insertions (n=6) and deletions (n=31). The variant list was further filtered to give a final variant list (n=60) (Table 22). Of these, 33 were confirmed by IGV visualisation of germline material, to be somatic variants. Sequencing primers could not be designed for 4 of the variants because of the nature of the mutated regions (high GC content), but of the 29 variants which were sequenced, 23 were confirmed as real somatically acquired variants (Table 22 and Figure 42).

Table 22. Variants identified from the bioinformatical pipeline in SMZL cases

PID	Genomic location	Gene Name	Variant type	Genomic Variant	Coding Variant	VAF (%)	Confirmed by IGV	Confirmed by Sanger sequencing
22	chr1 234983939	*ACTN2*	Non-synonymous	c.C1909T	p.R637C	30	Yes	Yes
22	chr3 64511678	*ADAMTS9*	Non-synonymous	c.G4799A	p.R1600Q	39	Yes	Unable to design primers
22	chr15 82373078	*ADAMTSL3*	Non-synonymous	c.C1931T	p.T644M	47	Yes	Yes
14	chr4 109900510	*AGXT2L1*	Non-synonymous	c.G5A	p.G2E	33	Yes	No
14	chr2 232953716	*ALPP*	splicing			NA	No	NA
18	chr18 14838862	*ANKRD30B*	Non-synonymous	c.C2972T	p.A991V	28		Unable to design primers
14	chr5 142417442	*ARHGAP26*	Non-synonymous	c.T1475C	p.L492P	33	Yes	Yes
16	chr4 115042983	*ARSJ*	Non-synonymous	c.A1696C	p.K566Q	23	Yes	Yes
14	chr1 17189358	*ATP13A2*	Non-synonymous	c.C2263G	p.Q755E	NA	No	NA
18	chr10 93774483	*BTAF1*	splicing			NA	No	NA
22	chr11 73467138	*C2CD3*	Non-synonymous	c.G4273A	p.G1425S	35	Yes	Yes
22	chr3 54888408	*CACNA2D3*	splicing			NA	No	NA
22	chr12 6580911	*CHD4*	Non-synonymous	c.A604G	p.M202V	NA	No	NA
14	chr19 19516033	*CILP2*	Non-synonymous	c.T1679A	p.V560D	37	Yes	Yes
22	chr3 185557967	*CLCN2*	Non-synonymous	c.G775A	p.V259I	NA	No	NA
22	chr16 3728606	*CREBBP*	Non-synonymous	c.A4235G	p.Y1412	44	Yes	Yes

14	chr8 113766990	*CSMD3*	Non-synonymous	c.T2303A	p.L768Q	23	Yes	Yes
22	chr7 148085161	*CUL1*	Non-synonymous	c.T469G	p.Y157D	59	Yes	No
18	chr1 100434502	*DBT*	Non-synonymous	c.C1346T	p.S449L	19	Yes	No
14	chr12 122969230	*DNAH10*	Non-synonymous	c.C10933T	p.R3645W	23	Yes	Yes
18	chr5 140031078	*DND1*	Non-synonymous	c.C1046G	p.T349S	NA	No	NA
22	chr9 139387231	*EXD3*	Non-synonymous	c.A409C	p.S137R	NA	No	NA
14	chr6 146167691	*FBXO30*	Non-synonymous	c.A1544T	p.Q515L	36	Yes	Yes
14	chr18 32189584	*FHOD3*	stopgain	c.G250T	p.E84X	31	Yes	Yes
16	chr17 78138309	*FOXK2*	Non-synonymous	c.C1658T	p.P553L	25	Yes	Yes
22	chr2 49049370	*FSHR*	stopgain	c.C747A	p.C249X	15	Yes	Yes
18	chr2 121429371	*GLI2*	Non-synonymous	c.A538C	p.M180L	NA	No	NA
18	chr2 121460526	*GLI2*	Non-synonymous	c.G2159A	p.R720H	NA	No	NA
14	chr3 130744847	*H1FOO*	splicing			69	Yes	Yes
14	chr1 154981432	*HDGF*	Non-synonymous	c.G274A	p.G92S	NA	No	NA
16	chr15 26174696	*HERC2*	Non-synonymous	c.A2788G	p.M930V	NA	No	NA
16	chr21 39638930	*HMGN1*	splicing			NA	No	NA
16	chr15 63409130	*IGDCC3*	Non-synonymous	c.G1984A	p.G662S	53	Yes	Yes
14	chr11 133296255	*IGSF9B*	Non-synonymous	c.G2575C	p.V859L	19	Yes	Yes
14	chr7 50422612	*IKZF1*	Non-synonymous	c.G665C	p.R222P	NA	No	NA
20	chr7 50422612	*IKZF1*	Non-synonymous	c.G665C	p.R222P	NA	No	NA
18	chr15 88770478	*IQGAP1*	stopgain	c.T288A	p.Y96X	27	Yes	Yes
16	chr17 37577469	*KCNH4*	Non-synonymous	c.G1058A	p.R353Q	NA	No	NA
22	chr19 8909390	*MUC16*	Non-synonymous	c.G33241T	p.A11081S	25	Yes	Unable to design primers
22	chr12 111888133	*OAS3*	Non-synonymous	c.C2605T	p.R869C	29	Yes	Yes
14	chr14 19285512	*OR4Q3*	stopgain	c.T86G	p.L29X	24	Yes	No
18	chr2 205873511	*PARD3B*	Non-synonymous	c.A2012G	p.K671R	27	Yes	No
14	chr8 110604635	*PKHD1L1*	splicing			NA	No	NA
18	chr7 102094895	*POLR2J2*	Non-synonymous	c.C193T	p.P65S	NA	No	NA
22	chr11 110755094	*POU2AF1*	splicing			37	Yes	Yes
18	chr3 109958687	*RETNLB*	stopgain	c.36_37insCCCCTAAG	p.I12_P13ins*	NA	No	NA

20	chr15 39603646	*RPAP1*	Non-synonymous	c.T2174A	p.L725Q	NA	No	NA
18	chr17 18184256	*SHMT1*	Non-synonymous	c.C640G	p.L214V	38	Yes	Yes
18	chr20 1565002	*SIRPG*		c.A580G	p.T194A	NA	No	NA
20	chr7 122599193	*SLC13A1*	Non-synonymous	c.T230G	p.V77G	16	Yes	No
22	chr17 45980082	*SPATA20*	Non-synonymous	c.G80C	p.C27S	32	Yes	Yes
14	chr17 31612433	*TBC1D3C*	Non-synonymous	c.C208T	p.R70W	19	Yes	Unable to design primers
22	chr16 1968034	*TBL3*	Non-synonymous	c.G2033A	p.R678Q	NA	No	NA
22	chr1 214140081	*USH2A*	Non-synonymous	c.G7553C	p.S2518T	NA	No	NA
14	chr3 87100534	*VGLL3*	Non-synonymous	c.A833G	p.D278G	23	Yes	yes
16	chr7 150713142	*WDR86*	Non-synonymous	c.G827A	p.R276H	NA	No	NA
18	chr7 5222711	*WIPI2*	splicing			NA	No	NA
18	chr8 6666910	*XKR5*	Non-synonymous	c.G698C	p.S233T	NA	No	NA
14	chr3 14168846	*XPC*	Non-synonymous	c.T1997A	p.V666E	46	Yes	Yes
16	chr19 22366975	*ZNF98*	Non-synonymous	c.G902A	p.R301K	NA	No	NA

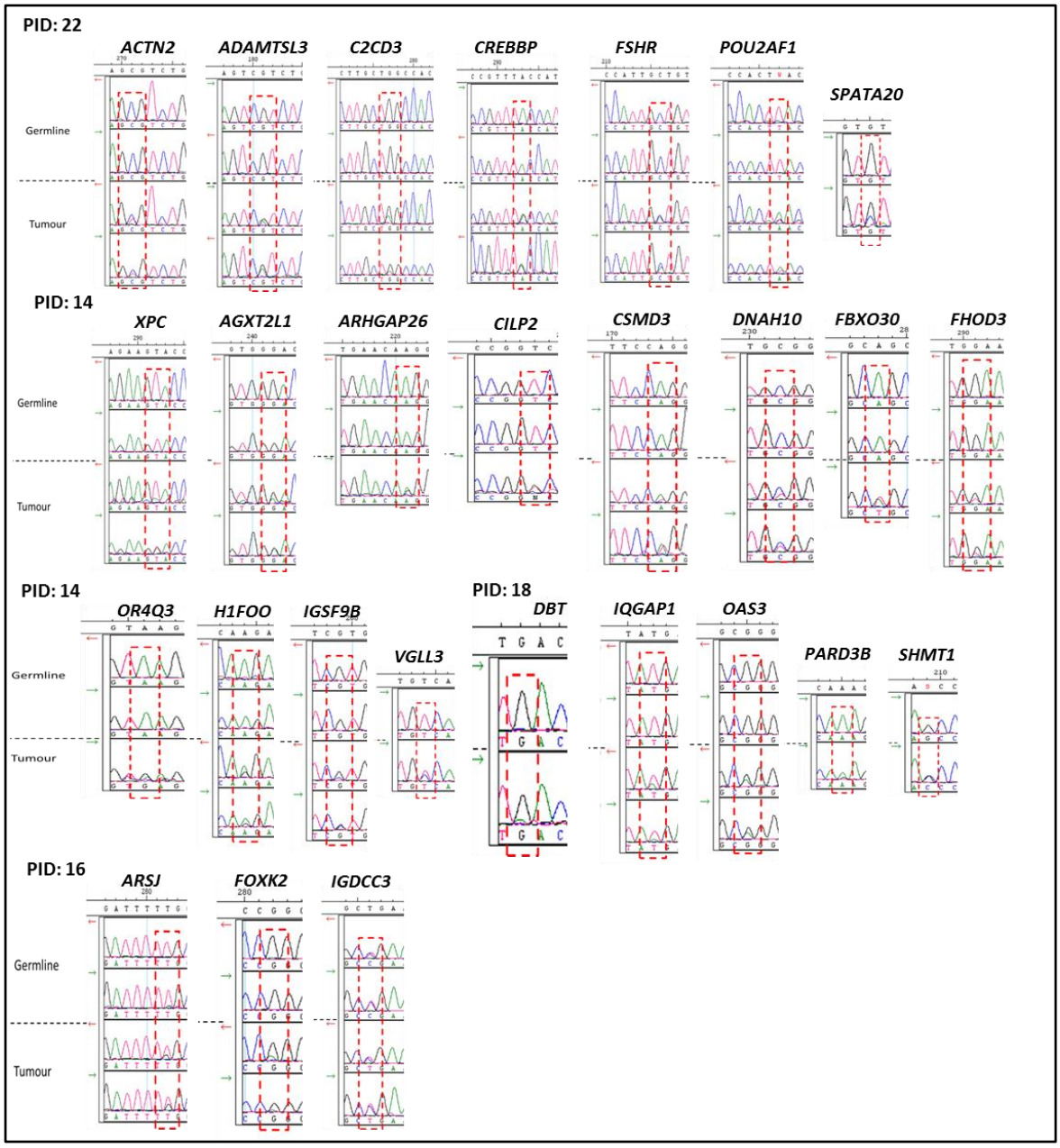


Figure 42. The somatic variants identified by the Varscan bioinformatical pipeline in SMZL samples, validated by both IGV and Sanger sequencing.
For each variant, the germline and tumour sanger sequencing traces are shown, with the peak representing the variant highlighted by a red dashed box. The corresponding IGV profiles are also shown

Discussion

SMZL is a very rare B cell lymphoma predominately affecting the elderly, and like CLL the disease is heterogeneous, with some patients needing little intervention whilst others can transform to a more aggressive lymphoma.[12] Unlike CLL there is no standard approach to therapy and as it is rarer, it is hugely understudied and a lot less is known about SMZL pathogenesis.[12] Biomarkers that are needed to help aid prognosis are scarce for SMZL and none are SMZL specific; they include well-known clinical markers such as lymphocyte count, serum albumin and lactase dehydrogenase.[11] This has meant that exploring the SMZL genome is essential to uncovering new pathways and gene mutations that could be used as a biomarker for prognosis or to target with a new therapy. Somatically acquired SMZL variants (n=23) were identified from the exome sequencing of 7 SMZL patients in this study. Some of the genes affected have been implicated in other cancers such as *CSMD3* in prostate and colorectal, *FHOD1* in ALL and *ADAMTSL3* in colorectal cancer, where variants and corresponding reduced gene expression have been identified.[133–135] *CREBBP* notably has links to AML, ALL and BCL, and functions as a histone acetyltransferase involved in chromatin remodelling and transcription recognition.[14] The *POU2AF1* gene is involved in activating immunoglobulin gene transcription in B cells.[136] *XPC* gene is part of the nucleotide repair pathway and has been implicated in solid tumours such as breast, prostate and bladder.[137] The *ARHGAP26* gene encodes a GTPase activating protein and variants have been identified in AML, CML and Juvenile myelomonocytic leukemia (JMML).[138] Other genes have less notable links to SMZL such as *H1FOO*, *FSHR* and *C2CD3.* The former gene functions as a regulator of centriole elongation, is essential for cilia formation and is a regulator of hedgehog signalling, which is essential during embryogenesis and for continual cell maintenance.[139,140] The data described in this chapter was recently published (Parry *et al)* and as part of that process additional analyses of these cases was performed; SNP 6.0 profiling of the cases identified 3 candidate genes; *EZH2*, *CUL1* and *FLNC,* in the region encompassed by the 7q deletion.[14] The variant data was also re-filtered to look for known 'SMZL' genes; *NOTCH2* [exon 34, n=2], *TNFAIP3* [n=3], *MAP3K14* [n=2], *MLL2* [n=1] and *SPEN* [n=1] mutations were present but these were originally filtered out.[14] Targeted resequencing in 175 SMZL patients using a Haloplex gene panel identified *TP53*, *KLF2*, *NOTCH2*, *MLL2*, *MYD88*, *ARID1A* and *TNFAIP3* genes as recurrently mutated in over 40 of the cases.[15] *KLF* gene mutations associated with the frequent 7q deletion as well as IGHV1-2*04 variant and most importantly significantly reduced the TTFT. *NOTCH2* mutations also significantly reduced the TTFT and *TP53* mutations impacted OS by significantly shortening survival.[15] In one of the pivotal SMZL NGS studies, Rossi *et al*. used WES to sequence 12 cases. The authors also identified *NOTCH2* mutations

in 21 of SMZL patients, and a further 11 with mutations to NOTCH pathway genes, as well as NFκB pathway gene mutations in 34 of cases.[13,19]

The characterisation of the SMZL genome has only just begun; major breakthroughs in this disease have already advanced the understanding of pathogensis. Variants in pathways such as NOTCH and NFκB signalling, chromatin remodelling and cytoskeleton have been implicated in SMZL, the next step is to relate these to clinical outcome and future clinical trials may begin to target recurrently mutated pathways with specific pathway inhibitors. The major hurdles of studying SMZL is the rarity of this B cell neoplasm and often studies have to collaborate with centres from around the world to get enough patient samples.[15]

Next generation sequencing has been pivotal in understanding the genes and pathways involved in the pathogenesis of CLL.[75,110,122,141,142] This study used whole exome sequencing on matched tumour-germline samples of CLL patients with good risk biomarkers but with progressive disease. This was an interesting group for gene discovery though no recurrent mutation was identified, there were variants (n=8) identified in genes implicated in other cancers, diseases and/or known CLL disrupted pathways. For example, *SPEN,* as previously discussed, is involved in NOTCH signalling and has been implicated in other B cell neoplasms such as SMZL.[13] The *NID1* gene encodes a membrane glycoprotein and is involved in cell interactions and has been identified in invasive endometrial cancers.[143] The *TNFRSF10C* gene is a p53 regulated DNA damage inducible gene that functions as an antagonistic receptor, protecting cells from TRAIL induced apoptosis; CNAs of this gene have been identified in colorectal cancer.[144] The *NAV2* gene encodes a neuron navigator protein, which is involved in cellular growth and migration, an upregulation of which has been identified in colorectal carcinomas.[145] The *DNAH9* gene encodes an axonemal dynein involved in the movement of cilia and flagella, and has been implicated in many cancer types including CML.[146] Other genes have less certain links to CLL and/or have limited information regarding their gene function such as *LOC646508*, *C9orf66* and *NOMO1*, the latter of which is implicated in a nodal signalling pathway essential for vertebrae development.[147] All the genes identified in this study have been cross referenced with the Puente NGS study to see if any of these genes are recurrently mutated in both studies. Two of the genes; *DNAH9* and *SPEN* were identified in both studies so these were further investigated in regards to CLL; discussed in later chapters.[75]

Further analysis of the CLL genome is becoming more refined; with the analysis of many CLL patients' genomes and exomes already sequenced the quest for novel, frequently mutated, candidate genes is coming to an end. The next step is to validate the lesser known genes in larger cohorts of patients and to relate this back to known biomarkers and impact upon survival. There is

also a need to study the mutations in more depth, for example, the impact of individual mutations and sub clonal mutations, to understand the dynamics of clonal evolution and the driving mutations behind this. The next chapter explores frequently mutated genes that have been identified from the literature.

Chapter 6: Assessing individual gene mutations in the context of CLL

6.1 Introduction

As previously discussed, next generation sequencing allows researchers to delve into the CLL genome, to discover novel candidate genes and pathways, in a quest to better understand and treat the disease. Novel and surprising discoveries included frequent mutations in the splicing factor *SF3B1* and *NOTCH1* that presented new pathways for investigation in CLL. Other candidate genes identified are implicated in survival pathways such as the NFκB genes *BIRC3* and *MYD88*, and the important telomeric gene *POT1*, and another splicing factor; *U2AF1*. All these genes will be further dissected and discussed in this chapter.

6.1.1 *NOTCH1* and gene mutation implications in CLL

The Notch signalling pathway is an important regulator of cellular differentiation, proliferation and apoptosis.[148] Rosati *et al.* implicated Notch signalling in CLL and postulated that it contributes to CLL cell survival and apoptosis resistance.[149] They found Notch signalling was constitutively activated in B-CLL cells but absent in normal B cells. In addition, Notch1 and Notch2 proteins, along with their ligands Jagged1 and Jagged2, were found to be constitutively expressed in CLL but again not in the normal B cells thus suggesting that it is CLL specific dysregulation.[149] The authors were unable to provide an explanation for the differences in protein levels, as the mRNA levels were found to be the same in both normal B cells and CLL cells; they suggested protein deregulation of components of the Notch signalling pathway.[149] The mechanism by which the pathway was activated in CLL was found to be a mutation within the *NOTCH1* gene, PEST domain.[124] This was further confirmed with the NGS studies in CLL which identified *NOTCH1* gene mutations as recurrent in CLL.[110] The *NOTCH1* gene encodes a class I transmembrane protein that functions as a ligand activated transcription factor.[148] When activated, the notch1 protein translocates to the nucleus, where it transcriptionally activates multiple target genes including c-Myc that drives cell proliferation and survival.[148] The Notch1 protein contains a PEST domain that is targeted for ubiquitination leading to degradation of the Notch protein.[148] Up to 90% of *NOTCH1* mutations in CLL have the same recurrent 2bp frame-shift deletion (p.P2515Rfs*4) in exon 34. This mutation results in a premature stop codon that truncates the Notch1 protein, which then lacks the PEST

domain and so the protein can no longer be ubiquitinated for degradation.[148] This confers increased Notch signalling and cellular survival. [148,149] Mutations in *NOTCH1* have been associated with a more progressive clinical disease course in CLL, with a reduced survival and transformation to RS.[34,111] This is reflected in the mutational burden at differing disease stages; only 3% of MBL patients harbour a mutation, increasing to around 10% of CLL patients at diagnosis, more than 20% of chemo refractory patients and up to 30% of RS transformed patients.[34,99,111,130,150] In the German CLL8 clinical trial, patients harbouring a *NOTCH1* mutation showed no improved survival after treatment with Rituximab. It is thought that these patients have lower levels of CD20.[151,152] Other members of the Notch signalling pathway have been implicated in CLL, such as *FBXW7*. This gene encodes a protein that targets Notch1 for degradation. Mutations have been identified at low frequency in CLL (~3%) though no adverse effects to patient survival have been shown to date.[126]

6.1.2 **The splicing pathway and *SF3B1* dysfunction in CLL**

Intron removal is essential for gene function and intron transcripts have been suggested to regulate gene expression.[153] Spliceosomes are proteins consisting of small nuclear ribonucleoproteins (snRNPs) and small nuclear RNA (snRNA) complexes that remove introns during mRNA transcription.[154] Splicing factor 3b, subunit 1 (*SF3B1*) gene, was found to be one of the most frequently mutated genes (15%) in CLL[110]. SF3B1 protein is an essential component of the spliceosome U2 small nuclear ribonucleoprotein subunit that is essential for RNA splicing .[155] U2 introns have highly conserved dinucleotides; GT and AG, at their 5' and 3' ends and rarer U12 introns can be found within some genes.[153] The SF3B1 protein has two defined regions; the N-terminal hydrophilic region that contains several protein binding motifs and the C-terminal domain, which is made up of 22 huntington Elongation Factor 3 PR65/A TOR (HEAT) repeats. All of the mutations detected in CLL are located within this highly conserved C-terminal domain and most of these mutations have been found within the 5^{th} and 8^{th} HEAT repeats that are encoded by exons 14 to 16[155]. The majority of the mutations are 'hotspot' mutations, the most common being found at codon 700 (p.K700E) in around half of cases, followed by p.G742 in around 19% of cases, p.K666 in 12% and p.H662 in 4% of cases [110–113]. Prior to the publications implicating *SF3B1* in CLL, WES data was published identifying frequent mutations in multiple genes involved in the RNA splicing mechanistic pathway, in myelodysplasia (MDS) and other haematological disorders.[156] The splicesome genes, which include *SF3B1*, *U2AF35*, *ZRSR2*, and *SRSF2* are also involved in the 3'-splice site recognition of pre-mRNA processing, and were found to be mutated in 45%-85% of MDS patients and were shown to cause abnormal splicing[156]. The recurrent K700E *SF3B1* mutation has

been shown to be present in both MDS and CLL, whilst the three other hotspot mutations affecting codons 742, 666 and 662, have only been detected in CLL suggesting a different pathobiological role of *SF3B1* within the two diseases[153]. Patients carrying a *SF3B1* mutation in MDS confer a better survival, whilst in CLL studies the opposite has been shown and generally patients have a poor outcome[157][111]. The prognostic impact of *SF3B1* mutations have been assessed in various studies within CLL. Wang *et al* and Quesada *et al*, were the first to observe a poor clinical outcome in patients carrying a *SF3B1* mutation and found the TTFT and OS to be markedly reduced and, most importantly, an independent prognostic marker[110,131]. These two studies used a heterogeneous cohort of patients which were not part of a clinical trial study so were not well characterised/ uniformly sampled. As such, further studies have validated the impact of the *SF3B1* gene mutations on survival in clinical trial data sets, such as the UK LRF CLL4 trial and the German CLL8 study.[111,114] Other studies, such as Rossi *et al.* have revealed that *SF3B1* mutations are associated with a more therapy resistant, aggressive and advanced disease course (17% of cases), whilst only 5% of patients at diagnosis carried a mutation.[112] This was recapitulated by Greco *et al*, where only 1.5 % of MBL patients carried an *SF3B1* mutation, showing that this mutation is associated with a more advanced disease.[158] *SF3B1* mutated patients have been shown to have an independent, functional response that is similar to *ATM*/*TP53* disrupted patients; increase in DNA damage, response to DNA damaging agents.[117] As well as abnormal splicing events, a truncated inactivated form of the FOXP1 protein was identified in *SF3B1* mutated CLL patients.[131] As briefly mentioned, other splicing genes have been implicated in other haematological malignancies.[156]

U2AF1 is a 35kDa subunit of the U2AF heterodimer, along with the U2AF2 protein, and is involved in 3' splice site recognition and the recruitment of U2 snRNPs; SF3A1 and SF3B1.[154] Recurrent hotspot mutations have been identified in the *U2AF1* gene, in codons 34 and 157, in up to 12% of MDS patients.[156] The mutations cover a zinc finger domain region and variants have been shown to affect splicing due to an increase in intron retention.[159] Prognostically this gene has relevance in MDS, as patients harbouring this mutant have a reduced survival.[159] The *SRSF2* gene encodes a pre-mRNA splicing factor and is critical for constitutive and alternative mRNA splicing. The SRSF2 protein is also thought to have a role in regulating DNA stability and in the acetylation and phosphorylation pathways.[154] The *SRSF2* gene has been identified as recurrently mutated in up to 47% CMML cases.[156] The most common variant seen is at codon 95 between an RNA recognition motif and a RS domain. It has been suggested that depletions of SRSF2 expression may lead to global genomic instability. [154,156] The recurrence (if any) of these splicing genes has yet to be determined in CLL, no exome study picked up either *SRSF2* or *U2AF1* as being recurrently mutated in CLL but no extended study has focused on these genes.

6.1.3 ***BIRC3***

NF-κB is a family of transcription factors that regulate genes that provide pro survival signals such as those involved in immune response, inflammation, cell growth, survival and development. Unsurprisingly deregulation of this pathway has been associated with autoimmunity, inflammation, diabetes, cardiovascular disease, neurodegenerative disease and cancers as well as many other disorders[160,161]. Constitutive NF-kB activation is seen in many lymphomas where it drives tumour proliferation and survival[162]. Activation of NF-κB signalling relies on a number of positive and negative regulatory components, with the core elements being the IKK complex (IKKα, IKKβ, NEMO), IκB proteins and NF-κB dimers. NF-κB proteins include RelA (p65), RelB, c-Rel, NF-κB1 (p105/50) and NF-κB2 (p100/52) [163]. Before activation of IκB signalling, NF-kB dimers form complexes with IκB proteins in the cytoplasm of cells. Once activated, a signalling cascade allows the IκB complex to be phosphorylated, ubiquitinated and then degraded, freeing the NF-kB dimers to translocate to the nucleus where they bind to the promotor regions of certain genes to stimulate transcription[161]. The classical pathway follows this route and may be induced through ligand binding to cytokine receptors such as TNF receptor (TNFR), IL-1 receptor (IL-1R) and Toll-like receptor 4 (TLR4). This causes the signalling cascade that activates IKKβ, which is in a complex with IKKα and NEMO that then phosphorylates the IκB proteins (Figure 43). Non canonical NF-κB signalling subsequently is activated through specific members of the TNF family such as CD40 ligand[161]. The non-canonical pathway does not rely on NEMO and is dependent upon IKKα activation by CD40 ligand that leads to p100 phosphorylation, which allows p52 –RelB complexes to form and translocate to the nucleus (Figure 44)[161]. The non-canonical NF-κB signalling pathway regulates the development of lymphoid organs, B-cell survival and maturation, dendritic cell activation, bone metabolism and thymic deletion of autoimmune T cells and deregulation of this alternate pathway is associated specifically with lymphoid malignanices[164]. The Inhibitor of apoptosis (IAP) family of proteins, were first discovered in baculoviruses and were thought to be involved in supressing the host cell death response when cells become infected by a virus. [165] Eight members of this protein have been discovered in mammalian cells and they are now known to play important roles in regulating NF-κB signalling as well as being involved in cell death pathways such apoptosis and necroptosis [165] (Figure 45). The IAP family includes; X-linked inhibitor of apoptosis (XIAP), cellular inhibitor of apoptosis 1 and 2 (cIAP1, cIAP2 but also known as BIRC2/3), melanoma inhibitor of apoptosis (ML-IAP) and survivin proteins.[165] They are characterised by a 70 amino acid baculovirus IAP repeat (BIR) domain, and up to 3 of these BIR domains may occur within a protein.[165] cIAP1 (BIRC3) is required for positive regulation of classical NF-κB signalling by TRAF2

dependant K63 linked ubiquitination of RIP1 in, whilst in the non-canonical pathway BIRC3 is needed for TRAF 2 mediated ubiquitination and degradation of NIK[161].

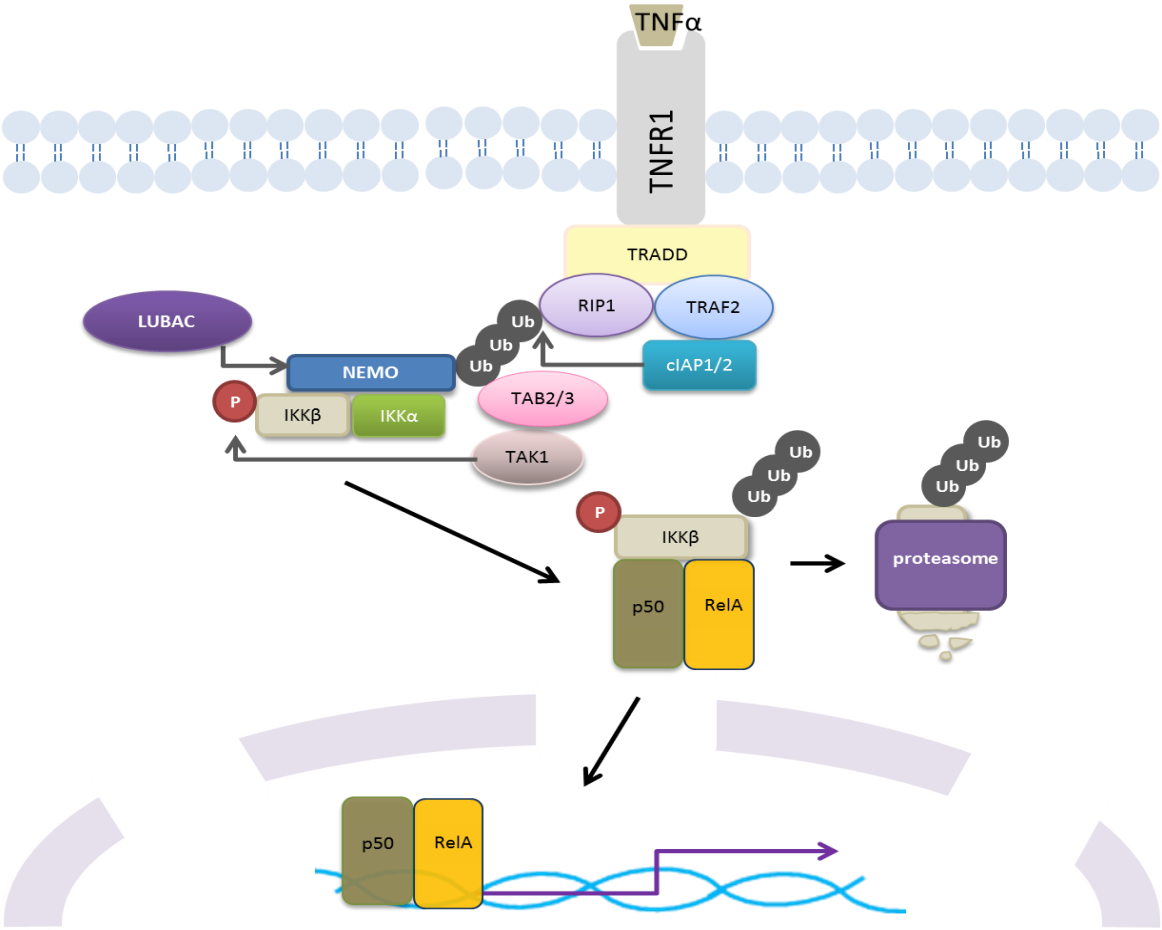


Figure 43. TNF induced canonical NF-κB signalling. Although BIRC3 has a major role in negatively regulating the non canonical pathway it also has a role in canonical signalling, ubiquinitating RIP1.

Ubiquinitation is a post translational modification of proteins by a covalent attachment of ubiquitin chains to lysine residues on a target protein through an enzyme cascade[161]. Many of these IAP proteins have been found to be overexpressed in many different cancers and are thought to help the cancer cells survival by evading apoptosis.[166] Somatic variations of *BIRC3* have been identified in several lymphoid malignancies including; Gastric mucosa associated lymphoid tissue (MALT) lymphoma, multiple myeloma (MM), splenic marginal zone lymphoma (SMZL) and mantle cell lymphoma (MCL).[19,99] In MALT lymphoma the tumour cell can acquire a fusion oncogene involving the *BIRC3* gene and *MALT1* gene, causing NF-κB signalling to become activated.[166] Where as in SMZL and MCL, loss of function *BIRC3* somatic mutations have been detected, which activate non canonical NF-κB signalling.[19] In MM the *BIRC3* gene is disrupted by copy number loss and nonsense or frameshift mutations.[166]

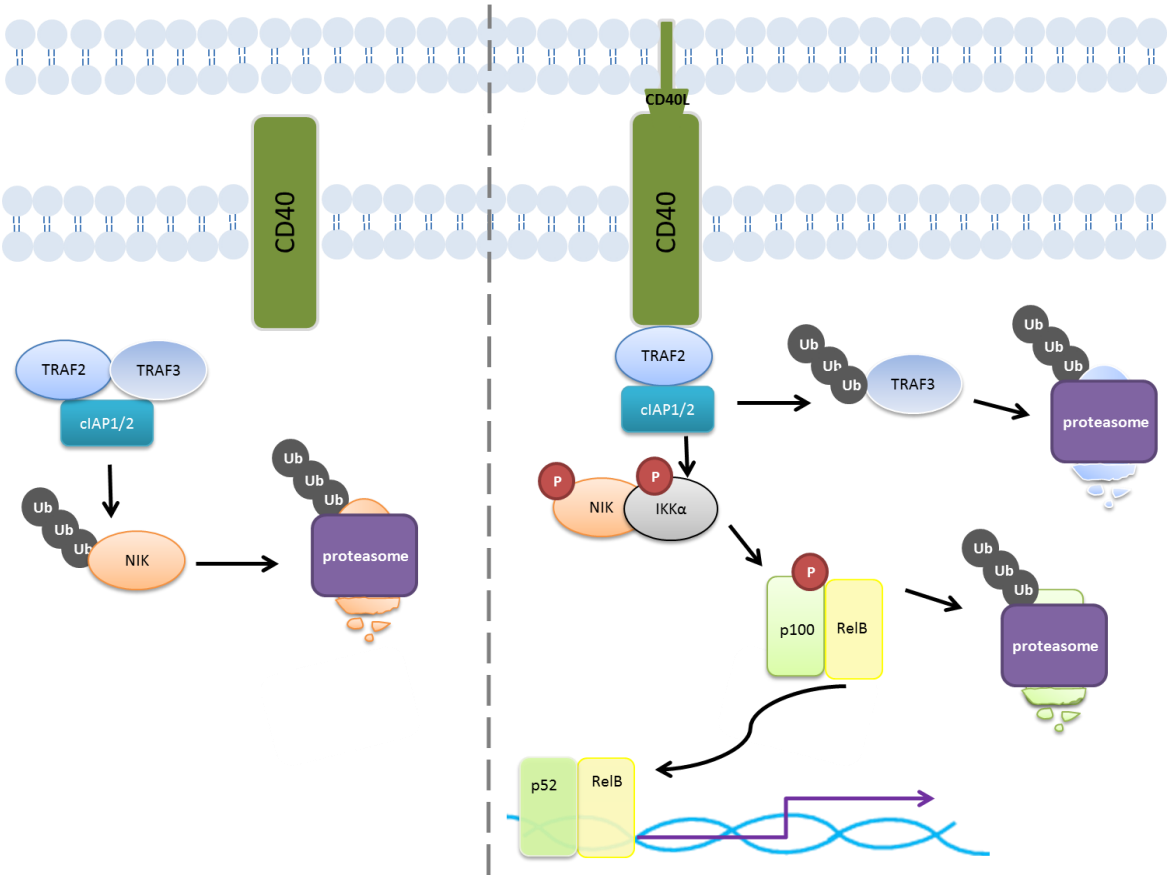


Figure 44. cIAP1/2 (BIRC3) is a negative regulator of non canonical NFκB signalling. This pathway is activated through the binding of a CD40 ligand that causes the recruitment of cIAP1/2 and TRAF2 that normally ubiquinate NIK to prevent activation of the pathway. Figure adapted from Hayden and Ghosh, 2012. Genes Dev. [161]

Sequencing of 20 genes involved in the NF-κB pathway in 101 SMZL patients revealed mutations in 20% patients, targeting 5 different genes; *TRAF3, MAP3K14, IKBKB, TNFAIP3 and BIRC3*[19]. NF-κB activation in canonical and/or non-canonical pathways was confirmed in 58% of patients[19]. In CLL both deletion and somatic mutations of *BIRC3* are found, causing activation of the non-canonical signalling pathway.[99] *BIRC3* was first identified in CLL in a pivotal NGS study by Fabbri *et al*, who used whole exome sequencing to find some novel genes including *BIRC3* only at a low frequency [130]. Rossi *et al* went on to identify *BIRC3* mutations in 4% of CLL patients at diagnosis, 24% of fludarabine-refractory CLL patients, and they were notably absent from monoclonal B-cell lymphocytosis patients, suggesting *BIRC3* gene disruption is associated with an aggressive phenotype [99]. Although *BIRC3* disruption, classed as either a deletion and/or mutation of the gene, was found to be mutually exclusive to *TP53* lesions, survival in these two patient sub-groups is equally poor. This study showed that *BIRC3* mutations alter protein function by removal of the C-terminal RING domain, which is essential for proteosomal degradation of MAP3K14 (NIK), resulting in constituative activation of non-canonical NF-κB signalling; [99]. In contrast to SMZL, mutations in

other genes involved in NF-κB signalling were not identified. Rossi *et al* also studied *BIRC3* mutations in the context of other recurrent genomic lesions [57]; cytogenetics (FISH) for established recurrent chromosomal aberations and the mutational burden in *TP53*, *NOTCH1*, *SF3B1*, *MYD88* and *BIRC3* (sanger sequencing) was analysed in 1274 CLL patients and correlated to survival data. A hierarchical model, stratifying patients according to four groups was developed; 1) a high risk group with *BIRC3* or *TP53* disruption; 2) an intermediate risk group, with *NOTCH1* and *SF3B1* mutations and/or 11q deletion; 3) a low risk group with trisomy 12 or normal genomes; 4) a very low risk group with del(13q) as the sole aberration. This model was shown statistically to be a better predictor of survival than the traditional cytogenetic 'dohner' model[57]. However, this data is contentious. *BIRC3* is located at 11q22.2, ~6Mb centromeric to the *ATM* gene locus, which is almost always deleted along with the *BIRC3* gene and other studies support *ATM* mutational status as the critical driver of the poor prognosis in 11q-deleted CLL [70].

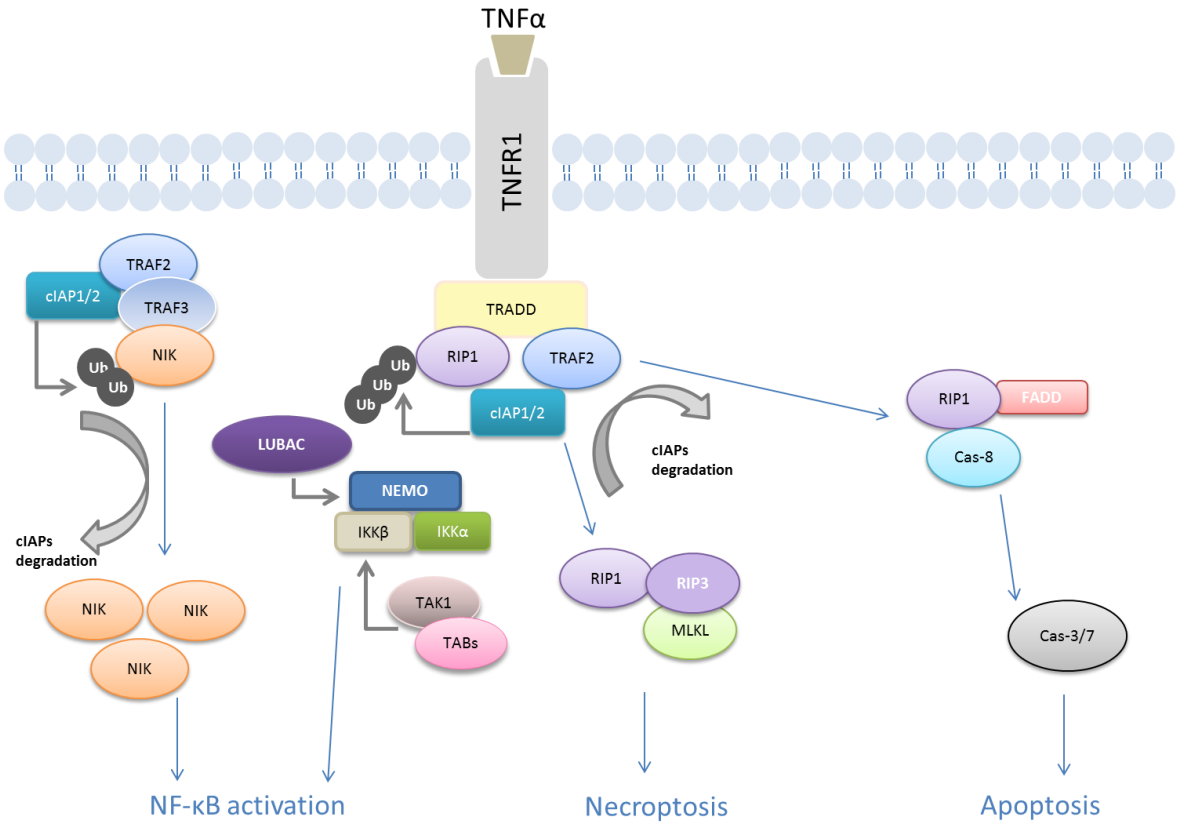


Figure 45. cIAP2 (BIRC3) proteins play important roles in NF-κB signalling and cell death pathways (Necroptosis/Apoptosis) adapted from Bai *et al*, 2014. [163]

6.1.4 ***MYD88***

As with BIRC3, the myeloid differentiation primary response 88 (MYD88) protein is involved in NF-κB signalling, in the interleukin-1 receptor/Toll like receptor activation pathway (Figure 46). This NF-κB pathway is involved in the immune/inflammatory response to bacterial lipopolysaccharides (LPS). There are 10 different TLRs that recognise microbial cells or pathogenic nucleic acids and form part of the innate immune response. These TLRs contain an extracellular leucine rich repeat domain that recognises pathogens and a cytoplasmic conserved Toll/IL-IR (TIR) domain, that is activated through recruiting adapter kinases[167]. MYD88 functions as an adapter protein, that forms part of the TLR2 and 4 receptor-associated signalling complex containing; TRAF6, cIAP1, cIAP2 and TRAF3[120]. Upon ligand (LPS) binding, MYD88 is recruited to the TLR4, were it is activated and phosphorylated. It then recruits IL-associated kinases (IRAKs), such as IRAK-4 that bind TRAF6, and once ubiquitinated recruits TAK1-TAB2/3 and IKK complexes leading to activation of canonical NF-κB signalling[161,166]. Furthermore, the MYD88-receptor associated complex can translocate to the cytosol were it activates TAK1 and then the MAPKs pathway. This process is inhibited by the TRAF3 protein, but is positively regulated by cIAP1/2 (BIRC2/3) ubiquitylation of TRAF3[166]. MYD88 contains a C-terminal TIR domain connected to an amino-terminal death domain and a short intermediate domain; the death domain allows the MYD88 protein to interact with the IRAK proteins [168,169]. The majority of *MYD88* mutations in CLL, are found in a hotspot in exon 5; L265P [170]. Other recurrent mutations identified in CLL include; V147L, S243N and S219C[120]. Mutations have been detected in 2-10% of CLL patients and the hotspot mutation has been detected in other lymphomas; 4% of Marginal Zone Lymphoma (MZL) and 96% of Waldenströms macroglobulinemia (WM) (also known as lymphoplasmacytic lymphoma (LPL)) patients. It is now used in the diagnosis of WM, which was notoriously difficult to distinguish from CLL and other MZL [121] [171]. In CLL, the L265P mutation causes constituative activation of NF-κB signalling without any ligand engagement[122]. The constituent activation of TLRs has been shown to promote proliferation and protect against spontaneous apoptosis in CLL cells[172]. *MYD88* mutations were first discovered in the pivotal Puente *et al* WGS study[122] and the furthur analysis of 363 patients confirmed an incidence of 2.9% and an association with a M-CLL status, low expression of ZAP70 and CD38 [120,122], a younger age at diagnosis and improved survival[120]. The findings of subsequent larger studies are not concordant; some found no correlation between *MYD88* mutation and age or improved survival [57,95,126,170] and attribute the improved survival to the M-IGHV status seen in most *MYD88* mutated. The prognostic significance of *MYD88* mutations and the impact upon survival in CLL has

been the subject of much debate in recent times; could CLL patients harbouring a *MYD88* mutation actually be misdiagnosed WM?

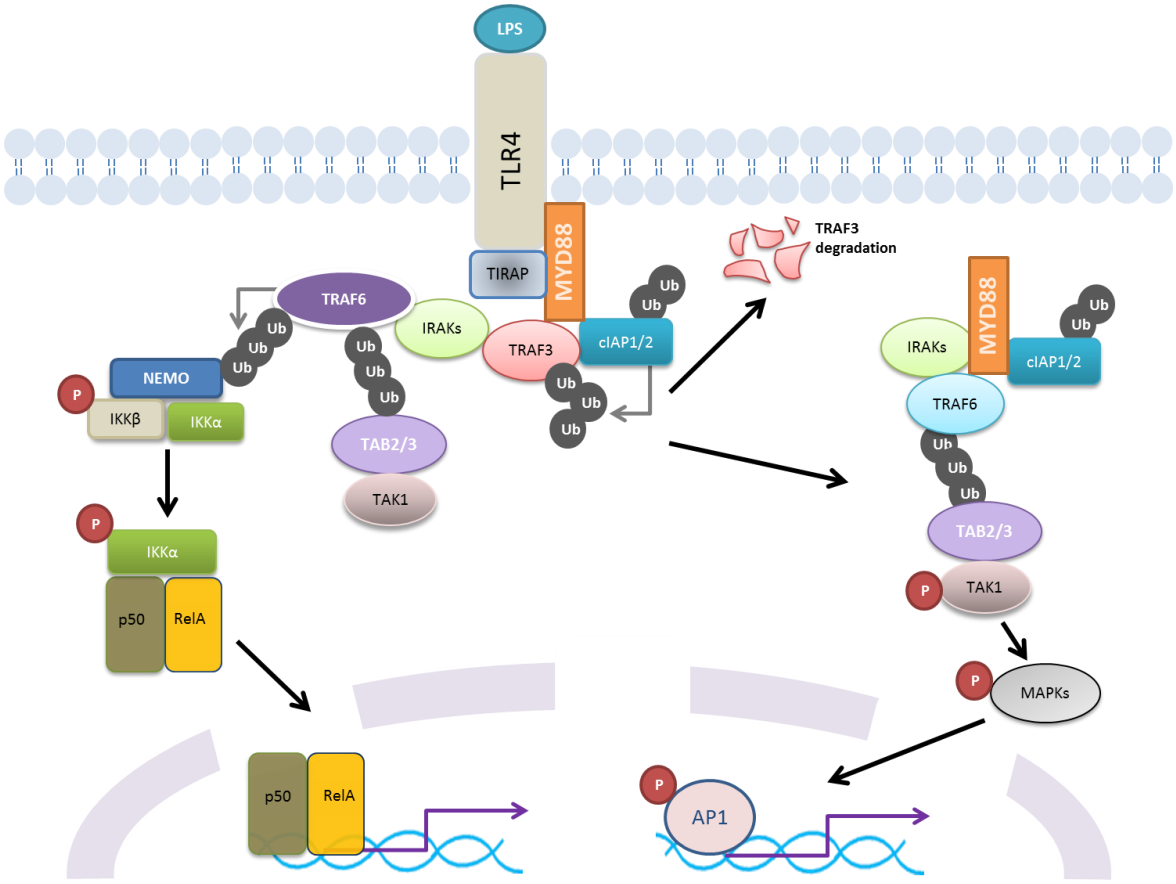


Figure 46. MYD88 interaction within the TLR4 activated NFkB pathway. MYD88 functions as a adapter protein, part of the TLR4 signalling complex. This complex translocates to the cytosol when TRAF3 is ubiquinated by BIRC2/3, where the complex can then activate the MAPK/ERK pathway, leading to cell survival and proliferation.

6.1.5 ***POT1***

Telomeres are comprised of specialised DNA and related proteins, found at the terminal ends of chromosomes. They are composed of 5-15kb non-coding double-stranded repeats of TTAGGG tandem sequence, that end in a 50-300bp 3' single strand overhang. The overhang serves as a protective measure and has a 'capping' function; it folds back the double stranded DNA telomere helix and forms a 'T' loop, providing chromosome integrity and genome stability. This is largely dependent upon the interaction of the DNA with telomere proteins such as the shelterin complex. This complex consists of TRF1, TRF2, TIN2, RAP1, and TPP1 and POT1 proteins, that protect the single stranded telomeres from DNA damage-repair recognition, that could lead to TP53 mediated cell senescence or apoptosis[173] (Figure 47).

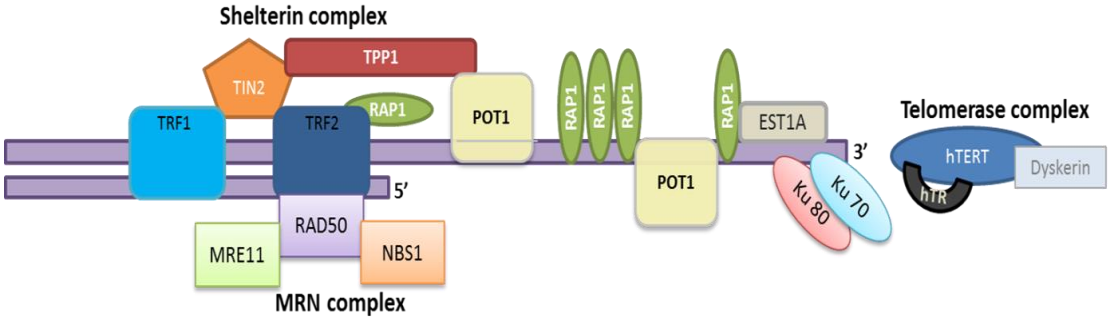


Figure 47. Telomeres and their related protein structures and complexes.
The Shelterin complex includes TRF1, TRF2, TIN2, RAP1, TPP1 and POT1 and protects the telomeres. The MRN complex includes MRE11, RAD50 and NBS1 that bind to TRF2. The telomerase complex includes hTERT; a catalytic subunit, template RNA hTR and Dyskerin. Adapted from Poncet *et al*, 2008.[174]

DNA polymerase cannot fully replicate the 3' end of DNA and so, with each cell division the telomeres shorten by 30 to 150bp until they reach a critical stage, known as the 'hayflick limit'. At this point telomeres are unable to recruit shelterin proteins or form the 'T' loop and the cell is put into a senescent state or apoptosis occurs. Telomere shortening can also be caused by external environmental factors, such as reactive oxygen species. If the cell by-passes the hayflick limit, then genomic stability can be compromised; in many cancers and cell lines telomeres have reached this critical point and activation of telomerase maintains telomere length and counteracts shortening. Telomerase is a complex containing the enzymatic component hTERT, Dyskerin and the telomerase RNA component (TERC), which can produce additional TTAGGG repeats at the ends of telomeres[173]. Other telomere related complexes include the RAD50-MRE11-NBS1 (MRN) complex, as well as multifunctional proteins including RPA1, hEST1A KU70 and KU80, which are also thought to interact with DNA damage response proteins [174]. Telomerase is active in germ cells, stem cells, stimulated lymphocytes and 85% of human cancers, although this does not always prevent erosion, as seen in lymphocytes and stem cells.[175] Telomere length has been investigated in many types of cancers, including CLL[176–180]; telomeres are abnormally short compared to healthy cells and telomere length is thought to associate with survival outcome. A pivotal study, by Hultdin *et al* [179], revealed that there was an association between telomere length and *IGHV* status in CLL, where long and short telomeres (<6kb) associate with M-CLL and UM-CLL, respectively. Consistent with this finding, was an association between short telomeres and a worse survival. The authors speculated that as telomerase becomes activated during the germinal centre (GC) reaction, and elongation of the telomeres occurs in GC B cells, a difference between the telomere length and the two CLL subsets is expected[179]. Grabowski *et al* suggest that telomere length is better at predicting survival than *IGHV* status [180]. Additional studies have confirmed these findings and also associated short telomeres with genomic complexity, poor risk markers such as high ZAP-70/CD38 expression, 17p

deletion and *TP53* abnormalities, 11q deletion and mutations in *NOTCH1* and *SF3B1* [128] [178]. Lin *et al*. investigated dysfunctional telomeres in CLL using the single telomere length analysis (STELA) technique [181]. They demonstrated that in patients with a poor prognosis, telomere length was similar to *in vitro* cells undergoing crisis; a cellular phase of high genome instability. Cell crisis has been associated with defects in *ATM* and *TP53*, which allow cell cycle checkpoints to be bypassed, leading to high telomere attrition and dysfunction [182]. Augereau *et al* studied stage A CLL patients, finding that telomere dysfunction, in the absence of telomere shortening was associated with a significant down regulation of the *TPP1* and *TIN2* shelterin genes, suggesting that genome integrity in CLL is affected by protein complex structure as well as telomere length [183]. Poncet *et al*., carried out expression analysis on telomerase components, shelterin proteins and other proteins thought to be involved in telomere maintenance in CLL patients [174]. They found dysregulation of most telomeric proteins, including reduced expression of hTERT, Dyskerin, TRF1, hRAP1, POT1, hEST1A, MRE11, RAD50 and ku80, and increased expression of TPP1 and RPA1, suggesting a role for telomere biology in the pathogenesis of CLL. These observations and investigations in CLL have shown that telomere length can be used as an accurate prognostic marker, however, it is yet to be used extensively due to the lack of clinically applicable methods for measuring telomeres. Protection of telomeres 1 protein (POT1) forms a critical component of the telomeric shelterin protein complex that restricts the access of telomerase to the telomeric single strand overhang at the ends of each chromosome, providing a capping function for chromosomal integrity and genomic stability[184]. POT1 proteins contain two N terminal oligonucleotide/oligosaccharide-binding (OB) domains, and a C terminus that binds to TPP1[185]. Deletion of the *POT1* gene lead to the rapid loss of telomeric and subtelomeric DNA, followed by chromosome end fusions in fission yeast. Although these effects were not seen in knockdown experiments of human *POT1*; indeed, telomeres were elongated and telomere fusions were rare, decreased proliferation and genome instability were shown[184]. POT1 can act as either an inhibitor or an activator of telomerase, dependant on the proteins' location on the telomeric DNA. If bound to the telomeric 3' single-stranded DNA overhang, then telomerase cannot gain access to the DNA for replication [186]. However, POT1 can be recruited to form a heterodimer with TTP1 that promotes telomerase recruitment therefore activating telomere elongation[174,186]. The TTP1 heterodimer binds to TRF1 and TRF2-RAP1 via TIN2, completing the shelterin complex[186]. Mutations in the *POT1* gene, identified by exome sequencing of CLL patients, are one of the first mutations to be identified in shelterin complex components in any human cancer[131,186]. Ramsay *et al*. analysed 127 CLL patients' exomes and further Sanger sequenced 214 patients for *POT1* mutations, finding a mutational incidence of 3.5% in the total cohort, which were enriched to 9% in UM-CLL cases[185]. All mutations were heterozygous and the majority (9/12) occurred in the region encoding the two OB folds that affect

the ability of the POT1 protein to bind to telomeric DNA. These findings are reflected in the annotated database of all human cancers, COSMIC, with 19/25 mutations occurring in this region[185]. Cell line studies confirm that with POT1 unable to bind to the DNA, telomerase is able to gain access more readily and lengthen telomeres [185,186] (Figure 48). Accounting for gene size, number of mutations and codon composition, *POT1* is stated to be the second most frequently mutated gene in CLL (*SF3B1* being the most frequent)[185]. In a study of clonal and subclonal evolution throughout CLL disease progression, Landau *et al* identify *POT1* mutations as early events and likely driver mutations in CLL pathogenesis[132,186].

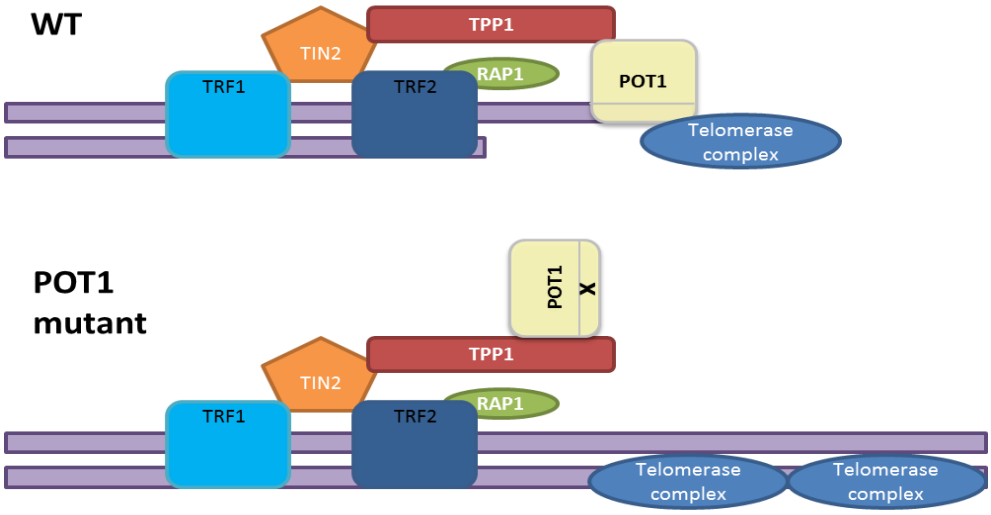


Figure 48. *POT1* mutations can affect protein interaction to promote telomere extension and chromosome stability. Figure adapted from Chang 2013.[186]

6.2 Aims

As no novel, recurrently mutated candidate genes were identified in the discovery study (previous chapter), under studied target genes were chosen from the literature that seemed to have value in CLL.

- To screen candidate genes currently described in the literature, in large cohorts of CLL patients. Genes chosen include:
    - BIRC3
    - MYD88
    - POT1
    - U2AF1
- To perform functional validation of the importance of selected genes.
    - mRNA analysis
    - Protein analysis

6.3 Methodology

6.3.1 **Patient cohorts**

Cohort 1 was chosen to explore the mutations in early disease. This cohort (n=206) included previously untreated patients whom had a diagnosis of either Binet stage A CLL [n=116] or cMBL [n=-90] according the standard morphological and immunophenotypic criteria. The samples were taken within 6 months of diagnosis and a second time point was analysed in 84 patients with a median 72 months (range 24-145 months) between samples. Follow up ranged from 1 to 35 years (median 10 years). A large proportion of the patients (53%) remained in stable Binet Stage A or cMBL, while 32% of patients with cMBL evolved to Binet stage A. 33% of cases progressed to Binet stage B or C disease and required treatment; 23% of these from cMBL and 40% from Binet stage A CLL.

Cohort 2 was chosen to explore the relationship between *BIRC3*, 11q and *ATM*: This cohort of CLL patients (n=293), diagnosed according to standard morphological and immunophenotypic criteria, included patients from the UK LRF CLL4 trial with an enrichment for (n=80) 11q deleted patients. This allowed more significant associations between the 11q deletion and the mutational status of *ATM* and *BIRC3* genes to be studied. The del11q patients were sampled again, subsequent to the development of progressive disease, with a median time from diagnosis of 3.2±5.2 years (±1 standard deviation (SD); range 1 month-27 years).

Cohort 3 was chosen to explore the frequency of splicing gene mutations in a clinical trial setting. This cohort of CLL patients (n=100) were from the UK LRF CLL4 trial and as before were diagnosed according to standard morphological and immunophenotypic criteria.

All cohort studies were implemented in accordance with the Declaration of Helsinki and were ethically approved by the Regional Ethics Committee (REC).

6.3.2 High resolution melting analysis

Whole genome amplified (WGA) DNA (Chapter 3.7) from 200 CLL patients was screened for mutations in the *BIRC3, MYD88* and *POT1* genes, using HRM (Chapter 3.9). Sanger sequencing was then used to validate any potential mutations identified by this high-throughput screening technique. Although *BIRC3* has 10 exons, mutations have only been reported in exons 7 and 10, therefore screening was restricted to these exons[99]. Exons 4 to 9 of the *POT1* gene, which encode the telomere binding domain were screened and the mutational hotspots on exons 3, 4 and 5 of the *MYD88* gene were screened. Two HRM assays were designed to cover the hotspot mutations in *U2AF1* (codons 34, 157) and one was designed to cover the codon 95 mutation within the *SRSF2* gene. Primers were designed, using Primer 3 software (Appendix x). For HRM the optimum annealing temperature was approx. 60°C, and the product length was between 150-250bp. Longer product may decrease assay sensitivity as multiple melting domains can occur. The optimum temperature for Sanger primers was 55°C, and product length was between 450-500bp long. Appendix G shows a list of all the primers, for both HRM and Sanger sequencing, designed for use in this study.

6.3.3 Sanger sequencing

PCR Primers for sequencing were designed (Chapter 3.7) to confirm abnormal HRM melting curves. Genomic DNA was used to confirm that variants identified by HRM were real and not an artefact of WGA. Additionally, a proportion (n=35) of the HRM normal melt profiles were Sanger sequenced to check no mutations were missed by HRM. No mutations were identified in these cases.

6.3.4 Copy number analysis

Archival SNP 6.0 data (Affymetrix, Santa Clara, CA, USA), aligned onto the human genome sequence (GRCh37) and analysed in Partek Genomics Suite (Partek Inc, Missouri, USA) as reported previously, was available for a proportion of the CLL patients screened.

6.3.5 RT-QPCR

A qRT-PCR -UPL probe assay was used to assess the gene expression of *BIRC3* (#80, cat. no. 04689038001, ROCHE) in a cohort of the LRF UK CLL4 clinical trial patients (n=80) (see methods). The reference assays used was the housekeeping gene; 18s. RNA was extracted and the quantity and quality was assessed using the Agilent RNA chip and nanodrop (see Chapter 3:).

6.3.6 Statistical analysis of expression data

Statistical analysis of the *BIRC3* expression data was performed using IBM SPSS Statistics (V.21). The expression data was not normally distribution (Figure 49), therefore the Mann-Whitney statistical test was used to compare means between different groups and assign significance ($p \leq 0.05$). Patients were stratified into high vs low expression of *BIRC3* using various measures of *BIRC3* expression (mean, median, mean plus 1 SD). Univariate survival analysis was conducted using Kaplan-Meier analysis, and the log rank test was used to determine significance for overall survival (OS) and progression-free survival (PFS). Overall survival (OS) was defined as the time from randomisation to death, or to the last follow-up date (August 2012) for survivors. Progression free survival (PFS) was defined as time from randomisation to relapse needing further therapy, progression or death; or to the last follow-up date (Oct 2010; final LRF CLL4 PFS update) for those with no progression or death.

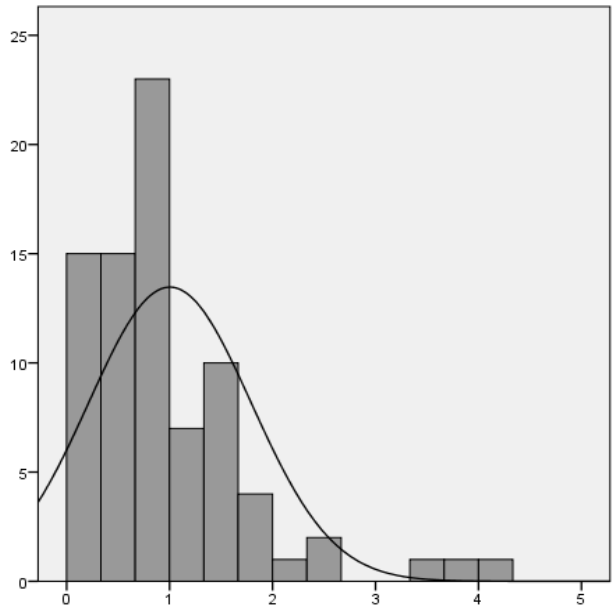


Figure 49. Data distribution of BIRC3 expression from rtqPCR assay of 80 CLL4 patients. Data has a skewed distribution, so must be analysed using non-parametric tests.

6.3.7 **Statistical analysis of survival data (cohort 2)**

Analysis was kindly performed by Dr Matthew Rose-Zerilli. Survival analysis was conducted for PFS and OS using Kaplan meier analysis and the log rank statistical test for univariate analysis and cox regression. A p value of ≤0.05 was considered significant.

6.3.8 **Protein analysis**

Protein was extracted from B-cells stored in DMSO (-80°C) from 9 CLL patients. Selected patients had either an 11q deletion (n=3), a mutation of *BIRC3* with concomitant 11q deletion (n=2), a breakpoint deletion within the *BIRC3* gene (n=1) or were wildtype for 11q abnormalities (n=3). Along with the patient samples, a positive multiple myeloma control cell line was used (KMS-12-PE).

Western blots were run as described in the methodology section. The c-IAP2 (cat# 3130P) and NF-κB2 p100/p52 (cat#4882P) antibodies (New England Biolabs) were used to detect the BIRC3 and NFκB2 proteins.

6.4 Results

6.4.1 The mutational landscape

6.4.2 Cohort 1

Cohort 1 was composed of previously untreated Binet stage A/cMBL patients (n=206), of which 177 were successfully screened for *MYD88*, *BIRC3* and *POT1* mutations using HRM analysis, and Sanger sequencing was used to validate the mutations identified. For 30 patients a second sample was also available for screening. Table 23 details the mutations detected and confirmed through this analysis.

Table 23. Gene mutations confirmed by Sanger sequencing in the diagnostic CLL cohort 1, including the available sequential samples.

PID	IGHV	Date sample	gene	exon	mutation	genomic	coding	amino acid
568	U	24/4/06	BIRC3	exon 10	indel/frameshift	g.19464delC	c.1639delc	p.Q547Nfs*21
248	M	20/8/96	MYD88	exon 3	missense	g.2057G>T	c.649G>T	p.V217F
248	M	7/3/05	MYD88	exon 3	missense	g.2057G>T	c.649G>T	p.V217F
335	U	28/3/00	MYD88	exon 3	indel (in frame)	g.2002_2004delAAA	c.594_596delAAA	p.N199del
339	M	26/3/02	MYD88	exon 5	*	*	*	
339	M	03/01/96	MYD88	exon 5	missense	g.2673T>C	c.794T>C	p.L265P
2414	M	28/3/01	MYD88	exon 5	Missense	g.2673T>C	c.794T>C	p.L265P
50	M	27/07/05	POT1	intron 18-19	splice site	g.104749T>G		
100	U	21/1/93	POT1	exon 10	*	*	*	
100	U	18/7/97	POT1	exon 10	Missense	g.77006T>C	c.1462T>C	p.L288P
335	U	28/3/00	POT1	exon 10	Missense	g.76868A>G	c.1321A>G	p.Y242C
568	U	24/4/06	POT1	exon 5	UTR	g.32806A>G	c.595A>G	.
568	U	1/4/97	POT1	exon 5	UTR	g.32806A>G	c.595A>G	.
2355	U	25/08/04	POT1	intron 5-6	Splice site	g.37603G>C		.
2355	U	24/01/01	POT1	intron 5-6	Splice site	g.37603G>C		.
2938	U	30/1/03	POT1	exon 6	missense	g.37710A>T	c.116A>T	p.K39I
3873	U	15/08/06	POT1	exon 5	UTR	g.32777A>G	c.566A>G	.

*no genomic DNA to confirm mutatations

POT1 mutations were found in 8 patients (4%). No recurrent or hotspot mutations were seen and the mutations targeted both intronic and exonic (exons 5- 19) regions (Figure 50). Two mutations targeted the important telomere binding domain, and 2 were found in the untranslated region (UTR) of exon 5. These UTR mutations may have an effect on translation of mRNA and protein production; the g.32806A>G mutation (PID 568) is expected to cause an out of frame AUG to ACG transition, resulting in increased protein production and the g.23777T>G mutation (PID 3873) causes a small change to the context of the initiation codon of the POT1 protein and may also have an effect on protein expression. The splice site mutation (g.37603G>C (PID 2355)) in the intron 5-6 sequence may affect splicing of the protein, resulting in expression of alternative isoforms of POT1. The remaining coding mutations are also expected to have a damaging effect on protein function as predicted by the SIFT scores (0.01). No deletions of chromosome 7q31.33, the genomic location of the *POT1* gene, were found in patients with archival SNP 6.0 data. A proportion of the *POT1* patients (5/8 cases) had IGHV unmutated genes and 4/8 progressed to needing treatment.

Figure 50. Gene map of *POT1* showing all 19 exons (shaded boxes) and highlighting the location of the important protein (telomere) binding domain.
The mutations confirmed in CLL patients are indicated by red lollypop

A single *BIRC3* mutation (g.19464delC (PID 568)) was found in this cohort, within the RING protein domain encoded by exon 10. This frameshift deletion is predicted to result in a truncated form of the BIRC3 protein. Of the 4 *MYD88* mutations (2.3%) identified, 2 (PIDs 339 and 2414) were the recurrent p.L265P mutation in exon 5, previously discussed. Second time-point samples were available for 5 patients with a *MYD88*, *BIRC3* or *POT1* mutation; all mutations were found in both samples (Table 23).

6.4.3 Cohort 2-*BIRC3* mutations

Screening for *BIRC3* mutations, as previously described, was also performed on 293 patients with progressive disease (cohort 2). *BIRC3* mutations were identified in 4 patients (<1%)

Figure 51 and Table 24). Two of the mutations (p.R434Sfs*2, PID 73 and 78) were a recurrent 2bp insertion within exon 7, which causes a frameshift and is predicted to result in truncated protein transcripts leading to loss of gene function. These mutations are located in a highly conserved

region, the caspase activation and recruitment domain (CARD), and result in loss of the RING domain. The 2 other mutations are also in highly conserved regions of exon 10, encoding for the RING domain of the protein, which is essential for the proteasomal degradation of *MAP3K14*; the g.19549delT mutation (PID 263) results in a premature stop codon, truncating the *BIRC3* protein; the g.19533_19535delATT (PID 176) mutation results in loss of codon 570. This codon codes for isoleucine, a hydrophobic amino acid, which is involved in the binding and recognition of hydrophobic lipid ligands. Structural profiling showed that codon 570, is part of a binding site on the *BIRC3* protein (Figure 52). Loss of this binding site is likely to affect protein function.

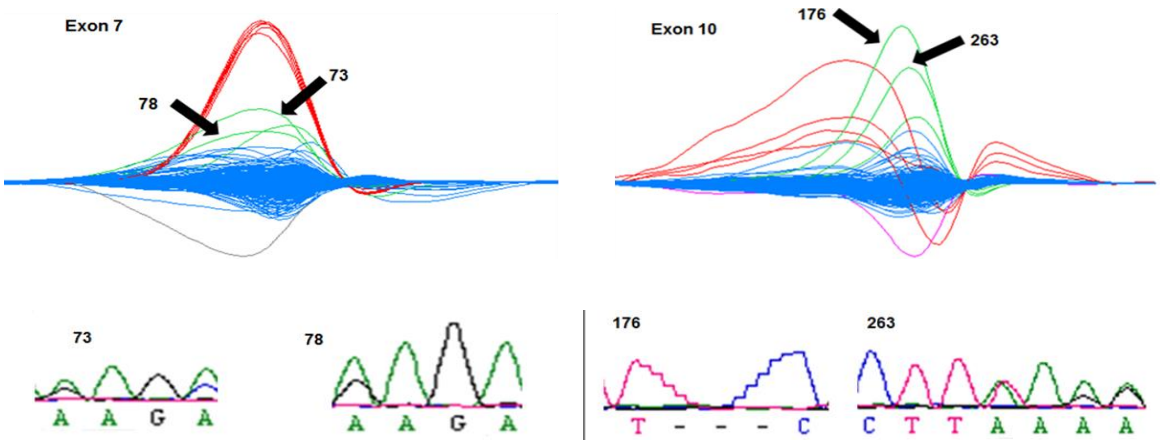


Figure 51. Raw HRM data for the 293 progressive patients screened with the BIRC3 Assay.
Blue lines indicate normal samples, whereas red and green lines represent patients with a different melting temperature, and therefore possible mutations. In this case the red melting profiles were caused by a normal SNP. Sanger sequencing profiles show the confirmation of the variants in patients 73,78,176 and 263.

Table 24. Cohort 2: *BIRC3* mutational nomenclature that were detected using HRM and confirmed by Sanger sequencing

ID	Exon	Nucleotide change	Amino acid change	Type of mutation	Predicted functional change
73	7	g.13755_13756delAG	p.R434Sfs*2	frameshift deletion	Truncated protein
78	7	g.13755_13756delAG	p.R434Sfs*2	frameshift deletion	Truncated protein
176	10	g.19533_19535delATT	p.I570del	inframe deletion	Loss of binding site
263	10	g.19549delT	p.L585*	Stop	Truncated protein

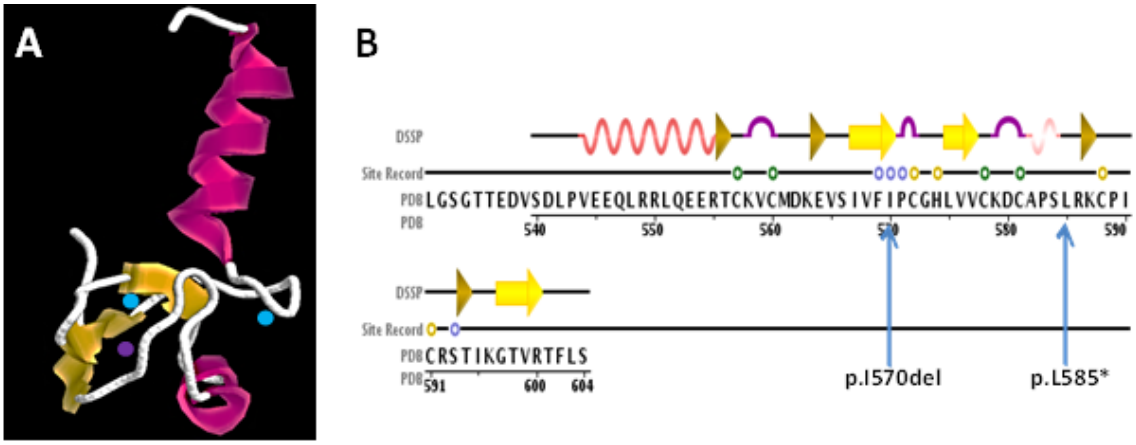


Figure 52. Protein structure of BIRC3 and the location of mutations within structure.
Mutations (blue dot) are found in the important RING domain and (codon 57) binding site (purple dot). A) Protein structure of RING domain B. Liner protein structure including amino acid sequence and mutations found in this study.

6.4.4 *BIRC3* mutations accompanied deletions of the retained allele

Analysis of the archival SNP 6.0 data revealed that all 4 patients with a *BIRC3* mutation had deletion of the remaining allele, suggesting biallelic inactivation of the gene, resulting in complete loss of *BIRC3* protein function. These deletions also encompass many other genes on 11q including *ATM*. A further 48 patients with copy number profiling had an 11q deletion that encompasses the *BIRC3* gene and four more had a proximal deletion breakpoint within the *BIRC3* gene itself (5% of the 11q deleted patient cohort (Figure 53), which potentially results in truncation of the protein and loss of the crucial RING domain activity.

These 4 patients all experienced progression of their disease, requiring treatment (either chlorambucil or fludarabine/chlorambucil) with 179 months of diagnosis (mean Treatment Free Survival (TFS)=67 months, range=2-179 months), and did not have the poor prognostic del(17p) or *TP53* mutation. However, 3 of these 4 patients also had an *ATM* mutation and 3 had a 13q deletion.

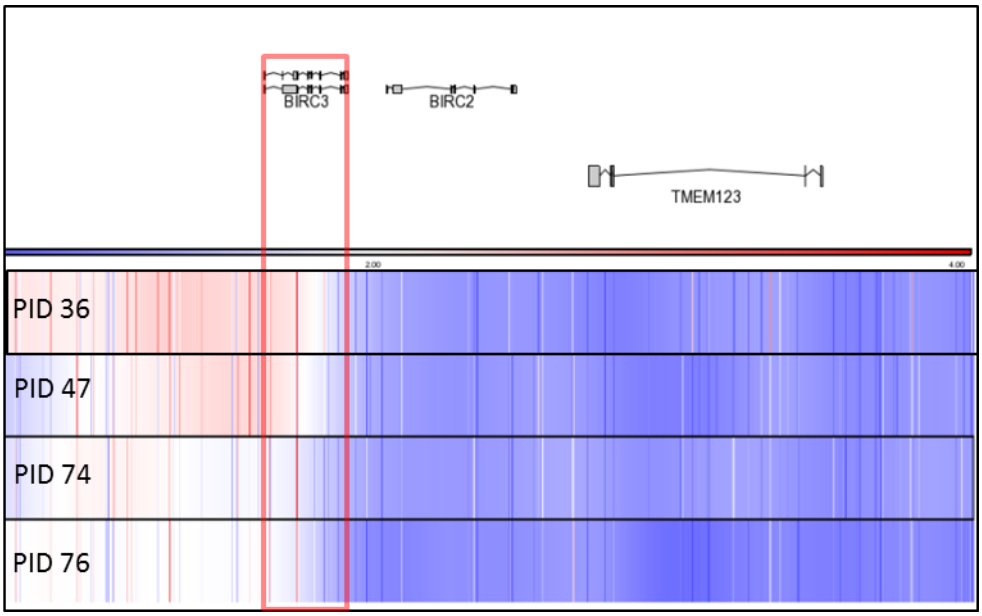


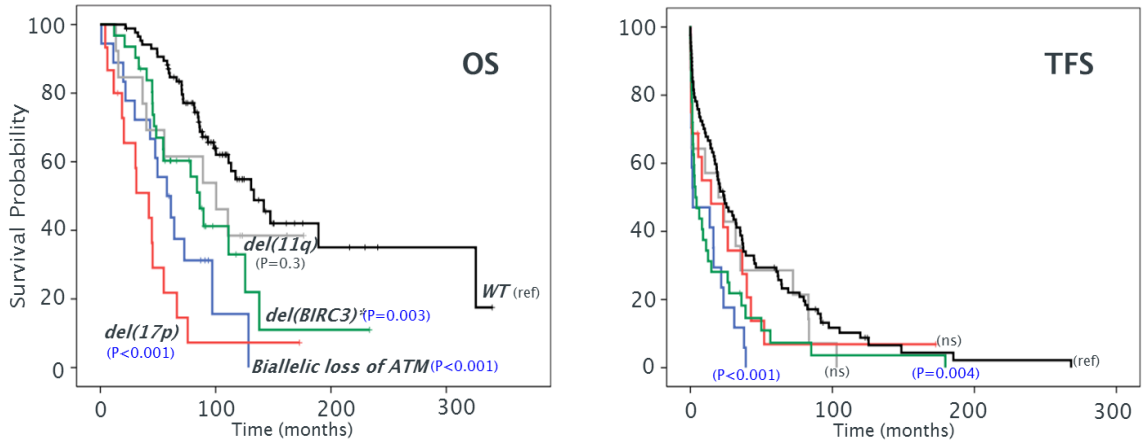
Figure 53 SNP6.0 heatmap profiles for 4 patients with a deletion breakpoint within the BIRC3 gene.
Each horizontal bar represents an individual patient; blue, red and white indicates loss, gain and normal copy number, respectively. The location of the BIRC3 gene is indicated by the red box.

6.4.5 *ATM* mutations rather than *BIRC3* impact survival

Subsequent survival analysis (Courtesy of Dr Matthew Rose-Zerilli) was performed on a subset (n=133) of the cohort 2 patients. Of these patients, 36 had loss of both the *BIRC3* and *ATM* genes, 2 also had a *BIRC3* mutation and 12 had an *ATM* mutation. The deletions of *BIRC3* always co-occurred with *ATM* deletion (Table 25). The 11q deletion reduced OS and PFS to 53 and 20 months, respectively, though this was not significant (Figure 54). *BIRC3* mutated and deleted patients further reduced survival to 53 months OS (p=0.003) and 17 months for PFS (p=0.004). The *ATM* mutated and deleted cases further reduced survival, with a reduced PFS 10 months (p<0.001) and 42 months for OS (p<0.001). 17p deletion gave the most reduced OS for patients harbouring this abnormality (p<0.001). Comparing cases with a *BIRC3* deletion/mutation with *ATM* mutation/deletion showed that the *BIRC3* patients had significantly longer survival than the *ATM* patients; OS was 76 vs. 42 months (p=0.05) and PFS was 28 vs. 10 months (p=0.01).

Table 25. *BIRC3* disruption versus *ATM* status in CLL4 patients.

	BIRC3 WT	*BIRC3* deleted	*BIRC3* del+mut	Total
ATM mutated	15	0	0	15
ATM deleted	5	22 (55%)	3	30
ATM del+mut	5	18 (45%)	0	23
Total	25	40 (100%)	3	68



Abnormality	Events/Cases	Median OS Sur. (mths)	95% CI	P-value
Wildtype	38/86	133	103-164	ref
del(11q)	8/13	100	36-165	0.30
del(*BIRC3*)*	19/31	86	71-101	0.003
Biallelic loss of *ATM*	14/18	58	35-80	<0.001
del(17p)	13/15	42	21-63	<0.001

Abnormality	Events/Ca ses	Median TFS Sur. (mths)	95% CI	P-value
Wildtype	81/87	24	14-34	ref
del(11q)	14/14	20	0-44	0.41
del(17p)	14/16	15	0-42	0.31
del(*BIRC3*)*	31/32	4	0-8	0.004
Biallelic loss of *ATM*	17/17	2	0-19	<0.001

Figure 54. Biallelic *ATM* disruption rather than *BIRC3* disruption impacts more negatively upon overall and progression free survival.
***del(*BIRC3*) includes the *BIRC3* deleted and mutated cases.**

6.4.6 *BIRC3* gene expression analysis

BIRC3 expression was analysed in a subset of cohort 2 patients (n=80). *BIRC3* (and *ATM*) was deleted in 24 of the patients, and a further 5 patients had *ATM* loss without concurrent *BIRC3* deletion. Two patients had *BIRC3* mutation and deletion. The relative fold change, using the housekeeping gene, *18s*, as a reference, ranged from 0.15 to 4.33 (mean 1.003, median 0.84, standard deviation 0.79) (Figure 55). As expected, those patients with a *BIRC3* deletion, had a reduction in *BIRC3* expression (Figure 56), although this was not quite significant (p=0.055). When *BIRC3* expression was correlated to other clinically relevant biomarkers an interesting association with *ATM* was identified; patients with a biallelic inactivation (deletion and mutation) of *ATM* showed a reduction in *BIRC3* expression compared to wild type *ATM* patients (p=0.053) and *ATM* deleted patients (p=0.060) (Figure 56). Although these observations were not quite significant,

there was a trend towards *BIRC3* loss in the *ATM* deleted patients. Patients were then stratified as having high or low expression of *BIRC3* based on the mean, median and mean +1SD. There was no significant effect on survival (PFS or OS) using any of these groupings.

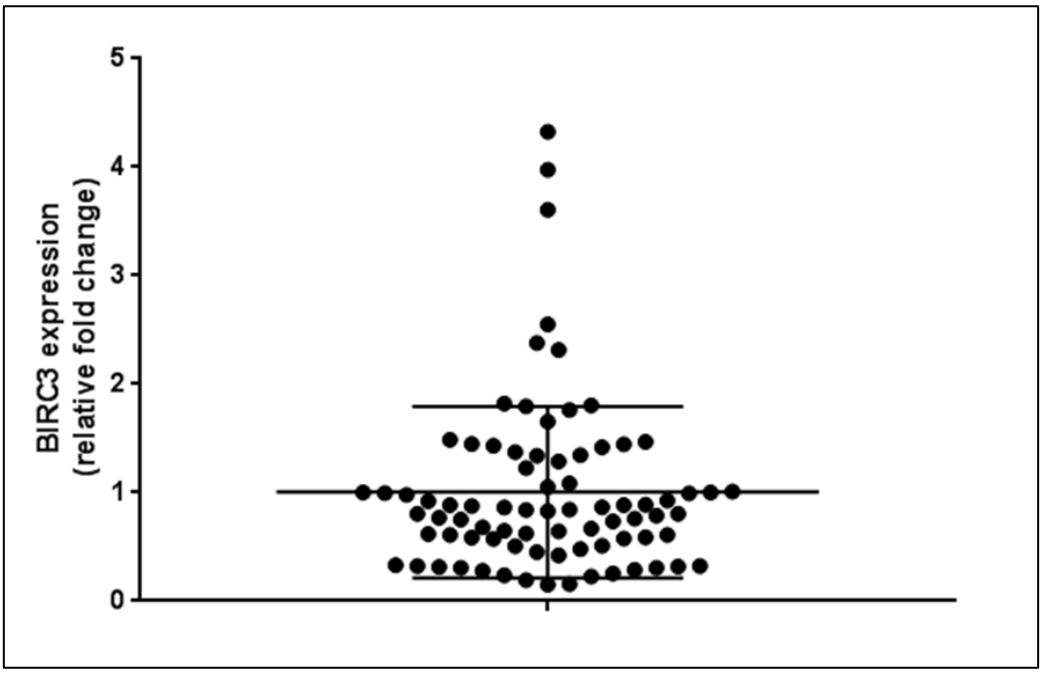


Figure 55. mRNA *BIRC3* expression in 80 CLL4 patients, relative to the housekeeping gene, 18S. The relative fold change ranged from 0.15 to 4.33 (mean 1.003, median 0.84, standard deviation 0.79).

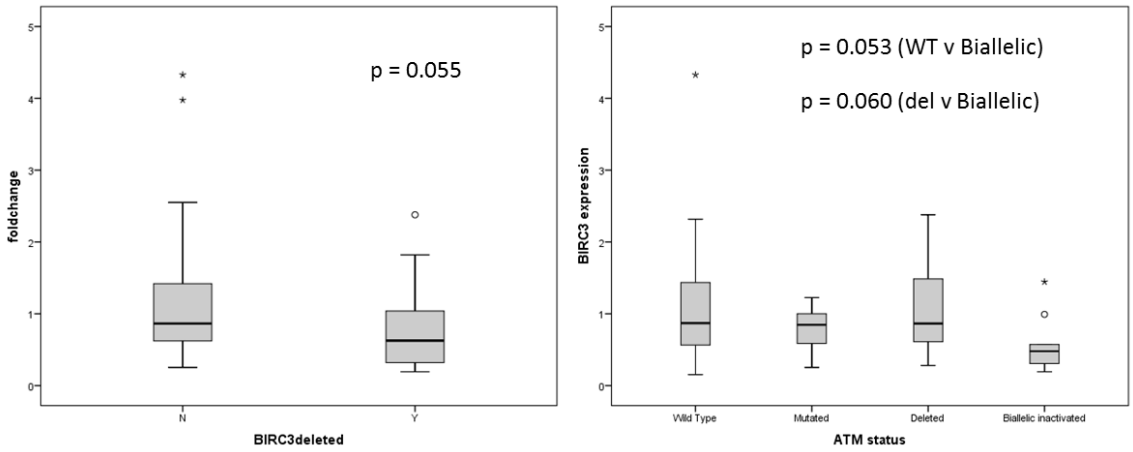


Figure 56. mRNA *BIRC3* expression categorised according to 11q status.
***BIRC3* expression was reduced in *BIRC3* deleted patients vs WT patients (p=0.055). *BIRC3* mRNA expression was also reduced in patients with biallelically inactivated *ATM* vs WT patients (p=0.053).**

6.4.7 BIRC3 protein analysis in 11q abnormal patients

Protein analysis of the BIRC3 and NFKB2 proteins using western blotting was carried out on 9 CLL patient samples that all had either WT, deleted and/or mutated *BIRC3* gene status. The western blot showed the expected 70KDa band but also had additional 52-55 KDa and 42-45 KDa bands; these were seen in the majority of the samples. In addition, the NFKB2 protein analysis did not show the expected p100 band (Figure 57), though the active form was present (p52). The expected molecular weight (MW) of the mutant proteins were calculated using an online protein MW calculator (http://www.bioinformatics.org/sms/prot_mw.html). The expected MW of the mutant proteins were; (PID 176) 68.27k Da, (PID 263) 66.12k Da, and (PID 73/78) 49.29k Da, whilst the normal (BIRC3 001 transcript) is 68.38k Da.

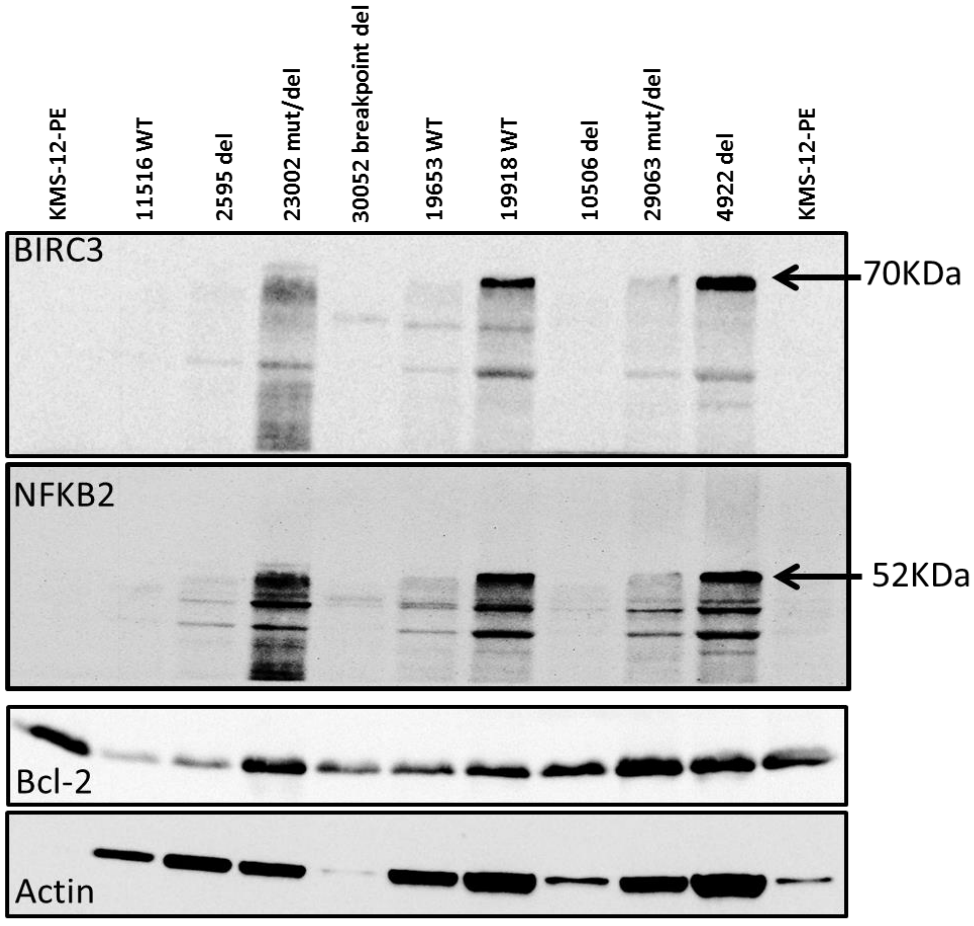


Figure 57. Protein (western blot) analysis of 9 CLL patients harbouring genomic abnormalities implicating the NF-kB pathway. Patient samples included; 11q deletion (n=3), *BIRC3* mutated with an 11q deletion (n=2), *BIRC3* breakpoint deletion (n=1), and wildtype patients (n=3). The KMS-12-PE cell line was used as a control sample. Actin and Bcl-2 were used as loading controls. p52 (active form of NFKB2) was detected but no p100, this could indicate that non canonical NF-kB signalling was active, though the presence of additional low MW bands suggest sample degradation. The 70KDa band indicates BIRC3 protein is present, again additional low MW bands suggest protein degradation.

6.4.8 Splicing gene mutations occur rarely in CLL

HRM analysis for the mutational hotspots in *U2AF1* and *SRSF2* (codon 34 and 157, and codon 9, respectively) was performed in 100 CLL4 patients (cohort 3). An abnormal melting curve was identified in a single patient (1%) in codon 34 of *U2AF1*. This patient was Sanger sequenced and the variant confirmed (Figure 58). The mutation was a missense, Serine to Phenylalanine amino acid substitution (p.S34F). This is a highly conserved residue and the variant has a low SIFT score (0) and a high PolyPhen score (0.998), suggesting that the variant has a detrimental impact to protein function. The patient was male, aged 65 years and binet stage C at diagnosis, with unmutated *IGHV* genes and a trisomy 12, whom progressed to needing chlorambucil treatment (PFS 2 months) and subsequently died (OS 65 months).

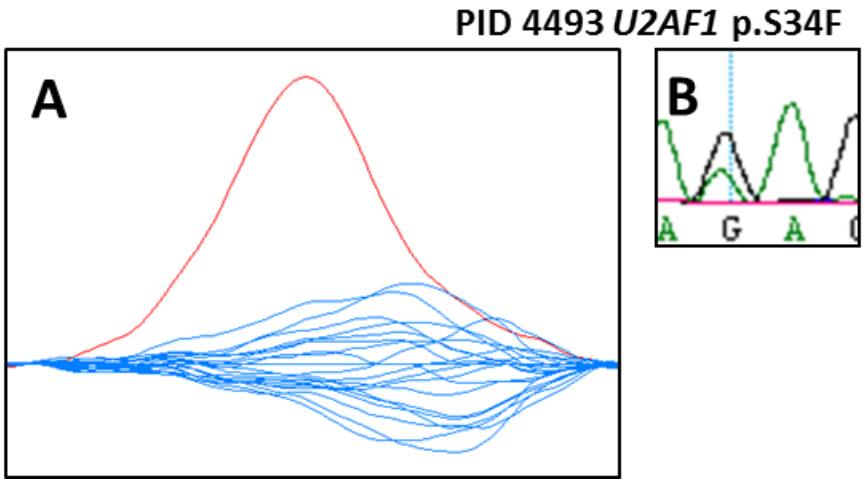


Figure 58. High resolution melting analysis was used to identify mutations in the *U2AF1* gene. Genomic DNA from patients with abnormal curves (red line) were then sequenced. One variant was identified in one patient (PID 4493).

6.5 Discussion

Splicing, the NF-κB pathway and telomere regulation are known to be disrupted in many cancer types and key components have been implicated in CLL and other related B cell neoplasms such as SMZL.[19] As members of these key pathways, with recurrent mutations discovered by recent next generation sequencing studies; *POT1*, *BIRC3*, *MYD88*, *SRSF2* and *U2AF1* were chosen for this study, as interesting candidate genes in CLL pathology [130,131] High resolution melting analysis, Sanger sequencing and archival SNP 6.0 data was used to explore these genes in an extended cohort of CLL patient's, with vital clinical data available. As well as trying to establish the clinical impact of the mutations, the functional consequences of mutations in *BIRC3* were assessed.

Splicing is the process of removing introns during RNA processing. In MDS and CLL mutations of the splicing machinery are common, up to 76% and 18%, respectively.[118] These have a detrimental impact upon survival and show that splicing may be important in the pathogensis of these diseases.[155] The *U2AF1* gene encodes a protein that forms part of the spliceosome complex, and functions as a U2 auxiliary factor, recognising the AG splice acceptor dinucleotide at the 3' end of introns.[159] One patient out of 100 screened (1%) harboured a *U2AF1* mutation and no mutations were seen in *SRSF2*. The *U2AF1* mutation affects codon 34 and is a recurrent variant identified frequently in MDS patients (~9%). This codon forms part of the zinc finger domain and the variant identified in this study has been shown to impact OS negatively in MDS patients and has been associated with an increased risk of disease transformation to AML. [154,156,159] Cell lines expressing this *U2AF1* variant have shown a reduction in cell proliferation but an increase in splicing and exon skipping. [154,156,159] In CLL the pathogenesis is less clear as *U2AF1* is very rarely mutated. Indeed, no mutations were identified in the largest known WES/WGS study in CLL.[75] This discrepancy may be due to cohort composition; the Puente *et al.* cohort included patients with MBL (pre-CLL), whilst the CLL4 cohort screened here represents a more progressive group of patients. Though more recently in a targeted deep sequencing study investigating 13 of the CLL driving genes (*TP53*, *ATM*, *SF3B1*, *NOTCH1*, *POT1*, *BIRC3*, *MYD88*, *KRAS*, *NRAS*, *U2AF1*, *BRAF*, *BCOR* and *SETBP1*) in 180 untreated CLL patients, and identified 1.1% of their cohort harbouring a *U2AF1* mutation.[187] Therefore, *U2AF1* mutations, although very rare, may well be associated with CLL and possibly a later disease stage though this will need to be investigated in a larger, enriched cohort of patients.[75]

The MYD88 protein is involved in the interleukin-1 receptor/Toll like receptor activation of canonical NF-κB signalling. Whilst mutations of the *MYD88* gene are detected in 2-10% of CLL patients the clinical impact remains controversial as evidence for independent prognostic

significance is lacking. In conjunction with published data, this study identified mutations in 2.3% of cases, including previously described hotspot mutations (pL265P n=2 and V147L n=1).

The POT1 protein is a critical component of the telomeric shelterin protein complex, found at the ends of each chromosome, that provides an essential capping function, protecting chromosomal integrity and providing genomic stability [184]. Published studies have only described mutations in exons 5, 6, 7 and 9, and splice site mutations between exons 4 and 5 in CLL. In accordance with this, mutations were detected in exons 5 and 6 in this cohort. In a large NGS CLL study, the *POT1*gene was identified as being frequently mutated (4.8%) and only in IGHV unmutated patients [131]. Similarily in this data *POT1* mutations were identified in 4% of the cohort, though these were not restricted to IGHV UM cases. The impact of *POT1* mutations in a CLL clinical trial have yet to be published, although a study in gastric cancer found that increased *POT1* expression was associated with an advanced tumour grade (increased tumour invasion and metastasis). RNAi experiments on gastric cancer cells had a decrease in cell proliferation, invasion and increased apoptosis [188]. *POT1* expression is associated with telomerase activation and inhibition of telomerase leads to increased telomere erosion and cell death. As such, telomerase has been identified as a promising target for therapy [188] [178].

The BIRC3 protein negatively regulates MAP3K14, a key activator of the non-canonical NF-κB signalling pathway, an important regulatory pathway involved in immune response and development. Along with TRAF2 and TRAF3, BIRC3 forms a vital part of a protein complex that allows the proteosomal degradation of MAP3K14. Disruption of this negative regulatory protein complex is sufficient to up regulate the signalling pathway, as discovered in multiple myeloma and cell lines [164]. *BIRC3* was first identified as a potential target gene in CLL in an exome sequencing study, and subsequently found to be disrupted by either a mutation or deletion in 4% of cases, and associated with chemorefractoriness (in wild type *TP53*) [99,130] [26]. The mutations that were found in this study were within the coding region and are predicted to occur within the RING and CARD domains, causing truncation of the protein by loss of the C-terminal RING domain, which is essential for the ubiquitin ligase activity and proteosomal degradation of MAP3K14. These findings are concordant with other studies that also found mutations in the same exonic regions, with similar consequences [19,99]. The impact of *BIRC3* mutations was examined in relation to *ATM* mutations, as both of these genes lie within the 11q deleted loci and have been implicated in CLL pathogenesis, and no study has yet investigated the impact of both genes. The SNP 6.0 data confirmed that *BIRC3* deletions were concurrent with *ATM* loss, though *BIRC3* mutations were independent of *ATM* mutations. This has been further validated recently in another published study exploring *BIRC3* alterations in CLL and ALL, using iFISH, MLPA and aCGH technologies,although this study did not take into account of gene mutations.[189] Despite that,

frequent *BIRC3* alterations were found in CLL (~20%) and all of the *BIRC3* deletions also had a concordant *ATM* deletion, additionally 3 *BIRC3* duplications with 2 of these being a MALT-ATM duplication, were also detected [189] The survival data from this chapter showed that *ATM* mutations (with or without deletion) impact more negatively upon survival than *BIRC3* mutations (with deletion). Although the *BIRC3* mRNA expression analysis did not yield any statistically significant results, the experiment did provide interesting leads to follow up, such as the reduction of *BIRC3* expression in patients with biallelically inactivated *ATM* (both a deletion and mutation of the gene). The poor quality of the RNA available and small number of patients in each grouping likely impacted this finding and further work to explore this possible correlation is warranted. Many studies have provided evidence to show that ATM is involved in NF-κB signalling [190–194]. A recent study suggested that ATM is a modulator of nuclear damage response in TNF induced NF-κB activation, and has a direct involvement in IκBα degradation through the recruitment of E3 ubiquitin ligases, as well as activation of RELA by phosphorylation of the Serine 276 residue[191] (Figure 59). One of key roles of the BIRC3 protein within the NF-κB pathway is as an E3 ubiquitin ligase, so one would presume that ATM could act as a modulator of BIRC3 and therefore loss of ATM would mean a reduction in BIRC3. The role of ATM in NF-κB signalling and indeed its role in BIRC3 regulation have yet to be fully elucidated, providing future avenues to explore. Protein analysis was performed on 9 CLL4 samples and cell lines used as controls in the literature[99]. Unfortunately, several problems were encountered; expected protein bands were missing or resolved at the incorrect weight, and multiple BIRC3 bands were resolved. These bands may represent different isoforms of the BIRC3 protein or the protein samples may have been extracted from poor quality, degraded material. Ideally proteins need to be extracted from viable cells stored in liquid nitrogen, that are protected from damage and degradation. Unfortunately, as the *BIRC3* mutated samples were from patients enlisted in the CLL4 trial (recruited between 1999 and 2004), only samples stored at -80°C were available. This would have compromised cell viability, and consequently protein quality and quantity. In addition to poor quality starting material, the housekeeping genes, BCL-2 and Actin, which were used as a loading controls for the western blot in this experiment, both gave variable results. This could have been caused by incorrect concentrations of the standards or else poor choice of controls. BCL-2, which in CLL has been shown to have variable expression across patients;it tends to be overexpressed leading to cell survival due to Bcl-2 having an anti-apoptotic function.[195–197] So as a loading control this is not ideal, future experiments would have to use a different control loading protein.

The protein analysis needs refining, but due to lack of quality samples, lack of informative samples and expertise this was not possible in the timeframe. It would be interesting to relate BIRC3 protein expression back to RNA expression and the *ATM* disruption status to see if the effect seen at RNA level is extended to protein level.

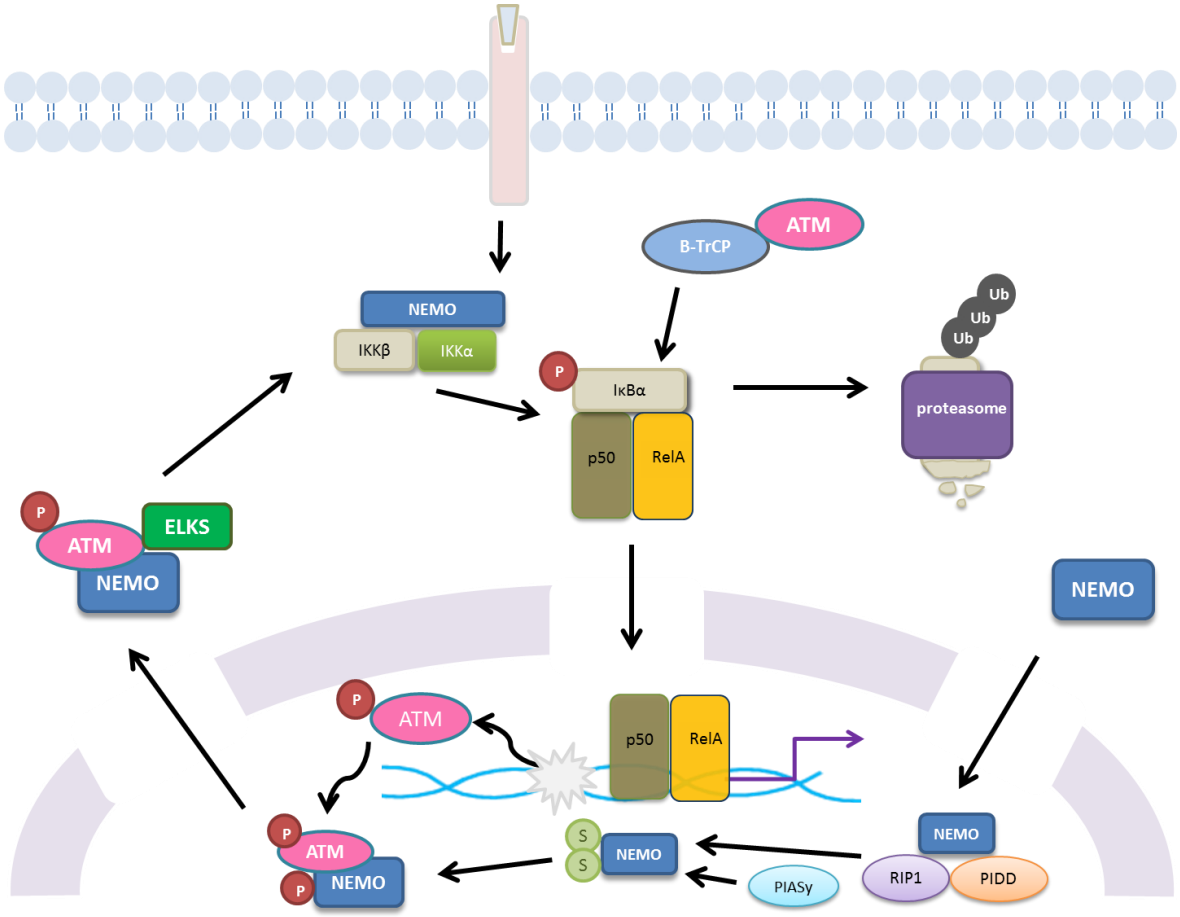


Figure 59. The proposed involvement of ATM within the NF-κB signalling pathways180–182. The ATM protein activates RELA to a response to DNA damage in TNF induced NF-κB activation as well as recruiting of E3 ubiquitin ligases (BIRC3) to degrade IκBα

Initial screening for *POT1*, *MYD88* and *BIRC3* mutations was performed in a cohort of 177 diagnostic CLL cases (cohort 1). Whilst 6 *POT1* and 7 *MYD88* mutations were identified only a single *BIRC3* mutation was seen. However, 4 *BIRC3* mutations were detected in the progressive cohort (cohort 2). This supports the published concept that as *POT1* and *MYD88* are typically clonal mutations, they occur early in the disease course and are driver mutations, whereas mutations that are sub-clonal at disease presentation, such as *BIRC3*, expand during disease progression. This description of CLL as a dynamic disease in which the mutation landscape can alter over time[94] is further corroborated by PID 568; a *POT1* mutation was detected in the

diagnostic sample, and a sequential sample taken 9 years later. However, a *BIRC3* mutation was present at the later time point only, suggesting clonal evolution.

With the genomes and exomes of many CLL patients already sequenced, the quest for novel, frequently mutated genes is coming to an end. Future analysis of the CLL genome is becoming more refined and the next step is to validate the occurance of less frequent mutations in large, clinically relevant cohorts to understand their impact on biological processes, their association with known biomarkers and effect on survival. In addition, the relevance of sub-clonal mutations in the context of clonal evolution needs further study.

Chapter 7: Using targeted deep sequencing technology in Chronic Lymphocytic Leukaemia

7.1 Introduction

Next generation sequencing as previously discussed has been widely applied to CLL due to the natural longevity of the disease and ease of accessibility to tumour material. This has allowed researchers to get a snapshot into disease heterogeneity at the differing stages, from diagnosis and low grade CLL through to progressive and refractory disease. Investigating the genomes of the disease and the competitive subclonal environment has led to a major understanding of CLL pathogenesis, most importantly the impact that therapy has in driving disease to a more aggressive phenotype.[132,142,198]

Schuh *et al*. used WGS to investigate the molecular heterogeneity of subclonal evolution of 3 IGHV-UM CLL patients over several years and through disease progression.[198] Samples were taken pre-treatment, post treatment as well as relapsed stage over (on average) 7 years, to evaluate the genetic changes and also the clonal architecture over time and the impact of therapy. Mutation profiles were shown to be heterogeneous between patients and alter over time and after treatment. Copy number alterations, such as 11q and 13q deletions and trisomy 12 remained constant at all sampling points, whilst 8p, 8q CNAs and 6q deletions were identified at relapse. The authors also used targeted deep sequencing of mutations to further define subclonal evolution and branching.[198] They identified 'founder' mutations (present in all cells at all time points), as drivers initiating tumorigenesis, as well as additional passenger mutations. The initiators, defined as being seen in CLL or other cancers before, are; *SF3B1*, *SAMHD1* and *MED12*, whilst secondary mutations (not in all cells) were identified in *ATM*, *PLEKHGS* and *IRF4* genes. The subclonal evolution of CLL was then compared to other haematopoietic neoplasms and shown to be different to ALL and AML; were a single subclone causes relapse. Whilst in CLL, similarly to MDS/AML, multiple, heterogeneous subclones exist.[198–201] The data from the study advocates that CLL is subject to subclonal expansion, dependent upon time and other selective pressures such as therapy, as well as competition between sub clones.[155,198] Further to this study Landau *et al*. also studied sub clonal evolution of CLL, through WES of 160 CLL patients.[132] This study also integrated copy number data with the WES data to provide an accurate calculation of cells carrying the variant in 149 of those sequenced. Further, 18 of these patients had two time points analysed, providing an assessment of clonal evolution in relation to clinical outcome. Using a recurrence analysis that detects enriched (mutated) genes above the background mutation rate

as well as those that have already been reported on COSMIC, the authors identified 20 putative CLL genes (*TP53*, *ATM*, *MYD88*, *SF3B1*, *NOTCH1*, *DDX3X*, *ZMYM3*, *FBXW7, MAPK1, XPO1*, *CHD2*, *POT1*, *NRAS*, *KRAS*, *BCOR*, *EGR2*, *MED12*, RIPK1, SAMHD1, *ITPKB* and *HIST1H1E*).[132] Additionally to the gene variants, copy number alterations of 8p, 13q, 11q, and 17p as well as trisomy 12 were identified as putative CLL drivers. The driving mutations were shown to target several key pathways including; DNA repair and cell- cycle control, Notch signalling, inflammatory pathways, Wnt signalling, RNA splicing, and processing, B cell receptor signalling and chromatin modification. Further analysis on the evolution of subclones showed that *MYD88* mutations, trisomy 12 and (monoallelic) 13q deletions were all early disease driving events occurring in 80-100% of all cells. *ATM*, *TP53*, and *SF3B1* mutations were shown to be predominately sub clonal and drivers most associated with disease progression. Longitudinal analysis to investigate the impact of chemotherapy treatment on subclonal evolution in CLL showed that the expansions of their subclonal populations are enriched with driver mutations. This means whereas in untreated patients the sub clonal drivers may expand over a longer period of time as they are competing with existing clonal populations, in the treated patients the dominant clone is removed paving the way for the subclone harbouring the survival advantage (driving mutation).[132] The authors also showed that the presence of a driver as a subclone prior to treatment can predict disease progression and reduced survival.[132]

Further to this study Landau *et al*. published a follow on study in line with Puente *et al*. that investigated 538 CLL exomes including 278 that were from the German CLL8 clinical trial.[75,132,142] Similarly to their previous study, the role of sub clones in relation to disease progression and survival was built upon in a larger cohort, a limitation of their previous publication to detect low frequency driver mutations, and validated using the clinical trial cohort. The large cohort size was powered to detect candidate CLL genes mutated in 5% of patients, whilst 94% powered to detect 3% variants and 61% for 2% mutations. In the study 59 of the patients also had matched pre-treatment and relapse samples. [142] This study identified 44 driver mutations, 18 of which were also identified in their previous study plus an additional 26 putative CLL drivers, [132,142] Of these additional drivers there were 3 *MYC* regulatory genes including *MGA*, *PTPN11* and *FUBP1*, as well as an additional MAPK-ERK pathway genes such as *MAP2K1* (as well as *NRAS*, *KRAS* and *BRAF*). Novel drivers included a ribosomal subunit protein *RPS15* and a B cell transcription factor *IKZF3*. Pathway analysis of these additional drivers showed that the genes are involved in RNA processing and export (*FUBP1*, *XPO4*, *EWSR* and *NXF1*), also DNA damage (*CHEK2*, *BRCC3*, *ELF4* and *DYRK1A*) and chromatin modification (*ASXL1*, *HIST1H1B*, *BAZ2B* and *IKZF3*) as well as B-cell-activity-related pathway genes (*TRAF2*, *TRAF3* and *CARD11*).[142] Looking at the temporal relationship of clonal and subclonal mutations, identified

CNAs as early events followed by gene mutations, for instance 11q deletions predominately preceded *ATM* or *BIRC3* mutations. Whereas 17p deletions and *TP53* mutations tended to be subclonal and were selectively enhanced post therapy due to the survival advantage these aberrations provide.[142] Analysing the CLL8 clinical trial cohort, the authors identified that patients whom harboured a subclonal gene mutation of *TP53*, *SF3B1* and *RPS15* also had a reduced PFS although this did not remain significant when looking in a multivariate model with IGHV status included. The authors also confirmed in the CLL8 trial their earlier observations in relapsed cases, the expansion of a more aggressive subclones post therapy.[132,142]

These large scale WGS and WES studies help to uncover a number of novel genes driving disease in CLL as well as starting to understand the relationship of subclonal mutations especially in relation to therapy. Following on from these pivotal large scale NGS sequencing investigations, more targeted approaches have been utilised to study the impact of these putative CLL genes to further understand CLL aetiology as well as possible utility as biomarkers.[75,110,122,130,131,142,202] One such investigation used single end deep sequencing technology to deep sequence the exonic and regulatory regions of 301 genes that are associated with BCR signalling or BCR related pathways, in 10 CLL patients. Four genes were identified as being recurrently mutated; *KRAS*, *SMARCA2*, *NFKBIE* and *PRKD3*.[203] Another study used a TruSeq Custom Amplicon panel that covered 13 CLL driver genes including; *TP53*, *ATM*, *SF3B1*, *NOTCH1*, *POT1*, *BIRC3*, *MYD88*, *KRAS*, *NRAS*, *U2AF1*, *BRAF*, *BCOR* and *SETBP1*, on 180 diagnostic CLL patients.[187] 22% of the patients had at least one of the mutations present, with *ATM* being the most frequent mutation (9.4%), followed by *NOTCH1* and *TP53* (7.7%), *SF3B1* (5.5%), *BCOR* (1.6%), *BIRC3*, *KRAS*, *U2AF1* and *POT1* (1.1%) and finally *MYD88* and *SETBP1* (0.6%). No somatic mutations were found for the *BRAF* and *NRAS* genes. Clonality was determined by gene using FISH data to measure heterozygosity and then adapting the VAF accordingly. The authors reported that 71% of *ATM* and *NOTCH1* were clonal mutations, whilst 80% *TP53* variants and 100% of the Ras-family genes (*KRAS*/*NRAS*) and *U2AF1* were sub clonal mutations. Although not all patients had reached the medium TTFT or OS, survival analysis could be carried out and the study found that patients carrying at least one of the mutations from the panel had a significantly worse TTFT than those without, as did patients harbouring a *TP53*, *NOTCH1*, *SF3B1* and *ATM* mutation. Further sub analysis of a double hit patient carrying an *ATM* deletion and 11q deletion also showed a significantly reduced TTFT. Using all the significant variables from univariate analysis, multivariate models were created and only the presence of a *TP53* mutation predicted an independent poorer TTFT and OS. This model was then recapitulated using sub clonal mutations (<20% clone) and again these sub clonal *TP53* mutations were found to be a significant independent indicator for TTFT but not for OS. Of note, interestingly, β2-microglobulin levels; a

traditional clinical prognostic marker, was still found to be a significant independent indicator in both the multivariate models for TTFT and OS.[187] This study also found that subclonal variants of *TP53* had a significantly reduced TTFT and was an independent risk factor in a multivariate model that included *ATM*, *NOTCH1*, IGHV and other clinical variables.[187] Targeted resequencing of 9 genes (*BRAF*, *EZH2*, *KRAS*, *MYD88*, *NOTCH1*, *NRAS*, *PIK3CA*, *SF3B1*, and *TP53*) in another CLL study (n=166) provided an insight into subclonal mutations, which were found in 67.5% of patients.[204] Interestingly the authors did not include the *ATM* gene within the panel of genes, due to the size and nature of mutations (no hotspots). No mutations were found for *EZH2* and *PIK3CA* genes, the most frequently mutated genes were; *TP53*, *SF3B1*, and *NOTCH1* (16.3%, 16.9%, 10.7% respectively), followed by the less frequent *BRAF* (4.2%), *MYD88* (1.8%), *NRAS* (1.2%) and *KRAS* (0.6%) mutations.[204] Another study investigated the frequency, prognostic impact and clonal evolution of three genes; *SF3B1*, *NOTCH1* and *BIRC3* on a cohort of 304 newly diagnosed CLL patients at consecutive time points throughout the disease.[205] Ultra deep next generation sequencing was used to detect the mutations as well as using a bioinformatical algorithm that can detect with a high sensitivity, very small mutated subclones, up to 0.3% (3 mutated reads in 1000 wildtype reads) with a 95% confidence interval of 0.2-0.5%. Allele specific PCR was then employed to confirm those clone sizes that fell below Sanger sequencing sensitivity. *NOTCH1*, *SF3B1* and *BIRC3* mutations were detected in 14%, 11% and 8% of the cohort, respectively. 14, 25 and 30 of the *NOTCH1*, *SF3B1* and *BIRC3* mutation, correspondingly, were missed by conventional Sanger sequencing but detected by ultra-deep NGS sequencing.[205] Even though the authors found the spectrum of sub clonal mutations similar to that of clonal mutations, the impact on survival was not found to be significant nor as great as the clonal events, though the authors state that the study was under powered so this may be why no significance was seen. Subsequently the authors found that *NOTCH1*, *SF3B1* and *BIRC3* mutations should have at least a 25%, 35% and 1% clone size, respectively to have an impact on overall survival in CLL. When looking at the dynamics and clonal evolution of specific mutations within the patients disease course, only *SF3B1* sub clonal mutations were consistently selected for and enriched post therapy, whilst *NOTCH1* and *BIRC3* tended to decrease or stay the same over time.[205] This is in contrast to what was seen by the same authors in a previous, but similar study investigating the sub clonal status of the *TP53* gene, were it was found that a low clone size had a survival advantage over other clones in the same patient, especially in regard to therapy and these *TP53* subclones also had a similar impact upon survival as a large clonal population.[94] Guièze *et al*, used targeted next generation sequencing to assess the prognostic impact of *TP53*, *SF3B1*, *ATM*, *NOTCH1*, *XPO1*, *SAMHD1*, *MCD12*, *BIRC3* and *MYD88* mutations in 114 relapsed or refractory CLL patients.[115] The authors chose an advanced disease cohort as the majority of previous published studies had only used diagnostic or early stage CLL cohorts. *TP53* and *BIRC3*

both were shown to have an independent impact upon OS, whilst no other mutation was found to have an adverse effect upon survival. Also in contrast to other studies they found *SF3B1* and *NOTCH1* did not have a poor outcome on survival and in fact found similarly to Schnaiter *et al*, that *NOTCH1* mutations had a better PFS.[115,125] This demonstrates that the impact of mutations at different disease stages may influence patient outcome differently. Further analysis clustered the mutations into number (per patient) and reoccurrence of mutations present. Patients with >1 gene mutations had a significantly poorer survival than those without any mutations; this was then confirmed in a multivariate analysis. They also found many of the patients carried concurrent mutations of *TP53*, *ATM* and *SF3B1* genes, these were shown to have significantly reduced survival for both PFS and OS than those patients with just one or no mutations.[115] As most of the mutations are not mutually exclusive, it is difficult to define the exact effects of specific mutations upon survival as they could be confounding the results. A targeted re sequencing study from Nadeu *et al*, assessed the clinical impact of both clonal and sub clonal mutations in CLL.[206] The study assessed down to a 0.3% allele frequency and focused on 5 specific genes (*TP53*, *SF3B1*, *BIRC3*, *NOTCH1*, and *ATM*) in 406 untreated CLL patients, though had sequential sampling of only 48 patients. *TP53* mutations were detected in 10.6% of patients (6.4% clonal, 4.2% subclonal), *ATM* mutations in 11.1% (7.8% clonal, 1.3% subclonal, 2% germ line mutations considered pathogenic), *SF3B1* mutations in 12.6% (7.4% clonal, 5.2% subclonal), *NOTCH1* mutations in 21.8% (14.2% clonal, 7.6% subclonal), and BIRC3 mutations in 4.2% (2% clonal, 2.2% subclonal).[206] In terms of survival *TP53* clonal and subclonal mutations whilst only clonal *NOTCH1* mutations provided a reduced OS for patients harbouring these abnormalities, though subclonal NOTCH1 mutations did impact upon TTT. Clonal *SF3B1* and *ATM* mutated patients had a reduced TTT but not subclonal mutations, although the authors add that subclonal *ATM* mutations (without 11q deletion) was a rare event.[206] Their investigation of clonal evolution, showed similar results to other studies in that subclones harbouring *TP53* mutations are selected for by therapy, although they did observe de novo subclonal expansion without therapy along with *NOTCH1* and *SF3B1* mutations.[206]

All these studies, although using differing cohorts at different disease stages and focusing on variable genes with varying sequencing technologies, all provide a similar answer; that the CLL genome is a dynamic, heterogeneous and constantly evolving disease. The study of clonal evolution has shown that NGS could be used in patient management to detect low subclones harbouring an unfavourable mutation (i.e. *TP53*) at diagnosis to avoid selecting this clone for expansion (with chemo toxic regimes). This study hopes to expand on previous studies by including several putative CLL genes using a custom panel on a local cohort, as well as actually investigating the utility of NGS amplicon technology as a clinical tool.

7.2 Aims

- To design an Illumina TruSeq custom amplicon panel to detect mutations in 25 genes implicated in CLL pathogenesis, as described in the literature.
- To apply this custom panel to a large diagnostic cohort of CLL patients, determining the frequency of mutations, their associations with other genomic aberrations and impact on survival.
- To develop and assess the utility of this technology as a clinically relevant screening tool.

7.3 Methodology

7.3.1 **Patient samples**

The cohort used in this study was a previously untreated diagnostic cohort (n=68) of CLL patients recruited in the lymphoproliferative disorder study at time of initial evaluation at the Department of Haematology of the Southampton University Hospital Trust from January 2001 to May 2015. Diagnosis of CLL was according to the 2008 International Workshop on Chronic Lymphocytic Leukaemia (IWCLL2008)/National Cancer Institute (NCI) criteria. All patients provided informed consent prior to inclusion in the study. Restricted access to clinical data meant that no survival analysis was able to be conducted.

The second cohort in this study was in collaboration with The European research initiative on CLL (ERIC) consortium, which in collaboration with 9 Pan-European research centres aimed to assess the applicability of NGS in a clinical setting. Two technologies were assessed, TruSeq custom amplicon (Illumina) and Haloplex (Agilent). This study received the Illumina TruSeq Custom Amplicon kit. The patient samples (n=48) were collected from all the centres with the caveat of having at least one clonal mutation present in the sample. The DNA samples were then aliquoted (250ng) and sent blind to all the centres to analyse by their respective technology. The remaining library was returned to the main centre (Uppsala) along with the Fastq files generated for comparison with the other centres.

7.3.2 **TruSeq Custom Panel**

Two TruSeq custom amplicon gene panels were designed, one consisting of 25 genes found mutated in CLL (Table 26), this custom panel included well known CLL driving genes as well as additional genes thought to play a role in CLL.

The other custom panel (for the ERIC study) consists of a 12 gene panel of genes (Table 27), fewer genes were chosen for this study as all these genes have been shown to have a clinical significance within CLL. For both of these panels the 250bp sequencing chemistry was used as amplicons were designed to roughly cover this length.

Table 26 Gene list of 25 genes thought to be involved in CLL pathogenesis. These 25 genes were used in the custom TruSeq panel that was designed for this study, to be used on a local (Southampton) cohort.

ATM	*DDX3X*	*MED12*	*NOTCH1*	*SETD2*
BIRC3	*EGR2*	*MGA*	*NRAS*	*SF3B1*
BRAF	*FBXW7*	*MYD88*	*POT1*	*TP53*
CHD2	*HIST1H1E*	*NBEAL2*	*RPS15*	*XPO1*
CTBP2	*KRAS*	*NFKBIE*	*SAMHD1*	*ZFPM2*

Table 27. The 12 gene panel of CLL genes (ERIC) known to have prognostic relevance and/or follow up validation as a CLL gene

ATM	*NOTCH1*	*EGR2*	*NFKBIE*
BIRC3	*SF3B1*	*POT1*	*BRAF*
MYD88	*TP53*	*SAMHD1*	*FBXW7*

7.3.3 TruSeq Custom Amplicon Library preparation

DNA was prepared to 250ng (10µl) prior to TruSeq protocol, an agarose gel and Qubit were performed to check DNA quantity and quality (Material and methods3.5). The TruSeq protocol was followed accordingly.

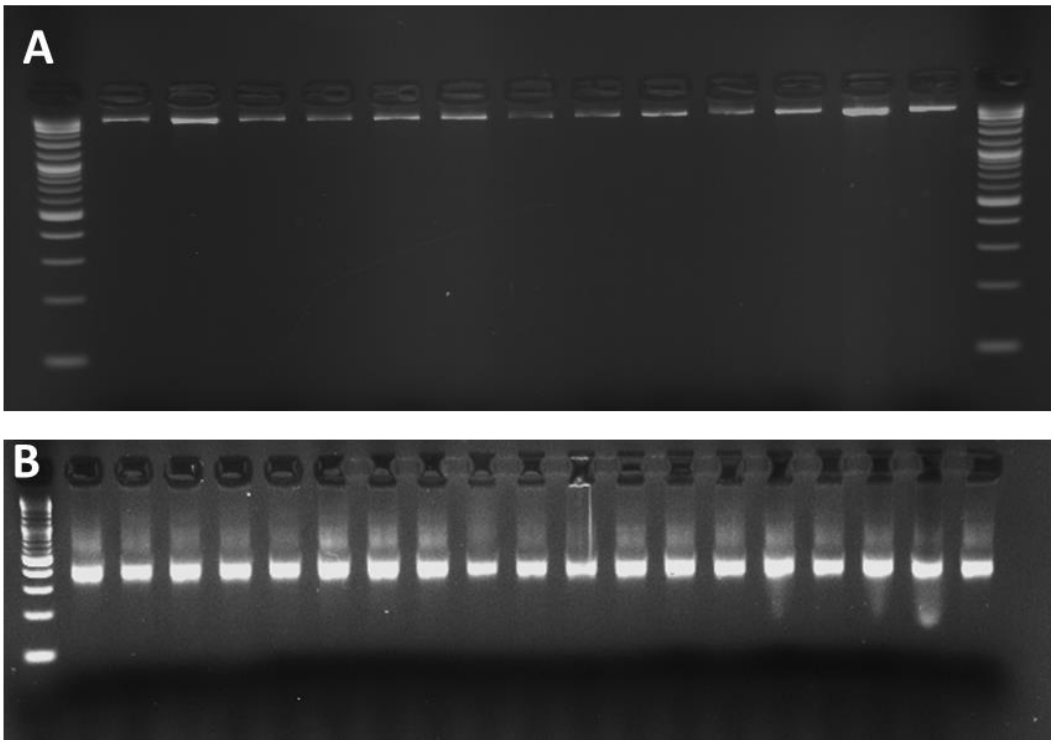


Figure 60 A) QC prior to TruSeq, DNA was electrophoresed on an agarose gel to assess DNA quality. High molecular weight band of DNA (as shown) with little to no smearing is required B) After amplification step of TruSeq protocol, PCR product is resolved on a 3% agarose gel to check amplification worked. Qubit was also done at this step to quantify DNA.

7.3.4 Data analysis

Sequencing data was put through a bioinformatical pipeline were it was trimmed, QC, aligned, to a reference genome to generate a variant list. The fastq sequence was first trimmed to remove the adapter sequence with support from the Faculty of Medicine Bioinformatics Group. The trimmed fastq sequence was aligned to the human reference genome sequence (hg19/GRCh37) using Novoalign (v2.08.02), and unmapped or poor-quality reads were removed. Variant calling was performed using samtools mpileup (v0.1.18) and filtering was based on a minimum read depth of 4 and a minimum phred scaled quality score of 20. All the variants identified were annotated against dbSNP (build 135). The variant list was manually curated to filter and further remove synonymous variants, low frequency variants (<5 VAF) and SNPs. Integrated genomics viewer (IGV) was used to visualise and confirm the variants.

7.3.5 Sanger sequencing

Sanger sequencing was used to confirm a small number of the variants that had >20% VAF (Appendix R). Assays were designed in line to previous and a tdPCR programme was used as before.

Table 28 Assays designed to cover variants detected by TruSeq and confirmed in IGV

Gene	Exon	Assay	Oligonucleotide sequence	Tm (°C)	Product size (bp)
TP53	6	F	ACCCATTTACTTTGCACATC	55	612
TP53	6	R	TACGCATGTTTGTTTCTTTG	55	
EGR2	2	F	CAAGACTGCTGCTGTTACTG	55	459
EGR2	2	R	CCTATAACCCACACCACCT	55	
ATM	50	F	GCTGGAAATTATGATGGAGA	55	362
ATM	50	R	AGAACCCAAAAGACCAAGAT	55	
ATM	35	F	CAAAGTGCTGGGATTACAGT	55	395
ATM	35	R	TGTGAAGTATCATTCTCCATGA	55	
ATM	51	F	TGGCTTTTGTGTTTTACCTT	55	450
ATM	51	R	AAGGAACTTTTGTAGAAACAAGAC	55	
SAMHD1	7	F	GCCTCTAAAATAGCCAAATG	55	400
SAMHD1	7	R	TGTTCCTAAGATTGTGGTTAGA	55	
XPO1	16	F	CAGAGTAAGACCCTGTCTCAA	55	361
XPO1	16	R	ATGGAGTCCAGGACAAAATA	55	
RPS15	4	F	ATGATCATCCTACCCGAGAT	55	443
RPS15	4	R	TCCACACAACACTACCACAC	55	
BIRC3	10	F	CCACAGAAGATGTTTCAGGT	55	546
BIRC3	10	R	GTGCTACCTCTTTTTCGTTC	55	
ATM	42	F	GTGGAGGGAAGATGTTACAA	55	453
ATM	42	R	ACCTGTCTGACTTTCCTGTG	55	

7.3.6 Gene expression analysis

Taqman assays from Thermofisher (Table 29) were used to study the RNA expression in genes of interest; *BIRC3*, *POT1*, *TP53* and *ATM* genes. These genes were chosen as a follow on from previous gene expression data. Two reference genes were used; 18s and PGK1 and a Raji cell line RNA was used as a control calibrator to normalise the gene expression data to. A small cohort of the CLL patients already sequenced by the (Southampton) TruSeq custom amplicon panel (n=60) were chosen based solely on accessibility to viable cells. RNA was exacted and cDNA synthesised as stated previously. Standard curves were first generated for all genes and reference genes, these were saved as an external standard curve and used to normalise data to in each run.

Table 29 Taqman assays used to analyse the gene expression in the Southampton cohort

Taqman assays	cat no	supplier
Hs01112355_g1 ATM	4331182	Applied biosystems
Hs00209984_m1 POT1	4331182	Applied biosystems
Hs01034249_m1 TP53	4331182	Applied biosystems
Hs00943178_g1 PGK1	4331182	Applied biosystems
Hs99999901_s1 18S	4331182	Applied biosystems
Hs00985031_g1 BIRC3	4331182	Applied biosystems

7.3.7 Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics (V.21). Fisher's exact test was used to look for significant associations between genomic variants and clinically relevant biomarkers such as IGHV status and CNAs (11q, 13q, 17p deletions and trisomy 12).

The expression analysis used Pearson's test to identify an significant correlations between the different gene measured (POT1, BIRC3, ATM, TP53) expression along with telomere length (Pearsons correlations of expression data and telomere length. *POT1* and *BIRC3* expresion show a positive significant correlation, whilst *TP53* and *ATM* expression are also positively, significantly correlated.). Mann-Whitney statistical test was used to identify gene expression changes with genomic associations.

In all tests, the significance was identified if the p value was ≤0.05.

7.4 Results

7.4.1 ERIC project-coverage data

All 48 samples were ran in one TruSeq run, the coverage was comparable to the other TruSeq data previously discussed; with a total amplicon coverage (average) of 97.9% coverage x100 reads and x1000 reads 80.5%. The minimum total reads was 27474 and the maximum was 1447636 whilst the average was 756675 total reads. Only one sample had particularly low read coverage (sample 3) at 33.7% total amplicon coverage compared to >98% for the rest of the samples. This sample only had a total of 27474 reads compared to the next lowest with 562280 reads (sample 7). There was a total of 205 amplicons covering the 12 genes, though 92 of these were solely for the *ATM* gene (Table 30). The gene coverage showed a high level of average coverage with all genes having >1500 reads (Figure 62).

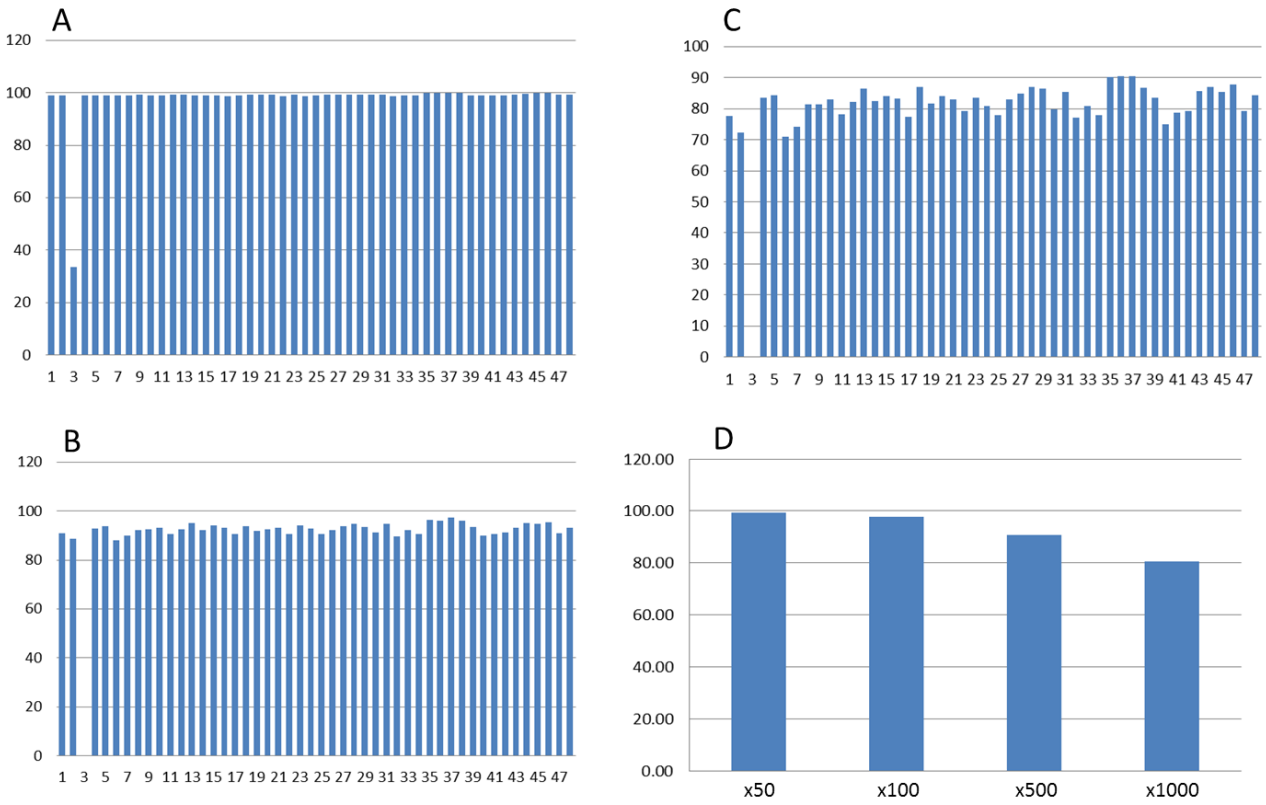


Figure 61. Read coverage for the 48 sample 12 gene custom TruSeq gene panel A-C Read coverage across all samples (1-48) A. Total amplicon coverage (%) with more than 100 reads across all amplicons, only one sample had low coverage (sample 3) B. Total amplicon coverage (%) with more than 500 reads C. Total amplicon coverage (%) with more than 1000 reads D. The average total amplicon coverage percentage of all samples >97% had x100 reads whilst >80% had x1000 reads.

Table 30 Average of all amplicons (per gene) coverage from the ERIC study. *ATM* as a large gene, with no re occuring hotspot locations has the most amplicons

Gene	number of amplicons	Average coverage
ATM	92	3894
BIRC3	17	4224
EGR2	8	1540
FBXW7	19	3596
MYD88	8	2519
NFKBIE	12	2246
NOTCH1	10	2276
POT1	18	4191
SF3B1	7	4606
TP53	11	2802
XPO1	3	1783

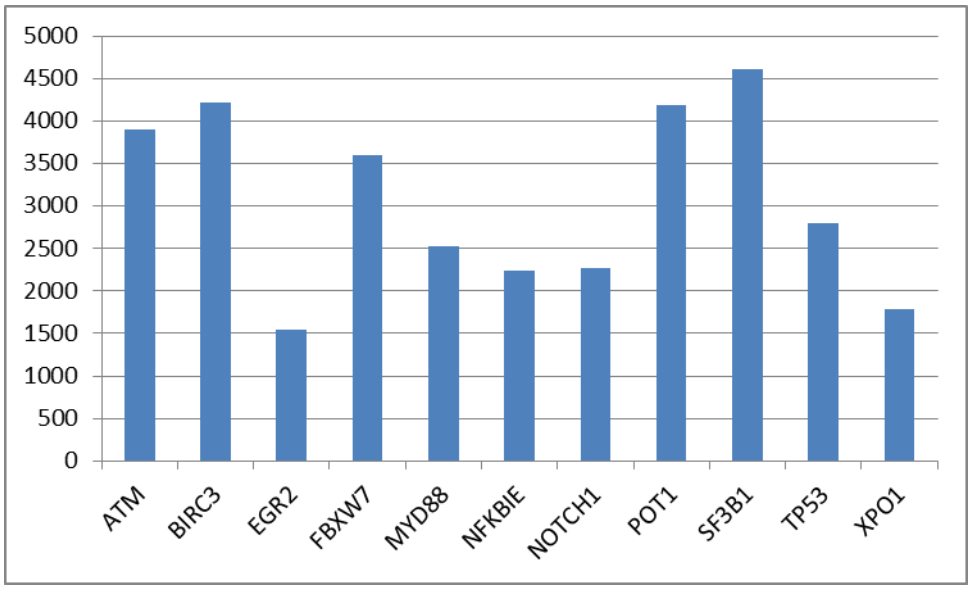


Figure 62 Average coverage per gene from the ERIC 12 gene custom TruSeq panel sequenced on 48 patients.

7.4.2 Mutational variants- ERIC samples

After the data was put through the bioinformatical pipeline the mutational list was extensive (n=5538 variants) so further manual curating was needed to reduce the mutational list down to a list of potentially real variants. Genes not on the panel were removed as were all known SNPs, and only non-synonymous variants that had >5% VAF were kept. This gave a total of 186 variants from all but 1 patient (PID 8). This patient when removing some of the strict filtering had a *TP53* mutation, so this was included in the final list (Appendix). A large proportion (n=45) had an *ATM* variant (I105N) and a (n=47) *TP53* variant (T245P) that were identified as false positive through IGV. An additional 27 of the variants were also false positive variants. These were removed from the final list (Table 31), so overall in the IGV confirmed variants there was a total of 66 variants in 41 of the patients. Variants were identified in *ATM* (n=11), *BIRC3* (n=5), *FBXW7* (n=3), *POT1* (n=5), *SF3B1* (n=14), *TP53* (n=14), *XPO1* (n=6).

The 11 *ATM* mutations included 3 frameshift variants, 5 missense mutations, 2 splicing mutations and 1 additional truncating mutation, all affecting various amino acids. The 5 *BIRC3* mutations were identified in exons 7 (n=2), 9 (n=2), these were truncating mutations and exon 10 (n=1) a missense mutation. The 3 *FBXW7* variants were missense mutations, effecting codons 347 (Arginine to Leucine), codon 360 (Serine to Threonine), and codon 427 (Tyrosine to Cysteine). 6 of the 8 *MYD88* variants affected codon 160, changing the stop codon to a Arginine. The remaining 2 mutations affected codons 172 (Valine to Phenylalanine) and codon 6 (Alanine to Proline). There were 5 variants identified across the *POT1* gene, affecting missense mutations affected codons

35, 96, 196 and 273 whilst a glutamine to stop codon (codon 488) was identified. The *SF3B1* variants (n=12) were identified as missense variants; p.K700E (n=3), p.G740E (n=1), p.G742D (n=3), p.I704F (n=1), p.K666N(n=2)/Q (n=1), p.N626D (n=1)/S (n=1) and p.T663I (n=1). Of the 14 *TP53* variant, several (n=9) were missense mutations ranging from codons 44 to 245. There were also 3 splicing mutations and an Arginine to stop (77) codon truncating variant. All 6 of the *XPO1* variants affected codon 571, altering the amino acid from glutamic acid to a Glycine (n=2), Lysine (n=3) or Valine (n=1).

The most prevalent *NOTCH1* variant at codon 2515 was not identified by the bioinformatical pipeline, this meant that all patients had to be identified by eye in IGV (Figure 63). In total 12 *NOTCH1* (p.P2515fs4) variants were identified in IGV ranging from 11% to 63% VAF, two patients (PIDs 46 & 48) had 4% VAF but these were not included due to the stated 5% cut off.

No statistical analysis was conducted on the ERIC variant data as; 1) no Sanger validation was done due to time constraints; 2) all samples had a known variant and were selected for this investigation for this reason, so any associations with another mutation may not be representative to CLL as a whole and; 3) all samples were blinded to this study so no clinical data was available at the time of writing.

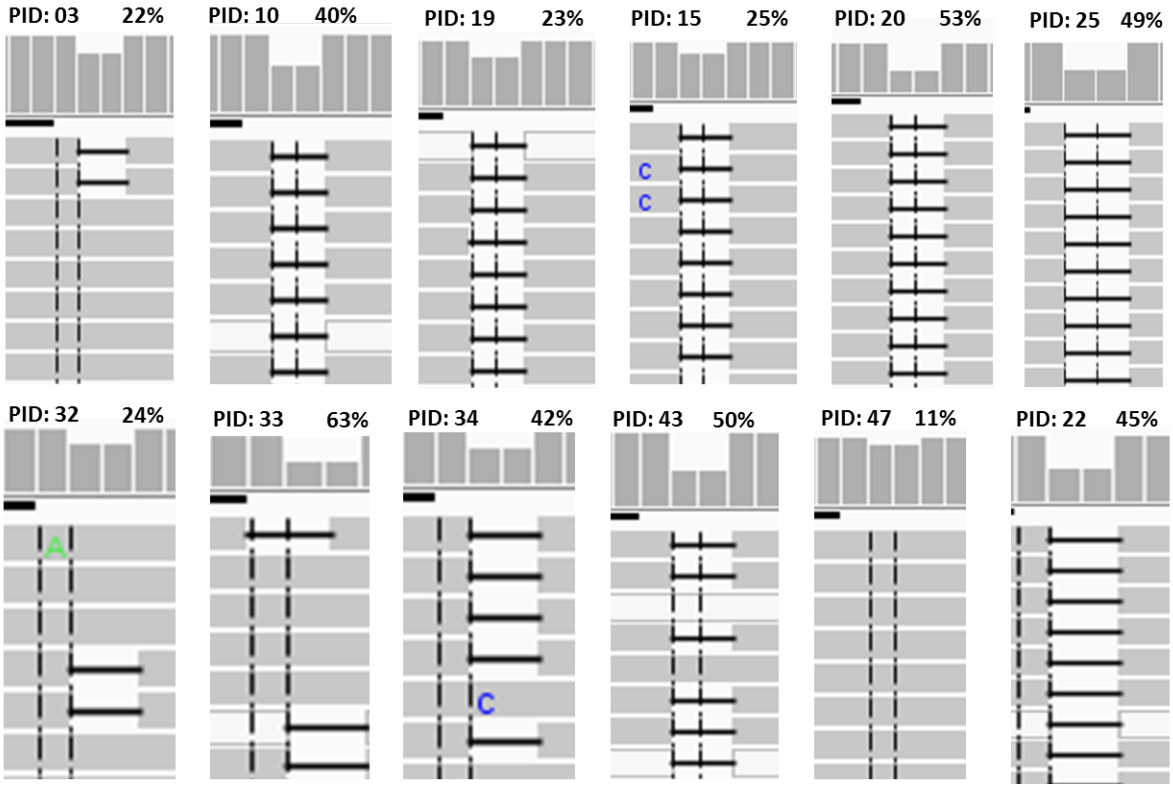


Figure 63. *NOTCH1* (p.P2515fs4) gene variant, IGV confirmation in 12/48 cases (25%). The bioinformatical pipeline did not identify this recurrent *NOTCH1* mutation so all cases had to be identified manually by eye in IGV.

Table 31 ERIC variants identified from a filtered variant list, all false positive results have been removed (full n=187 variant list can be found in the appendix)

PID	variant	Gene	coding	protein	QDEPTH	VAF
2	nonsynonymous	*POT1*	c.A587G	p.E196G	3867	7.20%
2	nonsynonymous	*SF3B1*	c.C1988T	p.T663I	810	44.60%
4	nonsynonymous	*SF3B1*	c.A2098G	p.K700E	4276	11.50%
5	splicing	*ATM*			1738	42.80%
5	stopgain	*ATM*	c.T8907G	p.Y2969X	1887	9.30%
8	nonsynonymous	*TP53*	c.A733C	p.T245P	38	81.60%
9	stoploss	*MYD88*	c.T478C	p.X160R	1529	37.40%
10	nonsynonymous	*XPO1*	c.A1712G	p.E571G	2230	29.90%
11	nonsynonymous	*XPO1*	c.G1711A	p.E571K	2205	54.30%
12	stoploss	*MYD88*	c.T478C	p.X160R	1403	39.40%
13	nonsynonymous	*ATM*	c.C2933T	p.S978F	4629	10.80%
13	frameshift insertion	*ATM*	c.9010dupA	p.N3003fs	7487	15.20%
13	nonsynonymous	*SF3B1*	c.G2219A	p.G740E	6762	7.20%
14	nonsynonymous	*SF3B1*	c.A1877G	p.N626S	7599	11.50%
16	nonsynonymous	*SF3B1*	c.G2225A	p.G742D	3805	23.70%
16	nonsynonymous	*TP53*	c.T337G	p.F113V	1070	28.60%
17	nonsynonymous	*ATM*	c.A5611C	p.T1871P	349	5.10%
17	stoploss	*MYD88*	c.T478C	p.X160R	1370	46.80%
19	nonsynonymous	*SF3B1*	c.A2098G	p.K700E	3731	5.20%
19	nonsynonymous	*SF3B1*	c.G1998T	p.K666N	1559	23.10%
19	nonsynonymous	*XPO1*	c.A1712G	p.E571G	2410	48.00%
21	frameshift insertion	*BIRC3*	c.1616dupT	p.V539fs	2114	25.60%
22	frameshift insertion	*BIRC3*	c.1288dupG	p.E429fs	3442	10.90%
22	frameshift insertion	*BIRC3*	c.1613_1614insTGTT	p.D538fs	1941	10.70%
23	nonsynonymous	*ATM*	c.G8776T	p.V2926F	1452	49.20%
24	frameshift insertion	*ATM*	c.3981dupA	p.L1327fs	1313	48.50%
25	nonsynonymous	*SF3B1*	c.A2098G	p.K700E	3303	11.30%
26	nonsynonymous	*FBXW7*	c.G1040T	p.R347L	1910	42.10%
26	nonsynonymous	*ATM*	c.A7648G	p.M2550V	106	5.70%
26	nonsynonymous	*FBXW7*	c.T1078A	p.S360T	206	5.30%
26	nonsynonymous	*SF3B1*	c.G1998T	p.K666N	1532	43.90%
26	nonsynonymous	*TP53*	c.G131A	p.C44Y	1234	66.10%
26	nonsynonymous	*XPO1*	c.G1711A	p.E571K	2760	40.50%
27	stoploss	*MYD88*	c.T478C	p.X160R	1598	50.60%
28	stoploss	*MYD88*	c.T478C	p.X160R	1921	43.30%
29	nonsynonymous	*MYD88*	c.G514T	p.V172F	3463	12.30%
30	nonsynonymous	*POT1*	c.C104T	p.P35L	1950	50.60%
31	nonsynonymous	*POT1*	c.C817T	p.R273W	3356	50.30%
32	nonsynonymous	*ATM*	c.A5611C	p.T1871P	419	5.30%
32	stopgain	*POT1*	c.C1462T	p.Q488X	5725	19.40%
33	frameshift insertion	*BIRC3*	c.1292dupA	p.E431fs	3192	11.30%
34	nonsynonymous	*MYD88*	c.G16C	p.A6P	433	41.80%
35	nonsynonymous	*FBXW7*	c.A1280G	p.Y427C	2767	50.90%

35	.	.			1203	13.60%
36	frameshift insertion	*ATM*	c.6602dupT	p.V2201fs	5440	21.70%
36	nonsynonymous	*SF3B1*	c.A1876G	p.N626D	7702	46.60%
37	nonsynonymous	*SF3B1*	c.G2225A	p.G742D	6107	52.20%
37	nonsynonymous	*XPO1*	c.G1711A	p.E571K	4414	44.60%
38	nonsynonymous	*BIRC3*	c.G1772A	p.C591Y	2564	76.90%
39	nonsynonymous	*TP53*	c.C380T	p.S127F	2293	82.20%
40	nonsynonymous	*POT1*	c.A286T	p.I96F	3296	49.60%
41	nonsynonymous	*SF3B1*	c.A2110T	p.I704F	3676	13.40%
42	nonsynonymous	*TP53*	c.G191C	p.R64P	2047	99.80%
43	stopgain	*TP53*	c.A229T	p.R77X	2677	12.30%
43	Splicing	*TP53*	c.920-2A>G		2266	26.70%
43	Splicing	*TP53*	c.673-2A>T		1108	21.20%
44	stoploss	*MYD88*	c.T478C	p.X160R	2136	48.90%
45	nonsynonymous	*TP53*	c.C447G	p.D149E	908	45.30%
45	nonsynonymous	*TP53*	c.A377G	p.E126G	1066	34.40%
45	nonsynonymous	*TP53*	c.C134G	p.P45R	1933	20.80%
46	nonsynonymous	*SF3B1*	c.G2225A	p.G742D	5505	25.50%
46	frameshift insertion	*TP53*	c.351dupG	p.T118fs	607	91.10%
47	nonsynonymous	*TP53*	c.A140G	p.H47R	1255	43.40%
47	Splicing	*TP53*	c.96+1G>T		1343	43.00%
48	nonsynonymous	*SF3B1*	c.A1996C	p.K666Q	2295	27.70%
48	nonsynonymous	*XPO1*	c.A1712T	p.E571V	2749	51.50%

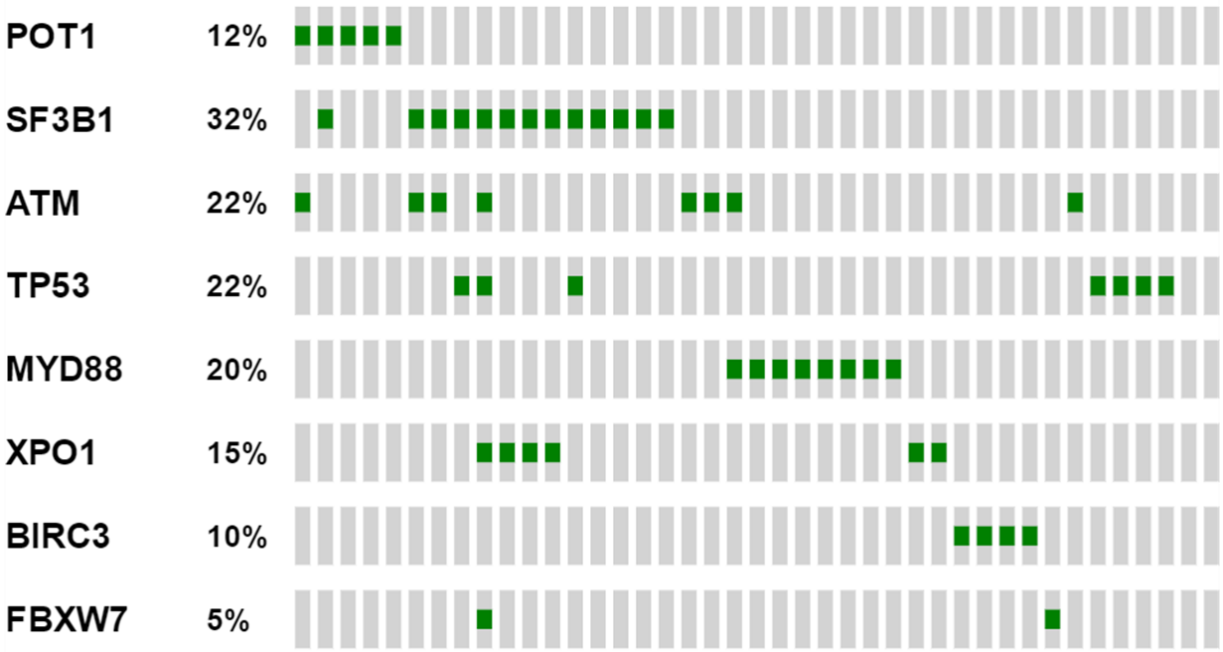


Figure 64 Gene mutations detected using TruSeq custom amplicon panel (12 genes) . All samples (n=48) were expected to have a mutation due to the nature of the experimental validation, so no further statistical analysis was undertaken. (http://www.cbioportal.org/oncoprinter.jsp#)

7.5 Coverage data-Southampton cohort

Overall the total amplicon coverage with more than 100 reads across all amplicons and samples was 95.6%, for more than 1000 reads this figure was 79.6%. The minimum read depth was 1293103 and the maximum read depth was 2250243 reads. An average was taken for the read coverage per amplicon and then combined per gene to show which genes had a poor coverage. The majority of genes had excellent coverage (range 998.94-4316.40) and only the *NOTCH1* gene had coverage just below 1000 reads.

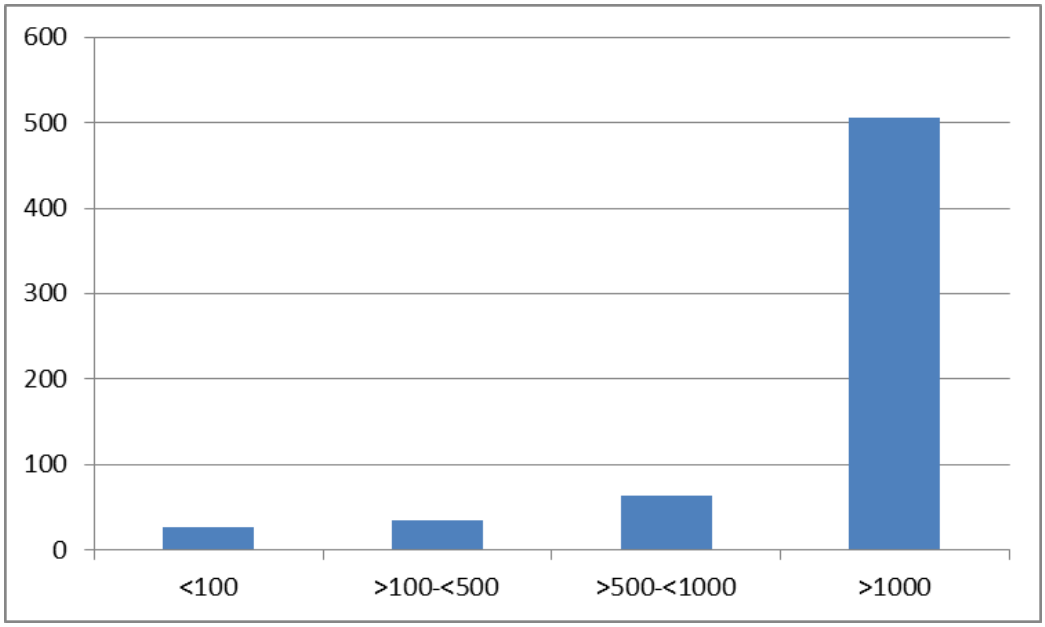
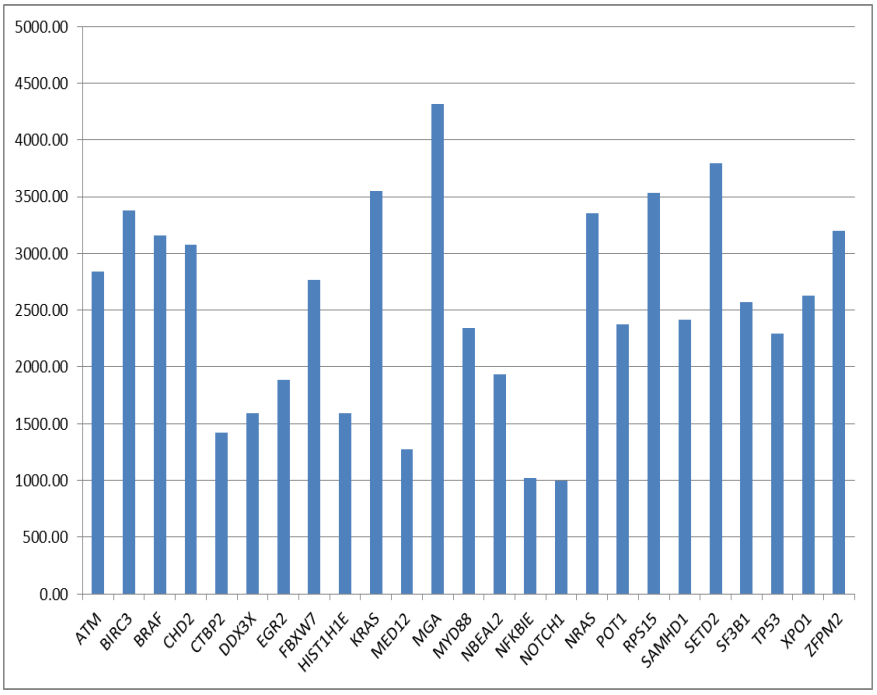


Figure 65: The average amplicon read depth across all patients. The majority of the amplicons had reads above 1000, a small number had a poor coverage of <100 (n=27). Only one amplicon failed to generate any reads across all samples (one of the CTBP2 amplicons)



Gene	Amplicon coverage
NOTCH1	998.94
NFKBIE	1020.11
MED12	1271.15
CTBP2	1422.42
DDX3X	1589.65
HIST1H1E	1589.87
EGR2	1883.32
NBEAL2	1934.22
TP53	2295.56
MYD88	2339.75
POT1	2372.76
SAMHD1	2415.18
SF3B1	2567.62
XPO1	2627.56
FBXW7	2764.70
ATM	2840.35
CHD2	3073.21
BRAF	3156.05
ZFPM2	3194.66
NRAS	3349.49
BIRC3	3380.51
RPS15	3528.99
KRAS	3546.21
SETD2	3796.11
MGA	4316.40

Figure 66: Combined average amplicon coverage per gene. Amplicon coverage for the majority of the genes is good and well above 1000 reads. Only the *NOTCH1* gene has coverage below 1000 reads. This gene is notorious for being difficult to sequencing due to the high GC content in this gene.

7.6 Poorly performing amplicons

A number of amplicons were identified as having little to no read depth (Table 32). The *NOTCH1* amplicons that have a poor performance is probably due to the high GC content around this region (exon 34), this region is notoriously difficult to sequence. When the amplicon probe sequences were put into the UCSC blat (http://genome.ucsc.edu/cgi-bin/hgBlat) some of the expected amplicons actually targeted the wrong gene or chromosome. For instance the poorly performing *DDX3X* amplicons targeted chromosome Y (*DDX3Y* gene) instead of chromosome X, whilst *CTBP2* poorly performing amplicons also targeted chromosomes Y and 11 instead of chromosome 10.

Table 32 Poorly performing amplicons, with a read depth of <100. One amplicon failed to generate any reads in any of the samples. The DDX3X amplicon that on average has good coverage has been included as in 4 patients there was no or little read depth. The other amplicons with an average read depth >100 are also included as some of the samples deviated below 100.

Amplicon	Average read depth per amplicon
CTBP2_Exon (32453041)_52602747.CTBP2_UserDefined (32453025)_52603273.2	0
ATM_Exon (32442426)_52602866.1	0.12
DDX3X_UserDefined (32442453)_52602910.3	0.12
NBEAL2_UserDefined (32453016)_52603213.ATM_Exon (32083111)_52603133.2	0.24
CTBP2_ThreeUtrExon (32442533)_52602966.2	0.29
DDX3X_UserDefined (32442453)_52602910.2	2.06
BRAF_ThreeUtrExon (32452972)_52603093.2	2.41
CTBP2_ThreeUtrExon (32442533)_52602966.4	6
SAMHD1_Exon (32105522)_52603059.1	6.65
ATM_UserDefined (32408355)_52602849.1	6.94
CTBP2_ThreeUtrExon (32442533)_52602966.3	8.12
SETD2_FiveUtrExon (32105548)_52603247.1	8.35
Braf_UserDefined (32453043)_52603292.1	8.88
TP53_ThreeUtrExon (32401998)_52603082.1	11.29
HIST1H1E_ThreeUtrExon (32442552)_52603022.1	13.94
NBEAL2_UserDefined (32453015)_52602808.1	14.12
NBEAL2_UserDefined (32453017)_52602813.1	14.59
NBEAL2_UserDefined (32453016)_52603213.1	15.47
ZFPM2_FiveUtrExon (32129111)_52603025.1	19.71
NFKBIE_UserDefined (32442504)_52602933.1	22.94
HIST1H1E_ThreeUtrExon (32442552)_52603023.1	32.71
DDX3X_UserDefined (32442509)_52602944.2	43.24
NOTCH1_ThreeUtrExon (32452971)_52603089.1	55.94
CTBP2_UserDefined (32453025)_52603275.1	60
NOTCH1_ThreeUtrExon (32452971)_52603090.1	84.29
NFKBIE_UserDefined (32442504)_52602930.1	86.47
NBEAL2_UserDefined (32453017)_52602812.1	95.06
SAMHD1_UserDefined (32442559)_52603053.1	106.24
XPO1_Exon (32442438)_52602900.1	110.12
TP53_UserDefined (32452969)_52602853.1	119.12
FBXW7_Exon (32452998)_52603140.1	138.94
NBEAL2_FiveUtrExon (32453028)_52603278.1	191.94
DDX3X_UserDefined (32453024)_52603268.2	445.41

7.7 Variants identified from the Southampton cohort

There were 21,113 variants identified from the 25 gene TruSeq bioinformatical data in the 68 CLL cases. Synonymous variants were removed, along with any below 5% VAF and any genes not part of the panel. The final filtered variant list (n=104) were then checked in IGV to be confirmed real or not. Some variants were found to be lower than the suggested VAF or some were false positives due to poly A regions or poor sequencing depth. The false positive variants were removed from the variant list. *NOTCH1* mutations again had to be identified manually due to the pipeline not detecting this mutation, probably due to the poor coverage of this gene. IGV confirmed around half of the variants were true mutations (n=50), some of these with a VAF of more than 20% were Sanger sequenced, below this level Sanger sequencing may not detect the variant.

The final variant list (Table 33) involved 14/25 genes in 39 of the samples sequenced, including *ATM* (n=8), *BIRC3* (n=1) *EGR2* (n=3), *MYD88* (n=3), *NBEAL2* (n=1) *NOTCH1* (n=4), *POT1* (n=5), *RPS15* (n=3), *SAMHD1* (n=1), *SETD2* (n=2), *SF3B1* (n=13), *TP53* (n=3), *XPO1* (n=2) and *ZFPM2* (n=1) genes (Table 33).

The most prevalent gene mutated was *SF3B1* (n=13 variants); all were non synonymous variants within exons 14 (codons 662, 666 and 625), exon 15 (p.K700E) and codon 742 in exon 16. All three *MYD88* mutations were the same codon 160 (p.*160R) stop loss variant within exon 3 of the gene, VAFs were from 41% to 45%. 4/5 *POT1* mutations were exonic whilst one was a splicing mutation, the VAF ranged from 6% to 49%. The p.P2515fs *NOTCH1* variant had was identified manually in 3 cases and an additional *NOTCH1* exon 34 mutation (p.Q2459X) was identified by the bioinformatical pipeline. All 8 *ATM* variants were exonic and non synonymous mutations in exons 34, 41, 49 and 50, the VAF ranged from 11% to 55%. One patient (PID 504) had two samples sequenced and the *ATM* mutation (p.R2443Q) was identified in both samples; sample A at 15% VAF and in the later sample D this had increased to 19% VAF. Two of the patients (PID 598 and 644) also had a concurrent 11q deletion. The only *BIRC3* variant identified by the pipeline was a frameshift insertion exon 10 variant (Q547fs) with a VAF of 38%. All three non synonymous *EGR2* mutations were located within exon 3 of the gene, the VAF ranging from 6% to 56%. Only one *NBEAL2* exon 13 non synonymous mutation were identified affecting codon 538 changing arginine to histidine (47% VAF).*RPS15* variants were all non synonymous mutations within exon 4 of the gene. One *RPS15* mutated patient (PID 531) had two samples harbouring the mutation (p.P131R); sample A had a VAF of 52% whilst this was reduced to 28% in sample D. *SAMHD1* mutations were only seen in one patient; this was a 73% VAF, exon, 7 non synonymous (p.M254I) variant. Two non synonymous exon 3 *SETD2* gene variants were identified from the sequencing data. Also 3

TP53 mutations, one splicing and two exonic variants of which both also had a concurrent 17p deletion that is reflected in the VAFs (75% and 87% respectively). Two non-synonymous *XPO1* variants were identified in exons 15 (p.E571K) and 16 (p.D1871G), VAF of only 8% and 49%, respectively. The final mutation was identified in exon 8 of the *ZFPM2* gene (p.R574Q).

Table 33 IGV confirmed variants (n=50) from the TruSeq analysis of 68 CLL cases. Variants were first filtered removing any genes not on the panel, any known SNPs, synonymous mutations and variants with a low VAF <5%. The remaining variants were then visualised using IGV to confirm the variants are real. The *NOTCH1* mutation (p.P2514fs) was not identified by the bioinformatical pipeline so this variant had to be manually inspected in all patients by eye in IGV.

PID	gene	variant	Variant	exon	coding	protein	VAF
322	*SF3B1*	exonic	nonsynonymous	15	c.A2098G	p.K700E	48%
428	*SF3B1*	exonic	nonsynonymous	15	c.A2098G	p.K700E	49%
501	*NOTCH1*	exonic	Frame_Shift_Del	34	c.7541_7542delCT	p.P2515fs	57%
501	*POT1*	exonic	nonsynonymous	7	c.G133A	p.G45R	49%
542	*MYD88*	exonic	stoploss	3	c.T478C	p.*160R	45%
548	*SF3B1*	exonic	nonsynonymous	15	c.A2098G	p.K700E	34%
561	*SF3B1*	exonic	nonsynonymous	15	c.A2098G	p.K700E	27%
568	*NBEAL2*	exonic	nonsynonymous	13	c.G1613A	p.R538H	47%
578	*SF3B1*	exonic	nonsynonymous	14	c.G1874A	p.R625H	41%
585	*EGR2*	exonic	nonsynonymous	3	c.C1000A	p.H334N	31%
588	*MYD88*	exonic	stoploss	3	c.T478C	p.*160R	43%
588	*SETD2*	exonic	nonsynonymous	3	c.A3175G	p.S1059G	48%
590	*ATM*	exonic	nonsynonymous	41	c.A6056G	p.Y2019C	11.16%
590	*SF3B1*	exonic	nonsynonymous	15	c.A2098G	p.K700E	43%
598	*ATM*	exonic	nonsynonymous	49	c.T7280C	p.L2427P	47%
604	*SETD2*	exonic	nonsynonymous	12	c.A5987G	p.Q1996R	51%
619	*EGR2*	exonic	nonsynonymous	3	c.G916A	p.E306K	56%
632	*POT1*	exonic	nonsynonymous	12	c.A713G	p.Y238C	48%
632	*SF3B1*	exonic	nonsynonymous	14	c.G1874A	p.R625H	48%
635	*ATM*	exonic	nonsynonymous	34	c.G5036T	p.G1679V	53%
644	*ATM*	exonic	nonsynonymous	50	c.T7499G	p.V2500G	55%
644	*EGR2*	exonic	nonsynonymous	3	c.G1081C	p.D361H	6%
644	*XPO1*	exonic	nonsynonymous	15	c.G1711A	p.E571K	8%
661	*ZFPM2*	exonic	nonsynonymous	8	c.G1721A	p.R574Q	41%
662	*MYD88*	exonic	stoploss	3	c.T478C	p.*160R	41%
664	*SF3B1*	exonic	nonsynonymous	15	c.A2098G	p.K700E	25%
670	*RPS15*	exonic	nonsynonymous	4	c.T398G	p.I133S	43%
674	*NOTCH1*	exonic	stopgain	34	c.C7375T	p.Q2459*	20%
674	*POT1*	exonic	nonsynonymous	6	c.G119A	p.G40E	8%
674	*SAMHD1*	exonic	nonsynonymous	7	c.G762A	p.M254I	73%
695	*ATM*	exonic	nonsynonymous	41	c.T6055A	p.Y2019N	47%
704	*SF3B1*	exonic	nonsynonymous	14	c.C1986A	p.H662Q	32%
704	*TP53*	splicing	.	6	c.559+1G>A		44%
704	*XPO1*	exonic	nonsynonymous	16	c.A1871G	p.D624G	49%
727	*TP53*	exonic	nonsynonymous	2	c.A182C	p.H61P	75%
163B	*SF3B1*	exonic	nonsynonymous	14	c.G1998C	p.K666N	39%
343G	*ATM*	exonic	nonsynonymous	49	c.A7304T	p.N2435I	43%
346B	*BIRC3*	exonic	frameshift insertion	10	c.1642dupT	p.Q547fs	38%

389A	*SF3B1*	exonic	nonsynonymous	14	c.C1873T	p.R625C	44%
465A	*SF3B1*	exonic	nonsynonymous	15	c.A2098G	p.K700E	47%
468B	*NOTCH1*	exonic	Frame_Shift_Del	34	c.7541_7542delCT	p.P2515fs	28%
488A	*SF3B1*	exonic	nonsynonymous	16	c.G2225A	p.G742D	36%
489E	*POT1*	splicing	.	6	c.9+1G>A		49%
504A	*ATM*	exonic	nonsynonymous	50	c.G7328A	p.R2443Q	15%
504B	*ATM*	exonic	nonsynonymous	50	c.G7328A	p.R2443Q	19%
531A	*RPS15*	exonic	nonsynonymous	4	c.C392G	p.P131R	52%
531D	*RPS15*	exonic	nonsynonymous	4	c.C392G	p.P131R	28%
558B	*POT1*	exonic	nonsynonymous	6	c.T74A	p.V25D	6%
675A	*NOTCH1*	exonic	Frame_Shift_Del	34	c.7541_7542delCT	p.P2515fs	51%
675A	*TP53*	exonic	nonsynonymous	2	c.A182T	p.H61L	87%

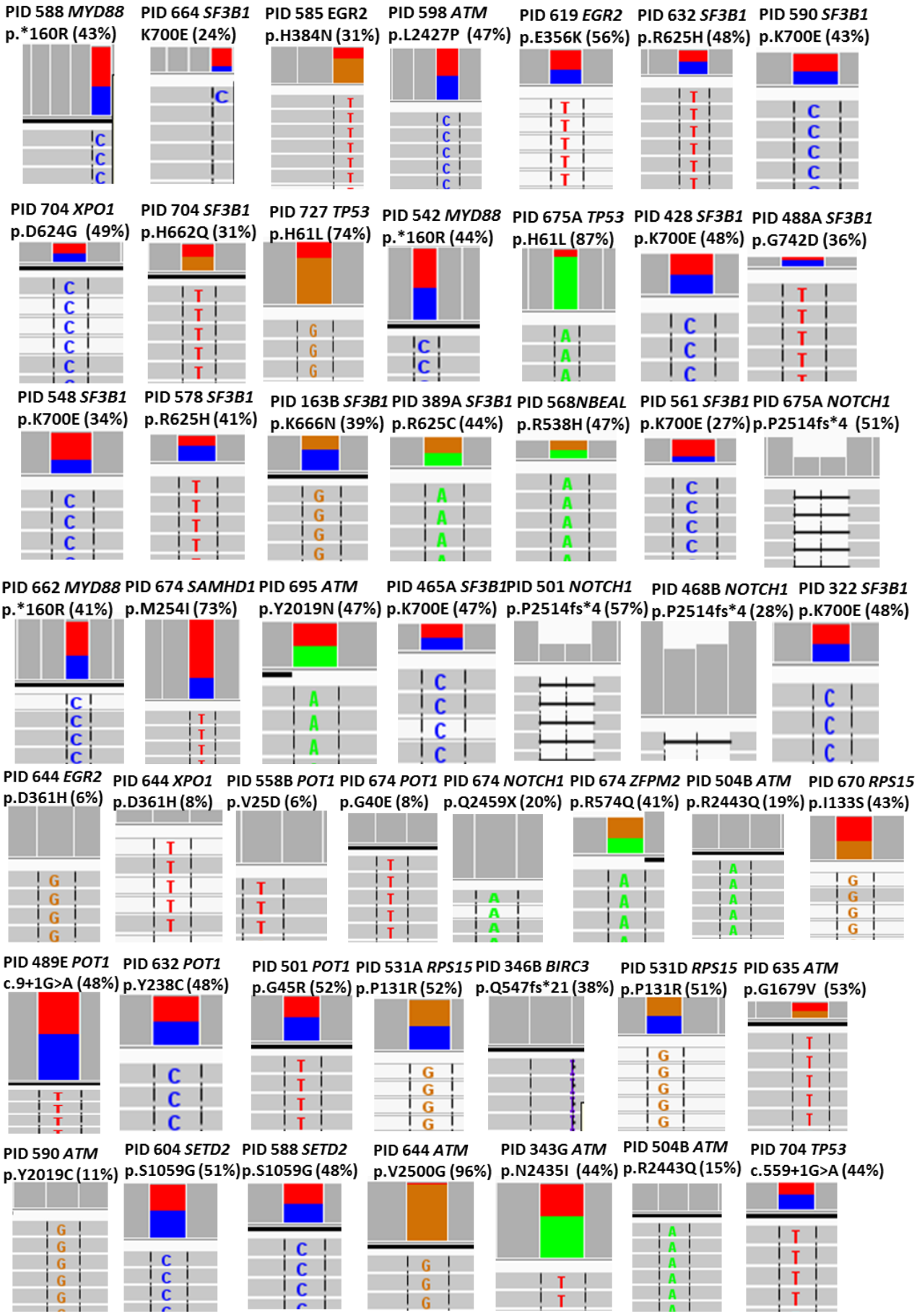


Figure 67. IGV confirmation of TruSeq variants (n=50) identified by the bioinformatical pipeline from the Southampton cohort (n=68 CLL patients).

7.8 Sanger sequencing confirms variants detected by TruSeq

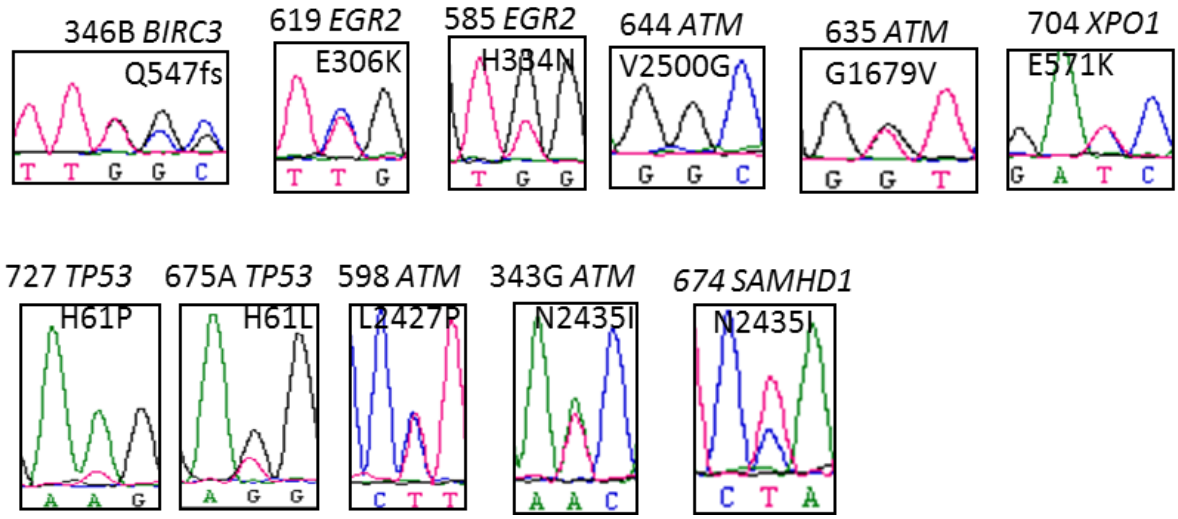


Figure 68. Sanger sequencing validation of TruSeq variants (Southampton cohort). A subset of variants identified from the TruSeq data with a VAF >20% were Sanger sequenced. This was to confirm and validate the variants identified in the TruSeq data. n=11 variants were confirmed in this manner.

Sanger sequencing was conducted on a small subset of the variants (Figure 68). The variants chosen to sequence (n=15) all had a VAF over 20%. Some of the assays failed to sequence (n=4) including the *RPS15* assay for exon 4 variants (n=2) and *ATM* assay for exon 41 variants (n=2), the sequencing data for these 2 assays on all 4 variants was too messy to interpret. This could be due to the nature of the mutation or else poorly performing primers, due to time constraints there was not time to re design or optimise these assays further. Overall 11/15 of the variants were confirmed by Sanger sequencing (Figure 68). The Sanger sequencing confirmation allows validation of the bioinformatical pipeline.

7.9 Genomic associations of variants

Fisher's exact test was conducted to see if any of the mutations associated with one another along with IGHV status and copy number alterations (Figure 69).

NOTCH1 associated with unmutated IGHV status (p=0.04) whilst *SF3B1* mutations associated significantly with mutated IGHV CLL (p=0.041). *EGR2* mutations associated with 11q deletions (p=0.042) and *RPS15* mutations with 17p deletions (p=0.019). *POT1* significantly associates with *NOTCH1* mutations (p=0.024), whilst there is a borderline association of *SAMHD1* mutations with *NOTCH1* (p=0.058).

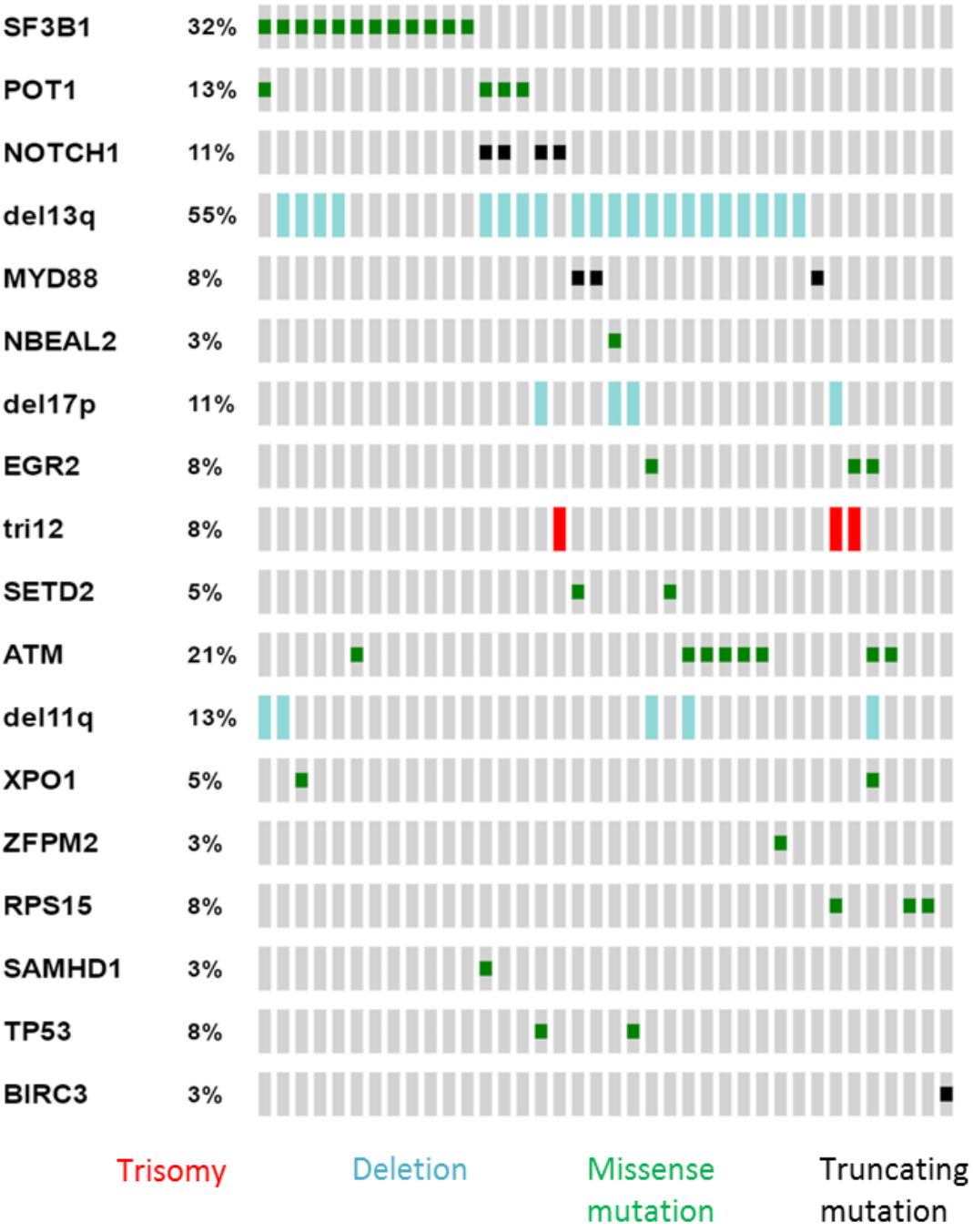


Figure 69 Gene mutations identified by the TruSeq custom panel in Southampton cohort associate with CNAs and each other, the number of mutations per gene was low, once more patients have been sequenced these may strengthen associations (http://www.cbioportal.org/oncoprinter.jsp#)

7.10 Gene expression analysis

Taqman technology (Thermofisher) was used to study the RNA expression in *BIRC3*, *POT1*, *TP53* and *ATM* genes, in a small cohort of the CLL patients already sequenced using TruSeq (n=60). These genes were chosen as a follow on from previous gene expression data, extended with additional genes of interest as well as in a different cohort of patients. Two reference genes were used *18s* and *PGK1*and a Raji cell line RNA was used as a control calibrator to normalise the gene expression data.

ATM gene expression was found to range from 1.03 to 61.12 and TP53 from 0.03 to 7.12, in all 60 cases. The mean expression was 12.78 (25^{th}, 50^{th} and 75^{th} quartiles 5.95, 9.18, 15.76 respectively) and for TP53 an average of 1.27 (25^{th}, 50^{th} and 75^{th} quartiles 0.52, 0.92, 1.55 respectively).

BIRC3 and *POT1* gene expression was lower than *ATM* and *TP53*, with the mean expressions 1.46 and 1.29, respectively. This meant a lot of the Cp values were above 35 so these could not be analysed further, the cohort was reduced to n=52 for POT1 while BIRC3 had n=37 patients with usable expression data.

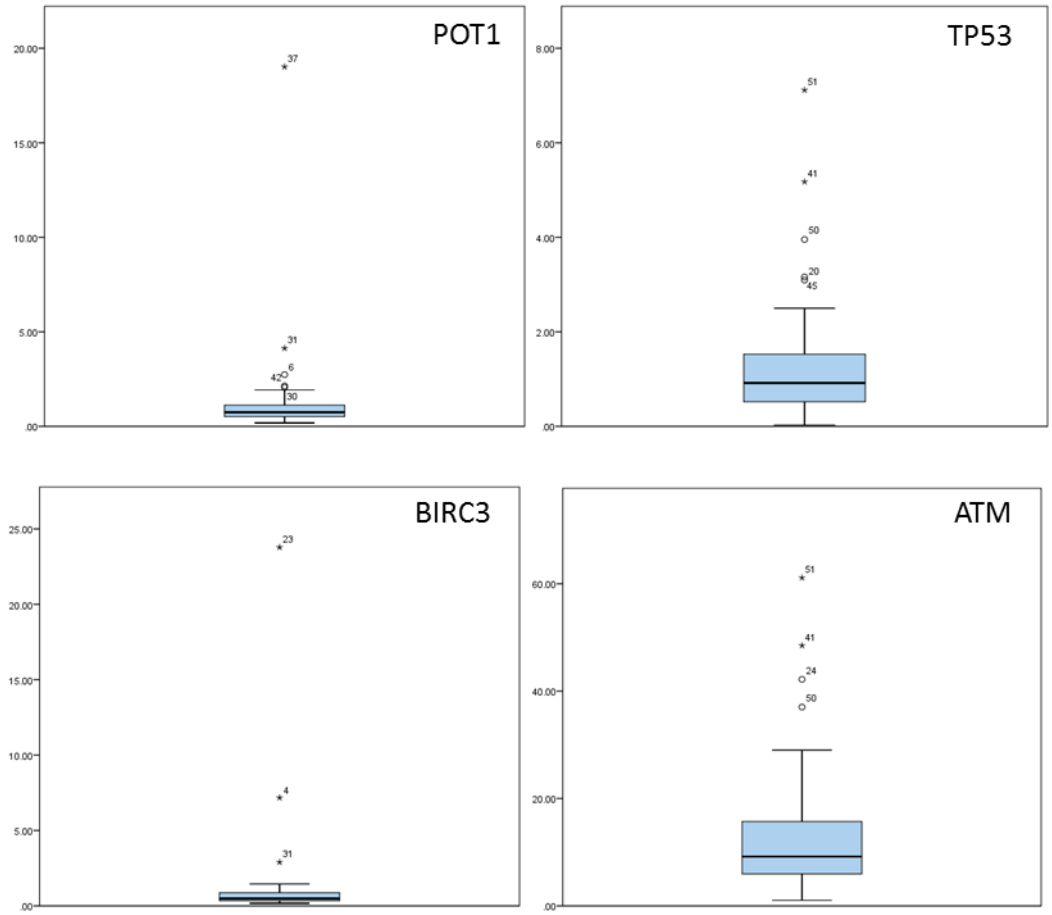


Figure 70 The range of gene expression for target genes; POT1, TP53, BIRC3 and ATM in a small cohort of TruSeq targeted sequenced patients. Reference genes 18s and PGK1 were used as well as a Raji cell line RNA as a calibrator to normalise the data against.

7.11 Correlations

A bivariate correlation test (Pearsons) was used to identify if the genes expression are associated with one another (Figure 71 and appendix).

POT1 and *BIRC3* gene expression as well as *TP53* and *ATM* gene expression both showed a positive correlation, with a Pearson's correlation coefficient value of 0.851 ($p<0.001$).

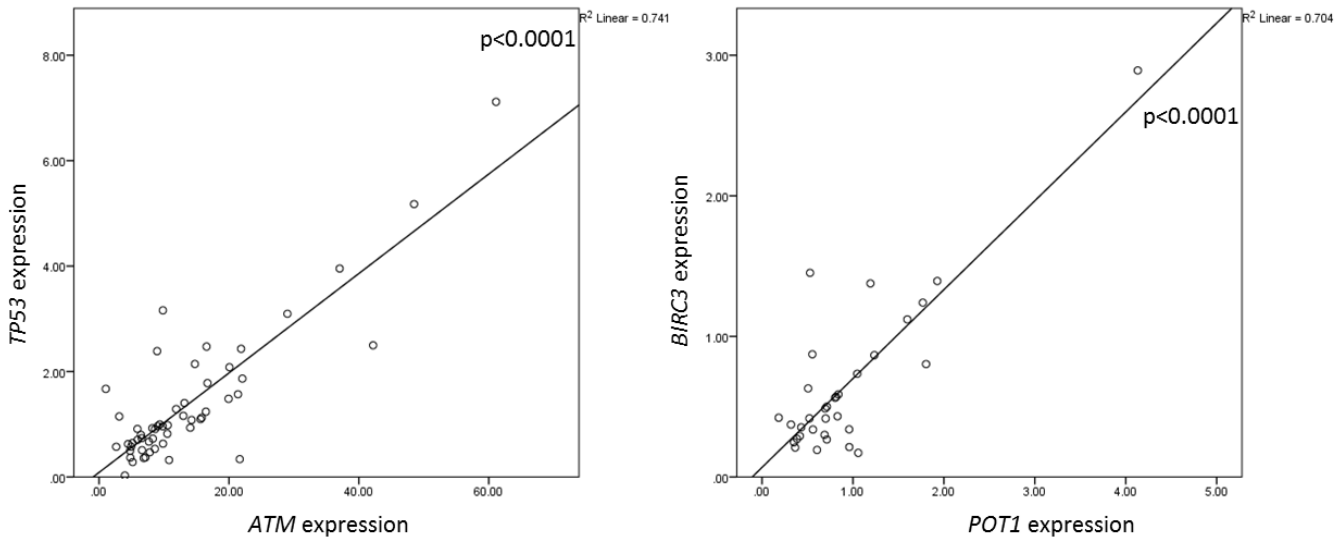


Figure 71 TP53 and ATM gene expression and POT1 and BIRC3 gene expression correlate positively and significantly, respectively.

7.12 Gene expression associates with genomic abnormalities

The Mann Whitney test was applied to test for any significance with other genomic abnormalities as detected by FISH (Figure 72, Figure 73 and Figure 74). *TP53* expression saw a significant increase in patients harbouring a trisomy 12 (p=0.049). *ATM* expression decreased in patients with an 11q deletion, whilst increased in patients with a trisomy 12 (p=0.038) or a 17p deletion (p=0.025). *POT1* expression significantly increased in patients carrying a *SF3B1* gene mutation and also had a borderline significant reduction in gene expression in *NOTCH1* mutated (p=0.054) or 17p deleted patients (p=0.091).

Low frequency gene mutations also drew significant associations with gene expression analysis (Figure 73). *EGR2* mutated patients saw a reduced *ATM* gene expression (p=0.003), whilst *ZFPM2* gene mutations identified a significantly increased *TP53* and *ATM* gene expression (p=0.037). Although there is only 1 case that is *ZFPM2* mutated, this case represents the highest expresser of both of these genes.

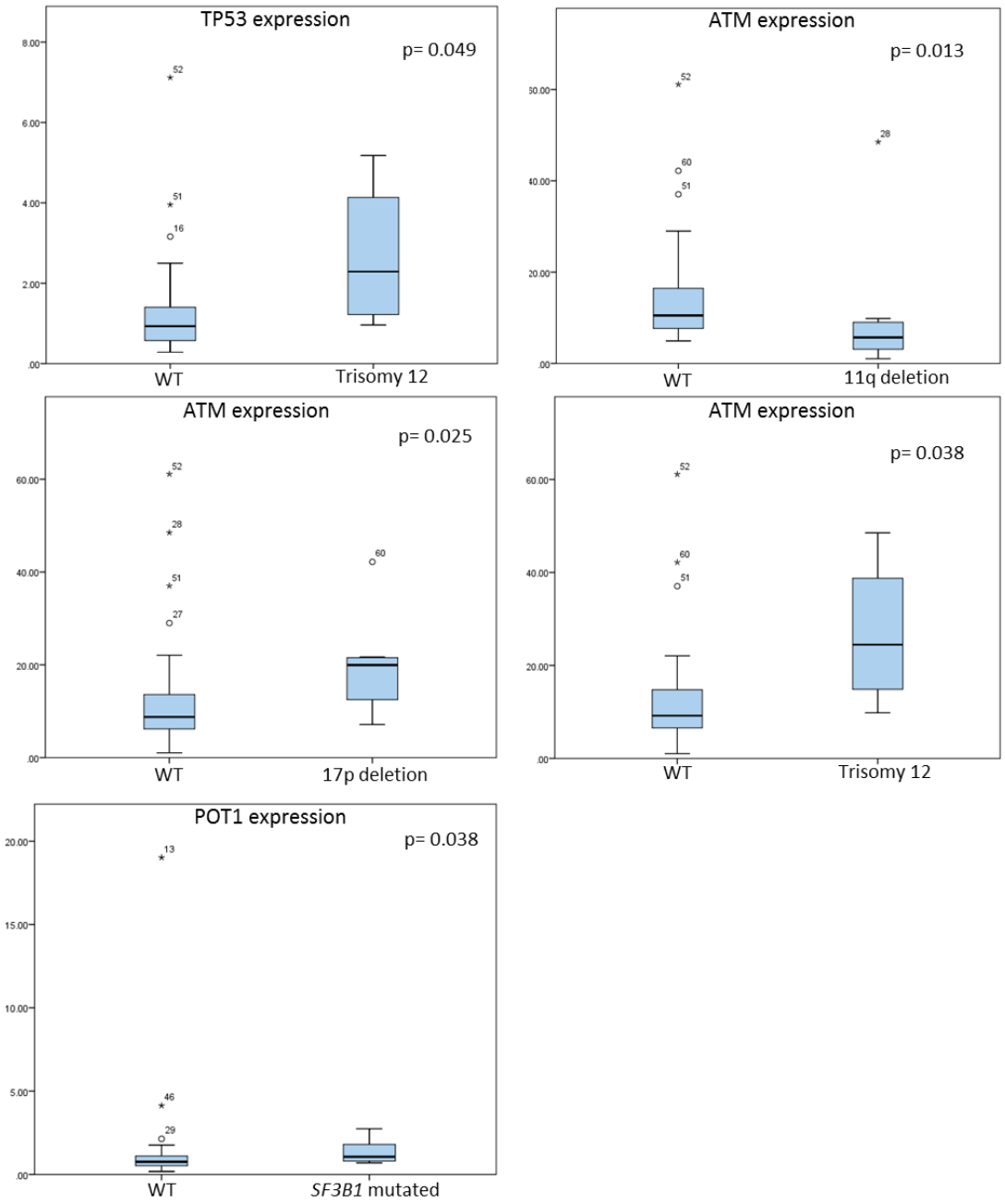


Figure 72 Significant genomic associations with gene expression data. *TP53*, *ATM* and *POT1* gene expression were all found to associate to one or more copy number alterations

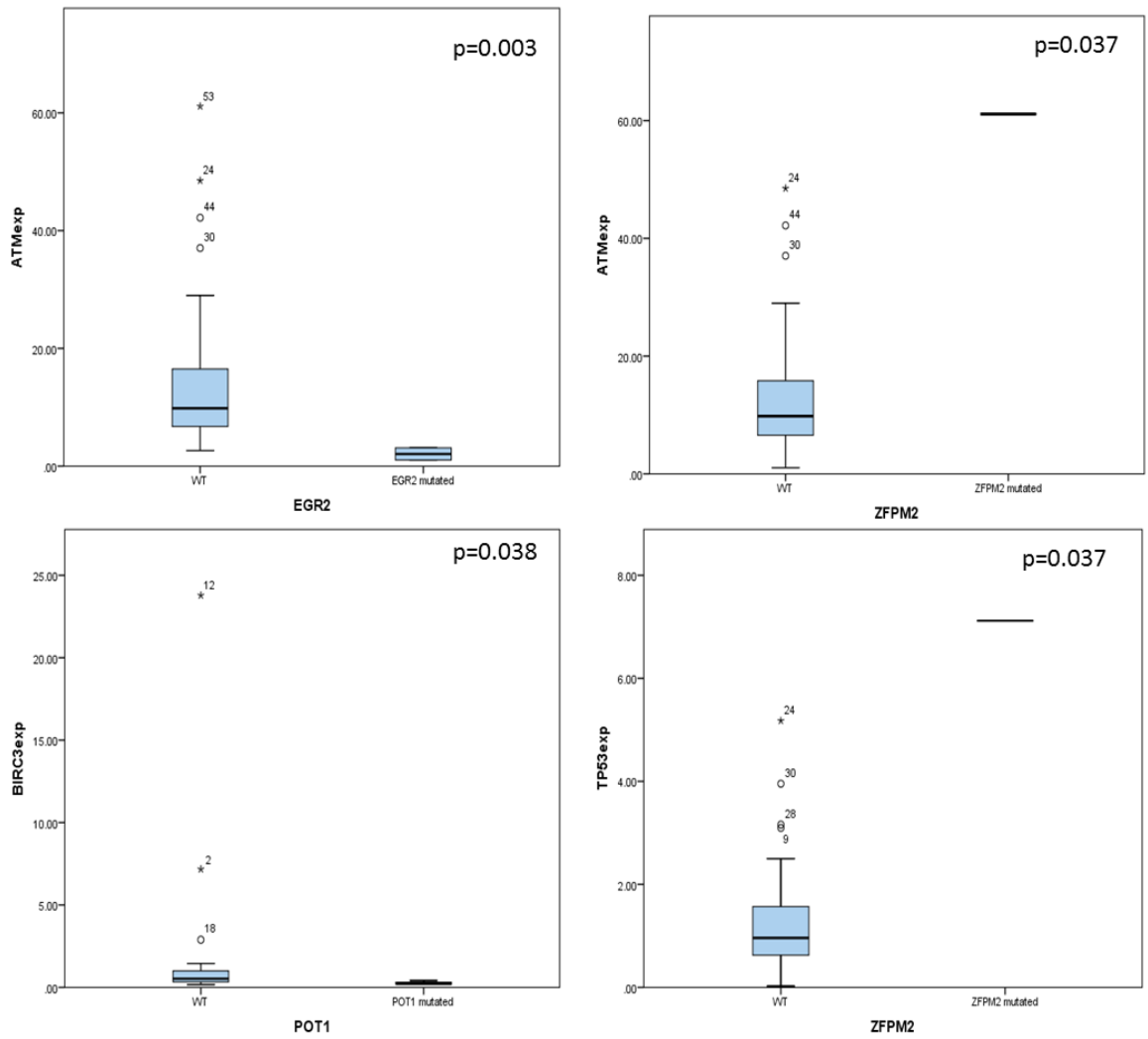


Figure 73 Low frequency gene mutations saw significant associations with gene expression analysis. *EGR2* mutated patients saw a reduced *ATM* gene expression (p=0.003), whilst *ZFPM2* gene mutations identified a significantly higher *TP53* and *ATM* gene expression (p=0.037).

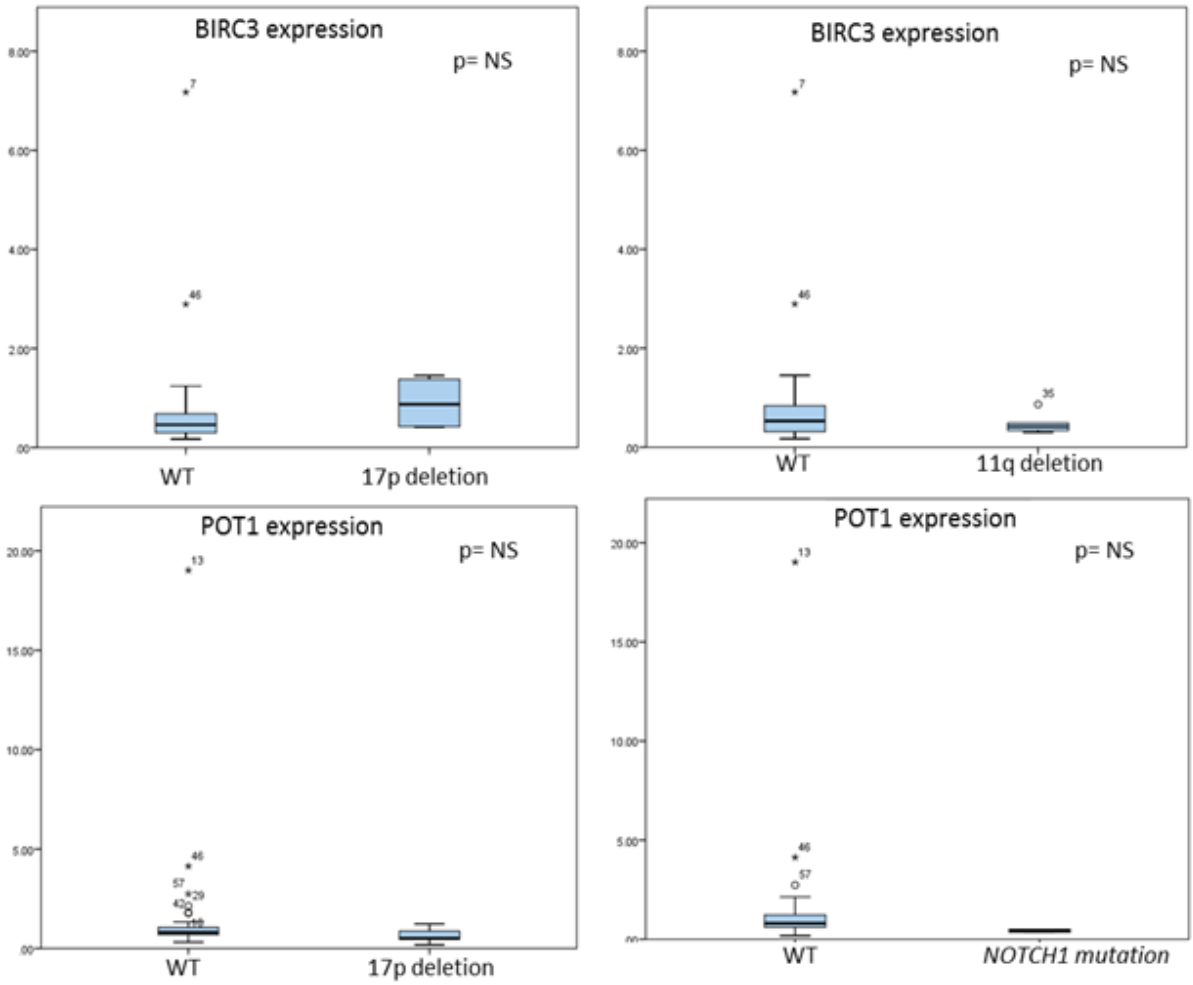


Figure 74 No significant associations were seen for *BIRC3* expression although a trend was seen with 11q and 17p deletions (reduced/increased, respectively). *NOTCH1* and 17p deleted patients had a borderline significance for reduced *POT1* expression.

7.13 Discussion

Next generation sequencing studies in CLL have been paramount to the discovery of genes and biological pathways and processes implicated in CLL. Targeted deep sequencing has furthered these discoveries by allowing focused studies on a select number of genes to be performed in large, informative cohorts of patients. As such, mutations in these genes have been described in the context of clinically relevant biomarkers and measures of clinical outcome, including overall survival and progression free survival.[115,205,206]

In this study, a custom designed 25 gene TruSeq panel (Illumina) was applied to the deep sequencing of 68 CLL untreated, diagnostic (Southampton) patient samples. Another TruSeq panel, this time with 12 CLL prognostically relevant genes was applied to 48 blinded CLL samples. These CLL patients all had a known mutation although their status was unknown to this study. As with many NGS studies, the management, processing and knowledge extraction from the huge amounts of data generated was a limiting factor. A new bioinformatical pipeline for the processing of raw TruSeq sequencing data had to be developed. The data received from this pipeline then had to be manually filtered and curated using IGV software; despite computational filtering, a large proportion (>50%) of the filtered variants were found to be false positives. Therefore, future analyses will require nuancing of the pipeline and more stringent filtering. In addition, the principal *NOTCH1* mutation (P2515Rfs*4), accounting for over 90% of all *NOTCH1* mutations, was not detected by the pipeline. This is likely due to the poor coverage of this gene and well established difficulty in sequencing caused by the high GC content. With approximately 13% of CLL patients harbouring this clinically relevant mutation, all sequenced patients had to be visually inspected for its presence; a lengthy process which identified 3 instances in the Southampton cohort and 12 in the ERIC cohort.

As already stated, during the filtering process to refine the variant list, all synonymous mutations were removed, as this type of mutation is assumed to be a silent event and will not change protein function. Synonymous mutations account for up to 40% of mutations detected in the cancer exomes and are frequently used by bioinformatical algorithms to estimate the background mutation frequency to confidently detect new cancer driving genes. [207,208] Supek *et al*. investigated over 3000 cancer exomes and more than 300 genomes and identified an enrichment of synonymous mutations on or near exon boundaries, they further showed that splicing is affected and so alternative protein transcripts may be expressed.[3] The study suggests that as many as 1 in 5 synonymous mutations may have been selected for pathogenesis and may represent up to 8% of driving mutations. It may be of interest in future studies in CLL to investigate the enrichment and the recurrence (as suggested by Supek *et al*.) of synonymous mutations, especially in light of no common

driving mutations identified in CLL and the implications of other splicing aberrations such as *SF3B1* mutations in the disease pathogenesis of CLL.[3,75,155]

Interestingly, the predominant recurrent *MYD88* mutation in CLL (P265) was not identified, although a p.*160R variant, which results in a premature stop codon, was detected in 3 patients in the Southampton cohort and 6 (out of 9) *MYD88* variants from the ERIC cohort. The significance of the codon 265 mutation in CLL has recently been hotly debated; this variant is a key factor in diagnosing WM and as this disease has notoriously been difficult to distinguish from CLL in the past, there is a suggestion that cases harbouring this mutation may be misdiagnosed WM cases.[170] Additionally the variant found in this study seems to be the more predominant variant and may be more CLL specific, although the variant has yet to be validated by Sanger sequencing, so this could just be an amplification artefact/ false positive variant.

The application of NGS to temporally or anatomically distinct samples from the same CLL patient have facilitated the investigation of sub clonal expansion and evolution of CLL throughout the disease course , providing interesting insights from both a biological and clinical perspective. Cancer cells are subject to the Darwinian concept of 'survival of the fittest', through dynamic sub-clonal evolutions. Of interest to clinicians and patients is the understanding that therapeutics are known to change clonal dynamics and may even result in the loss of a dominant low-risk clone, paving the way for therapy-resistant sub-clonal cells to thrive. In these cases, therapy may in fact drive disease progression. More targeted treatments are being designed and rolled out in clinical trials, such as the drug ibrutinib, which has already had a huge impact upon chemo-immunotherapy resistant CLL patients. Two of the patients from the Southampton study had sequential samples, one harbouring an *ATM* mutation whilst the other a *RPS15* mutation. The *ATM* mutation in the later sample had increased from 15% to 19% VAF, whilst the *RPS15* mutation had decreased from 52% to 28%. This would be an interesting feature to correlate with clinical outcomes such as disease progression, treatment response and other clinical biomarkers.

Even with the relatively small number of variants identified in the Southampton project, there were some genomic associations identified. Again relating the gene mutations to the expression data identified some interesting findings were identified. One such was that both *ATM* and *TP53* gene expression correlated significantly, both of these genes are involved in the response to DNA damage and ATM has a well-known role of phosphorylating and so activating TP53, so it is perhaps not too surprising that these genes correlate positively.[209] Though, more unexpectedly, *POT1* and *BIRC3* gene expression also showed the same positive correlation. The latter again was confirmed looking at the effects of *POT1* mutations on gene expression, with reduced *BIRC3*

expression in *POT1* mutated patients. This could be attributed to the interesting finding by Tong *et al*, whom showed that the damage response kinases ATM and ATR, are essential for the recruitment of telomerase at human telomeres.[210] This was then closely followed by Lee *et al*, whom also identified ATM as a crucial component of telomere elongation in both mouse and human cells and showed that by blocking ATM this inhibited telomere lengthening.[211] As previously discussed ATM has shown to also have a role in the NFκB signalling pathway along with BIRC3, so there may be a role of BIRC3 in telomere maintenance or else a feedback between POT1-ATM-BIRC3 in the NFκB pathway.[191,210,211]. There was also an increased *POT1* expression in *SF3B1* mutated cases. *SF3B1* mutated patients are known to have severely compromised and shortened telomeres.[128] Raa *et* al. identified that *SF3B1* mutations are associated with increased DNA damage and an aberrant response to DNA damage similarly to *ATM*/*TP53* disrupted patients.[117] So it could be interesting to look further at different isoforms of *POT1* expressed in *SF3B1* patients compared to other CLL patients to see if different POT1 isoforms are expressed differentially in these patients.

Another interesting result was that *ZFPM2* mutations show an increase in both *ATM* and *TP53* gene expression. Although only 1 mutation was identified, this case had the highest relative gene expression for both *ATM* and *TP53* in all the cases analysed. The zinc finger protein, FOG family member 2 (*ZFPM2)* gene is a protein that binds to the N-terminal zinc finger of the GATA4 protein, and dependent upon the cellular context and target gene, it acts as either a transcriptional coactivator or corepressor.[212] ZFPM2 has a role in the PI3K/AKT pathway and DNA damage response pathway, that could activate/ increase p53 levels, though this will need to be confirmed in more patients and examined at both the RNA and protein levels.[213,214]

EGR2 is a transcription factor and plays an important role in early B cell differentiation, it is also a downstream target of BCR signalling, known to be deregulated in CLL.[215] *EGR2* has recently been identified as being recurrently mutated in advanced CLL (8%) and proposed as a mutation that precedes CLL and may alter BCR signalling driving pathogenesis.[215] In this study, *EGR2* mutations associated significantly with 11q deletions. This was further validated in the expression analysis as there was a significant decrease in *ATM* gene expression seen in the *EGR2* mutated patients.

RPS15 has recently been identified as being recurrently mutated using WES, in up to 19.5% of therapy (FCR) relapsed CLL cases.[141] The RPS15 protein has been shown to interact with MDM2 and mutant RPS15 was shown to induce deregulation of p53 protein. [141] The WES study also highlighted that a third of the *RPS15* mutated cases, also carried a concurrent *TP53* abnormality. This was also seen in this study; with *RPS15* mutated cases showing a significant association with 17p

deletions, though no significant dysregulation of *TP53* expression was seen in the expression analysis.

This TruSeq (Southampton) study identified several genomic associations in line with the literature, even with relatively few variants identified. As this study continues, the number of patients sequenced will be increased and so the statistical analysis will advance, these genomic associations may become more apparent and will be validated appropriately. Tying the genetic data, identified by the TruSeq custom panel in this study with the biological data from the same cohort, investigated in other Southampton research groups would be hugely interesting and beneficial in understanding disease pathogenesis. Though due to time constraints and not receiving all the bioinformatical data has meant that this was outside of the scope of this project at this current point.

The ERIC study is a pan European project involving 9 centres, including Southampton. The project is assessing the clinical applicability of using targeted resequencing technologies and sent out blinded samples with known clonal gene mutations, to all 9 centres to run Haloplex (Agilent) or TruSeq (Illumina) custom gene panels. All the data generated (fastq and variant lists) and left over prepared libraries have been sent back to the ERIC consortium for further analysis to validate this technology. Though from the variant analysis only 41 of the 48 patients sequenced had a mutation in the final list. This means that the filtering may have removed some of the real variants, so the variant analysis may need more refinement for future projects. As more clinicians wish to incorporate next generation sequencing into the general practice to help refine prognosis, treatment choices and predict disease course, it is imperative that more studies are done to 1) understand the individual effects of each specific gene mutation and 2) implementing the most useful technology into a clinical setting.

As next generation sequencing is becoming cheaper, quicker and easier to perform on bench top sequencers and is therefore more readily available to the wider scientific community. Further cataloguing and characterising these gene mutations in CLL is vital to identify their clinical and biological importance across the entire disease course, as targeted therapy and personalised medicine becomes a reality.

Chapter 8: Developing Nextera XT technology for extended screening

8.1 Introduction

When the CLL exome data (previous discussed) was cross referenced with one of the largest NGS studies published in CLL, two genes stood out as recurrently mutated and found in both studies; these were *DNAH9* and *SPEN*. [75] Additionally, two other genes *BTK* and *PLCG2* seemed exceptionally important to study in more detail and to develop a sensitive assay to detect low level clones for reasons to be further discussed below.

8.2 BCR signalling in CLL and resistance of the small molecule inhibitor, ibrutinib

B cell receptor signalling (BCR) in normal cells can result in cell proliferation, apoptosis or anergy. BCR signalling has been shown to be critical to cell proliferation and survival in CLL.[42] A recent alternative to the standard chemo-immunotherapy regime for relapsed or resistant CLL patients has been the use of a small molecule inhibitor Ibrutinib.[42] Ibrutinib irreversibly covalently binds to the cysteine-481 amino acid, inactivating the kinase and preventing BCR signalling that in turn stops proliferation and in CLL induces apoptosis.[42] Bruton's tyrosine kinase (BTK) is a critical component of the BCR signalling pathway. BCR activation triggers the formation of a signalosome that includes the BTK protein that is phosphorylated by LYN and SYK. [216] BTK and SYK then phosphorylate phospholipase Cγ2 (PLCG2) that leads to calcium mobilisation activation of ERK, AKT and nuclear factor-κB that then lead to cell proliferation and survival, differentiation and antibody production.[216] From recent clinical trial studies in clinically relapsed CLL patients, ibrutinib has been shown to induce some promising results for controlling this high risk group. One such study, a phase 1b-2 multicentre study assessed ibrutinib in 85 high risk CLL patients.[42] These patients were divided into two separate treatment groups; one with a high dose of ibrutinib and the other with a lower dose. The overall response rate was 71% with a complete or partial response and an additional 15-20% in both of the treatment groups with a partial response but also with lymphocytosis. No difference was seen between the two ibrutinib dosages, with the low dose group just as effective as a high one. The OS rate was 83% and a PFS rate of 75%.[42] As more patients are treated with ibrutinib, it has become more apparent that some patients as the treatment progresses are starting to develop resistance to the kinase inhibitor and the relapse.

Woyach *et al*. analysed 6 CLL patient exomes, whom had all become resistant to ibrutinib treatment.[217] They found that 5 patients all carried the same *BTK* gene mutation at the cysteine 481 amino acid changing it to a serine molecule. This mutation prevents the ibrutinib molecule from binding to the BTK protein so ibrutinib is therefore ineffective at preventing BCR signalling.[217] Additional mutations have been identified in ibrutinib refractory patients in the phospholipase C, g2(*PLCG2*) gene, these mutations allow the BTK independent activation of the BCR pathway.[218] In the final patient no *BTK* gene mutation was identified through the exome data but an arginine to tryptophan mutation in the *PLCG2* gene at amino acid 665 was detected. One of the patients that carried a low level *BTK* mutation also had 3 additional *PLCG2* mutations, one the same 665 codon mutation, another at position 845 and the final mutation on codon 707. The mutations were all validated using Ion torrent sequencing. [217] Samples taken prior to the ibrutinib therapy were also analysed using whole exome sequencing but no mutations were detected in either *BTK* or *PLCG2* genes. Using Ion torrent sequencing again the authors detected in 3 of the patients, mutations in 0.2% of the pre-treatment samples reads. [217] It is therefore essential for patients being enrolled onto an ibrutinib regime to be screened for mutations in the *BTK* and *PLCG2* genes, before and during treatment to monitor resistant clones.

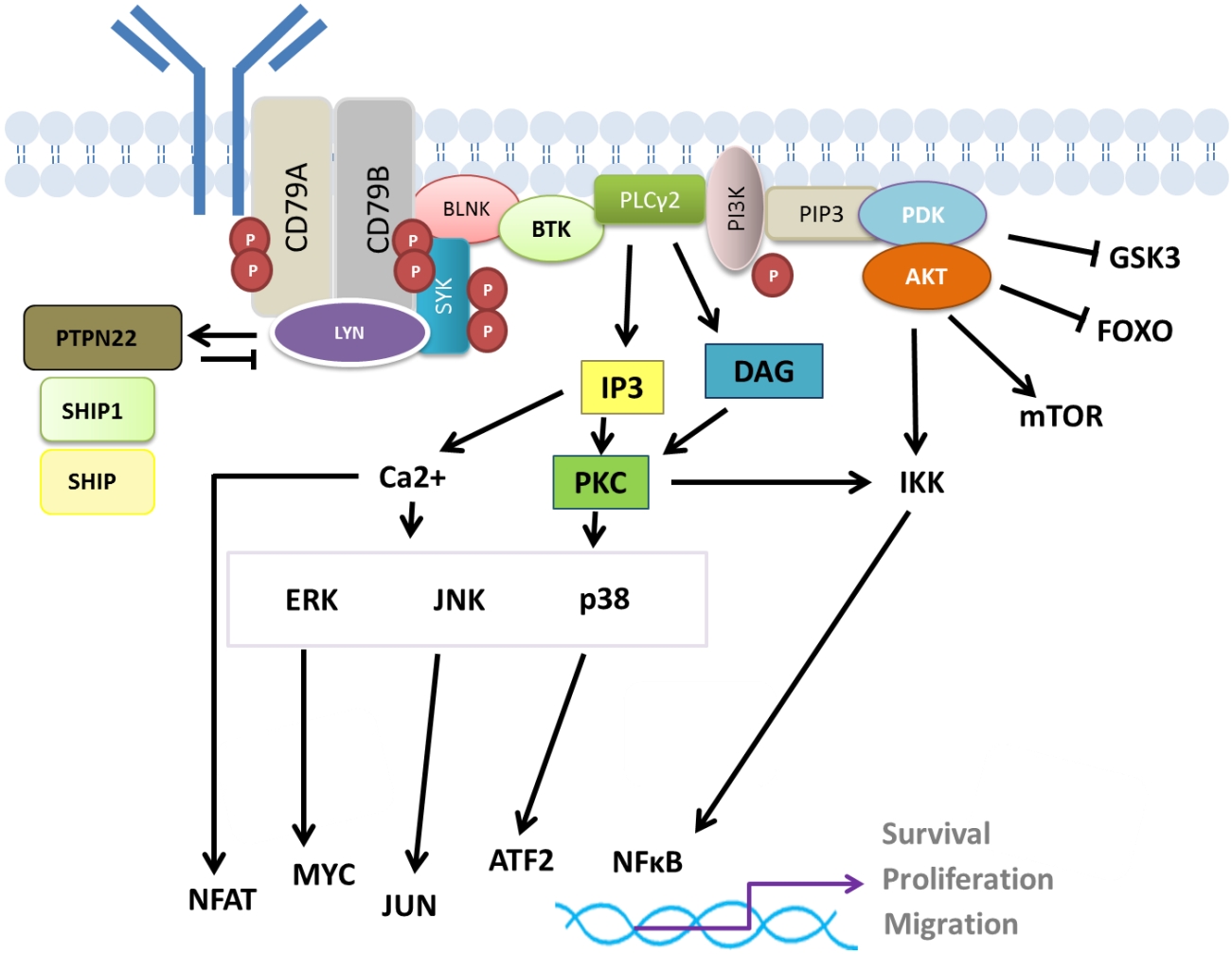


Figure 75 Figure adapted from Stevenson *et al*. 2011. Blood. 118. 16. The B cell receptor (BCR) is a membrane immunoglobulin that is bound to CD79A and CD79B. Activation of the BCR by antigen causes spleen tyrosine kinase (SYK) to be recruited to CD79A/B . BTK activation causes the phosphorylation of phospholipase C-gamma 2 (PLCγ2) that then promotes production of diacylglycerol (DAG), inositol triphosphate (IP3), which activate protein kinase C (PKC) that leads to NFκB activation. An uptake of calcium can also upregulate the transcription factor; nuclear factor of activated T cells (NFAT).[231]

8.3 DNAH9

The dynein, axonemal, heavy chain 9 (*DNAH9*) gene encodes a heavy chain subunit of axonemal dynein, which is a large multi-subunit molecular motor. The DNAH9 protein attaches to microtubules and hydrolyses ATP that mediates the movement of cilia and flagella.[146] Recently *DNAH9* was identified as a recurrently mutated in a WES study (Ion torrent) of CML with a sensitivity to tyrosine kinase inhibitor treatments.[146] *DNAH9* was identified as a gene that carried novel splice site mutation, affecting the donor splice site of intron 52 and this variant was found to be frequently mutated (52%) in atypical chronic myeloid leukaemia patients using whole exome sequencing.[219] The DNAH9 protein has also been shown to associate and interact with the B cell Lymphoma 6 (BCL6) protein in a non-Hodgkin's Lymphoma cell line by tandem mass spectrometry .[146,220] BCL6 functions as a transcriptional repressor, it is a member of the Zinc finger and Broad complex, Tramtrack, Bric-a-brac (ZBTB) family that is comprised of an N-terminal BTB/POZ (BTB/Poxvirus and zinc fingers) domain as well as Krüppel-type zinc fingers at the C terminus.[221] The zinc finger domain allows for the protein to bind to DNA at TTCCT(A/C)GAA consensus sequence.[221] BCL6 is required for B cell maturation and differentiation, with studies showing that BCL6 is essential in the survival of early B cells that emerge after successful selection of the pre-BCR checkpoint and thus subsequently allowing the early B cells to develop further via immunoglobulin light chain recombination. [146,221] BCL6 protects germinal centre B cells against p53 and Arf DNA damage induced apoptosis during somatic hypermutation and class switching, through the negative regulation of Arf. [222] BCL6 expression is only found during normal B cell development in the germinal centre, but it has also been implicated in malignancies and was first discovered as a frequent translocation involving the Ig heavy or light chain region; 3q27, in diffuse large B cell lymphoma.[221,222] Similarly BCL6 has been implicated in other neoplasms such as ALL, were it is thought cell survival is brought about by BCL6 repressing TP53 expression in patients carrying the Philadelphia chromosome (BCR-ABL1 fusion gene) abnormality.[221]

8.4 SPEN

Split Ends (SPEN) protein is a large nuclear protein of 402 kDa with four N-terminal RNA recognition motifs and a C-terminal SPOC (Spen Paralog and ortholog C-terminal) domain. SPEN has been shown to be critical during embryogenesis and throughout development due to being involved in important signalling pathways such as TCF/LEF, EGFR and NOTCH.[223] SPEN negative regulates NOTCH signalling in marginal zone B cells, through the inhibition of the transcription factor RBPJ. SPEN mutations have been identified in SMZL at a low frequency (~5%), inactivating

mutations are predicted to cause a truncated protein without a C-terminal domain that is unable to interact with RBPJ and so is incapable of inhibiting NOTCH signalling.[13] A recent breast cancer study identified LOH of the SPEN locus in 23% of breast cancers as well as a small fraction carrying a mutation of this gene.[223] Using estrogen receptive breast cancer cell lines the authors demonstrated that SPEN regulates proliferation, tumour growth and survival. The study further investigated the clinical significance of SPEN expression in 1784 breast cancer patients and identified a significantly better survival in those with high levels of SPEN. This study identified SPEN as a novel biomarker for tamoxifen response in estrogen receptive breast cancers.[223]

The *DNAH9* and *SPEN* genes have been identified in several studies as being involved directly/indirectly in CLL, SMZL and/or similar B cell diseases. These genes as previously discussed, were identified from exome sequencing in this study and one of the largest whole genome and whole exome sequencing studies as recurrently mutated in CLL but have yet to be investigated further in an extended cohort of patients.[75] Instead of using the same mutational screening technique (HRM and Sanger sequencing) as previously utilised for chapter the development of a new targeted sequencing NGS technique; Nextera (Illumina) would be optimised and developed. This was to keep up to date with current screening methods and this method would also be crucial to detect low frequency mutations. Nextera uses a uses a transposase that simultaneously cleaves and tags the DNA fragment with an adapter sequence, followed by a short PCR step that amplifies the DNA and inserts indexing sequences. This allows for a vastly quicker and easier protocol, prepared libraries can then be sequenced on an Illumina sequencing platform.

8.5 Aims

- Develop and optimise assays for 4 genes of interest; *DNAH9, SPEN, BTK* and *PLCG2*
- These assays will then be used with Nextera technology to deep sequence a small cohort of CLL patients for mutations in the respective genes

8.6 Methodology

8.6.1 **Assay design-BTK and PLCG2**

Four assays in total were designed to cover the mutational hotspots for the BCR genes BTK and PLCG2. The one BTK assay covered codon 481, whilst the other 3 PLCG2 assays covered codons 665, 707 and 848. They were all designed to similar parameters; similar but differing lengths (381-392bp) and a Tm of 55°C to allow for multiplexing of the assays. The individual primers were evaluated using oligo calc (http://biotools.nubic.northwestern.edu/OligoCalc.html) to check for

primer self-annealing sites/hairpin formations, as well as using the *in silico* PCR blat software (http://genome.ucsc.edu/cgi-bin/hgPcr?hgsid=512243145_X9dPBvPcM8VRXnRsEfNaUuV1snnf) to check the assay binds to the correct position and only that position on the genome.

Table 34 Assays designed to cover the mutational hotspots in the *BTK* and *PLCG2* genes. Assays were designed with multiplexing in mind

Gene	Chr	exon	Assay name	oligonucleotide sequence	Primer	Tm (°C)	product size
BTK	X	15	BTK_C481Sf	ATCTTCCACTGCTACTTCCA	Forward	55	381
BTK	X	15	BTK_C481Sr	TCTGATGCTCTACTCCTAGG	Reverse	55	
PLCg2	16	19	PLCg2_R665Wf	GTGATTTCCATGGTCGTTTC	Forward	55	388
PLCg2	16	19	PLCg2_R665Wr	TGGTGGTTGTTGTTGTCG	Reverse	55	
PLCg2	16	20	PLCg2_S707Yf	ACCCTGTTTATTTACCCACTT	Forward	55	392
PLCg2	16	20	PLCg2_S707Yr	TCAGAGGTTTGCTGAATTCA	Reverse	55	
PLCg2	16	24	PLCg2_L845Ff	GCTTCATTATTTAGCTCTCTCC	Forward	55	383
PLCg2	16	24	PLCg2_L845Fr	CCAGCCTGACCTAAATACTC	Reverse	55	

8.6.2 Assay design-SPEN

Assays were designed to cover exon 11 of the SPEN gene, just this exon was chosen as the SPEN gene is very large and all non-synonymous somatic mutations that were uncovered by Puente *et al* were distributed along this exon. As it is, exon 11 is the largest exon of SPEN, at 8000kb long. To get maximum coverage of the exon using Nextera technology, around 100bp will be lost due to the transposase cleavage of the product so the 8 assays were designed to have overlapping amplicons with an additional 100bp included in the design. All assays were designed around the same Tm (55°C) but with differing product size (Table 35) so as to aid the multiplex reaction and to be able to easily distinguish each amplicon on an agarose gel, respectively.

Table 35 Primers designed for Nextera assays for the SPEN gene. Primer3 software was used for primer design. Assay 8 has two reverse primers as the first assay could not be optimised.

Gene	exon	Assay name	oligonucleotide sequence	Primer	Tm	product size
SPEN	11	Assay 1	GCTACTACAAACTAATCATTTCACA	Forward	55	1297
SPEN	11	Assay 1	CTCTGGTTTTTGCTTTTCTG	Reverse	55	
SPEN	11	Assay 2	AAATCTTCTGCCCTGGAC	Forward	55	1246
SPEN	11	Assay 2	CCATATCTGCCATCTGTTCT	Reverse	55	
SPEN	11	Assay 3	CGAGAACGGAATTACAGAAG	Forward	55	1265
SPEN	11	Assay 3	GTTTTGTCACCAGTGGTCTT	Reverse	55	
SPEN	11	Assay 4	ACAGCACAACTGATTCCATT	Forward	55	1259
SPEN	11	Assay 4	AGTTTTCTGGGACCTTGG	Reverse	55	
SPEN	11	Assay 5	GTAAGGAGCGTCTATGCAAC	Forward	55	1305
SPEN	11	Assay 5	TGATCGGGTTGTTTTCTG	Reverse	55	
SPEN	11	Assay 6	CCCAGACAGATCTGCAAC	Forward	55	1319
SPEN	11	Assay 6	CAAACTTTTCAGTGTGGTCA	Reverse	55	
SPEN	11	Assay 7	CAAACAACTCTGAGATACAAGC	Forward	54	1276

SPEN	11	Assay 7	AAATGGGAGACAGTTCGAT	Reverse	54	
SPEN	11	Assay 8	GAGGCACAGTGAAGGTTCT	Forward	55	1377
SPEN	11	Assay 8	GCTGGGATTACAGGCATA	Reverse	55	
SPEN	11	Assay 8_2	TTAAGCAGTCCTCCCATCT	Reverse	55	1410

8.6.3 DNAH9

As with the SPEN assay design, all assays for DNAH9 were designed in the same manner. The amplicons covered all the exons which had previously been implicated in CLL from the Puente *et al* study, including exons 15, 20, 22, 31, 32, 35, 42, 48 and 56. In total, 10 assays had to be created to cover these regions.

Table 36 Assays designed to cover the DNAH9 gene. Exons chosen to screen had been previously identified as having mutations.

Gene	exon	oligonucleotide sequence	Primer	Tm	product size (bp)
DNAH9	15	AGCAAATTCTCCACAAGAGA	F	55°C	
		GCATCATTCACAAAAGGTCT	R	55°C	387
DNAH9	20_1	AGAATGTTTTTGGTGAAAGC	F	55°C	
		ACCTCATCCTCATAGTGGTG	R	55°C	654
DNAH9	20_2	ATATCCGAAACCTGGACAA	F	55°C	
		AGGTAAGGTGAGCATCTGTG	R	55°C	741
DNAH9	22	AAGTGTTTCTTGTGCTAATGC	F	55°C	
		ACACAGACAGGTCTGAAGTACA	R	55°C	373
DNAH9	31	TTTTCTAAGTCATCCCTTCC	F	55°C	
		AAGAGTGGTTGAAATGAATTG	R	55°C	579
DNAH9	32	TTCAGACCTCCTGACATCTT	F	55°C	
		CTCCCTTTCTACTGGCTTTT	R	55°C	509
DNAH9	35	GCCAAATCTCCTTTTCTTTT	F	55°C	
		ATCGCTCTGGGAACATCT	R	55°C	394
DNAH9	42	GTTTAGCAAGACAAGCCCTA	F	55°C	
		ACAGAATTTCCTTCCTCACC	R	55°C	415
DNAH9	48	TCCAGGATGTTAGGGTTAAA	F	55°C	
		CTGATGACATGGGAGTCTCT	R	55°C	506
DNAH9	56	CTCATGGTCTGAAGCTCTTT	F	55°C	
		AAGACAGAAAAGCCATGTGT	R	55°C	431

8.6.4 Optimisation of assays

All assays were first tested separately in a standard 30µl PCR reaction using a standard touchdown PCR programme. If amplified well with a product length matching to the design then the assays were then tested in a multiplex reaction. Again if the PCR worked, then a 1:100 dilution of the PCR product was tested using the assays separately again to check for amplification of each assay. Assays were then optimised in four multiplex reactions (BTK-PLCG2, SPEN; SPEN1, 3, 5 and 7 and SPEN2, 4, 6, 8, and DNAH9).

Table 37 PCR conditions for developing the assays to be used for Nextera

Standard PCR reaction:	**volume**
2x PCR mastermix (promega)	25µl
Primer (10µM)	2.5µl each
dH2O	18µl
Multiplex PCR reaction:	
2x PCR mastermix (promega)	25µl
Primer (10µM)	1.25µl each assay pair
dH2O	18µl

8.6.5 Samples

57 CLL samples were chosen for the BTK-PLCG2 assays, these included 21 CLL patients that are currently undergoing an ibrutinib treatment regime, and include sequential sampling during the course of treatment. Two of the patients (PID 346 and 665) are known to have become resistant to ibrutinib therapy, for unknown causes.

Table 38 The 17 CLL patients currently undergoing ibrutinib treatment. 57 samples from pre-treatment and various time points, were chosen to analyse for *BTK* and *PLCG2* mutations using Illumina Nextera sequencing technology

PID	Pre-treatment	Week 1	Time point 1	Time point 2
343	343G	343H	343P	
346	346B	346D	346N	346Q
409	409F	409G		409N
489	489E	489F		489P
495	495D	495E	495M	
531	531D	531E	531N	531Q
551	551C	551E	551N	551P
558	558B	558C	558L	558P
601	601B	601D	601L	
632	632D	632E	632M	632AB
644	644B	644C		644L
727	727A	727B	727I	727L
222	222E	222F		222L
665	665A	665B		665L
619	619A	619B		619H
530	530C	530D		530K
573	573C		573L	573N

37 CLL samples that had also been used in the TruSeq experiment and/or the BTK assays were chosen to be screened for *DNAH9* and *SPEN* mutations.

All samples were extracted using the standard Qiagen mini prep kit and diluted to working concentration (~20ng/µl). The four multiplex PCRs (BTK-PLCG2, DNAH9, SPEN1 and SPEN2) were performed and PCR samples quantified using HS DNA Qubit and then the DNA was diluted to 0.2ng/µl ready for the Nextera protocol. The DNAH9 and SPEN assays were pooled together before the Nextera XT steps were performed and all samples from both sets of assays were loaded into one plate.

8.6.6 **Nextera XT**

The Nextera XT kit (Illumina) was used and the protocols followed accordingly to tagment and amplify the PCR products so that these can then be sequenced on the Illumina NGS platform, MiSeq. To be able to sequence all 96 samples at once the 96 indices kit (Illumina) was used.

Manifest files and sample sheets were created for these amplicons.

The Nextera protocol was first tested on control corriel DNA (4 mixed/diluted samples) the sensitivity could also be assessed with known SNPs within the sequencing regions of the four assays. Three SNPs were identified within the BTK (rs3747288) and two in the PLCG2 gene assays; exon 19 rs4243222 and exon 24 rs4366702. Corriel DNA samples were mixed together with one of the samples having a different genotype to the others. The DNA was run on an agilent high sensitivity chip (Figure 78) to check the average product size after tagmentation (326bp)this was to use to calculate the library to 4nM (supp data!?). Then 1µl of diluted library was then spiked into a TruSeq run (1%) and this was then sequenced on the MiSeq. Using the fastq toolkit app (Illumina basespace) to first trim the sequence, then the data was converted to BAM files using BWA aligner. This was then put into IGV to visually inspect that the data aligned to the correct locations on the reference genome.

8.6.7 **SNP 6.0 data**

Archival CLL4 SNP 6.0 data, was used to focus on the 17p deletion (n=16) and on the *DNAH9* gene, which is found in this loci.

8.6.8 Data analysis

8.6.8.1 Optimisation

Once the fastq files were generated by the MiSeq, these were first manipulated using the FASQ Toolkit APP on BaseSpace. This APP allows the user to remove adapter sequences and small read sequences. For this data the Nextera adapter sequences were removed and a minimum read length of 32bp was selected. Next the BWA aligner basespace App was used to create a BAM file. The BAM files were then uploaded to IGV to inspect manually for the

8.6.8.2 96 Samples

Due to the number of samples sequenced (n=96) and the development of a bioinformatical pipeline, all the sequences were trimmed and aligned to a reference genome using the pipeline and then the BAM files were opened in IGV to identify any variants. Due to the bioinformatics pipeline being optimised for TruSeq and not Nextera, no variants were identified for BTK or PLCG2 so all samples had to be checked by eye in IGV identified

8.6.8.3 Calculating copy number from Nextera product

The total copy number of template (or number of molecules) imputed into the Nextera reaction could be calculated using a formula to convert to ng then multiplying by the amount of template. (http://cels.uri.edu/gsc/cndna.html) The number of moles per gram are calculated using avagadros number and one mole is based upon the average weight of a base pair, which is 650 Daltons.

Number of copies= (Template (ng) * 6.022 x10^{23}) / (Length of template (bp)) * 1x10^{9} * 650

Using this formula the copy number in both the Nextera reactions were able to be calculated; only 1ng of PCR template was added to the Nextera reaction:

BTK-PLCG2: 2.84 x 10^{9}

DNAH9-SPEN: 2.05 x 10^{9}

8.7 Results

8.7.1 Breakpoint deletion detected in *DNAH9*

The CLL4 SNP 6.0 archived data was analysed at the 17p loci, one of the patients carried a breakpoint deletion within the *DNAH9* gene. The breakpoint deletion is between the 18th and 24th

exon, the exact location cannot be determined from the SNP 6.0 data as it is interpreted between two probes but this region contains a dynein heavy chain, domain 2.

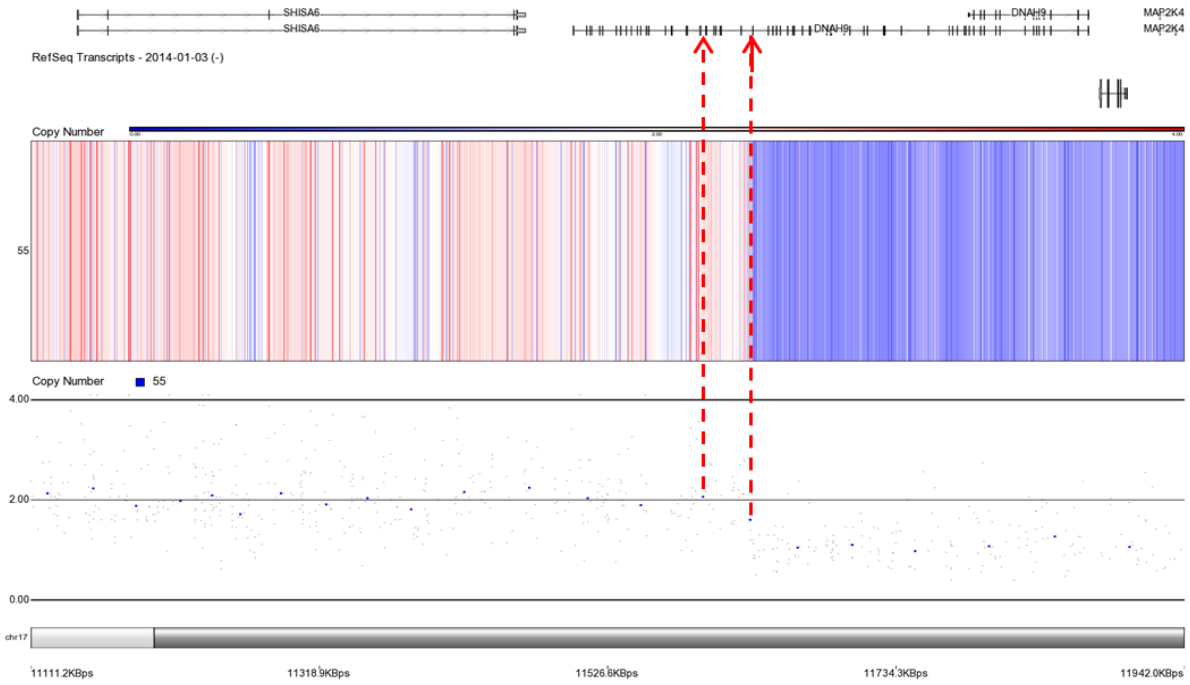


Figure 76 Breakpoint deletion identified within the *DNAH9* gene in 1 patient, as detected from archivial SNP 6.0 data. Data interpreted using Partek software.

8.7.2 Assay optimisation

All assays where optimised individually to find the best annealing temperature, all the DNAH9 and BTK and PLCG2 assays worked fine at 55°C but SPEN assays 2 and 8 had additional non specific bands and had to be further optimised at a higher Tm (Figure 77). Assay 2 (SPEN) the additional products disappeared at 62°C Tm whilst assay 8 still had multiple products so the reverse primer was redesigned (assay 8_2) that again had additional bands, but with further optimisation was found to work with a Tm of 60°C. The next step was to optimise assays to multiplex together. The BTK and PLCG2 assays were multiplexed successfully as these gave similar length products and annealed at the same temperature (55°C). The SPEN assays were split into two alternate multiplex reactions one with assays 1,3,5 and 7 (55°C) and the other with 2, 4, 6 and 8 (60°C). All 10 assays for DNAH9 were multiplexed successfully in one reaction (Tm 55°C).

8.7.3 Testing the Nextera protocol with *BTK* and *PLCG2* assays

To check the protocol performed with the custom design assays, and that these sequenced and then aligned correctly, a pilot run was done. The pilot test utilised the BTK-PLCG2 assays with the Nextera XT test DNA (Corriel DNA), and the resulting product was spiked into a TruSeq run (1%). The sequencing data was trimmed and aligned manually (basespace) and visualised using IGV. The sequencing data aligned to BTK exon 15 (covering codon 481) as well as aligning to PLCG2 exon 19, 20 and 24. This meant that the assays were successful in demonstrating deep sequencing covering the correct regions of interest (Figure 80). Using the Corriel DNA meant that known SNP frequency was utilised as a test for sensitivity (Table 39). The expected frequencies did not match with the sequenced read frequency, suggesting PCR bias during the amplification.

Table 39 Expected SNP frequency was not seen in the SNPs that fell within the PLCG2 genes, for SNP rs4243222 there was a bias seen, where the T was preferentially amplified

SNP	Allele	Expected	Actual
rs4243222	C	63%	48%
	T	37%	51%
rs4366702	C	63%	60%
	A	37%	40%

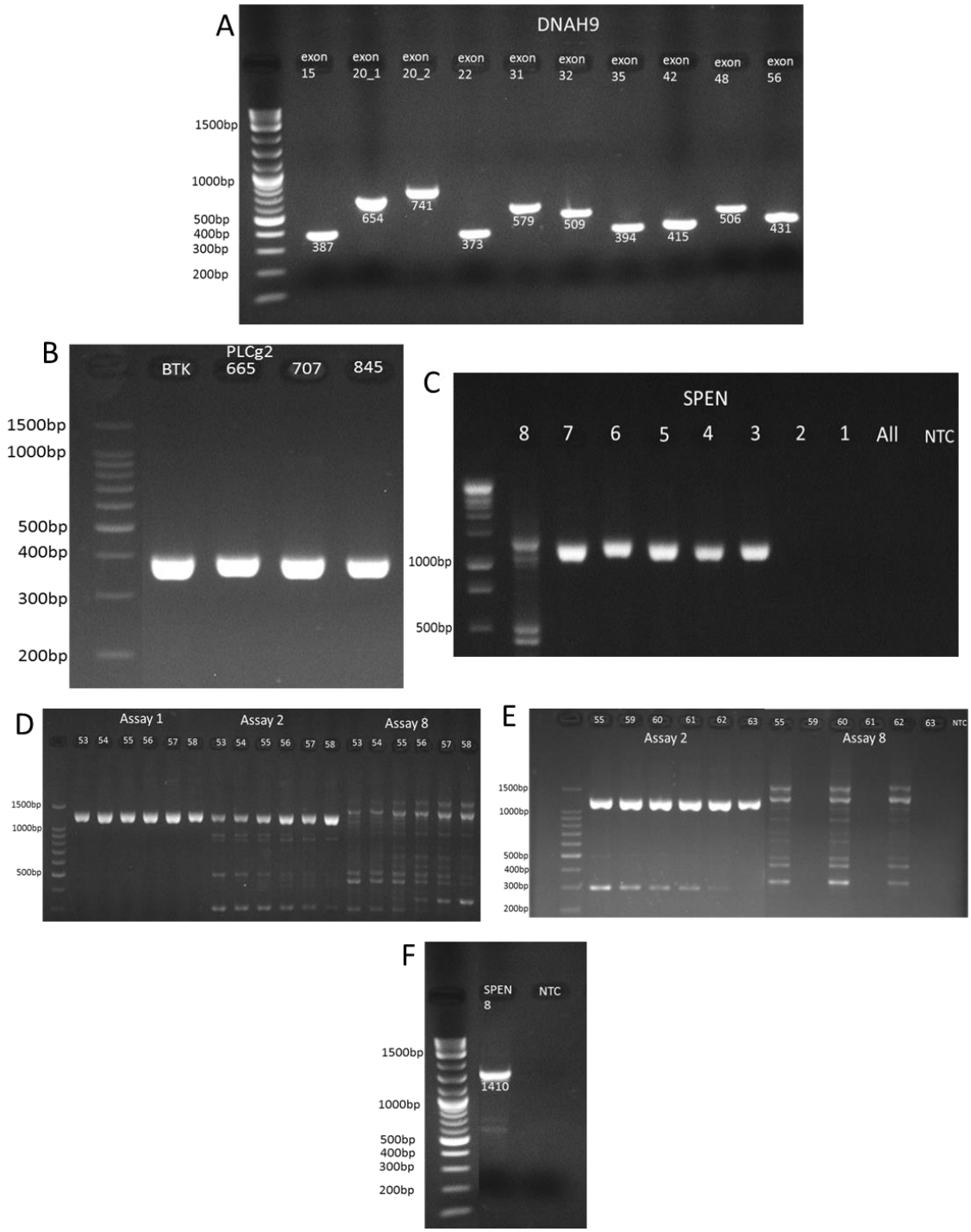


Figure 77 Assay optimisation of Nextera assays. All assays where first tested individually in a PC reaction at 55°C. A-B. DNAH9 and BTK and PLCG2 assays all amplified well at this temperature. C. SPEN assays 3 to 8 amplified but assay eight had multiple bands suggesting off target amplification. D-E. Further optimisation of SPEN assays at a higher temperatures. Assay 8 could not be amplified so the reverse primer had to be re designed (F).

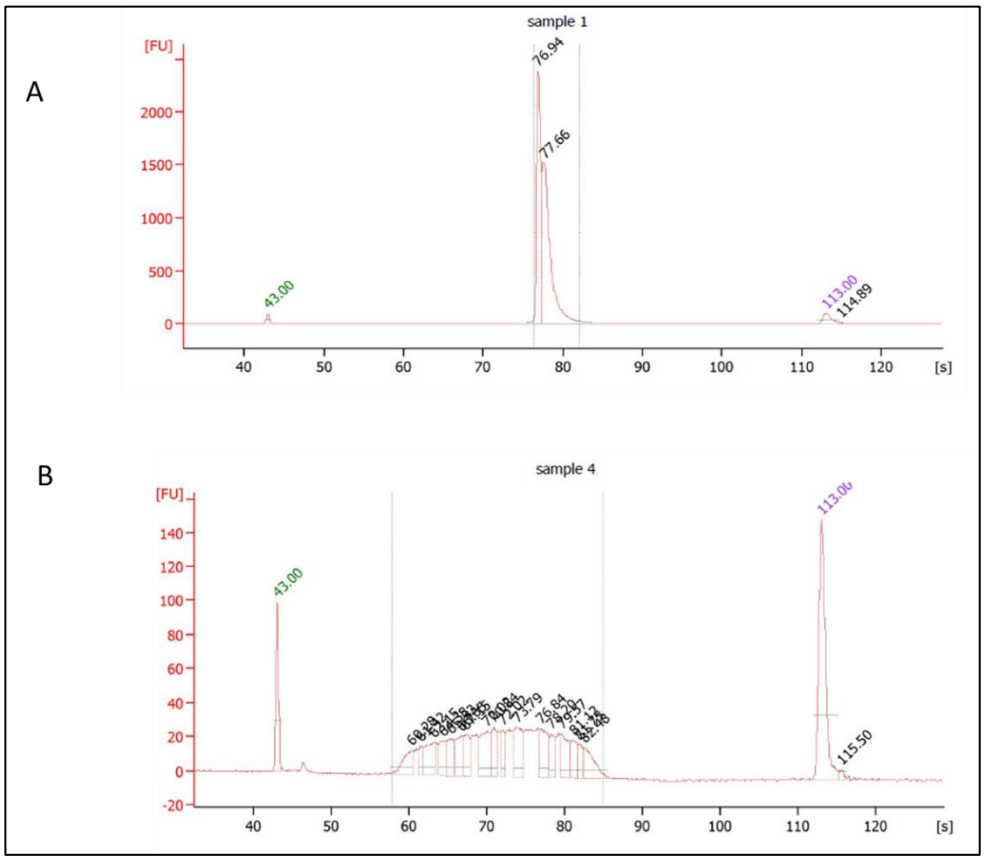


Figure 78: Agilent high sensitivity DNA assay to assess the Nextera XT tagmentation. A is the BTK and PLCG2 multiplex PCR products before the Nextera protocol. B shows the PCR product after tagmentation has occurred.

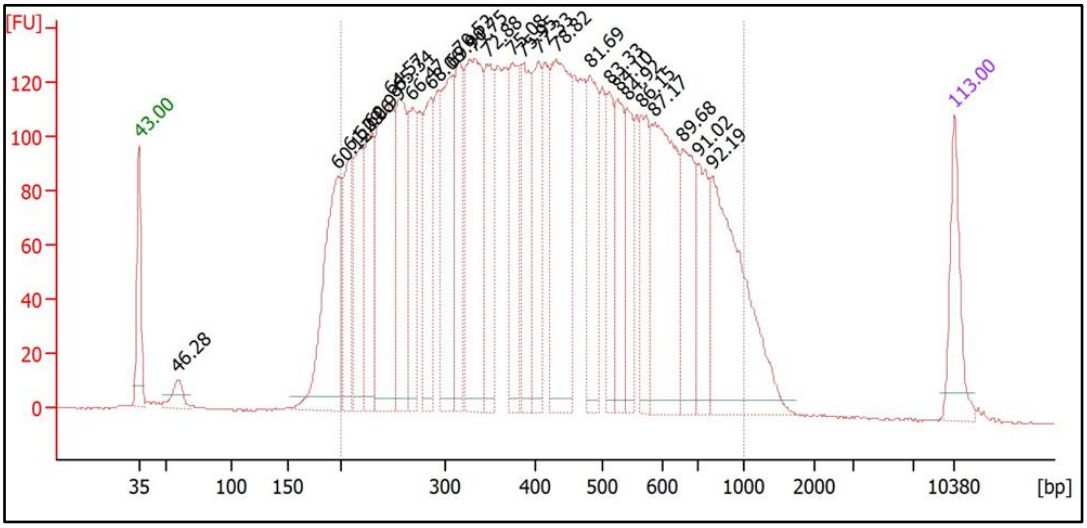


Figure 79 DNAH9 and SPEN multiplex PCR product after Nextera reaction. The product was run on an agilent chip to check Nextera tamentation had occurred. The figure shows the expected fragmentation with an average fragment size of 453bp.

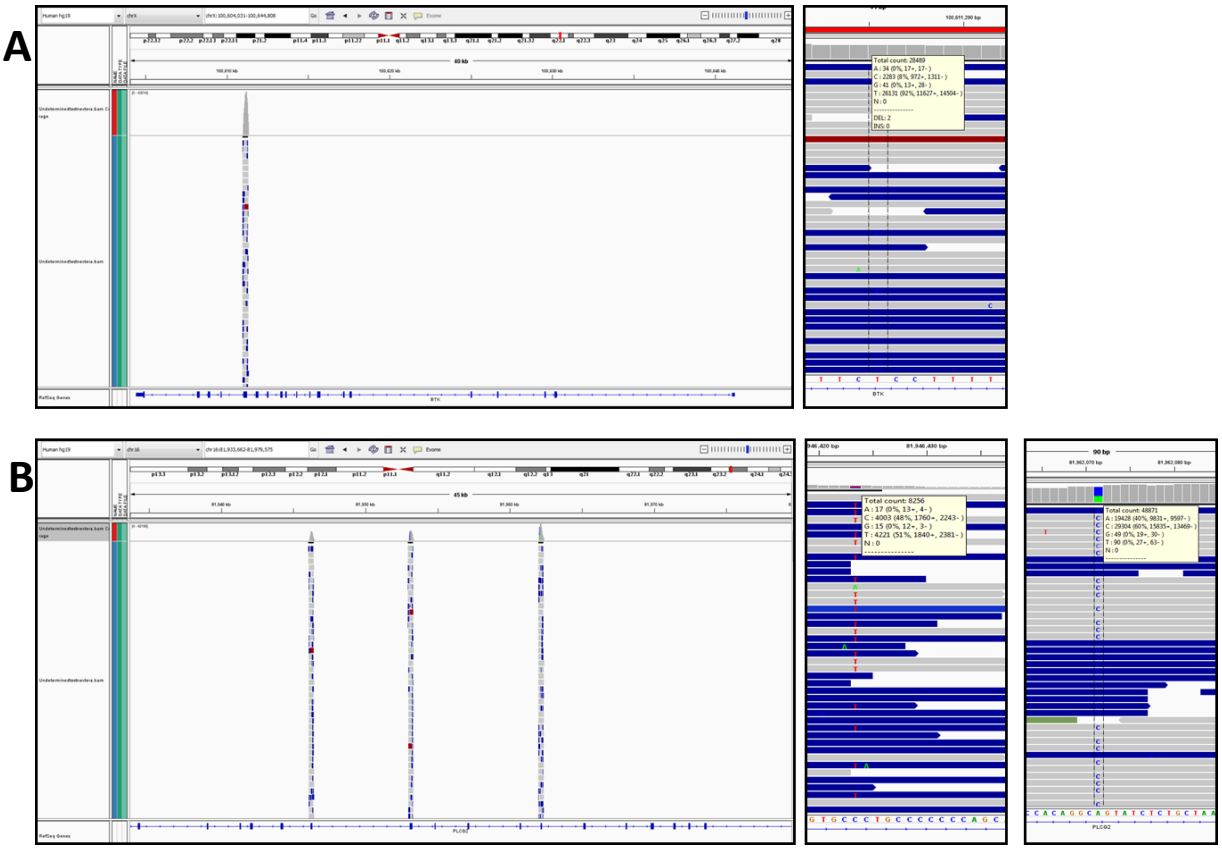


Figure 80 Sequencing of the BTK-PLCG2 multiplex PCR Nextera product on the MiSeq platform, spiked into a TruSeq run. IGV screenshots of the Nextera fragments aligned to reference genome. A The correct alignment to exon 15 to the *BTK* gene B. correctly aligns to the *PLCG2* gene: exons 19, 20 and 24. All amplicons have good coverage. The SNPs BTK rs3747288, PLCg2 exon 19 rs4243222 and exon 24 rs4366702 SNP were used to test sensitivity.

8.7.4 *DNAH9* and *SPEN* assays aligned to correct locations

The MiSeq sequenced DNAH9 and SPEN assays aligned to a reference genome successfully and were visualised using IGV. DNAH9 assays all conformed to the correct exons; 15, 20, 22, 31, 32, 35, 42, 48 and 56 whilst the coverage in SPEN was to exon 11 (Figure 81).

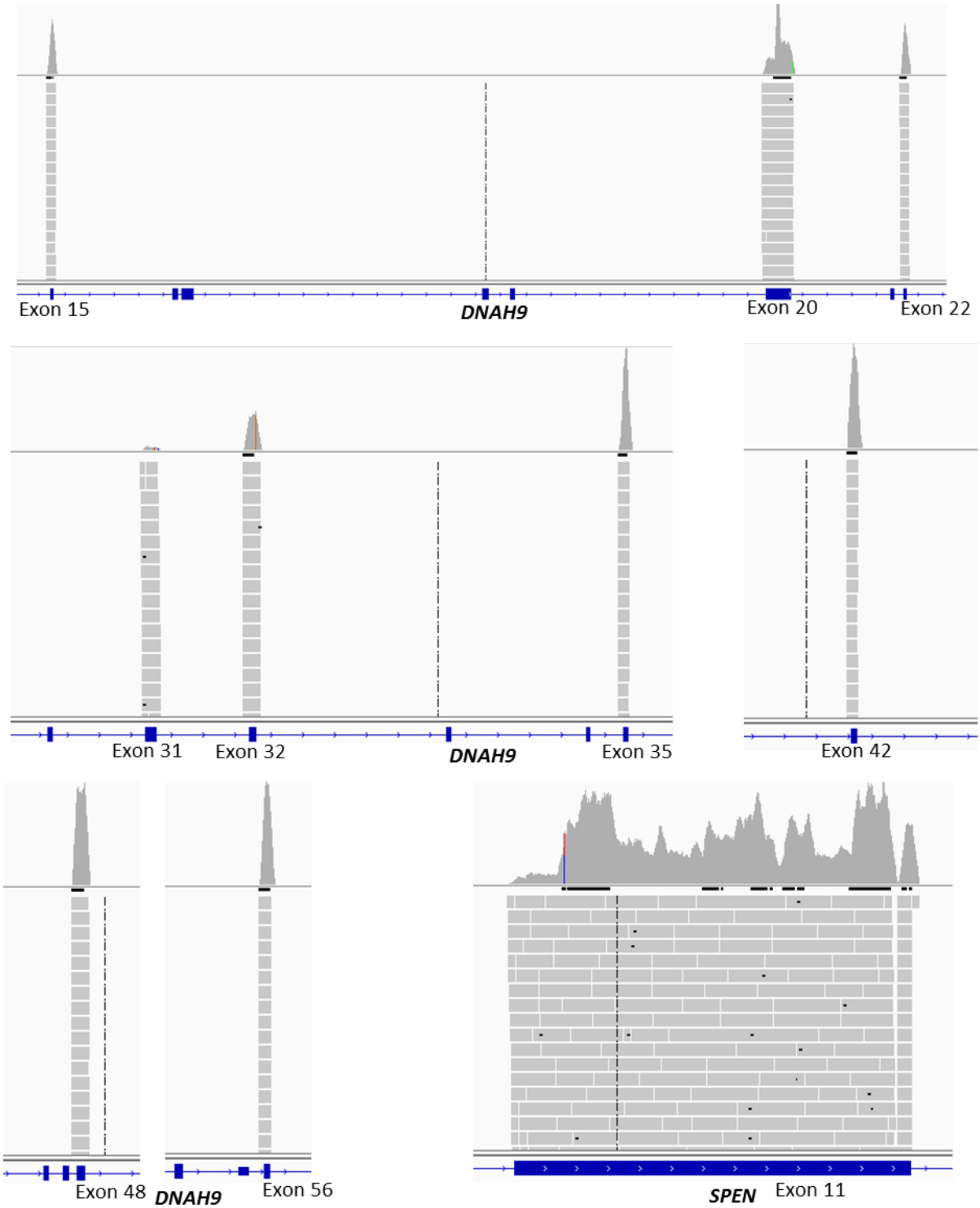


Figure 81. Sequencing of Nextera product on MiSeq then aligned to a reference genome and viewed in IGV. The fragments aligned to the correct locations on *DNAH9* gene exons 15, 20, 22, 31, 35, 42, 48 and 56, and also to the *SPEN* gene exon 11.

8.7.5 No *BTK* or *PLCG2* gene variants identified in Ibrutinib treated patients

The Nextera sequencing data was put through a bioinformatical pipeline to detect variants. No variants were identified by the bioinformatics pipeline in any of the samples sequenced for *BTK*

and *PLCG2* (n=57). As the variants may have been missed by the pipeline due to the low frequency of the mutations each sample was uploaded to IGV and viewed by eye at the corresponding codons to check if any mutations were present. No mutations were identified.

8.7.6 Variants identified in *DNAH9* and *SPEN* genes

There was a total of 7839 variants detected by the bioinformatics pipeline, which was then further filtered to remove synonymous variants with a VAF <5% and read depth <100. The final variant list gave only 7 variants, these were then checked in IGV to be real. Four of the seven variants were confirmed in IGV (Table 40). PID 675A had two confirmed *DNAH9* mutations; a stop-gain mutation in exon 48 (p.E3033*) and a non synonymous variant in exon 20 (p.R1517L). Both of these *DNAH9* mutations were low frequency mutations only found in 6% and 5% VAF respectively. Two *SPEN* variants were identified in PID 695, both non synonymous mutations within exon 11; p.R1411G was a clonal mutation identified in 43% of reads whilst the other (p.R1905S) was a low frequency (5%) mutation.

Table 40 *SPEN* and *DNAH9* gene variants identified and then filtered from the bioinformatical pipepline 4/7 were confirmed to be real variants using IGV to visualise the mutations within the sequencing data

PID	chr	genomic	Variant	Gene	exon	coding	Protein	VAF	IGV
675A	1	16262679	Non synonymous	SPEN	11	c.A9944C	p.H3315P	6%	No
675A	17	11593689	Non synonymous	DNAH9	20	c.G4550T	p.R1517L	5%	Yes
675A	17	11726202	Stop-gain	DNAH9	48	c.G9097T	p.E3033*	6%	Yes
1150	1	16257588	Non synonymous	SPEN	11	c.A4853G	p.K1618R	6%	No
695	1	16256966	Non synonymous	SPEN	11	c.A4231G	p.R1411G	43%	Yes
695	1	16258450	Non synonymous	SPEN	11	c.G5715T	p.R1905S	5%	Yes
695	1	16262679	Non synonymous	SPEN	11	c.A9944C	p.H3315P	7%	No

8.7.7 Predicting the effect of mutations to protein function

SIFT and PolyPhen were used to predict the effects of the mutations identified upon the protein function (Table 41). The scores for the *DNAH9,* 1517 Arginine to Leucine variant indicate that this mutation would probably be damaging to protein function as the SIFT score was 0 and the PolyPhen score was 0.96 as this is a highly conserved amino acid. The clonal 1411 Arginine to Glycine *SPEN* mutation according to SIFT would be tolerated (score 0.06) but from PolyPhen could possibly be damaging to protein function score (0.70). The impact of this variant is less certain. The *SPEN* Arginine to Serine variant has the least damaging scores from both SIFT and PolyPhen with tolerated (0.37) and possibly damaging (0.879), respectively. As the *DNAH9* 3033 variant Glutamic acid to a stop codon cannot be assessed, but as the stop codon would prematurely stop the amino acid sequence this would truncate the protein from this codon on exon 48.

Table 41 The variants were analysed *in silico* to see the impact upon protein function. SIFT and PolyPhen were used to see if the mutations would be expected to be damaging or not to the protein function. Only one of the variants had a low SIFT score (p.R1517L) and a high polyphen score; expected to have a damaging effect to protein function.

PID	Gene	Protein	SNP Type	SIFT Prediction	SIFT Score	Poly-Phen	Poly-phen score
675A	*DNAH9*	R1517L	Nonsynonymous	DAMAGING	0	Probably damaging	0.96
675A	*DNAH9*	E3033*	Nonsynonymous	N/A	N/A	N/A	N/A
695	*SPEN*	R1411G	Nonsynonymous	TOLERATED	0.06	Possibly damaging	0.70
695	*SPEN*	R1905S	Nonsynonymous	TOLERATED	0.37	Possibly damaging	0.879

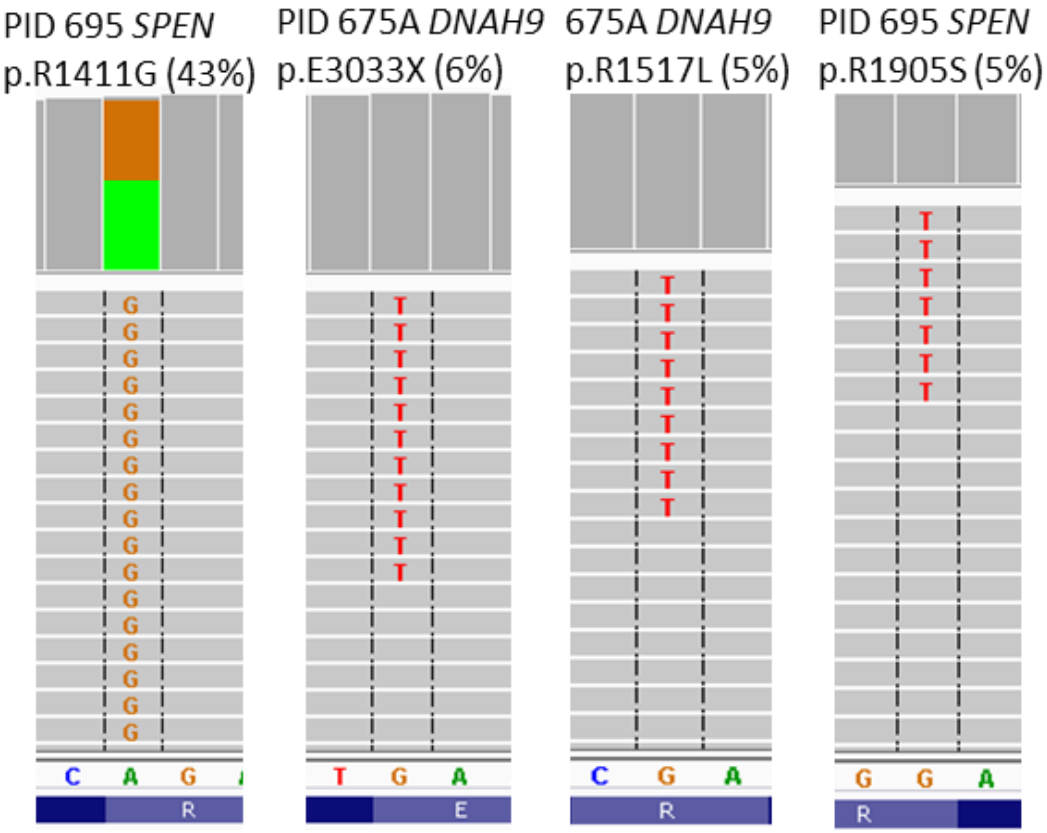


Figure 82 IGV confirmed *SPEN* and *DNAH9* variants sequenced using Nextera XT technology to create a sequencable library on the MiSeq platform. Four variants were identified from a bioinformatics pipeline.

8.8 Discussion

The use of whole genome and exome sequencing for gene discovery has been especially important in CLL at detecting new driver mutations and key pathways implicated in the disease.[75,110,122,131] A major caveat of these studies are the cost of sequencing whole exomes and genomes, although this has significantly reduced over the years it is still expensive to run a lot of samples as well as storing and analysing the huge amount of data obtained through these methods. As well as this, novel genes are less frequently discovered and understanding known genes and how they lie within CLL disease aetiology is moving forward. As sequencing technology develops, the use of NGS for screening purposes has become more readily accessible and available for wider use. Gene panels have been designed to cover cancer associated genes, also termed clinical exomes as well as being able to design custom amplicon panels to cover genes of interest.

Recently, the BTK inhibitor, ibrutinib has been used as a treatment for advanced CLL with success. This drug targets a BCR kinase BTK.[42] Some patients have become resistant to ibrutinib and subsequently relapse due to the emergence of gene mutations involving the *BTK* gene or another BCR signalling gene, *PLCG2*.[217] These gene mutations are very rare and have only been identified in patients treated with ibrutinib. So creating assays that screen these genes for use in the clinic may be critical and their applicability could be essential for screening patients before and during treatment with ibrutinib to detect the emergence of resistant clones. Two other genes identified from our own CLL exome data previously discussed, *DNAH9* and *SPEN* genes, were also interesting candidate genes to follow up. As technology has developed it seemed that using a targeted deep sequencing technique was a good option for screening, especially as *BTK* and *PLCG2* mutations can be sub clonal. But with the custom panel kits being rather expensive a cheaper alternative was to design assays covering the regions where gene mutations have previously been identified in the literature and using an Illumina Nextera technology to sequencing libraries quickly and cheaply. A transposase simultaneously fragments and tagments the DNA with adapter sequences along with a PCR step to incorporate identifying barcodes that can then be sequenced on a platform (Miseq). In this manner assays were designed and optimised to multiplex in four PCR reactions to cover the *BTK*, *PLCG2*, *DNAH9* and *SPEN* genes. During the optimisation the BTK and PLCG2 assays all amplified and multiplexed together fine, as did the DNAH9 assay, but the SPEN assays would not. In the end two multiplex reactions consisting of assays 1, 3 7 and 9 and 2, 4, 6 and 8_2 with differing Tm (55 and 60°C, respectively) had to be used. In retrospect, as the SPEN assays covered the same overlapping region (exon 11), assays could have been designed on alternating alleles to prevent this overlap and may have made multiplexing all assays easier. The DNAH9 and SPEN assay were combined before the Nextera

protocol. Libraries were generated for 96 samples, 57 of which were for the *BTK* and *PLCG2* genes and 37 samples for *DNAH9* and *SPEN*. These libraries were successfully sequenced on the MiSeq platform and aligned to a reference genome.

No mutations were identified in the *BTK* or *PLCG2* genes due to the rarity of this mutation this was not unexpected, though as two of the patients were known to have some resistance to the ibrutinib therapy it was hoped that one or both may have one of the published *BTK* or *PLCG2* mutations. This may mean that these patients have mutations outside of the exons screened in this study or else in a different gene within the BCR pathway preventing the effectiveness of ibrutinib. A recent study identified a *BTK* codon 316 variant in an ibrutinib relapsed CLL patient.[224] This mutation affected cells in a functional analysis, similarly to the codon 481 mutation. Sharma *et al*. suggest that all exons should be screened using only a deep targeted sequencing technology.[224] This is as Komarova *et al*. used *in silico* models to predict that mutations may be present at a frequency of $1/10^6$ to $1/10^8$ prior to ibrutinib therapy but then undergo clonal selection and expansion.[225] The copy number from our study using Nextera technology meant that only 1ng of starting DNA was used in the Nextera reaction, although a copy number of at least $2.02/10^9$ was calculated for the amplicons in this study. Though the technology has shown to have amplification bias throught the SNP sensitivity test. This is a major problem with other amplicon based NGS methods. Third generation sequencing aims to remove this bias by sequencing a single molecule of DNA or by incorporating a unique molecular identifier (UMI) onto the starting DNA molecules to allow the removal of duplicate reads. This will allow the sequencing data to be interpreted with more accuracy, especially when interpreting low level sub clones (<5%).

Reterospectively, Nextera may not have been appropriate NGS technology for a sensitive deep screening technique, such as for use in a clinical setting. The process is rather laborious and with multiple dilution steps as well as numerous opportunities for potential contamination prior to indexing samples. The standard process also is not suitable for a high throughput setting, as it is still costly and as mentioned, time consuming, although another study has streamlined the process to be quicker and reducing the volume of reagents used, managed to reduce the cost by 75%.[226]

Although *SPEN* and *DNAH9* are unfavourable genes in regard to the fact they are both large genes large genes with large exons and/or introns. The dynein genes were highlighted as a false positive result of passenger mutations.[207] Though some studies have shown direct correlations with the association with BCL6 protein and its genomic location upon 17p, a chromosome defect often associated with poor survival in CLL and further more the discovery in the CLL4 SNP 6.0 data of a

breakpoint deletion within *DNAH9,* suggests that maybe this gene could well be involved in CLL pathogenesis but to what extent will need further investigation, preferably in a 17p deleted cohort.[220]

The cohort for the *DNAH9* and *SPEN* gene screening was small (n=37) and undefined, the study was more of about optimising the Nextera technology for use as a NGS screening tool. The optimisation of the protocol with the custom designed assays was successful and all assays aligned to the correct gene locations, only two patients had gene mutations within these genes.

SPEN has been shown to regulate Notch signalling, a pathway known to be constituently activated in CLL patients carrying a *NOTCH1* mutation. In CLL this needs further investigation as our study was inconclusive, although mutations seem to be a rare event in CLL unlike in SMZL, and Notch upregulation is probably through *NOTCH1* gene dysfunction.[13,20,30,111,130]

The BTK and PLCG2 assays could be further developed for use in the clinic prior to BTK inhibitor therapy, to screen for sub clonal populations carrying the respective mutations. But a more high read depth maybe covering all exons of BTK and PLCG2 as well as maybe other BCR related genes could give a more comprehensive understanding of ibrutinib resistance. TruSeq (Illumina) could be a better alternative to using Nextera and this protocol uses more template DNA (250ng) so will have a better copy number and the protocol is well established, as well as the ability to screen several genes.

Since the advent of next generation sequencing, the world of genomics has been thrown wide open and researchers have utilised this technology to extensively study CLL. The slow nature of this disease and the ease of material access has allowed for numerous studies to investigate this disease throughout the different stages and in vast numbers of patients over time. The intricate details of how CLL develops; primarily in the context of evolution and how competitive pressures drive disease and induce more aggressive clones has been thoroughly an exciting step in understanding pathogenesis and applying novel therapies. There is exciting promise with small molecule inhibitors, though these still select for resistant clones. Using next generation sequencing has started to be, and will soon be at the forefront of patient care and tailorised treatments. Utilising this technology to be as effective and low cost to replace standard technology will be key in their implementation and developing the field of genomics in the clinic and treating CLL.

Chapter 9: General discussion

Haematopoesis is the development of a pluripotent haematopoietic stem cells, in the bone marrow into any one of the various blood cells types. B cells undergo a number of stages from a naive B cell state, to antigen stimulation, that then leads to hyper mutation and class switching, and then differentiation to a plasma or memory B cell.[5] B cell neoplasms are thought to arise and occur from these B cell stages. SMZL and CLL arise from a marginal zone B cell and CD5 positive or CD5 and CD27 positive pro germinal center B cell, respectively.[9,10] SMZL is a rare B cell neoplasm, so rare that only 2% of all B cell lymphomas are SMZL cases. Typically benign, though up to 70% of patients do progress to requiring treatment intervention. Although treatment response rates can be up to 90%, 40% of these patients do go on to further relapse. [11] CLL is a common B cell neoplasm prominent among the elderly, and the disease course of which can be highly heterogeneous with some patients leading a normal life whilst others may succumb quickly to disease progression. MBL is a benign precursor disease to CLL whilst Richters syndrome is a higher grade lymphoma that some CLL can develop into.[28] To help predict a patient's disease course, biomarkers have been developed to aid prognosis. Copy number alterations are found in over 80% CLL patients and more than 72% of SMZL patients. Frequent alterations in CLL include; 13q, 17p and 11q deletions and trisomy 12. These genomic abnormalities can be used to risk stratify patients, with 13q giving a good prognosis and 17p a poor risk identifier. Though according to this model a lot of the 'good' risk group still progress to needing treatment. So better genomic identifiers are needed. More recently with the advent of next generation sequencing, recurrently mutated genes have been identified in both CLL and SMZL. Though the importance of many of which in the respective diseases still remains unclear. Whole exome and genome sequencing in recent years have identified *NOTCH1* and *SF3B1* as recurrently mutated in CLL, these have also been shown to impact upon survival and may have clinical applicability in regards to treatment choice. These two genes along with another, *BIRC3* have also been integrated into an updated risk stratifying model that has improved upon the first cytogenetic model.[53,57] Similarly in SMZL, WGS and WES, identified *NOTCH2* as being recurrently mutated with some studies suggesting a link with clinical survival.[13,15]

To assess the genomic landscape in B cell malignancies, predominately in CLL (and MBL) and SMZL, various technologies were applied in this study. MLPA was employed to investigate rare and common chromosomal changes and 3 of the recurrently mutated genes in CLL; *SF3B1*, *NOTCH1* and *MYD88*. This study also investigated the clinical applicability of the technology by a comparison to FISH, which is the current gold standard. MLPA identified a number of common and rare chromosomal changes in line with the literature. Some of these were validated using

FISH, Sanger sequencing, digital PCR, qPCR and SNP 6.0 data. The clinical applicability of MLPA was also assessed by comparing to current cytogenetic practices; FISH and the standard mutational approach of Sanger sequencing. In terms of sensitivity there was good concordance to FISH and Sanger sequencing, though the technology is not sensitive at picking up small clones harbouring a CNA (<30%) but seems to be more sensitive at detecting mutations than Sanger sequencing. Due to the low cost and high output of data for numerous common and rare CNA as well as mutational data, MLPA is an attractive alternative to the gold standard FISH. Using a combination of MLPA then FISH if no 17p deletions are detected, prior to therapy could save money as well as providing a wealth of information that could be used to aid understanding of rarer CNAs. Through using a well characterised clinical trial cohort many novel clinical associations were made in this study. A novel extremely benign risk group was identified harbouring a biallelic 13q deletion with IGHV mutated genes as well as an opposing novel very poor risk group with a 9p deletion. The 9p deletion was only identified in 4 patients and 3 of these also had a concurrent *TP53* aberration, though as this abnormality is associated to RS transformation it could suggest a marker for transformation and poor outcome. Other rare CNA such as, 8p, 19p and 6q deletions and 2p and 8q gains were shown to impact negatively on survival. Though in a multivariate model only 9p deletions and 2p gains were shown to be independent predictors of progression and death out of the rarer CNAs. This highlights the need for more studies to investigate rare copy number alterations in CLL as well as an integrated approach when stratifying patients. Telomere analysis was also carried out on the CLL4 clinical trial. The majority of these cases also had MLPA data, which gave the opportunity to study rare CNA in relation to telomere length, which to current understanding has never been done before. Previous associations to IGHV status, 17p, 11q and 13q deletions were seen, as well as novel associations with 2p, 8p and 6q deletions and trisomy 19. Telomere length has been shown to be one of the most powerful predictive markers in CLL, though it has yet to be implemented in a clinical setting due to a to the lack of standard methodology of measuring TL, with many being difficult and/or expensive.[128,175,179,227] The TL data was generated using an adapted qPCR based method, MMQPCR, which was developed by Cawthon.[85] This method is high throughput, quick and relatively low cost method of analysing TL. Using this MMQPCR based method 384 CLL4 patients telomeres were analysed (data not shown previously) and a relative telomere measurement (T/S) was identified. This data was statistically analysed (Author kindly acknowledges Dr Latha Kadalayil) and compared to another TL measurement technique, STELA, were an excellent correlation between the two methods was seen.[228] Three groups for TL were generated; a short (<50 percentile), an intermediate (50–75 percentile) and a long (>75 percentile) and these were analysed by risk and survival analysis. The risk of progression was increased twofold for the intermediate group (HR: 2.07, 95% CI: 1.52–2.82, P<0.001) and by 2.7 times for the short group (HR: 2.67, 95% CI: 2.03–3.53, P<0.001) when

compared with the long group. The median PFS and OS for the long TL group was 4.0 and 9.9 years, respectively, and these patients were 63% (HR: 0.37, 95% CI: 0.28–0.49, P<0.001) and 72% (HR: 0.28, 95% CI: 0.21–0.39, P<0.001) less likely to progress or die compared with patients within the short TL group. In a multivariate model short TL patients were identified as an independent risk for progression (short vs long HR: 2.10, 95% CI: 1.37–3.21). TL has been shown to be an important, predictive marker of progression in CLL and MMQPCR validated as a method of measuring TL that is easily adaptable to a clinical setting.

Whole exome sequencing was applied in this study to a small number of CLL and SMZL patients to identify candidate genes. Though there were no recurrent genes identified from this study, several candidate genes were chosen from the exome data, as well as from the literature to screen in extended cohorts of CLL patients using HRM analysis and Sanger sequencing and also using NGS technologies. *BIRC3*, *MYD88* and *POT1* were all screened in CLL cohorts and low frequencies of mutations identified. *BIRC3* is a gene involved in the 11q deletion loci and it is an important regulator of alternative NF-κB signalling.[99] *BIRC3* was compared with the other 11q loci gene *ATM* to determine which gene influences the negative impact on survival associated with the 11q CNA. *ATM* mutation was associated with the worst survival but *BIRC3* also had a poor survival outcome.[71] The functional effects of *BIRC3* mutations in CLL were investigated through mRNA and protein analysis. Though protein analysis was fraught with sample integrity, mRNA analysis provided an interesting link to *ATM* status. *BIRC3* expression was almost significantly reduced in *ATM* biallelically inactivated patients; again showing these two genes may both be of importance to CLL pathogenesis and provides an insight into a largely unknown role of ATM in NF-κB signalling.[191] This work improved the characterisation and current understanding of *BIRC3* in relation to CLL disease course.

BCR signalling in CLL is critical to cell proliferation and survival.[229] A BTK inhibitor, ibrutinib that targets the BCR pathway has recently had excellent results in trials in treatment resistant CLL cases.[230] Though some patients have been identified as becoming resistant to ibrutinib through harbouring a *BTK* gene mutation that directly affects ibrutinib from binding to the BTK protein.[216] Another downstream BCR gene, *PLCG2* has also been identified as causing ibrutinib resistance through gene mutations that allows BCR signalling to continue without BTK involvement.[216,218] Further screening was extended to include these ibrutinib resistant genes, to screen in a small ibrutinib treated cohort, as these gene mutations have only been identified in this cohort of patients. This study custom designed and optimised NGS technology, Nextera, for use on these regions as well as two other genes, *SPEN* and *DNAH9* that were identified from the CLL exome study as being mutated in the Puente study.[75] TShe study successfully sequenced the *BTK*, *PLCG2*, *SPEN* and *DNAH9* genes using the Nextera XT technology and MiSeq platform

that showed all the amplicons mapped to the correct locations on a reference genome proving that NGS can be easily adapted into custom designed assays. Though the technique may need further adaptation to prove useful as a high throughput method. Two patients were identified as having *SPEN* and *DNAH9* gene mutations, but due to the small sample size and lack of clinical information no further associations could be made. There was also a patient from the CLL4 trial whom had a breakpoint deletion within the *DNAH9* gene suggesting this gene could be relevant in CLL pathogenesis. Unfortunately no mutations were identified in the *BTK* or *PLCG2* regions screened, this is not surprising as only two of the patients were starting to show signs of resistance. The lack of mutations may also suggest that mutations may lie within exons outside of these assays or in other BCR pathway genes. As more patients are put on ibrutinib, more patients with resistance are bound to occur so future studies investigating these genes will be critical to the guidance and care of this treatment regime.

The final part of this study was using TruSeq to investigate a panel of genes by targeted deep sequencing. Although 129 libraries were prepared and sequenced on the MiSeq platform, only 68 had been analysed by the bioinformatical pipeline (kind regards Dr Mathew Rose-Zerilli and Dr Rueben Pengelly) for this study. As with most NGS studies the main caveat is the bioinformatical processing of the data. That said a variant list was identified from these patients and some were confirmed by Sanger sequencing. Functional analysis of the same patients was undertook using Taqman assays, were a novel association of increased *POT1* expression in *SF3B1* mutated patients was seen. Another project this study was involved in was an ERIC consortium driven, multicentre, pan European effort to assess the clinical applicability of NGS targeted deep sequencing panels. 48 CLL samples with known mutation were sent blinded to the 9 centres (including this study) and the 12 gene TruSeq panel was applied. Although the sequencing and bioinformatical analysis was complete and variants identified, the comparative analysis between centres is still ongoing.

Chapter 10: Conclusion

In conclusion, this study has evaluated the genomic landscape of predominately CLL and to a lesser extent MBL and SMZL, utilising technology available at the time. The study adapted and incorporated new and cutting edge next generation sequencing technology as well as using more widely adapted techniques such as Sanger sequencing, MLPA and SNP 6.0. As NGS developed; from whole genome and exome sequencing to more targeted technologies, these methodologies were developed in house and utilised appropriately. New prognostic groups have been identified using genomic status. CNA analysis identified a benign biallelic 13q with IGHV mutated genes and conversely poor prognostic group harbouring a 9p deletion. Further dissection of the 11q deletion in relation to *ATM* and *BIRC3* genes, identified the biallelic inactivation of both genes as impacting worse on survival than 11q deletion (without mutation), though *ATM* mutation impacted more negatively then *BIRC3*. These prognostic markers may want to be included in future risk stratification models.

Chapter 11: Future studies

The results from this study may want to be further investigated in future studies as indicated in the discussion sections of relevant chapters. Though possible interesting lines of investigation may be:

- Due to the low costs and high throughput, MLPA may wish to be used in a clinical setting and in combination with FISH. This is so the rarer genomic regions can be investigated more thoroughly in larger cohorts of patients.
- The biological/function validation of the biallelic 13q-IGHV mutated cases and why these patients seem to do so well; is it through an upregulation of *TP53* or another pathway?
- A similar functional study assessing the 9p deletion in relation to *TP53* patient status and links to RS in a larger cohort would be interesting to see if the same independent effects were seen.
- Validation of the multivariate model; are the same independent risk factors identified?
- The functional validation of *BIRC3* mutations in larger cohorts of 11q deleted patients at the mRNA level may validate the results identified in this study with a biallelic *ATM* status. Future protein analysis including ATM would also be an interesting line of follow up.
- Further developing the Nextera assays optimised in this work for *BTK* and *PLCG2* to cover a wider gene coverage or else utilise another NGS technology such as a gene panel to include these and other BCR genes for screening ibrutinib patients prior and during treatment.

Bibliography

1. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell [Internet]. 2011;144(5):646–74. Available from: http://dx.doi.org/10.1016/j.cell.2011.02.013

2. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481(7381):306–13.

3. Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. Synonymous mutations frequently act as driver mutations in human cancers. Cell. 2014;156(6):1324–35.

4. Ceredig R, Rolink AG, Brown G. Models of haematopoiesis: seeing the wood for the trees. Nat Rev Immunol. 2009;9(4):293–300.

5. Lebien TW, Tedder TF. B lymphocytes: How they develop and function. Blood. 2008;112(5):1570–80.

6. García-Muñoz R, Galiacho VR, Llorente L. Immunological aspects in chronic lymphocytic leukemia (CLL) development. Ann Hematol. 2012;91(7):981–96.

7. Korsmeyer SJ, Arnold A, Bakhshi A, Ravetch J V, Siebenlist U, Hieter PA, Sharrow SO, LeBien TW, Kersey JH, Poplack DG, Leder P, Waldmann TA. Immunoglobulin gene rearrangement and cell surface antigen expression in acute lymphocytic leukemias of T cell and B cell precursor origins. J Clin Invest [Internet]. 1983 Feb;71(2):301–13. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC436868/

8. Coco F Lo, Ye BH, Lista F, Corradini P, Offit K, Knowles DM, Chaganti RS, Dalla-Favera R. Rearrangements of the BCL6 gene in diffuse large cell non-Hodgkin's lymphoma. Blood. 1994;83(7):1757–9.

9. Thieblemont C, Davi F, Noguera ME, Briere J, Bertoni F, Zucca E, Traverse-Glehen A, Felman P, Berger F, Salles G, Coiffier B. Splenic Marginal Zone Lymphoma: Current Knowledge and Future Directions. Oncology. 2012;26(2):194–202.

10. Seifert M, Sellmann L, Bloehdorn J, Wein F, Stilgenbauer S, Durig J, Kuppers R. Cellular origin and pathophysiology of chronic lymphocytic leukemia. J Exp Med. 2012;209(12):2183–98.

11. Baliakas P, Strefford JC, Bikos V, Parry M, Stamatopoulos K, Oscier D. Splenic marginal-zone lymphoma: ontogeny and genetics. Leuk Lymphoma [Internet]. 2015 May 6 [cited 2014 Aug 13];56(2):301–10. Available from:

http://informahealthcare.com/doi/abs/10.3109/10428194.2014.919636

12. Kalpadakis C, Pangalis G a, Vassilakopoulos TP, Sachanas S, Angelopoulou MK. Treatment of splenic marginal zone lymphoma: should splenectomy be abandoned? Leuk Lymphoma [Internet]. 2013;55(August 2013):1–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24050506

13. Rossi D, Trifonov V, Fangazio M, Bruscaggin A, Rasi S, Spina V, Monti S, Vaisitti T, Arruga F, Famà R, Ciardullo C, Greco M, Cresta S, Piranda D, Holmes A, Fabbri G, Messina M, Rinaldi A, Wang J, Agostinelli C, Piccaluga PP, Lucioni M, Tabbò F, Serra R, Franceschetti S, Deambrogi C, Daniele G, Gattei V, Marasca R, Facchetti F, Arcaini L, Inghirami G, Bertoni F, Pileri SA, Deaglio S, Foà R, Dalla-Favera R, Pasqualucci L, Rabadan R, Gaidano G. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. J Exp Med [Internet]. 2012 Aug 15 [cited 2012 Aug 27];209(9):1537–51. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3428941&tool=pmcentrez&rendertype=abstract

14. Parry M, Rose-Zerilli MJJ, Gibson J, Ennis S, Walewska R, Forster J, Parker H, Davis Z, Gardiner A, Collins A, Oscier DG, Strefford JC. Whole exome sequencing identifies novel recurrently mutated genes in patients with splenic marginal zone lymphoma. PLoS One [Internet]. 2013 Jan [cited 2014 Aug 8];8(12):e83244. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3862727&tool=pmcentrez&rendertype=abstract

15. Parry M, Rose-Zerilli MJJ, Ljungström V, Gibson J, Wang J, Walewska R, Parker H, Parker A, Davis Z, Gardiner A, McIver-Brown N, Kalpadakis C, Xochelli A, Anagnostopoulos A, Fazi C, De Castro DG, Dearden C, Pratt G, Rosenquist R, Ashton-Key M, Forconi F, Collins A, Ghia P, Matutes E, Pangalis G, Stamatopoulos K, Oscier D, Strefford JC. Genetics and prognostication in splenic marginal zone lymphoma: Revelations from deep sequencing. Clin Cancer Res [Internet]. 2015;21(18):4174–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25779943

16. Salido M, Baró C, Oscier D, Stamatopoulos K, Dierlamm J, Matutes E, Traverse-glehen A, Berger F, Felman P, Thieblemont C, Athanasiadou A, Davis Z, Gardiner A, Milla F, Ferrer A, Calasanz MJ, Florensa L, Espinet B, Luño E, Wlodarska I, Verhoef G, García-granero M, Salar A, Papadaki T, Serrano S, Piris A, Solé F, Dc W, Salido M, Baro C, Oscier D, Stamatopoulos K,

Dierlamm J, Matutes E, Traverse-glehen A, Berger F, Felman P, Thieblemont C, Gesk S, Athanasiadou A, Davis Z, Gardiner A, Milla F, Ferrer A, Mollejo M. Cytogenetic aberrations and their prognostic value in a series of 330 splenic marginal zone B-cell lymphomas : a multicenter study of the Splenic B-Cell Lymphoma Group Cytogenetic aberrations and their prognostic value in a series of 330 splenic marginal . 2013;116(9):1479–88.

17. Vega F, Cho-Vega JH, Lennon PA, Luthra MG, Bailey J, Breeden M, Jones D, Medeiros LJ, Luthra R. Splenic marginal zone lymphomas are characterized by loss of interstitial regions of chromosome 7q, 7q31.32 and 7q36.2 that include the protection of telomere 1 (POT1) and sonic hedgehog (SHH) genes. Br J Haematol [Internet]. 2008 Jun [cited 2014 Aug 13];142(2):216–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18492102

18. Watkins AJ, Hamoudi RA, Zeng N, Yan Q, Huang Y, Liu H, Zhang J, Braggio E, Fonseca R, de Leval L. An integrated genomic and expression analysis of 7q deletion in splenic marginal zone lymphoma. PLoS One. 2012;7(9):e44997.

19. Rossi D, Deaglio S, Dominguez-Sola D, Rasi S, Vaisitti T, Agostinelli C, Spina V, Bruscaggin A, Monti S, Cerri M, Cresta S, Fangazio M, Arcaini L, Lucioni M, Marasca R, Thieblemont C, Capello D, Facchetti F, Kwee I, Pileri SA, Fo?? R, Bertoni F, Dalla-Favera R, Pasqualucci L, Gaidano G. Alteration of BIRC3 and multiple other NF-kB pathway genes in splenic marginal zone lymphoma. Blood [Internet]. 2011 Nov 3 [cited 2012 Aug 2];118(18):4930–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21881048

20. Kiel MJ, Velusamy T, Betz BL, Zhao L, Weigelin HG, Chiang MY, Huebner-Chan DR, Bailey NG, Yang DT, Bhagat G, Miranda RN, Bahler DW, Medeiros LJ, Lim MS, Elenitoba-Johnson KSJ. Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. J Exp Med [Internet]. 2012 Aug 27 [cited 2012 Sep 5];209(9):1553–65. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3428949&tool=pmcentrez&rendertype=abstract

21. Martínez N, Almaraz C, Vaqué JP, Varela I, Derdak S, Beltran S, Mollejo M, Campos-Martin Y, Agueda L, Rinaldi A, Kwee I, Gut M, Blanc J, Oscier D, Strefford JC, Martinez-Lopez J, Salar A, Sole F, Rodriguez-Peralto JL, Diez-Tascón C, García JF, Fraga M, Sebastián E, Alvés J, Menárguez J, González-Carreró J, Casado LF, Bayes M, Bertoni F, Gut I, Piris M a. Whole-exome sequencing in splenic marginal zone lymphoma reveals mutations in genes involved in marginal zone differentiation. Leukemia [Internet]. 2014;28(28):1334–40. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/24296945

22. Desai S, Pinilla-Ibarz J. Front-line therapy for chronic lymphocytic leukemia. Cancer Control [Internet]. 2012 Jan;19(1):26–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22143060

23. Allendorf D, Davis R. Unraveling the Molecular Pathogenesis of Chronic Lymphocytic Leukemia. JAMA J Am … [Internet]. 2011 [cited 2012 Sep 6];305(1):95–7. Available from: http://jama.ama-assn.org/content/305/1/95.short

24. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, Hillmen P, Keating MJ, Montserrat E, Rai KR, Kipps TJ. 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 [Internet]. 2008 Jun 15 [cited 2012 Jul 22];111(12):5446–56. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2972576&tool=pmcentrez&rendertype=abstract

25. Rawstron a. C. Monoclonal B lymphocytes with the characteristics of “indolent” chronic lymphocytic leukemia are present in 3.5% of adults with normal blood counts. Blood [Internet]. 2002 Jun 28 [cited 2012 Sep 5];100(2):635–9. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood.V100.2.635

26. Rawstron AC, Mulligan CS, Thomas ME, Mulligan SP. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. N Engl J Med [Internet]. 2008 Nov 6;359(19):2065–6; author reply 2066. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22262125

27. Landgren O, Albitar M, Ma W. B-cell clones as early markers for chronic lymphocytic leukemia. … Engl J … [Internet]. 2009 [cited 2012 Sep 6];659–67. Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:new+england+journal#1

28. Hallek M. Chronic lymphocytic leukemia: 2015 Update on diagnosis, risk stratification, and treatment. Am J Hematol. 2015;90(5):446–60.

29. Ferrajoli A, Shanafelt TD, Ivan C, Shimizu M, Rabe KG, Ikuo M, Ghosh AK, Lerner S, Rassenti LZ, Xiao L, Hu J, Reuben JM, Calin S, You J, Manning JT, Wierda WG, Estrov Z, Brien SO, Kipps TJ, Keating MJ, Kay NE, Calin G a. Prognostic value of miR-155 in individuals with monoclonal B-cell lymphocytosis and patients with B-chronic lymphocytic leukemia. Blood. 2013;122(11):1891–900.

30. Lionetti M, Fabris S, Cutrona G, Agnelli L, Ciardullo C, Matis S, Ciceri G, Colombo M, Maura F, Mosca L, Gentile M, Recchia AG, Ilariucci F, Musolino C, Molica S, Di Raimondo F, Cortelezzi A, Rossi D, Gaidano G, Morabito F, Ferrarini M, Neri A. High-throughput sequencing for the identification of NOTCH1 mutations in early stage chronic lymphocytic leukaemia: Biological and clinical implications. Br J Haematol. 2014;165(March):629–39.

31. Ghia P, Hallek M. Management of chronic lymphocytic leukemia. Haematologica. 2014;99(6):965–72.

32. Eichhorst B, Robak T, Montserrat E, Ghia P, Hillmen P, Hallek M, Buske C. Chronic lymphocytic leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2015;26(August 2003):v78–84.

33. Rai KR, Jain P. Chronic lymphocytic leukemia (CLL)-Then and now. Am J Hematol. 2016;91(3):330–40.

34. Fabbri G, Khiabanian H, Holmes AB, Wang J, Messina M, Mullighan CG, Pasqualucci L, Rabadan R, Dalla-Favera R. Genetic lesions associated with chronic lymphocytic leukemia transformation to Richter syndrome. J Exp Med [Internet]. 2013;210(11):2273–88. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24127483

35. Rai KR, Sawitsky a, Cronkite EP, Chanana a D, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. Blood [Internet]. 1975 Aug;46(2):219–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7417943

36. Binet JL, Auquier a, Dighiero G, Chastang C, Piguet H, Goasguen J, Vaugier G, Potron G, Colona P, Oberling F, Thomas M, Tchernia G, Jacquillat C, Boivin P, Lesty C, Duault MT, Monconduit M, Belabbes S, Gremy F. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. Cancer [Internet]. 1981 Jul 1;48(1):198–206. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7237385

37. Rosenquist R, Cortese D, Bhoi S, Mansouri L, Gunnarsson R. Prognostic markers and their clinical applicability in chronic lymphocytic leukemia: where do we stand? Leuk Lymphoma [Internet]. 2013;54(11):2351–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23480493

38. Rossi D, Cerri M, Capello D, Deambrogi C, Rossi FM, Zucchetto A, De Paoli L, Cresta S, Rasi S, Spina V, Franceschetti S, Lunghi M, Vendramin C, Bomben R, Ramponi A, Monga G, Conconi A, Magnani C, Gattei V, Gaidano G. Biological and clinical risk factors of chronic lymphocytic

leukaemia transformation to Richter syndrome. Br J Haematol [Internet]. 2008 Jun [cited 2015 Nov 22];142(2):202–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18492108

39. Hallek M, Wanders L, Ostwald M, Busch R, Senekowitsch R, Stern S, Schick HD, Kuhn-Hallek I, Emmerich B. Serum beta(2)-microglobulin and serum thymidine kinase are independent predictors of progression-free survival in chronic lymphocytic leukemia and immunocytoma. Leuk Lymphoma [Internet]. 1996 Aug;22(5–6):439–47. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8882957

40. Montserrat E, Sanchez-Bisono J, Viñolas N, Rozman C. Lymphocyte doubling time in chronic lymphocytic leukaemia: analysis of its prognostic significance. Br J Haematol [Internet]. 1986 Mar [cited 2015 Nov 22];62(3):567–75. Available from: http://doi.wiley.com/10.1111/j.1365-2141.1986.tb02969.x

41. Strefford JC. The genomic landscape of chronic lymphocytic leukaemia: Biological and clinical implications. Br J Haematol [Internet]. 2015 Dec 11 [cited 2014 Dec 17];169(1):14–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25496136

42. Byrd JC, Furman RR, Coutre SE, Flinn IW, Burger JA, Blum KA, Grant B, Sharman JP, Coleman M, Wierda WG, Jones JA, Zhao W, Heerema NA, Johnson AJ, Sukbuntherng J, Chang BY, Clow F, Hedrick E, Buggy JJ, James DF, O'Brien S. Targeting BTK with Ibrutinib in Relapsed Chronic Lymphocytic Leukemia. N Engl J Med [Internet]. 2013 Jun 19;369(1):32–42. Available from: http://dx.doi.org/10.1056/NEJMoa1215637

43. Hamblin BTJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig VH Genes Are Associated With a More Aggressive Form of Chronic Lymphocytic Leukemia. Blood. 1999;94(6):1848–54.

44. Damle RN, Wasil T, Fais F, Ghiotto F, Valetto a, Allen SL, Buchbinder a, Budman D, Dittmar K, Kolitz J, Lichtman SM, Schulman P, Vinciguerra VP, Rai KR, Ferrarini M, Chiorazzi N. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood. 1999;94(6):1840–7.

45. Damle RN, Temburni S, Calissano C, Yancopoulos S, Banapour T, Sison C, Allen SL, Rai KR, Chiorazzi N. CD38 expression labels an activated subset within chronic lymphocytic leukemia clones enriched in proliferating B cells. Blood [Internet]. 2007 Nov 1 [cited 2012 Sep 6];110(9):3352–9. Available from:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2200908&tool=pmcentrez&rendertype=abstract

46. Wiestner A, Rosenwald A, Barry TS, Wright G, Davis RE, Henrickson SE, Zhao H, Ibbotson RE, Orchard J a, Davis Z, Stetler-Stevenson M, Raffeld M, Arthur DC, Marti GE, Wilson WH, Hamblin TJ, Oscier DG, Staudt LM. ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. Blood [Internet]. 2003 Jun 15 [cited 2012 Aug 9];101(12):4944–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12595313

47. Magnac C, Porcher R, Davi F, Nataf J, Payelle-Brogard B, Tang RP, Oppezzo P, Levy V, Dighiero G, Ajchenbaum-Cymbalista F. Predictive value of serum thymidine kinase level for Ig-V mutational status in B-CLL. Leukemia [Internet]. 2003;17(1):133–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12529670

48. Lamanna N, Weiss M, Dunleavy K. Chronic Lymphocytic Leukemia and Hairy-Cell Leukemia. Cancer Netw. 2014;1–12.

49. Gahrton G, Robèrt K-H, Friberg K, Zech L, Bird AG. Extra chromosome 12 in chronic lymphocytic leukaemia. Lancet. 1980;315(8160):146–7.

50. Hurley JN, Fu SM, Kunkel HG, Chaganti RSK, German J. Chromosome abnormalities of leukaemic B lymphocytes in chronic lymphocytic leukaemia. 1980;

51. Fitchett M, Griffiths MJ, Oscier DG, Johnson S, Seabright M. Chromosome abnormalities involving band 13q14 in hematologic malignancies. Cancer Genet Cytogenet. 1987;24(1):143–50.

52. Stankovic T, Weber P, Stewart G, Bedenham T, Murray J, Byrd PJ, Moss PAH, Taylor AMR. Inactivation of ataxia telangiectasia mutated gene in B-cell chronic lymphocytic leukaemia. Lancet. 1999;353(9146):26–9.

53. Döhner H, Stilgenbauer S, Benner a, Leupolt E, Kröber a, Bullinger L, Döhner K, Bentz M, Lichter P. Genomic aberrations and survival in chronic lymphocytic leukemia. N Engl J Med [Internet]. 2000 Dec 28;343(26):1910–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11136261

54. Edelmann J, Holzmann K, Miller F, Winkler D, Bühler A, Zenz T, Bullinger L, Kühn MWM, Gerhardinger A, Bloehdorn J, Radtke I, Su X, Ma J, Pounds S, Hallek M, Lichter P, Korbel J,

Busch R, Mertens D, Downing JR, Stilgenbauer S, Döhner H. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. [Internet]. Blood. 2012 [cited 2012 Oct 23]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23047824

55. Van Dyke DL, Werner L, Rassenti LZ, Neuberg D, Ghia E, Heerema NA, Dal Cin P, Dell Aquila M, Sreekantaiah C, Greaves AW, Kipps TJ, Kay NE. The Dohner fluorescence in situ hybridization prognostic classification of chronic lymphocytic leukaemia (CLL): The CLL Research Consortium experience. Br J Haematol [Internet]. 2016;173(1):105–13. Available from: http://doi.wiley.com/10.1111/bjh.13933

56. Gunnarsson R, Mansouri L, Rosenquist R. Exploring the genetic landscape in chronic lymphocytic leukemia using high-resolution technologies. Leuk Lymphoma [Internet]. 2013 Nov 21 [cited 2012 Nov 26];54(8):1583–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23167608

57. Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C, Deambrogi C, Khiabanian H, Serra R, Bertoni F, Forconi F, Laurenti L, Marasca R, Dal-bo M, Rossi FM, Bulian P, Nomdedeu J, Poeta G Del, Gattei V, Pasqualucci L, Rabadan R, Foa R. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. Blood. 2013;121(8):1403–12.

58. Puiggros A, Delgado J, Rodriguez-Vicente A, Collado R, Aventín A, Luño E, Grau J, Hernandez JÁ, Marugán I, Ardanaz M, González T, Valiente A, Osma M, Calasanz MJ, Sanzo C, Carrió A, Ortega M, Santacruz R, Abrisqueta P, Abella E, Bosch F, Carbonell F, Solé F, Hernández JM, Espinet B. Biallelic losses of 13q do not confer a poorer outcome in chronic lymphocytic leukaemia: analysis of 627 patients with isolated 13q deletion. Br J Haematol [Internet]. 2013 Oct [cited 2015 Jan 13];163(1):47–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23869550

59. Puiggros A1, Venturas M, Salido M, Blanco G, Fernandez-Rodriguez C, Collado R, Valiente A, Ruiz-Xivillé N, Carrió A, Ortuño FJ, Luño E, Calasanz MJ, Ardanaz MT, Piñán MÁ, Talavera E, González MT, Ortega M, Marugán I, Ferrer A, Gimeno E, Bellosillo B, Delg EB. Interstitial 13q14 deletions detected in the karyotype and translocations with concomitant deletion at 13q14 in chronic lymphocytic leukemia: different genetic mechanisms but equivalent poorer clinical outcome. Genes Chromosomes Cancer. 2014;53(9):788–97.

60. Chena C, Avalos JS, Bezares RF, Arrossagaray G, Turdó K, Bistmans A, Slavutsky I. Biallelic deletion 13q14.3 in patients with chronic lymphocytic leukemia: Cytogenetic, FISH and clinical

studies. Eur J Haematol [Internet]. 2008 Aug [cited 2015 Jan 13];81(2):94–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18462257

61. Pekarsky Y, Croce CM. Role of miR-15/16 in CLL. Cell Death Differ [Internet]. 2014;1–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24971479

62. Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T, Ambesi-Impiombato A, Califano A, Migliazza A, Bhagat G, Dalla-Favera R. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. Cancer Cell [Internet]. 2010 Jan 19 [cited 2012 Jul 17];17(1):28–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20060366

63. Lia M, Carette A, Tang H, Shen Q, Mo T, Bhagat G, Dalla-Favera R, Klein U. Functional dissection of the chromosome 13q14 tumor-suppressor locus using transgenic mouse lines. Blood [Internet]. 2012 Mar 29 [cited 2012 Sep 5];119(13):2981–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22174151

64. Parker H, Rose-zerilli MJJ, Parker A, Chaplin T, Wade R, Gardiner A, Griffiths M, Collins A, Young BD, Oscier DG. 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. Leukemia [Internet]. 2010;25(3):489–97. Available from: http://dx.doi.org/10.1038/leu.2010.288

65. Ouillette P, Collins R, Shakhan S, Li J, Li C, Shedden K, Malek SN. The prognostic significance of various 13q14 deletions in chronic lymphocytic leukemia. Clin Cancer Res. 2011;17(21):6778–90.

66. Van Dyke DL, Shanafelt TD, Call TG, Zent CS, Smoley S a., Rabe KG, Schwager SM, Sonbert JC, Slager SL, Kay NE. A comprehensive evaluation of the prognostic significance of 13q deletions in patients with B-chronic lymphocytic leukaemia. Br J Haematol. 2010;148(4):544–50.

67. Ibbotson R, Athanasiadou a, Sutton L, Davis Z, Gardiner a, Baliakas P, Gunnarsson R, Anagnostopoulos a, Juliusson G, Rosenquist R, Oscier D, Stamatopoulos K. Coexistence of trisomies of chromosomes 12 and 19 in chronic lymphocytic leukemia occurs exclusively in the rare IgG-positive variant. Leukemia. 2012;26:170–2.

68. Sellmann L, Gesk S, Walter C, Ritgen M, Harder L, Martín-Subero JI, Schroers R, Siemer D, Nückel H, Dyer MJS, Dührsen U, Siebert R, Dürig J, Küppers R. Trisomy 19 is associated with trisomy 12 and mutated IGHV genes in B-chronic lymphocytic leukaemia. Br J Haematol. 2007;138:217–20.

69. Puiggros A, Blanco G, Espinet B. Genetic abnormalities in chronic lymphocytic leukemia: where we are and where we go. Biomed Res Int [Internet]. 2014 Jan;2014:435983. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4054680&tool=pmcentrez&rendertype=abstract

70. Stankovic T, Skowronska A. The role of ATM mutations and 11q deletions in disease progression in chronic lymphocytic leukemia. Leuk Lymphoma [Internet]. 2014 Jul [cited 2014 Jul 31];55(6):1227–39. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23906020

71. Rose-Zerilli MJJ, Forster J, Parker H, Parker A, Rodríguez AE, Chaplin T, Gardiner A, Steele AJ, Collins A, Young BD, Skowronska A, Catovsky D, Stankovic T, Oscier DG, Strefford JC. ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q-deleted chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. Haematologica [Internet]. 2014 Apr [cited 2014 Aug 13];99(4):736–42. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3971084&tool=pmcentrez&rendertype=abstract

72. Jamroziak K, Tadmor T, Robak T, Polliack A. Richter syndrome in chronic lymphocytic leukemia: updates on biology, clinical features and therapy. Leuk Lymphoma [Internet]. 2014 Oct 30 [cited 2014 Nov 2];1–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25356923

73. Parikh S a, Shanafelt TD. Risk factors for Richter syndrome in chronic lymphocytic leukemia. Curr Hematol Malig Rep [Internet]. 2014 Sep [cited 2015 Jan 12];9(3):294–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25218362

74. Ernst A, Jones DTW, Maass KK, Rode A, Deeg KI, Jebaraj BMC, Korshunov A, Hovestadt V, Tainsky MA, Pajtler KW, Bender S, Brabetz S, Gr??bner S, Kool M, Devens F, Edelmann J, Zhang C, Castelo-Branco P, Tabori U, Malkin D, Rippe K, Stilgenbauer S, Pfister SM, Zapatka M, Lichter P. Telomere dysfunction and chromothripsis. Int J Cancer. 2016;138(12):2905–14.

75. Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI, Munar M, Rubio-Pérez C, Jares P, Aymerich M, Baumann T, Beekman R, Belver L, Carrio A, Castellano G, Clot G, Colado E, Colomer D, Costa D, Delgado J, Enjuanes A, Estivill X, Ferrando AA, Gelpí JL, González B, González S, González M, Gut M, Hernández-Rivas JM, López-Guerra M, Martín-García D, Navarro A, Nicolás P, Orozco M, Payer ÁR, Pinyol M, Pisano DG, Puente DA, Queirós AC, Quesada V, Romeo-Casabona CM, Royo C, Royo R, Rozman M, Russiñol N, Salaverría I,

Stamatopoulos K, Stunnenberg HG, Tamborero D, Terol MJ, Valencia A, López-Bigas N, Torrents D, Gut I, López-Guillermo A, López-Otín C, Campo E. Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature [Internet]. 2015;526(7574):519–24. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26200345%5Cnhttp://www.nature.com/doifinder/10.1038/nature14666

76. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of Next-Generation Sequencing Systems. J Biomed Biotechnol [Internet]. 2012 [cited 2012 Jul 17];2012:1–11. Available from: http://www.hindawi.com/journals/jbb/2012/251364/

77. Woollard PM, Mehta N a L, Vamathevan JJ, Van Horn S, Bonde BK, Dow DJ. The application of next-generation sequencing technologies to drug discovery and development. Drug Discov Today [Internet]. 2011 Jun [cited 2012 Aug 7];16(11–12):512–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21440664

78. Jeremy W. Dale MVS. From Genes to Genomes - Concepts and Applications of DNA Technology [Internet]. Wiley; 2002. 372 p. Available from: http://lib.myilibrary.com?ID=340484

79. Rothberg BEG, Rothberg JM. Massively parallel ("next-generation") DNA sequencing. Clin Chem. 2015;61(7):997–8.

80. Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30(9).

81. Parker H, Strefford JC. The mutational signature of chronic lymphocytic leukemia. Biochem J [Internet]. 2016;473(21):3725–40. Available from: http://biochemj.org/cgi/doi/10.1042/BCJ20160256

82. Zhao Z, Goldin L, Liu S, Wu L, Zhou W, Lou H, Lin X, Wang Z, Xu L, Marti G, Li G, Wu K. Evolution of multiple cell clones over a 29-year period of a CLL patient. 2016;(May).

83. Stevens-Kroef MJ, van den Berg E, Olde Weghuis D, Geurts van Kessel A, Pfundt R, Linssen-Wiersma M, Benjamins M, Dijkhuizen T, Groenen PJ, Simons A. Identification of prognostic relevant chromosomal abnormalities in chronic lymphocytic leukemia using microarray-based genomic profiling Identification of prognostic relevant chromosomal abnormalities in chronic lymphocytic leukemia using microarray-based. Mol Cytogenet [Internet]. 2014;7(1):3. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3905918&tool=pmcentrez&ren

dertype=abstract

84. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc [Internet]. 2009 Jan [cited 2012 Jul 16];4(7):1073–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19561590

85. Cawthon RM. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. Nucleic Acids Res [Internet]. 2009 Feb [cited 2012 Jul 18];37(3):e21. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2647324&tool=pmcentrez&rendertype=abstract

86. White TB, McCoy AM, Streva VA, Fenrich J, Deininger PL. A droplet digital PCR detection method for rare L1 insertions in tumors. Mob DNA [Internet]. 2014;5(1):30. Available from: http://www.mobilednajournal.com/content/5/1/30

87. Hömig-Hölzel C, Savola S. Multiplex Ligation-dependent Probe Amplification (MLPA) in Tumor Diagnostics and Prognostics. Diagnostic Mol Pathol. 2012;21(4):1.

88. Véronèse L, Tournilhac O, Combes P, Prie N, Pierre-Eymard E, Guièze R, Veyrat-Masson R, Bay JO, Vago P, Tchirkov A. Contribution of MLPA to routine diagnostic testing of recurrent genomic aberrations in chronic lymphocytic leukemia. Cancer Genet [Internet]. 2013 [cited 2014 Sep 22];206(1–2):19–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23313109

89. Alhourani E, Rincic M, Othman MA, Pohle B, Schlie C, Glaser A, Liehr T. Comprehensive chronic lymphocytic leukemia diagnostics by combined multiplex ligation dependent probe amplification (MLPA) and interphase fluorescence in situ hybridization (iFISH). Mol Cytogenet [Internet]. 2014 Jan [cited 2015 Jan 8];7(1):79. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4247644&tool=pmcentrez&rendertype=abstract

90. Abdool A, Donahue AC, Wohlgemuth JG, Yeh C-H. Detection, analysis and clinical validation of chromosomal aberrations by multiplex ligation-dependent probe amplification in chronic leukemia. PLoS One [Internet]. 2010 Jan [cited 2014 Sep 22];5(10):e15407. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2963645&tool=pmcentrez&rendertype=abstract

91. Coll-Mulet L, Santidrián AF, Cosialls AM, Iglesias-Serret D, de Frias M, Grau J, Menoyo A,

González-Barca E, Pons G, Domingo A, Gil J. Multiplex ligation-dependent probe amplification for detection of genomic alterations in chronic lymphocytic leukaemia. Br J Haematol [Internet]. 2008 Sep [cited 2015 Jan 8];142(5):793–801. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18564355

92. Al Zaabi E a, Fernandez L a, Sadek I a, Riddell DC, Greer WL. Multiplex ligation-dependent probe amplification versus multiprobe fluorescence in situ hybridization to detect genomic aberrations in chronic lymphocytic leukemia: a tertiary center experience. J Mol Diagn [Internet]. 2010 Mar [cited 2014 Dec 21];12(2):197–203. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2871726&tool=pmcentrez&rendertype=abstract

93. Fabris S, Scarciolla O, Morabito F, Cifarelli RA, Dininno C, Cutrona G, Matis S, Recchia AG, Gentile M, Ciceri G, Ferrarini M, Ciancio A, Mannarella C, Neri A, Fragasso A. Multiplex Ligation-Dependent Probe Amplification and Fluorescence In Situ Hybridization to Detect Chromosomal Abnormalities in Chronic Lymphocytic Leukemia : A Comparative Study. genes Chromosom cancer. 2011;734(March):726–34.

94. Rossi D, Khiabanian H, Spina V, Ciardullo C, Bruscaggin A, Famà R, Monti S, Deambrogi C, Paoli L De, Wang J, Gattei V, Guarini A, Foà R, Rabadan R, Gaidano G. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. Blood. 2014;123(APRIL):2139–47.

95. Baliakas P, Hadzidimitriou a, Sutton L, Rossi D, Minga E, Villamor N, Larrayoz M, Kminkova J, Agathangelidis a, Davis Z, Tausch E, Stalika E, Kantorova B, Mansouri L, Scarfò L, Cortese D, Navrkalova V, Rose-Zerilli MJJ, Smedby KE, Juliusson G, Anagnostopoulos a, Makris a M, Navarro a, Delgado J, Oscier D, Belessi C, Stilgenbauer S, Ghia P, Pospisilova S, Gaidano G, Campo E, Strefford JC, Stamatopoulos K, Rosenquist R. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. Leukemia [Internet]. 2014 Jun 19 [cited 2014 Jul 23];(April):1–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24943832

96. Gunnarsson R, Isaksson a, Mansouri M, Göransson H, Jansson M, Cahill N, Rasmussen M, Staaf J, Lundin J, Norin S, Buhl a M, Smedby KE, Hjalgrim H, Karlsson K, Jurlander J, Juliusson G, Rosenquist R. Large but not small copy-number alterations correlate to high-risk genomic aberrations and survival in chronic lymphocytic leukemia: a high-resolution genomic screening of newly diagnosed patients. Leuk Off J Leuk Soc Am Leuk Res Fund, UK. 2010;24(1):211–5.

97. Liu J, Chen G, Feng L, Zhang W, Pelicano H, Wang F, Ogasawara M a, Lu W, Amin HM, Croce

CM, Keating MJ, Huang P. Loss of p53 and altered miR15-a/16-1MCL-1 pathway in CLL: insights from TCL1-Tg:p53(-/-) mouse model and primary human leukemia cells. Leukemia [Internet]. 2014;28(October 2012):118–28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23608884

98. Fabbri M, Bottoni A, Shimizu M, Spizzo R, Nicoloso MS, Rossi S, Barbarotto E, Cimmino A, Adair B, Wojcik SE, Valeri N, Calore F, Sampath D, Fanini F, Vannini I, Musuraca G, Dell'Aquila M, Alder H, Davuluri R V, Rassenti LZ, Negrini M, Nakamura T, Amadori D, Kay NE, Rai KR, Keating MJ, Kipps TJ, Calin G a, Croce CM. Association of a microRNA/TP53 feedback circuitry with pathogenesis and outcome of B-cell chronic lymphocytic leukemia. JAMA. 2011;305(1):59–67.

99. Rossi D, Fangazio M, Rasi S, Vaisitti T, Monti S, Cresta S, Chiaretti S, Del Giudice I, Fabbri G, Bruscaggin A, Spina V, Deambrogi C, Marinelli M, Famà R, Greco M, Daniele G, Forconi F, Gattei V, Bertoni F, Deaglio S, Pasqualucci L, Guarini A, Dalla-Favera R, Foà R, Gaidano G. Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. Blood [Internet]. 2012 Mar 22 [cited 2012 Aug 11];119(12):2854–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22308293

100. Chigrinova E, Rinaldi A, Kwee I, Rossi D, Rancoita PM V, Jonathan C, Oscier D, Stamatopoulos K, Papadaki T, Berger F, Ken H, Murray F, Rosenquist R, Greiner TC, Chan WC, Orlandi EM, Lucioni M, Marasca R, Inghirami G, Ladetto M, Forconi F, Votavova H, Swerdlow SH, Stilgenbauer S, Piris M a, Spagnolo D, Nikitin E, Zamò A, Gattei V, Bhagat G, Zucca E, Gaidano G, Bertoni F, Strefford JC, Young KH. Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome. Blood. 2014;122(15):2673–82.

101. Balatti V, Bottoni A, Palamarchuk A, Alder H, Rassenti LZ, Kipps TJ, Pekarsky Y, Croce CM. NOTCH1 mutations in CLL associated with trisomy 12. Blood. 2012;

102. Forconi F, Rinaldi A, Kwee I, Sozzi E, Raspadori D, Rancoita PM V., Scandurra M, Rossi D, Deambrogi C, Capello D, Zucca E, Marconi D, Bomben R, Gattei V, Lauria F, Gaidano G, Bertoni F. Genome-wide DNA analysis identifies recurrent imbalances predicting outcome in chronic lymphocytic leukaemia with 17p deletion. Br J Haematol [Internet]. 2008;(August):532–6. Available from: http://doi.wiley.com/10.1111/j.1365-2141.2008.07373.x

103. Tsirigotis P, Pappa V, Labropoulos S, Papageorgiou S, Kontsioti F, Dervenoulas J, Papageorgiou

E, Panani A, Mantzios G, Economopoulos T, Raptis S. Mutational and methylation analysis of the cyclin-dependent kinase 4 inhibitor (p16INK4A) gene in chronic lymphocytic leukemia. Eur J Haematol [Internet]. 2006;76(3):230–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16412137

104. Hannou SA, Wouters K, Paumelle R, Staels B. Functional genomics of the CDKN2A/B locus in cardiovascular and metabolic disease: what have we learned from GWASs? Trends Endocrinol Metab [Internet]. 2015;26(4):176–84. Available from: http://www.sciencedirect.com/science/article/pii/S1043276015000235

105. Haidar BMA, Cao X, Manshouri T, Chan LL, Glassman A, Kantarjian HM, Keating MJ, Beran MS, Albitar M. pl6 and p15 Gene Deletions in Primary Leukemias. Blood. 1995;86:311–5.

106. Arjan Buijs, Pieter Jaap Krijtenburg EM. Detection of risk-identifying chromosomal abnormalities and genomic profiling by multiplex ligation-dependent probe amplification in chronic lymphocytic leukemia We performed genomic profiling using multiplex. Haematologica. 2006;

107. Chapiro E, Leporrier N, Radford-Weiss I, Bastard C, Mossafa H, Leroux D, Tigaud I, De Braekeleer M, Terré C, Brizard F, Callet-Bauchu E, Struski S, Veronese L, Fert-Ferrer S, Taviaux S, Lesty C, Davi F, Merle-Béral H, Bernard OA, Sutton L, Raynaud SD, Nguyen-Khac F. Gain of the short arm of chromosome 2 (2p) is a frequent recurring chromosome aberration in untreated chronic lymphocytic leukemia (CLL) at advanced stages. Leuk Res [Internet]. 2010 Jan [cited 2015 Nov 12];34(1):63–8. Available from: http://www.sciencedirect.com/science/article/pii/S0145212609001660

108. Lapalombella R, Sun Q, Williams K, Tangeman L, Jha S, Zhong Y, Goettl V, Mahoney E, Berglund C, Gupta S, Farmer A, Mani R, Johnson AJ, Lucas D, Mo X, Daelemans D, Sandanayaka V, Shechter S, McCauley D, Shacham S, Kauffman M, Chook YM, Byrd JC. Selective inhibitors of nuclear export show that CRM1/XPO1 is a target in chronic lymphocytic leukemia. Blood [Internet]. 2012;120(23):4621–34. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3512237&tool=pmcentrez&rendertype=abstract

109. Jain P. Clinical and molecular characteristics of XPO1 mutations in patients with chronic lymphocytic leukemia. Am J Hematol. 2016;1–29.

110. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A,

DeLuca DS, Zhang L, Zhang W, Vartanov AR, Fernandes SM, Goldstein NR, Folco EG, Cibulskis K, Tesar B, Sievers QL, Shefler E, Gabriel S, Hacohen N, Reed R, Meyerson M, Golub TR, Lander ES, Neuberg D, Brown JR, Getz G, Wu CJ. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med [Internet]. 2011 Dec 29;365(26):2497–506. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22150006

111. Oscier DG, Rose-Zerilli MJJ, Winkelmann N, De Castro DG, Gomez B, Forster J, Parker H, Parker A, Gardiner A, Collins A, Else M, Cross NCP, Catovsky D, Strefford JC. The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. Blood. 2013;121(3):468–75.

112. Rossi D, Bruscaggin A, Spina V, Rasi S, Khiabanian H, Messina M, Fangazio M, Vaisitti T, Monti S, Chiaretti S, Guarini A, Del Giudice I, Cerri M, Cresta S, Deambrogi C, Gargiulo E, Gattei V, Forconi F, Bertoni F, Deaglio S, Rabadan R, Pasqualucci L, Foà R, Dalla-Favera R, Gaidano G. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. Blood [Internet]. 2011 Dec 22 [cited 2012 Aug 10];118(26):6904–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22039264

113. Xia Y, Fan L, Wang L, Gale RP, Wang M, Tian T, Wu W, Yu L, Chen Y, Xu W, Li J. Frequencies of SF3B1 , NOTCH1 , MYD88 , BIRC3 and IGHV mutations and TP53 disruptions in Chinese with chronic lymphocytic leukemia : disparities with Europeans. Oncotarget. 2014;1–9.

114. Stilgenbauer S, Schnaiter A, Paschka P, Zenz T, Rossi M, Döhner K, Bühler A, Böttcher S, Ritgen M, Kneba M, Winkler D, Tausch E, Hoth P, Edelmann J, Mertens D, Bullinger L, Bergmann M, Kless S, Mack S, Jäger U, Patten N, Wu L, Wenger MK, Fingerle-Rowson G, Lichter P, Cazzola M, Wendtner CM, Fink AM, Fischer K, Busch R, Hallek M, Döhner H. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. Blood [Internet]. 2014 May 22 [cited 2014 Nov 15];123(21):3247–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24652989

115. Guieze R, Robbe P, Clifford R, de Guibert S, Pereira B, Timbs a., Dilhuydy M-S, Cabes M, Ysebaert L, Burns a., Nguyen-Khac F, Davi F, Veronese L, Combes P, Le Garff-Tavernier M, Leblond V, Merle-Beral H, Alsolami R, Hamblin a., Mason J, Pettitt a., Hillmen P, Taylor J, Knight SJL, Tournilhac O, Schuh a. Presence of multiple recurrent mutations revealed by targeted NGS confers poor trial outcome of relapsed/refractory CLL. Blood [Internet]. 2015; Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2015-05-647578

116. Dreger P, Schnaiter A, Zenz T, Böttcher S, Rossi M, Paschka P, Bühler A, Dietrich S, Busch R,

Ritgen M, Bunjes D, Zeis M, Stadler M, Uharek L, Scheid C, Hegenbart U, Hallek M, Kneba M, Schmitz N, Döhner H, Stilgenbauer S. TP53, SF3B1, and NOTCH1 mutations and outcome of allotransplantation for chronic lymphocytic leukemia: six-year follow-up of the GCLLSG CLL3X trial. Blood. 2013;121(16):3284–8.

117. te Raa GD, Derks I a M, Navrkalova V, Skowronska a, Moerland PD, van Laar J, Oldreive C, Monsuur H, Trbusek M, Malcikova J, Lodén M, Geisler CH, Hüllein J, Jethwa a, Zenz T, Pospisilova S, Stankovic T, van Oers MHJ, Kater a P, Eldering E. The impact of SF3B1 mutations in CLL on the DNA-damage response. Leukemia [Internet]. 2014 Nov 5 [cited 2014 Nov 7];(August 2014):1133–42. Available from: http://www.nature.com/doifinder/10.1038/leu.2014.318

118. Wan Y, Wu CJ. SF3B1 mutations in chronic lymphocytic leukemia. Blood. 2013;121(23):5–8.

119. Foà R, Del Giudice I, Guarini A, Rossi D, Gaidano G. Clinical implications of the molecular genetics of chronic lymphocytic leukemia. Haematologica [Internet]. 2013 May [cited 2013 May 8];98(5):675–85. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3640109&tool=pmcentrez&rendertype=abstract

120. Martínez-Trillos A, Pinyol M, Navarro A, Aymerich M, Jares P, Juan M, Rozman M, Colomer D, Delgado J, Gine E, González-Díaz M, Hernández-Rivas JM, Colado E, Rayón C, Payer AR, Terol MJ, Navarro B, Quesada V, Puente XS, Rozman C, López-Otín C, Campo E, López-Guillermo A, Villamor N. Mutations in the Toll-like receptor/MYD88 pathway in chronic lymphocytic leukemia identify a subset of young patients with favorable outcome. Blood [Internet]. 2014;123(24):3790–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24782504

121. Wang JQ, Jeelall YS, Ferguson LL, Horikawa K. Toll-like receptors and cancer: MYD88 mutation and inflammation. Front Immunol. 2014;5(JUL):1–10.

122. Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, Escaramis G, Jares P, Beà S, González-Díaz M, Bassaganyas L, Baumann T, Juan M, López-Guerra M, Colomer D, Tubío JMC, López C, Navarro A, Tornador C, Aymerich M, Rozman M, Hernández JM, Puente D a, Freije JMP, Velasco G, Gutiérrez-Fernández A, Costa D, Carrió A, Guijarro S, Enjuanes A, Hernández L, Yagüe J, Nicolás P, Romeo-Casabona CM, Himmelbauer H, Castillo E, Dohm JC, de Sanjosé S, Piris M a, de Alava E, San Miguel J, Royo R, Gelpí JL, Torrents D, Orozco M, Pisano DG, Valencia A, Guigó R, Bayés M, Heath S, Gut M, Klatt P, Marshall J, Raine K, Stebbings L a, Futreal PA, Stratton MR, Campbell PJ, Gut I, López-Guillermo A, Estivill X,

Montserrat E, López-Otín C, Campo E. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature [Internet]. 2011 Jul 7 [cited 2012 Jul 12];475(7354):101–5. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3322590&tool=pmcentrez&rendertype=abstract

123. Rossi D, Rasi S, Fabbri G, Spina V, Fangazio M, Forconi F, Marasca R, Laurenti L, Bruscaggin A, Cerri M, Monti S, Cresta S, Famà R, De Paoli L, Bulian P, Gattei V, Guarini A, Deaglio S, Capello D, Rabadan R, Pasqualucci L, Dalla-Favera R, Foà R, Gaidano G. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. Blood [Internet]. 2012 Jan 12 [cited 2013 Jan 29];119(2):521–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3257017&tool=pmcentrez&rendertype=abstract

124. Sportoletti P, Baldoni S, Cavalli L, Papa B Del, Bonifacio E, Ciurnelli R, Bell AS, Tommaso A Di, Rosati E, Barbara C, Mecucci C, Isabella S, Marconi P, Martelli MF, Ianni DM, Falzetti F. NOTCH1 PEST domain mutation is an adverse prognostic factor in B-CLL. Br J Haematol. 2010;151(4):402–6.

125. Schnaiter A, Paschka P, Rossi M, Zenz T, Andreas B, Winkler D, Cazzola M, Konstanze D, Edelmann J, Mertens D, Kless S, Mack S, Busch R, Hallek M. patients treated with alemtuzumab : results from the CLL2H trial of the GCLLSG. Blood. 2016;122(7):1266–71.

126. Jeromin S, Weissmann S, Haferlach C, Dicker F, Bayer K, Grossmann V, Alpermann T, Roller a, Kohlmann a, Haferlach T, Kern W, Schnittger S. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. Leukemia [Internet]. 2014;28(1):108–17. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24113472

127. Röth A, de Beer D, Nückel H, Sellmann L, Dührsen U, Dürig J, Baerlocher GM. Significantly shorter telomeres in T-cells of patients with ZAP-70+/CD38+ chronic lymphocytic leukaemia. Br J Haematol [Internet]. 2008 Nov [cited 2013 Apr 9];143(3):383–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18759763

128. Mansouri L, Grabowski P, Degerman S, Svenson U, Gunnarsson R, Cahill N, Smedby KE, Geisler C, Juliusson G, Roos G, Rosenquist R. Short telomere length is associated with NOTCH1/SF3B1/TP53 aberrations and poor outcome in newly diagnosed chronic lymphocytic leukemia patients. Am J Hematol [Internet]. 2013 Apr 26 [cited 2013 May 3]; Available from:

http://www.ncbi.nlm.nih.gov/pubmed/23620080

129. Sutton L-A, Rosenquist R. Deciphering the molecular landscape in chronic lymphocytic leukemia: time frame of disease evolution. Haematologica [Internet]. 2015 Jan [cited 2015 Jan 3];100(1):7–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25552678

130. Fabbri G, Rasi S, Rossi D, Trifonov V, Khiabanian H, Ma J, Grunn A, Fangazio M, Capello D, Monti S, Cresta S, Gargiulo E, Forconi F, Guarini A, Arcaini L, Paulli M, Laurenti L, Larocca LM, Marasca R, Gattei V, Oscier D, Bertoni F, Mullighan CG, Foá R, Pasqualucci L, Rabadan R, Dalla-Favera R, Gaidano G. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. J Exp Med [Internet]. 2011 Jul 4 [cited 2012 Aug 10];208(7):1389–401. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3135373&tool=pmcentrez&rendertype=abstract

131. Quesada V, Conde L, Villamor N, Ordóñez GR, Jares P, Bassaganyas L, Ramsay AJ, Beà S, Pinyol M, Martínez-Trillos A, López-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Giné E, Hernández JM, González-Díaz M, Puente D a, Velasco G, Freije JMP, Tubío JMC, Royo R, Gelpí JL, Orozco M, Pisano DG, Zamora J, Vázquez M, Valencia A, Himmelbauer H, Bayés M, Heath S, Gut M, Gut I, Estivill X, López-Guillermo A, Puente XS, Campo E, López-Otín C. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet [Internet]. 2012 Jan [cited 2012 Jul 20];44(1):47–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22158541

132. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, Sougnez C, Stewart C, Sivachenko A, Wang L, Wan Y, Zhang W, Shukla SA, Vartanov A, Fernandes SM, Saksena G, Cibulskis K, Tesar B, Gabriel S, Hacohen N, Meyerson M, Lander ES, Neuberg D, Brown JR, Getz G, Wu CJ. Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. Cell [Internet]. 2013 Feb [cited 2013 Feb 14];152(4):714–26. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0092867413000718

133. Koo B-H, Hurskainen T, Mielke K, Aung PP, Casey G, Autio-Harmainen H, Apte SS. ADAMTSL3/punctin-2, a gene frequently mutated in colorectal tumors, is widely expressed in normal and malignant epithelial cells, vascular endothelial cells and other cell types, and its mRNA is reduced in colon cancer. Int J cancer [Internet]. 2007 Oct 15 [cited 2016 Sep 21];121(8):1710–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17597111

134. Gylfe AE, Sirkiä J, Ahlsten M, Järvinen H, Mecklin J-P, Karhu A, Aaltonen LA. Somatic

mutations and germline sequence variants in patients with familial colorectal cancer. Int J Cancer [Internet]. 2010;127(12):2974–80. Available from: http://dx.doi.org/10.1002/ijc.25529

135. French D, Yang W, Cheng C, Raimondi SC, Mullighan CG, Downing JR, Evans WE, Pui C-H, Relling M V. Acquired variation outweighs inherited variation in whole genome analysis of methotrexate polyglutamate accumulation in leukemia. Blood [Internet]. 2009 May 7;113(19):4512 LP-4520. Available from: http://www.bloodjournal.org/content/113/19/4512.abstract

136. Massa S, Junker S, Schubart K, Matthias G, Matthias P. The OBF-1 gene locus confers B cell-specific transcription by restricting the ubiquitous activity of its promoter. Eur J Immunol [Internet]. 2003 Oct [cited 2016 Sep 21];33(10):2864–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14515270

137. Qiu L, Wang Z, Shi X, Wang Z. Associations between XPC polymorphisms and risk of cancers: A meta-analysis. Eur J Cancer. 2008;44(15):2241–53.

138. Qian Z, Qian J, Lin J, Yao D, Chen Q, Ji R, Li Y, Xiao G, Li J. GTPase regulator associated with the focal adhesion kinase (GRAF) transcript was down-regulated in patients with myeloid malignancies. J Exp Clin Cancer Res [Internet]. 2010 Aug 12;29(1):111. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2927506/

139. Thauvin-Robinet C, Lee JS, Lopez E, Herranz-Perez V, Shida T, Franco B, Jego L, Ye F, Pasquier L, Loget P, Gigot N, Aral B, Lopes CAM, St-Onge J, Bruel A-L, Thevenon J, Gonzalez-Granero S, Alby C, Munnich A, Vekemans M, Huet F, Fry AM, Saunier S, Riviere J-B, Attie-Bitach T, Garcia-Verdugo JM, Faivre L, Megarbane A, Nachury M V. The oral-facial-digital syndrome gene C2CD3 encodes a positive regulator of centriole elongation. Nat Genet [Internet]. 2014 Aug;46(8):905–11. Available from: http://dx.doi.org/10.1038/ng.3031

140. Hoover AN, Wynkoop A, Zeng H, Jia J, Niswander LA, Liu A. C2cd3 is required for cilia formation and Hedgehog signaling in mouse. Development [Internet]. 2008 Nov 21;135(24):4049–58. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=19004860&retmode=ref&cmd=prlinks&holding=FindIt@Stanford%5Cnpapers3://publication/doi/10.1242/dev.029835

141. Ljungstrom V, Cortese D, Young E, Pandzic T, Mansouri L, Plevova K, Ntoufa S, Baliakas P,

Clifford R, Sutton LA, Blakemore SJ, Stavroyianni N, Agathangelidis A, Rossi D, H??glund M, Kotaskova J, Juliusson G, Belessi C, Chiorazzi N, Panagiotidis P, Langerak AW, Smedby KE, Oscier D, Gaidano G, Schuh A, Davi F, Pott C, Strefford JC, Trentin L, Pospisilova S, Ghia P, Stamatopoulos K, Sj??blom T, Rosenquist R. Whole-exome sequencing in relapsing chronic lymphocytic leukemia: Clinical impact of recurrent RPS15 mutations. Blood [Internet]. 2016;127(8):1007–16. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2015-10-674572

142. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, Kluth S, Bozic I, Lawrence M, Böttcher S, Carter SL, Cibulskis K, Mertens D, Sougnez CL, Rosenberg M, Hess JM, Edelmann J, Kless S, Kneba M, Ritgen M, Fink A, Fischer K, Gabriel S, Lander ES, Nowak MA, Döhner H, Hallek M, Neuberg D, Getz G, Stilgenbauer S, Wu CJ. Mutations driving CLL and their evolution in progression and relapse. Nature [Internet]. 2015;526(7574):525–30. Available from: http://www.nature.com/doifinder/10.1038/nature15395%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/26466571

143. Pedrola N, Devis L, Llauradó M, Campoy I, Martinez-Garcia E, Garcia M, Muinelo-Romay L, Alonso-Alconada L, Abal M, Alameda F, Mancebo G, Carreras R, Castellví J, Cabrera S, Gil-Moreno A, Matias-Guiu X, Iovanna JL, Colas E, Reventós J, Ruiz A. Nidogen 1 and Nuclear Protein 1: novel targets of ETV5 transcription factor involved in endometrial cancer invasion. Clin {&} Exp Metastasis [Internet]. 2015;32(5):467–78. Available from: http://dx.doi.org/10.1007/s10585-015-9720-7

144. Tanenbaum DG, Hall WA, Colbert LE, Bastien AJ, Brat DJ, Kong J, Kim S, Dwivedi B, Kowalski J, Landry JC, Yu DS. TNFRSF10C copy number variation is associated with metastatic colorectal cancer. J Gastrointest Oncol [Internet]. 2016;7(3):306–14. Available from: http://jgo.amegroups.com/article/view/5802/6577

145. Ishiguro H, Shimokawa T, Tsunoda T, Tanaka T, Fujii Y, Nakamura Y, Furukawa Y. Isolation of HELAD1, a novel human helicase gene up-regulated in colorectal carcinomas. Oncogene. 2002;21(41):6387–94.

146. Lavrov A V., Chelysheva EY, Smirnikhina SA, Shukhov OA, Turkina AG, Adilgereeva EP, Kutsev SI. Frequent variations in cancer-related genes may play prognostic role in treatment of patients with chronic myeloid leukemia. BMC Genet [Internet]. 2016;17(S1):14. Available from: http://www.biomedcentral.com/1471-2156/17/S1/14

147. Haffner C, Frauli M, Topp S, Irmler M, Hofmann K, Regula JT, Bally-Cuif L, Haass C. Nicalin and its binding partner Nomo are novel Nodal signaling antagonists. EMBO J [Internet]. 2004;23(15):3041–50. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=514924&tool=pmcentrez&rendertype=abstract

148. Balatti V, Bottoni A, Palamarchuk A, Alder H, Rassenti LZ, Kipps TJ, Pekarsky Y, Croce CM. NOTCH1 mutations in CLL associated with trisomy 12. Blood. 2012;119(2):329–31.

149. Rosati E, Sabatini R, Rampino G, Tabilio A, Di Ianni M, Fettucciari K, Bartoli A, Coaccioli S, Screpanti I, Marconi P. Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. Blood. 2009;113(4):856–65.

150. Rasi S, Monti S, Spina V, Foa' R, Gaidano G, Rossi D. Analysis of NOTCH1 mutations in monoclonal B cell lymphocytosis. Haematologica [Internet]. 2011 Oct 11; Available from: http://www.haematologica.org/content/early/2011/09/20/haematol.2011.053090.abstract

151. Pozzo F, Bittolo T, Arruga F, Bulian P, Macor P, Tissino E, Gizdic B, Rossi FM, Bomben R, Zucchetto A, Benedetti D, Degan M, D'Arena G, Chiarenza A, Zaja F, Pozzato G, Rossi D, Gaidano G, Del Poeta G, Deaglio S, Gattei V, Dal Bo M. NOTCH1 mutations associate with low CD20 level in chronic lymphocytic leukemia: evidence for a NOTCH1 mutation-driven epigenetic dysregulation [Internet]. Vol. 30, Leukemia. Macmillan Publishers Limited; 2016. p. 182–9. Available from: http://dx.doi.org/10.1038/leu.2015.182

152. Stilgenbauer S, Schnaiter A, Paschka P, Zenz T, Rossi M, Döhner K, Bühler A, Böttcher S, Ritgen M, Kneba M, Winkler D, Tausch E, Hoth P, Edelmann J, Mertens D, Bullinger L, Bergmann M, Kless S, Mack S, Jäger U, Patten N, Wu L, Wenger MK, Fingerle-Rowson G, Lichter P, Cazzola M, Wendtner CM, Fink AM, Fischer K, Busch R, Hallek M, Döhner H. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. Blood [Internet]. 2014 May 22;123(21):3247–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24652989

153. Filip A a. New boys in town: prognostic role of SF3B1, NOTCH1 and other cryptic alterations in chronic lymphocytic leukemia and how it works. Leuk Lymphoma [Internet]. 2013 Sep [cited 2014 Aug 12];54(9):1876–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23343182

154. Visconte V, Makishima H, Maciejewski JP, Tiu R V. Emerging roles of the spliceosomal

machinery in myelodysplastic syndromes and other hematological disorders. Leukemia [Internet]. 2012;26(12):2447–54. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3645466&tool=pmcentrez&rendertype=abstract

155. Wan Y, Wu CJ. SF3B1 mutations in chronic lymphocytic leukemia. Blood [Internet]. 2013 Jun 6 [cited 2014 Nov 13];121(23):4627–34. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3674664&tool=pmcentrez&rendertype=abstract

156. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, Shiosaka M, Kawahata R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Ishiyama K, Mori H, Nolte F, Hofmann W-K, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih L-Y, Haferlach T, Chiba S, Nakauchi H, Miyano S, Ogawa S. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature [Internet]. 2011 Oct 6 [cited 2012 Jul 13];478(7367):64–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21909114

157. Malcovati L, Papaemmanuil E, Bowen DT, Boultwood J, Porta MG Della, Pascutto C, Travaglino E, Groves MJ, Godfrey AL, Ambaglio I, Gallì A, Già MC Da, Conte S, Tauro S, Keenan N, Hyslop A, Hinton J, Laura J, Wainscoat JS, Futreal PA, Stratton MR, Campbell PJ, Cazzola M. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic / myeloproliferative neoplasms Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic / myeloproliferative neoplasms The first two. Blood. 2011;118(24):6239–47.

158. Greco M, Capello D, Bruscaggin A, Spina V, Rasi S, Monti S, Ciardullo C, Cresta S, Fangazio M, Gaidano G, Foà R, Rossi D. Analysis of SF3B1 mutations in monoclonal B-cell lymphocytosis. Hematol Oncol [Internet]. 2013;31(1):54–5. Available from: http://dx.doi.org/10.1002/hon.2013

159. Graubert TA, Shen D, Ding L, Okeyo-owuor T, Cara L, Shao J, Krysiak K, Harris CC, Koboldt DC, David E, Mclellan MD, Dooling DJ, Abbott RM, Fulton RS, Schmidt H, Kalicki-veizer J, Laughlin MO, Grillot M, Baty J, Westervelt P, Dipersio JF, Mardis ER, Ley TJ. RECURRENT MUTATIONS IN THE U2AF1 SPLICING FACTOR IN MYELODYSPLASTIC SYNDROMES. Nat Genet. 2012;44(1):53–7.

160. Buggins a. GS, Pepper C, Patten PEM, Hewamana S, Gohil S, Moorhead J, Folarin N, Yallop D,

Thomas NSB, Mufti GJ, Fegan C, Devereux S. Interaction with Vascular Endothelium Enhances Survival in Primary Chronic Lymphocytic Leukemia Cells via NF- B Activation and De novo Gene Transcription. Cancer Res [Internet]. 2010;70(19):7523–33. Available from: http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-10-1634

161. Hayden MS, Ghosh S. NF-κB, the first quarter-century: remarkable progress and outstanding questions. Genes Dev [Internet]. 2012 Feb 1 [cited 2012 Oct 29];26(3):203–34. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3278889&tool=pmcentrez&rendertype=abstract

162. Beyaert R. An E3 ubiquitin ligase-independent role of LUBAC. Blood [Internet]. 2014;123(14):2131–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24700714

163. Bai L, Smith DC, Wang S. Small-molecule SMAC mimetics as new cancer therapeutics. Pharmacol Ther [Internet]. 2014;144(1):82–95. Available from: http://dx.doi.org/10.1016/j.pharmthera.2014.05.007

164. Sun S-C. Non-canonical NF-κB signaling pathway. Cell Res [Internet]. 2011 Jan [cited 2012 Aug 10];21(1):71–85. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3193406&tool=pmcentrez&rendertype=abstract

165. Deveraux QL, Reed JC. IAP family proteins - Suppressors of apoptosis. Genes Dev. 1999;13(3):239–52.

166. Estornes Y, Bertrand MJM. IAPs, regulators of innate immunity and inflammation. Semin Cell Dev Biol [Internet]. 2014;39:1–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24718315

167. Antosz H, Sajewicz J, Marzec-Kotarska B, Dmoszyńska A, Baszak J, Jargiełło-Baszak M. Aberrant TIRAP and MyD88 expression in B-cell chronic lymphocytic leukemia. Blood Cells Mol Dis [Internet]. 2013 Feb 15 [cited 2013 Feb 21];1–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23419703

168. Lin S-C, Lo Y-C, Wu H. Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling. Nature [Internet]. 2010;465(7300):885–90. Available from: http://dx.doi.org/10.1038/nature09121

169. Medzhitov R, Preston-Hurlburt P, Kopp E, Stadlen A, Chen C, Ghosh S, Janeway CA. MyD88 Is an Adaptor Protein in the hToll/IL-1 Receptor Family Signaling Pathways. Mol Cell [Internet]. 1998 Aug [cited 2015 Oct 22];2(2):253–8. Available from: http://www.sciencedirect.com/science/article/pii/S1097276500801367

170. Baliakas P, Hadzidimitriou a., Agathangelidis a., Rossi D, Sutton L -a., Kminkova J, Scarfo L, Pospisilova S, Gaidano G, Stamatopoulos K, Ghia P, Rosenquist R. Prognostic relevance of MYD88 mutations in CLL: the jury is still out. Blood [Internet]. 2015;126(8):1043–4. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2015-05-648634

171. Insuasti-Beltran G, Gale JM, Wilson CS, Foucar K, Czuchlewski DR. Significance of *MYD88* L265P Mutation Status in the Subclassification of Low-Grade B-Cell Lymphoma/Leukemia. Arch Pathol Lab Med [Internet]. 2015;139(8):1035–41. Available from: http://www.archivesofpathology.org/doi/10.5858/arpa.2014-0322-OA

172. Muzio M, Scielzo C, Bertilaccio MTS, Frenquelli M, Ghia P, Caligaris-Cappio F. Expression and function of toll like receptors in chronic lymphocytic leukaemia cells. Br J Haematol. 2009;144(4):507–16.

173. Jones CH, Pepper C, Baird DM. Telomere dysfunction and its role in haematological cancer. Br J Haematol [Internet]. 2012 Mar [cited 2012 Nov 20];156(5):573–87. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22233151

174. Poncet D, Belleville A, t'kint de Roodenbeke C, Roborel de Climens A, Ben Simon E, Merle-Beral H, Callet-Bauchu E, Salles G, Sabatier L, Delic J, Gilson E. Changes in the expression of telomere maintenance genes suggest global telomere dysfunction in B-chronic lymphocytic leukemia. Blood [Internet]. 2008 Feb 15 [cited 2012 Sep 17];111(4):2388–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18077792

175. Pepper C, Baird D, Fegan C. Telomere analysis to predict chronic lymphocytic leukemia outcome: a STELA test to change clinical practice? Expert Rev Hematol [Internet]. 2014;7(6):701–3. Available from: http://informahealthcare.com/doi/abs/10.1586/17474086.2014.969705

176. Lin TT, Norris K, Heppel NH, Pratt G, Allan JM, Allsup DJ, Bailey J, Cawkwell L, Hills R, Grimstead JW, Jones RE, Britt-Compton B, Fegan C, Baird DM, Pepper C. Telomere dysfunction accurately predicts clinical outcome in chronic lymphocytic leukaemia, even in patients with early stage disease. Br J Haematol [Internet]. 2014 Oct [cited 2014 Nov

18];167(2):214–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24990087

177. Véronèse L, Tournilhac O, Callanan M, Prie N, Kwiatkowski F, Combes P, Chauvet M, Davi F, Gouas L, Verrelle P, Guièze R, Vago P, Bay JO, Tchirkov a. Telomeres and chromosomal instability in chronic lymphocytic leukemia. Leuk Off J Leuk Soc Am Leuk Res Fund, UK [Internet]. 2012 Jul 13 [cited 2012 Nov 20];1(July):1–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22878603

178. Roos G, Kröber A, Grabowski P, Kienle D, Bühler A, Döhner H, Rosenquist R, Stilgenbauer S. Short telomeres are associated with genetic complexity, high-risk genomic aberrations, and short survival in chronic lymphocytic leukemia. Blood [Internet]. 2008 Feb 15 [cited 2012 Sep 5];111(4):2246–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18045969

179. Hultdin M, Rosenquist R, Thunberg U, Tobin G, Norrback K-F, Johnson a, Sundström C, Roos G. Association between telomere length and V(H) gene mutation status in chronic lymphocytic leukaemia: clinical and biological implications. Br J Cancer. 2003;88(4):593–8.

180. Grabowski P, Hultdin M, Karlsson K, Tobin G, Aleskog A, Thunberg U, Laurell A, Sundström C, Rosenquist R, Roos G. Telomere length as a prognostic parameter in chronic lymphocytic leukemia with special reference to VH gene mutation status. Blood [Internet]. 2005 Jun 15 [cited 2012 Sep 6];105(12):4807–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15746080

181. Lin TT, Letsolo BT, Jones RE, Rowson J, Pratt G, Hewamana S, Fegan C, Pepper C, Baird DM. Telomere dysfunction and fusion during the progression of chronic lymphocytic leukemia: evidence for a telomere crisis. Blood [Internet]. 2010 Sep 16 [cited 2012 Aug 23];116(11):1899–907. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20538793

182. Britt-Compton B, Lin TT, Ahmed G, Weston V, Jones R, Fegan CD, Oscier DG, Stankovic T, Pepper CJ, Baird DM. Extreme telomere erosion in ATM-mutated and 11q-deleted CLL patients is independent of disease stage [letter]. Leukemia [Internet]. 2011;826–30. Available from: http://dx.doi.org/10.1038/leu.2011.281

183. Augereau A, T'kint de Roodenbeke C, Simonet T, Bauwens S, Horard B, Callanan M, Leroux D, Jallades L, Salles G, Gilson E, Poncet D. Telomeric damage in early stage of chronic lymphocytic leukemia correlates with shelterin dysregulation. Blood [Internet]. 2011 Aug 4 [cited 2013 Apr 9];118(5):1316–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21355086

184. Baumann P, Price C. Pot1 and telomere maintenance. FEBS Lett [Internet]. 2010 Sep 10 [cited 2012 Jul 17];584(17):3779–84. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2942089&tool=pmcentrez&rendertype=abstract

185. Ramsay AJ, Quesada V, Foronda M, Conde L, Martínez-Trillos A, Villamor N, Rodríguez D, Kwarciak A, Garabaya C, Gallardo M, López-Guerra M, López-Guillermo A, Puente XS, Blasco M a, Campo E, López-Otín C. POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. Nat Genet [Internet]. 2013 May [cited 2014 Aug 12];45(5):526–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23502782

186. Chang S. Cancer chromosomes going to POT1. Nat Genet [Internet]. 2013 Apr 26 [cited 2013 Apr 27];45(5):473–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23619786

187. Hurtado a M, Chen-Liang T-H, Przychodzen B, Hamedi C, Muñoz-Ballester J, Dienes B, García-Malo MD, Antón a I, de Arriba F, Teruel-Montoya R, Ortuño FJ, Vicente V, Maciejewski JP, Jerez a. Prognostic signature and clonality pattern of recurrently mutated genes in inactive chronic lymphocytic leukemia. Blood Cancer J [Internet]. 2015;5(8):e342. Available from: http://www.nature.com/doifinder/10.1038/bcj.2015.65

188. Wan S-M, Tie J, Zhang Y-F, Guo J, Yang L-Q, Wang J, Xia S-H, Yang S-M, Wang R-Q, Fang D-C. Silencing of the hPOT1 gene by RNA inference promotes apoptosis and inhibits proliferation and aggressive phenotype of gastric cancer cells, likely through up-regulating PinX1 expression. J Clin Pathol [Internet]. 2011 Dec [cited 2012 Sep 12];64(12):1051–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21778296

189. Alhourani E, Othman MAK, Melo JB, Carreira IM, Grygalewicz B, Vujić D, Zecević Z, Joksić G, Glaser A, Pohle B, Schlie C, Hauke S, Liehr T. BIRC3 alterations in chronic and B-cell acute lymphocytic leukemia patients. Oncol Lett. 2016;11(5):3240–6.

190. Sabatel H, Di Valentin E, Gloire G, Dequiedt F, Piette J, Habraken Y. Phosphorylation of p65(ReLA) on ser 547 by ATM represses NF-κb-dependent transcription of specific genes after genotoxic stress. PLoS One. 2012;7(6).

191. Fang L, Choudhary S, Zhao Y, Edeh CB, Yang C, Boldogh I, Brasier AR. ATM regulates NF-κB-dependent immediate-early genes via RelA ser 276 phosphorylation coupled to CDK9 promoter recruitment. Nucleic Acids Res. 2014;42(13):8416–32.

192. Hadian K, Krappmann D. Signals from the nucleus: activation of NF-kappaB by cytosolic ATM

in the DNA damage response. Sci Signal. 2011;4(156):pe2.

193. Wu Z-H, Shi Y, Tibbetts RS, Miyamoto S. Molecular linkage between the kinase ATM and NF-kappaB signaling in response to genotoxic stimuli. Science. 2006;311(February):1141–6.

194. Wu Z-H, Miyamoto S. Many faces of NF-kappaB signaling induced by genotoxic stress. J Mol Med (Berl) [Internet]. 2007;85(11):1187–202. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17607554

195. Robertson LE, Plunkett W, McConnell K, Keating MJ, McDonnell TJ. Bcl-2 expression in chronic lymphocytic leukemia and its correlation with the induction of apoptosis and clinical outcome. Leukemia [Internet]. 1996 Mar [cited 2016 Dec 7];10(3):456–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8642861

196. Al-Harbi S, Hill B, Mazumder S. An antiapoptotic BCL-2 family expression index predicts the response of chronic lymphocytic leukemia to ABT-737. Blood [Internet]. 2011;118(13):3579–91. Available from: http://www.bloodjournal.org/content/118/13/3579.abstract

197. Pepper1 C, Hoy2 T, Bentley D. Bcl-2/Bax ratios in chronic lymphocytic leukaemia and their correlation with in vitro apoptosis and clinical resistance. Br Joumal Cancer. 1997;76(7):935–8.

198. Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R, Feller SM, Grocock R, Henderson S, Khrebtukova I, Kingsbury Z, Luo S, McBride D, Murray L, Menju T, Timbs A, Ross M, Taylor J, Bentley D. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. Blood [Internet]. 2012 Aug 22 [cited 2012 Sep 21];120(20):4191–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22915640

199. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, Kempski H, Moorman A V, Titley I, Swansbury J, Kearney L, Enver T, Greaves M. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature [Internet]. 2011 Jan 20;469(7330):356–61. Available from: http://dx.doi.org/10.1038/nature09650

200. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, McMichael JF, Wallis JW, Lu C, Shen D, Harris CC, Dooling DJ, Fulton RS, Fulton LL, Chen K, Schmidt H, Kalicki-Veizer J, Magrini VJ, Cook L, McGrath SD, Vickery TL, Wendl MC, Heath S, Watson MA, Link DC, Tomasson MH, Shannon WD, Payton JE, Kulkarni S, Westervelt P, Walter MJ, Graubert TA, Mardis ER, Wilson RK, DiPersio JF. Clonal evolution in

relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature [Internet]. 2012 Jan 26;481(7382):506–10. Available from: http://dx.doi.org/10.1038/nature10738

201. Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K, Larson DE, McLellan MD, Dooling D, Abbott R, Fulton R, Magrini V, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Fan X, Grillot M, Witowski S, Heath S, Frater JL, Eades W, Tomasson M, Westervelt P, DiPersio JF, Link DC, Mardis ER, Ley TJ, Wilson RK, Graubert TA. Clonal Architecture of Secondary Acute Myeloid Leukemia. N Engl J Med [Internet]. 2012 Mar 14;366(12):1090–8. Available from: http://dx.doi.org/10.1056/NEJMoa1106968

202. Rossi D, Gaidano G. The clinical implications of gene mutations in chronic lymphocytic leukaemia. Br J Cancer [Internet]. 2016;114(8):849–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27031852

203. Doménech E, Gómez-López G, Gzlez-Peña D, López M, Herreros B, Menezes J, Gómez-Lozano N, Carro A, Graña O, Pisano DG, Domínguez O, García-Marco J a., Piris M a., Sánchez-Beato M. New mutations in chronic lymphocytic leukemia identified by target enrichment and deep sequencing. PLoS One. 2012;7(6):2–7.

204. Jethwa A, Hüllein J, Stolz T, Blume C, Sellner L, Jauch A, Sill M, Kater AP, te Raa GD, Geisler C, van Oers M, Dietrich S, Dreger P, Ho AD, Paruzynski A, Schmidt M, von Kalle C, Glimm H, Zenz T. Targeted resequencing for analysis of clonal composition of recurrent gene mutations in chronic lymphocytic leukaemia. Br J Haematol [Internet]. 2013;163(4):496–500. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24032483

205. Rasi S, Khiabanian H, Ciardullo C, Terzi-di-Bergamo L, Monti S, Spina V, Bruscaggin A, Cerri M, Deambrogi C, Martuscelli L, Biasi A, Spaccarotella E, de Paoli L, Gattei V, Foà R, Rabadan R, Gaidano G, Rossi D. Clinical impact of small subclones harboring NOTCH1, SF3B1 or BIRC3 mutations in chronic lymphocytic leukemia [Internet]. Vol. 101, Haematologica. 2016. 101-135 p. Available from: http://www.haematologica.org/cgi/doi/10.3324/haematol.2015.136051

206. Nadeu F, Delgado J, Royo C, Baumann T, Stankovic T, Pinyol M, Jares P, Navarro A, Mart D, Su H, Gonz M, Alcoceba M, Jos M, Colado E, Puente XS, Carlos L, Enjuanes A, Martín-García D, Bè S, Salaverria I, Oldreive C, Aymerich M, Sú arez-Cisneros H, Rozman M, Villamor N, Colomer D, opez-Guillermo A, Gon alez M, Alcoceba M, Jo Terol M, Colado E, Puente XS, opez-Otín C, Enjuanes A, Campo E. Clinical impact of clonal and subcloal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. Blood. 2016;127(17):2122–30.

207. Lawrence MS, Stojanov P, Polak P, Kryukov G V, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts S a, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, Pugh TJ, Stransky N, Helman E, Kim J, Sougnez C, Ambrogio L, Nickerson E, Shefler E, Cortés ML, Auclair D, Saksena G, Voet D, Noble M, DiCara D, Lin P, Lichtenstein L, Heiman DI, Fennell T, Imielinski M, Hernandez B, Hodis E, Baca S, Dulak AM, Lohr J, Landau D-A, Wu CJ, Melendez-Zajgla J, Hidalgo-Miranda A, Koren A, McCarroll S a, Mora J, Lee RS, Crompton B, Onofrio R, Parkin M, Winckler W, Ardlie K, Gabriel SB, Roberts CWM, Biegel J a, Stegmaier K, Bass AJ, Garraway L a, Meyerson M, Golub TR, Gordenin D a, Sunyaev S, Lander ES, Getz G. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature [Internet]. 2013;499(7457):214–8. Available from: http://dx.doi.org/10.1038/nature12213

208. Zheng S, Kim H, Verhaak RGW. Silent mutations make some noise. Cell [Internet]. 2014;156(6):1129–31. Available from: http://dx.doi.org/10.1016/j.cell.2014.02.037

209. McCubrey JA, Lertpiriyapong K, Fitzgerald TL, Martelli AM, Cocco L, Rakus D, Gizak A, Libra M, Cervello M, Montalto G, Yang L V., Abrams SL, Steelman LS. Roles of TP53 in determining therapeutic sensitivity, growth, cellular senescence, invasion and metastasis. Adv Biol Regul [Internet]. 2016;October(16):1–17. Available from: http://linkinghub.elsevier.com/retrieve/pii/S2212492616300550

210. Tong AS, Stern JL, Sfeir A, Kartawinata M, de Lange T, Zhu XD, Bryan TM. ATM and ATR Signaling Regulate the Recruitment of Human Telomerase to Telomeres. Cell Rep [Internet]. 2015;13(8):1633–46. Available from: http://linkinghub.elsevier.com/retrieve/pii/S2211124715012115

211. Lee SS, Bohrson C, Pike AM, Wheelan SJ, Greider CW. ATM Kinase Is Required for Telomere Elongation in Mouse and Human Cells. Cell Rep [Internet]. 2015;13(8):1623–32. Available from: http://dx.doi.org/10.1016/j.celrep.2015.10.035

212. Bashamboo A, Brauner R, Bignon-Topalovic J, Lortat-Jacob S, Karageorgou V, Lourenco D, Guffanti A, McElreavey K. Mutations in the FOG2/ZFPM2 gene are associated with anomalies of human testis determination. Hum Mol Genet. 2014;23(14):3657–65.

213. Saito T, Sadoshima J. MicroRNA-200 promotes lung cancer cell growth through FOG2-independent AKT activation. 2016;116(9):1477–90.

214. Kang C, Xu Q, Martin TD, Li MZ, Demaria M, Aron L, Lu T, Yankner BA, Campisi J, Elledge SJ. The DNA damage response induces inflammation and senescence by inhibiting autophagy of

GATA4. Science [Internet]. 2015;349(6255):aaa5612. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26404840

215. Damm F, Mylonas E, Cosson A, Yoshida K, Della Valle V, Mouly E, Diop M, Scourzic L, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Kikushige Y, Davi F, Lambert J, Gautheret D, Merle-Béral H, Sutton L, Dessen P, Solary E, Akashi K, Vainchenker W, Mercher T, Droin N, Ogawa S, Nguyen-Khac F, Bernard OA. Acquired initiating mutations in early hematopoietic cells of CLL patients. Cancer Discov. 2014;4(9):1088–101.

216. Cheng S, Guo a, Lu P, Ma J, Coleman M, Wang YL. Functional characterization of BTK(C481S) mutation that confers ibrutinib resistance: exploration of alternative kinase inhibitors. Leukemia [Internet]. 2014;(April):1–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25189416

217. Woyach JA, Furman RR, Liu T-M, Ozer HG, Zapatka M, Ruppert AS, Xue L, Li DH-H, Steggerda SM, Versele M, Dave SS, Zhang J, Yilmaz AS, Jaglowski SM, Blum KA, Lozanski A, Lozanski G, James DF, Barrientos JC, Lichter P, Stilgenbauer S, Buggy JJ, Chang BY, Johnson AJ, Byrd JC. Resistance Mechanisms for the Bruton's Tyrosine Kinase Inhibitor Ibrutinib. N Engl J Med [Internet]. 2014 May 28;370(24):2286–94. Available from: http://dx.doi.org/10.1056/NEJMoa1400029

218. Liu T-M, Woyach J a., Zhong Y, Lozanski A, Lozanski G, Dong S, Strattan E, Lehman A, Zhang X, Jones J a., Flynn J, Andritsos L a., Maddocks K, Jaglowski SM, Blum K a., Byrd JC, Dubovsky J a., Johnson a. J. Hypermorphic mutation of phospholipase C, gamma 2 acquired in ibrutinib resistant CLL confers BTK independency upon BCR activation. Blood [Internet]. 2015;126(1):61–8. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2015-02-626846

219. Spinelli R, Pirola A, Redaelli S, Sharma N, Raman H, Valletta S, Magistroni V, Piazza R, Gambacorti-Passerini C. Identification of novel point mutations in splicing sites integrating whole-exome and RNA-seq data in myeloproliferative diseases. Mol Genet genomic Med [Internet]. 2013;1:246–59. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3865592&tool=pmcentrez&rendertype=abstract

220. Miles RR, Crockett D, Lim M, Elenitoba-Johnson KSJ. Analysis of BCL6-interacting Proteins by Tandem Mass Spectrometry. Mol Cell Proteomics [Internet]. 2005;4(12):1898–909. Available from: http://www.mcponline.org/cgi/doi/10.1074/mcp.M500112-MCP200

221. Swaminathan S, Duy C, Muschen M. BACH2-BCL6 balance regulates selection at the pre-B cell receptor checkpoint. Trends Immunol. 2014;35(3):131–7.

222. Duy CC, Yu JJJJ, Nahar RR, Swaminathan SS, Kweon S-MSM, Polo JMJM, Valls EE, Klemm LL, Shojaee SS, Cerchietti LL, Schuh WW, Jäck H-MHM, Hurtz CC, Ramezani-Rad PP, Herzog SS, Jumaa HH, Koeffler HPHP, de Alborán IMIM, Melnick AMAM, Ye BHBH, Müschen MM. BCL6 is critical for the development of a diverse primary B cell repertoire. J Exp Med [Internet]. 2010;207(6):1209–21. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=20498019&retmode=ref&cmd=prlinks%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC2882829/pdf/JEM_20091299.pdf

223. Legare S, Cavallone L, Mamo A, Chabot C, Sirois I, Magliocco A, Klimowicz A, Tonin PN, Buchanan M, Keilty D, Hassan S, Laperri??re D, Mader S, Aleynikova O, Basik M. The estrogen receptor cofactor SPEN functions as a tumor suppressor and candidate biomarker of drug responsiveness in hormone-dependent breast cancers. Cancer Res. 2015;75(20):4351–63.

224. Sharma S, Galanina N, Guo A, Lee J, Kadri S, Slambrouck C Van, Long B, Wang W, Ming M, Larissa V, Segal JP, Stock W, Venkataraman G, Tang W, Lu P, Wang YL. Identification of a structurally novel BTK mutation that drives ibrutinib resistance in CLL. Oncotarget. 2016;2:1–9.

225. Komarova NL, Burger JA, Wodarz D. Evolution of ibrutinib resistance in chronic lymphocytic leukemia (CLL). Proc Natl Acad Sci [Internet]. 2014 Sep 23;111(38):13906–11. Available from: http://www.pnas.org/content/111/38/13906.abstract

226. Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-Fahmawi B, Piazza P. Improved workflows for high throughput library preparation using the transposome-based Nextera system. BMC Biotechnol [Internet]. 2013;13(104):1–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24256843

227. Vallabhaneni H, Zhou F, Maul RW, Sarkar J, Yin J, Lei M, Harrington L, Gearhart PJ, Liu Y. Defective repair of uracil causes telomere defects in mouse hematopoietic cells. J Biol Chem. 2015;290(9):5502–11.

228. Strefford JC, Kadalayil L, Forster J, Rose-Zerilli MJJ, Parker A, Lin TT, Heppel N, Norris K, Gardiner A, Davies Z, Gonzalez de Castro D, Else M, Steele AJ, Parker H, Stankovic T, Pepper C, Fegan C, Baird D, Collins A, Catovsky D, Oscier DG. Telomere length predicts progression and

overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. Leukemia [Internet]. 2015;(August):1–3. Available from: http://www.nature.com/doifinder/10.1038/leu.2015.217

229. Herling M, Vasyutina E. Multilevel BCR signals toward CLL. lymphoid neoplasia. 2015;125(10):1510–3.

230. Byrd JC, Furman RR, Coutre SE, Burger JA, Blum KA, Coleman M, Wierda WG, Jones JA, Zhao W, Heerema NA, Johnson AJ, Shaw Y, Bilotti E, Zhou C, James DF, O'Brien S. Three-year follow-up of treatment-naive and previously treated patients with CLL and SLL receiving single-agent ibrutinib. Blood. 2015;125(16):2497–506.

231. Stevenson FKF, Krysov S, Davies AJ, Steele AJ, Packham G. B-cell receptor signaling in chronic lymphocytic leukemia. Blood [Internet]. 2011;118(16):4313–20. Available from: http://bloodjournal.hematologylibrary.org/content/118/16/4313.short%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/21816833

Appendices

Appendix A Genomic location of all probes in the MLPA SALSA P037 and P038 CLL kits, including expected fragment size. The genomic position is based on hg18 genome build.

Probe name	Genomic Position	KIT	Fragment length (nt)
*Ctrl_01p21.1	*01-103.234006	37	133.1
*Ctrl_01p21.1	*01-103.250538	38	452.6
*Ctrl_01q25.1	*01-172.145336	38	172.1
*Ctrl_01q41	*01-213.981368	37	153.4
*Ctrl_02p13.2	*02-071.430713	38	353.9
*Ctrl_02q12.3	*02-108.913043	37	269.5
*Ctrl_02q24.3	*02-166.607103	38	210.1
*Ctrl_02q32.2	*02-189.572828	37	449.4
*Ctrl_03p22.2	*03-038.624633	37	390.1
*Ctrl_03p12.2	*03-081.774613	37	496.6
*Ctrl_03p12.2	*03-081.774613	38	496.9
*Ctrl_04q35.2	*04-187.390325	37	183.6
*Ctrl_05p13.2	*05-036.997382	37	199.2
*Ctrl_05q31.1	*05-132.037610	37	127.3
*Ctrl_05q31.1	*05-132.037610	38	127.3
*Ctrl_06p22.3	*06-024.652972	37	424.8
*Ctrl_07p14.3	*07-030.609169	38	260.4
*Ctrl_07q11.23	*07-072.755115	38	294.7
*Ctrl_09q34.3	*09-136.799081	37	345.5
*Ctrl_10p13	*10-013.207521	38	191.5
*Ctrl_11p15.1	*11-017.406391	37	330.2
*Ctrl_14q11.2	*14-020.872609	38	485.5
*Ctrl_15q21.1	*15-042.646957	37	469.7
*Ctrl_15q24.1	*15-072.490183	38	434.5
*Ctrl_16p13.2	*16-008.715000	38	408.4
*Ctrl_22q12.3	*22-035.290414	38	151.8
MYCN_3	02-015.999649	37	251.2
MYCN_1	02-015.999768	37	178.1
MYCN_4	02-016.003135	37	365.5
MYCN_2	02-016.003390	37	205.2
ALK_2	02-029.273962	37	478
ALK_1	02-029.608415	37	415.5
REL	02-060.999129	37	224.2
REL_2	02-061.002527	37	292.6
SF3B1	02-197.975049	38	234
MYD88	03-038.157619	38	251
AIM1	06-107.075546	37	433.3
SEC63	06-108.321498	37	442.8
TNFAIP3_2	06-138.234189	37	488.4
TNFAIP3_1	06-138.243930	37	211.2
LATS1	06-150.046168	37	456
IGF2R	06-160.350050	37	237.2
PARK2	06-161.727843	37	352.5
TNFRSF10B	08-022.941980	37	244
TNFRSF10A	08-023.138420	37	313.2
EIF3H	08-117.837051	37	139.5

MYC	08-128.822148	37	160.2
MYC_2	08-128.822294	37	281.5
CDKN2A	09-021.961213	37	258.8
CDKN2B_1	09-021.995813	37	229.9
NOTCH1	09-138.510394	38	216
PTEN_1	10-089.682866	38	366.5
PTEN_2	10-089.715284	38	469.7
ATM_1	11-107.599044	38	184.2
ATM_4	11-107.629847	38	372.4
ATM_1	11-107.632169	37	173.4
ATM_2	11-107.648710	37	338.3
ATM_5	11-107.695883	38	460.1
ATM_2	11-107.710905	38	228.2
ATM_3	11-107.741279	38	287.6
RDX	11-109.613470	38	281.5
PPP2R1B	11-111.136866	38	340.1
CADM1	11-114.880363	38	360.4
CCND2	12-004.279427	37	358.3
CD27	12-006.429606	38	140.3
LRMP	12-025.152177	37	321.6
STAT6	12-055.787722	38	133.9
CDK4	12-056.428450	37	274.9
HMGA2	12-064.508064	38	415.9
IFNG	12-066.835439	37	218.1
IGF1	12-101.337462	38	309.9
PAH	12-101.795401	38	242.2
CHFR	12-131.958952	37	400.1
RB1_1	13-047.853385	38	160.1
RB1	13-047.935661	37	381.7
RB1	13-047.952486	38	477.2
FNDC3A	13-048.603395	37	189.6
DLEU2	13-049.454748	38	177.8
KCNRG	13-049.487591	38	222.8
KCNRG	13-049.492724	37	145.9
MIR15A	13-049.521254	37	166.6
DLEU2	13-049.554227	37	194.8
DLEU1_2	13-049.576840	38	400.7
DLEU1_1	13-049.782357	38	380.4
DLEU7	13-050.185014	37	462.8
DLEU7_2	13-050.315361	37	373.7
ATP7B	13-051.433985	37	305.9
ATP7B	13-051.440676	38	249.1
AKT1	14-104.317514	38	145.7
MTA1	14-104.982782	38	444.7
KIAA0125	14-105.382897	38	203.7
TP53_5	17-007.514674	38	346.7
TP53_2	17-007.517784	37	287.3
TP53_3	17-007.518277	38	275.6
TP53_4	17-007.518934	38	319.1
TP53_1	17-007.519217	37	264.6
TP53_3	17-007.520063	37	299.1
TP53_2	17-007.520626	38	197.4
TP53_1	17-007.531444	38	166.5
TP53_4	17-007.531621	37	408.1
CDKN2D	19-010.538899	38	266.9
LDLR	19-011.082353	38	300.6
CCNE1	19-035.006411	38	390.2
AKT2	19-045.439753	38	329.3
MIR498	19-058.869274	38	425.8

Appendix B List of samples that have discordant MLPA and FISH results. List of samples that have discordant results, MLPA v FISH. Samples with a negative FISH result and positive MLPA result may be due to false positive results or else the deletion may be too small for a FISH probe, e.g. PID 4121; 13q deleted, has a DLEU2-7 genes deletion detected by MLPA but is negative by FISH.

PID	genomic abnormality	FISH (%)	Discordant reason
4007	11q deletion	Yes	Only ATM probes deleted
4125	11q deletion	No	No FISH data
4153	11q deletion	21	Low clone size <30%
4233	11q deletion	No	No FISH data
4259	11q deletion	13	Low clone size <20%
4260	11q deletion	15	Low clone size <20%
4263	11q deletion	11	Low clone size <20%
4287	11q deletion	15	Low clone size <20%
4294	11q deletion	20	Low clone size <30%
4301	11q deletion	No	No FISH data
4338	11q deletion	12	Low clone size <20%
4365	11q deletion	19	Low clone size <20%
4367	11q deletion	11	Low clone size <20%
4377	11q deletion	17	Low clone size <20%
4379	11q deletion	Yes	Only ATM probes deleted
4418	11q deletion	No	No FISH data
4442	11q deletion	8	Low clone size <10%
4464	11q deletion	7	Low clone size <10%
4508	11q deletion	No	No FISH data
4691	11q deletion	9	Low clone size <10%
4713	11q deletion	8	Low clone size <10%
4724	11q deletion	6	Low clone size <10%
4740	11q deletion	10	Low clone size <20%
4751	11q deletion	10	Low clone size <20%
4756	11q deletion	5	Low clone size <10%
4002	13q deletion	No	No FISH data
4011	13q deletion	Yes	RB1 to DLEU7 probes deleted
4015	13q deletion	No	No FISH data
4033	13q deletion	No	No FISH data

4047	13q deletion	Yes	RB1 to DLEU7 probes deleted
4071	13q deletion	13	Low clone size <20%
4077	13q deletion	No	No FISH data
4088	13q deletion	No	No FISH data
4121	13q deletion	Yes	DLEU2 to DLEU7 probes deleted
4130	13q deletion	No	No FISH data
4133	13q deletion	16	Low clone size <20%
4137	13q deletion	No	No FISH data
4162	13q deletion	13	Low clone size <20%
4177	13q deletion	No	No FISH data
4198	13q deletion	No	No FISH data
4207	13q deletion	No	No FISH data
4268	13q deletion	Yes	DLEU2 to DLEU7 probes deleted
4275	13q deletion	14	Low clone size <20%
4281	13q deletion	Yes	DLEU2 to DLEU7 probes deleted
4289	13q deletion	20	Low clone size <30%
4301	13q deletion	No	No FISH data
4306	13q deletion	8	Low clone size <10%
4334	13q deletion	15	Low clone size <20%
4338	13q deletion	8	Low clone size <10%
4351	13q deletion	18	Low clone size <20%
4357	13q deletion	No	No FISH data
4359	13q deletion	10	Low clone size <20%
4368	13q deletion	No	No FISH data
4377	13q deletion	24	Low clone size <30%
4378	13q deletion	16	Low clone size <20%
4379	13q deletion	7	Low clone size <10%
4382	13q deletion	15	Low clone size <20%
4386	13q deletion	No	No FISH data
4387	13q deletion	No	No FISH data
4412	13q deletion	No	No FISH data
4425	13q deletion	Yes	FNDC3A to DLEU1 probes deleted
4426	13q deletion	15	Low clone size <20%
4431	13q deletion	No	No FISH data
4436	13q deletion	No	No FISH data
4449	13q deletion	7	Low clone size <10%
4477	13q deletion	No	No FISH data
4483	13q deletion	6	Low clone size <10%
4491	13q deletion	10	Low clone size <20%
4508	13q deletion	No	No FISH data
4510	13q deletion	No	No FISH data
4511	13q deletion	Yes	P038 data missing, DLEU1 probes only deleted
4531	13q deletion	No	No FISH data
4657	13q deletion	5	Low clone size <10%
4658	13q deletion	10	Low clone size <20%
4696	13q deletion	6	Low clone size <10%
4713	13q deletion	5	Low clone size <10%

4727	13q deletion	7	Low clone size <10%
4733	13q deletion	17	Low clone size <20%
4739	13q deletion	8	Low clone size <10%
4759	13q deletion	No	No FISH data
4764	13q deletion	8	Low clone size <10%
4769	13q deletion	14	Low clone size <20%
4771	13q deletion	16	Low clone size <20%
4073	17p deletion	14	Low clone size <20%
4081	17p deletion	No	No FISH data
4267	17p deletion	20	Low clone size <30%
4306	17p deletion	8	Low clone size <10%
4313	17p deletion	15	Low clone size <20%
4327	17p deletion	6	Low clone size <10%
4346	17p deletion	7	Low clone size <10%
4347	17p deletion	11	Low clone size <20%
4355	17p deletion	8	Low clone size <10%
4365	17p deletion	12	Low clone size <20%
4377	17p deletion	15	Low clone size <20%
4378	17p deletion	8	Low clone size <10%
4426	17p deletion	8	Low clone size <10%
4436	17p deletion	No	No FISH data
4438	17p deletion	7	Low clone size <10%
4442	17p deletion	9	Low clone size <10%
4451	17p deletion	20	Low clone size <30%
4485	17p deletion	6	Low clone size <10%
4502	17p deletion	6	Low clone size <10%
4526	17p deletion	No	No FISH data
4638	17p deletion	13	Low clone size <20%
4662	17p deletion	6	Low clone size <10%
4666	17p deletion	19	Low clone size <20%
4672	17p deletion	6	Low clone size <10%
4688	17p deletion	6	Low clone size <10%
4691	17p deletion	6	Low clone size <10%
4693	17p deletion	5	Low clone size <10%
4694	17p deletion	12	Low clone size <20%
4704	17p deletion	11	Low clone size <20%
4705	17p deletion	9	Low clone size <10%
4713	17p deletion	6	Low clone size <10%
4715	17p deletion	9	Low clone size <10%
4721	17p deletion	8	Low clone size <10%
4725	17p deletion	8	Low clone size <10%
4729	17p deletion	20	Low clone size <30%
4731	17p deletion	9	Low clone size <10%
4733	17p deletion	19	Low clone size <20%
4735	17p deletion	31	Borderline clone size
4740	17p deletion	22	Low clone size <30%
4747	17p deletion	27	Low clone size <30%

4749	17p deletion	9	Low clone size <10%
4767	17p deletion	8	Low clone size <10%
4078	6q deletion	Yes	AIM1 and SEC63 probes only deleted
4412	6q deletion	Yes	AIM1 to TNFAIP3 probes deleted
4674	6q deletion	Yes	AIM1 and SEC63 probes only deleted
4696	6q deletion	Yes	AIM1 and SEC63 probes only deleted
4724	6q deletion	Yes	All but 1 (borderline) probe deleted
4755	6q deletion	Yes	AIM1 to TNFAIP3 probes deleted
4263	Trisomy 12	9	Low clone size <10%
4264	Trisomy 12	17	Low clone size <20%
4275	Trisomy 12	12	Low clone size <20%
4289	Trisomy 12	10	Low clone size <20%
4357	Trisomy 12	No	No FISH data
4375	Trisomy 12	No	No FISH data
4389	Trisomy 12	No	No FISH data
4486	Trisomy 12	9	Low clone size <10%
4496	Trisomy 12	No	No FISH data
4526	Trisomy 12	No	No FISH data
4685	Trisomy 12	4	Low clone size <10%
4694	Trisomy 12	4	Low clone size <10%
4698	Trisomy 12	9	Low clone size <10%
4711	Trisomy 12	12	Low clone size <20%
4725	Trisomy 12	3	Low clone size <10%
4727	Trisomy 12	3	Low clone size <10%
4762	Trisomy 12	4	Low clone size <10%

Appendix C Comparison of MLPA to SNP 6.0 and FISH data. Discordant results are highlighted in bold.

PID	MLPA	SNP 6.0	FISH (%)
4006	13q deletion	13q deletion	13q deletion (78.5%)
4006	17p deletion	17p deletion	17p deletion (47%)
4006	not detected	19q deletion	NA
4011	13q deletion	not detected	Not detected
4011	not detected	14q	NA
4011	not called but present two probes deleted	6q deletion	Not detected
4011	not detected	6q gain	Not detected
4011	not detected	2p gain	NA
4023	not detected	6q deletion	Not detected
4033	13q deletion	13q deletion	NA
4038	11q deletion	11q deletion	11q deletion (76.5%)
4039	13q deletion	13q deletion	13q deletion (29.5%)
4039	Not detected-one *MYCN* probe deleted	2p gain	NA
4042	biallelic 13q deletion	biallelic 13q deletion	biallelic 13q deletion (41.7%)
4042	13q deletion	13q deletion	13q deletion (37%)
4062	11q deletion	11q deletion	11q deletion (95.5%)
4062	13q deletion	13q deletion	13q deletion (93.5%)
4065	not detected	12p gain	Not detected
4065	13q deletion	13q deletion	13q deletion (84%)
4072	11q deletion	11q deletion	11q deletion (93.5%)
4072	13q deletion	13q deletion	13q deletion (73%)
4073	biallelic 13q deletion	biallelic 13q deletion	biallelic 13q deletion (59%)
4073	13q deletion	13q deletion	13q deletion (31%)
4073	Not detected	17p deletion	17p deletion (14%)
4076	11q deletion	11q deletion	11q deletion(66%)
4076	13q deletion	13q deletion	13q deletion

			(22%)
4107	2p gain- *REL* probes only	2p gain	NA
4107	not detected-deletion not including CDKN2A/B genes	9p deletion	NA
4107	13q deletion	13q deletion	13q deletion (58%)
4107	Not detected	17p deletion	Not detected
4154	14q deletion	14q deletion	NA
4154	13q deletion	not detected	13q deletion (12%)
4168	8p deletion	8p deletion	NA
4168	11q deletion	11q deletion	11q deletion (94%)
4168	13q deletion	13q deletion	13q deletion (95%)
4175	13q deletion	13q deletion	13q deletion (38.5%)
4175	Not detected	13q gain	NA
4175	Not detected	LOH 11q	NA
4181	11q deletion	11q deletion	11q deletion (87%)
4181	Not detected	13q deletion	Not detected
4184	11q deletion	11q deletion	11q deletion (57%)
4184	not detected-two probes deleted	12p deletion	NA
4184	Not detected- two probes deleted	12q deletion	NA
4184	not detected-all probe values are zero	LOH 13q	NA
4184	biallelic 13q deletion	biallelic 13q deletion	biallelic 13q deletion (97%)
4184	17p deletion	17p deletion	17p deletion (91%)
4184	Not detected	19q deletion	NA
4197	17p deletion	17p deletion	17p deletion (82.5%)
4197	not detected	12p gain	NA
4209	Not detected	6q deletion	NA
4260	not detected	11q deletion	11q deletion (15%)
4260	trisomy 12	trisomy 12	Trisomy 12 (66%)
4266	11q deletion	11q deletion	11q deletion (22%)
4266	17p deletion	Not detected	Not detected
4266	Not detected-one probe deleted	12p deletion	NA
4289	11q deletion	11q deletion	11q deletion (93.5%)
4289	not detected	Not detected	Trisomy 12 (10%)
4289	Not detected	Not detected	6q deletion (25%)

4289	Not detected-6 non consecutive borderline value probes deleted	Not detected	13q deletion (20%)
4289	Not detected	12p deletion	NA
4294	Not detected-some probe values almost zero	LOH 13q	NA
4294	13q deletion	13q deletion	13q deletion (16%)
4294	biallelic 13q deletion	biallelic 13q deletion	biallelic 13q deletion (77%)
4295	17p deletion	17p deletion	17p deletion (88%)
4295	6q deletion	6q deletion	6q deletion (96%)
4295	Not detected	6q gain	NA
4310	11q deletion	11q deletion	11q deletion (41%)
4310	trisomy 12	trisomy 12	Trisomy 12 (70%)
4312	11q deletion	11q deletion	11q deletion (40.9%)
4312	13q deletion	13q deletion	13q deletion (31%)
4317	6q deletion	6q deletion	6q deletion (49%)
4317	13q deletion	13q deletion	13q deletion (93%)
4319	Not detected	6q gain	NA
4329	11q deletion	11q deletion	11q deletion (76%)
4329	13q deletion	13q deletion	13q deletion (79%)
4358	biallelic 13q deletion	biallelic 13q deletion	biallelic 13q deletion (71%)
4358	13q deletion	13q deletion	13q deletion (24%)
4358	Not detected	8q gain	NA
4365	trisomy 12	trisomy 12	Trisomy 12 (69%)
4365	No	13q deletion	Not detected
4365	Not detected	6q deletion	Not detected
4367	13q deletion	13q deletion	13q deletion (84%)
4367	Not detected	13q gain	NA
4377	trisomy 12	trisomy 12	Trisomy 12 (76%)
4377	Not detected	2p deletion	NA
4389	trisomy 12	trisomy 12	NA
4414	13q deletion	13q deletion	13q deletion (92.4%)
4429	11q deletion	11q deletion	11q deletion (19.7%)
4433	13q deletion	13q deletion	13q deletion

			(92%)
4433	11q deletion	11q deletion	11q deletion (75%)
4433	2p gain	2p gain	NA
4450	11q deletion	11q deletion	11q deletion (45%)
4450	13q deletion	13q deletion	13q deletion (87.6%)
4450	Not detected	6q deletion	Not detected
4450	Not detected	2p gain	NA
4450	Not detected	8q gain	NA
4451	Not detected	17p deletion	17p deletion (20%)
4451	Not detected-one probe is gained	17p gain	NA
4453	Not detected	trisomy 12	Not detected
4453	6q deletion	Not detected	6q deletion (47%)
4453	13q deletion	Not detected	13q deletion (92%)
4453	11q deletion	11q deletion	11q deletion (83%)
4476	Not detected	6q deletion	6q deletion (5.6%)
4476	11q deletion	11q deletion	11q deletion (54%)
4476	13q deletion	13q deletion	13q deletion (63.5%)
4476	13q deletion	13q deletion	biallelic 13q deletion (16%)
4476	Not detected	8p deletion	NA
4478	Not detected	11q deletion	Not detected
4478	Not detected-one probe deleted (CCNE1 gene)	19q deletion	NA
4478	6q deletion	6q deletion	6q deletion (57%)
4478	17p deletion	17p deletion	17p deletion (87%)
4479	11q deletion	11q deletion	11q deletion (83%)
4479	13q deletion	13q deletion	13q deletion (93%)
4479	2p gain	2p gain	NA
4485	11q deletion	11q deletion	11q deletion (83%)
4485	Not detected	Not detected	17p deletion (6%)
4485	No	13q deletion	Not detected
4485	Not detected	2p gain	NA
4490	13q deletion	Not detected	13q deletion (12%)
4490	11q deletion	11q deletion	11q deletion

			(86%)
4501	Not detected-one probe is deleted	3p deletion	NA
4507	13q deletion	13q deletion	13q deletion (14.8%)
4507	biallelic 13q deletion	biallelic 13q deletion	biallelic 13q deletion (80%)
4509	13q deletion	13q deletion	13q deletion (95%)
4517	Not detected- borderline	biallelic 13q deletion	biallelic 13q deletion (14.7%)
4517	13q deletion	13q deletion	13q deletion (78.3%)
4534	Not detected-one probe is deleted	8q deletion	NA
4534	2p gain	2p gain	NA
4534	11q deletion	11q deletion	11q deletion (89%)
4534	13q deletion	13q deletion	13q deletion (93.6%)
4534	17p deletion	not detected	Not detected
4536	8p deletion	8p deletion	NA
4536	Not detected	8q deletion	NA
4536	13q deletion	13q deletion	13q deletion (47.5%)
4536	Not detected	14q deletion	NA
4536	Not detected	17p deletion (not TP53)	Not detected
4583	Not detected	2p deletion	NA
4583	No	13q deletion	Not detected
4591	biallelic 13q deletion	biallelic 13q deletion	biallelic 13q deletion (90%)
4591	No	17p deletion (not TP53)	Not detected
4634	11q deletion	11q deletion	11q deletion (84%)
4634	13q deletion	13q deletion	13q deletion (43%)
4658	trisomy 12	trisomy 12	Trisomy 12 (63%)
4658	11q deletion	11q deletion	11q deletion 79%)
4658	Not detected	6q gain	NA
4662	13q deletion	13q deletion	13q deletion (28%)
4662	2p gain	2p gain	NA
4672	Not detected	8p deletion	NA
4672	Not detected	8q deletion	NA
4672	11q deletion	11q deletion	11q deletion (90%)
4672	Not detected	biallelic 13q deletion	Not detected

4672	13q deletion	13q deletion	13q deletion (52%)
4684	17p deletion	17p deletion	17p deletion (6%)
4688	Not detected	3p deletion	NA
4688	11q deletion	11q deletion	11q deletion (96%)
4688	13q deletion	13q deletion	13q deletion (92%)
4698	Not detected	3p deletion	NA
4698	8p deletion	8p deletion	NA
4698	Not detected	8q deletion	NA
4698	8q gain	8q gain	NA
4698	Not detected-one probe CDKN2D probe gained	19p gain	NA
4698	17p deletion	17p deletion	17p deletion (66%)
4698	biallelic 13q deletion	biallelic 13q deletion	Not detected
4711	Not detected	2p gain	NA
4711	11q deletion	11q deletion	11q deletion (86%)
4711	Not detected	Not detected	Trisomy 12 (12%)
4715	11q deletion	11q deletion	11q deletion (98%)
4727	8p deletion	8p deletion	NA
4727	19p deletion	19p deletion	NA
4727	9p deletion	9p deletion	NA
4727	8q deletion	8q deletion	NA
4727	17p deletion	17p deletion	17p deletion (90%)
4727	Not detected-2 probes gained	11q gain	NA
4727	Not detected	17p gain	NA
4741	11q deletion	11q deletion	11q deletion (80%)
4759	13q deletion	13q deletion	NA
4776	13q deletion	13q deletion	13q deletion (52%)
4776	13q deletion-borderline biallelic value in one probe	13q deletion	biallelic 13q deletion (17%)

Appendix D MLPA identified *NOTCH1* mutations were validation using TruSeq, Sanger sequencing and ddPCR

PID	TruSeq	VAF	Previous Sanger (blood paper)	ddPCR	ddPCR %	Mutation confirmed
4419	Mutated	9%	Mutated	NA		Yes
4705	Mutated	40%	Mutated	NA		Yes
4726	Mutated	52%	Mutated	NA		Yes
4014	NA		Mutated	NA		Yes
4099	NA		Mutated	NA		Yes
4128	NA		Mutated	NA		Yes
4145	NA		Mutated	NA		Yes
4159	NA		Mutated	NA		Yes
4186	NA		Mutated	NA		Yes
4200	NA		Mutated	NA		Yes
4447	NA		Mutated	NA		Yes
4454	NA		Mutated	NA		Yes
4475	NA		Mutated	NA		Yes
4536	NA		Mutated	NA		Yes
4711	NA		Mutated	NA		Yes
4043	Not mutated		Mutated	NA		Yes
4101	Not mutated		Mutated	NA		Yes
4130	Not mutated		Mutated	NA		Yes
4375	Not mutated		Mutated	NA		Yes
4396	Not mutated		Mutated	NA		Yes
4425	Not mutated		Mutated	NA		Yes
4435	Not mutated		Mutated	NA		Yes
4488	Not mutated		Mutated	NA		Yes
4489	Not mutated		Mutated	NA		Yes
4779	NA		NA	No DNA		No
4799	NA		NA	No DNA		No
4069	NA		Not mutated	No DNA		No
4731	Not mutated		Not mutated	No DNA		No
4762	Not mutated		Not mutated	No DNA		No
4717	Not mutated		NA	Yes	0%	No
4481	Not mutated		Not mutated	Yes	0%	No
4750	Not mutated		Not mutated	Yes	0%	No
4307	Mutated	6%	Not mutated	Yes	5%	Yes

4501	NA		Not mutated	Yes	7%	Yes
4146	NA		Not mutated	Yes	8%	Yes
4210	NA		Not mutated	Yes	8%	Yes
4023	Not mutated		Not mutated	Yes	8%	Yes
4032	Not mutated		Not mutated	Yes	8%	Yes
4267	Mutated	11%	Mutated	Yes	10%	Yes
4497	Mutated	9%	Not mutated	Yes	10%	Yes
4442	NA		Not mutated	Yes	10%	Yes
4165	NA		Not mutated	Yes	14%	Yes
4006	Mutated	43%	Mutated	Yes	25%	Yes
4629	Not mutated		NA	Yes	27%	Yes
4641	Not mutated		NA	Yes	36%	Yes
4362	Not mutated		NA	Yes	40%	Yes
4157	Mutated	52%	NA	Yes	46%	Yes
4309	Not mutated		Mutated	Yes	46%	Yes
4486	Mutated	55%	Not mutated	Yes	48%	Yes
4749	Not mutated		NA	Yes	49%	Yes
4366	Not mutated		NA	Yes	85%	Yes

Appendix E Comparitive table of MLPA identified *SF3B1* mutated patients to Sanger sequencing, Truseq and ddPCR.

PID	TruSeq	VAF	Sanger (archive)	Further Sanger	ddPCR	ddPCR %	Mutation confirmed
4691	NA		NA	No DNA	No DNA		No
4727	NA		NA	No DNA	No DNA		No
4647	Not mutated		NA	No DNA	No DNA		No
4660	Not mutated		NA	No DNA	No DNA		No
4740	Not mutated		NA	No DNA	No DNA		No
4209	Not mutated		NA	No DNA	No DNA		No
4273	Not mutated		Not mutated	No DNA	No DNA		No
4048	Not mutated		Not mutated	NA	Yes	0%	No
4011	Not mutated		Not mutated	Not mutated	Yes	0%	No
4338	Mutated	45%	NA	No DNA	No DNA		Yes
4576	Mutated	48%	NA	No DNA	No DNA		Yes
4603	Mutated	27%	NA	No DNA	No DNA		Yes
4777	Mutated	48%	NA	No DNA	No DNA		Yes
4028	Mutated	11%	NA	Mutated	Yes	12%	Yes
4130	Mutated	15%	NA	Mutated	Yes	13%	Yes
4166	Mutated	16%	Not mutated	Mutated	Yes	14%	Yes
4490	Mutated	7%	Not mutated	Mutated	Yes	5%	Yes
4032	Mutated	50%	Not mutated	NA	Yes	45%	Yes
4071	Mutated	21%	Not mutated	NA	Yes	19%	Yes
4133	Mutated	33%	NA	Mutated	NA		Yes
4330	Mutated	42%	NA	Mutated	NA		Yes
4449	Mutated	33%	NA	Mutated	NA		Yes
4634	Mutated	25%	NA	Mutated	NA		Yes
4713	Mutated	45%	NA	Mutated	NA		Yes

4725	Mutated	11%	NA	Mutated	NA	Yes
4623	NA		NA	Mutated	NA	Yes
4784	NA		NA	Mutated	NA	Yes
4005	NA		Not mutated	Mutated	NA	Yes
4099	NA		Not mutated	Mutated	NA	Yes
4128	NA		Not mutated	Mutated	NA	Yes
4261	NA		Not mutated	Mutated	NA	Yes
4485	NA		Not mutated	Mutated	NA	Yes
4104	Not mutated		NA	Mutated	NA	Yes
4033	Mutated	46%	Mutated	NA	NA	Yes
4038	Mutated	41%	Mutated	NA	NA	Yes
4266	Mutated	48%	Mutated	NA	NA	Yes
4319	Mutated	42%	Mutated	NA	NA	Yes
4334	Mutated	42%	Mutated	NA	NA	Yes
4393	Mutated	38%	Mutated	NA	NA	Yes
4419	Mutated	34%	Mutated	NA	NA	Yes
4451	Mutated	47%	Mutated	NA	NA	Yes
4535	Mutated	43%	Mutated	NA	NA	Yes
4662	Mutated	33%	Mutated	NA	NA	Yes
4734	Mutated	42%	Mutated	NA	NA	Yes
4008	NA		Mutated	NA	NA	Yes
4012	NA		Mutated	NA	NA	Yes
4046	NA		Mutated	NA	NA	Yes
4053	NA		Mutated	NA	NA	Yes
4132	NA		Mutated	NA	NA	Yes
4205	NA		Mutated	NA	NA	Yes
4501	NA		Mutated	NA	NA	Yes
4355	Not mutated		Mutated	NA	NA	Yes
4755	Not mutated		Mutated	NA	NA	Yes

Appendix F Sequencing primers designed to validate somatic mutations as confirmed by IGV for the CLL exome cases

Gene	Sequence (5'-3')	Primer length (bp)
SPEN	GCGTGCTTATGAACATAGTG	20
	GGCTGTTAACCTACCTGTCA	20
CHIT1	ATGAAGCTTGGGAAATTACA	20
	AAGACAAGGAGACTGCAAAA	20
NID1	ATACTTGGCCAAATTTTCAA	20
	GCCACAGTTATGCAGAAGAT	20
CPZ	CCTTCACATACACTGTGTCCT	21
	AAGACAGAAACAAAACTCACG	21
TNFRSF10C	GATTACAAGCTTGAGCCATC	20
	TAGCTTGATTTAGCCATTCC	20
C9orf66	GTCCTCCCCAAGAATCTC	18
	AGTTCTGTAAATCTCAAAATGTCA	24
KRTAP5-5	CCTGGACAGGGATATAAAGA	20
	GACACAGCCTCTCTTGGAG	19
DNHD1	GCCTATCCAAATTCTTCTCTC	21
	AATAAGCTGTGCCAACAAGT	20
NAV2	GCCTGTGATACCAATGAAGT	20
	AATGAAAAAGCCATACAGGA	20
SNX19	CTGGCTCACTGTAACTTAGGA	21
	AGAGCTACATTCAGGCAAAG	20
NOMO1	GTGCCAGTTTTTAACTTTGC	20
	GCTTTTACTGTGTCATGAGAGTCT	24
ABCC1	CCTTAAAGGAGTGTCTCTGG	20
	AGAGCGAGACCCTGTTTC	18
DNAH9	CCCTATATTACCGTCATGGA	20
	CTACAACGTTTTGGGAAAAG	20
BCL2	CCTTCAGCTTGAGAAACACT	20
	ATTGTGGCCTTCTTTGAGT	19
CACNA1A	TCTTGTTGGTGTTGTTGTTG	20
	AGGGTTCTTGAGCAGAGG	18
LOC646508	ACTGCAACTGGACTTGGAT	19
	CCTCAGAGGAGGAAGTGACC	20

Appendix G Sequencing primers designed to validate somatic mutations as confirmed by IGV for the SMZL exome cases

Gene	Sequence (5'-3')	Ta (°C)	Product (bp)
ACTN2	GCTAAATCAAGCTCATCCTG	55	498
	GTGATAACAGAACGGTGCTC	55	
ADAMTSL3	GACCCTCAGTGTCCTTTTCT	55	477
	TGAATGACTTCACCTCCTTT	55	
C2CD3	AAGTACCAAAGCAGTGATGG	55	575
	AAGGGATCACAGGATGGTAT	55	
CREBBP	GCTTTGGAGATTCTGAATTG	55	591
	ACATGCCTTGTCATCCTAAC	55	
FSHR	TTGGTATATGTTCTGCCACA	55	512
	AGCAGAAAGTTTTAGGTCCA	55	
POU2AF1	CAGTGATTCATGCACAGAAG	55	494
	GAGGTGGGAGGATGTGAT	55	
XPC	TTCTCTCCTCCTTGAATCCT	55	586
	AGGGTTTGTAAGTGGACACA	55	
AGXT2L1	CCATGAATCAAAAGGAAAGA	55	540
	GGAGTAAAGCTGACTTGGAA	55	
ARHGAP26	CCAGAAAGCACCATTAAAAG	55	519
	CTGCAATGAACATTTCACAC	55	
ARSJ	AGCAGGACAGTTTTCAGTGT	55	491
	CATATGAGAGGGTGGACCTA	55	
CILP2	GTGTTGTGGCTGCTGACT	55	552
	GAGAACATGCCGTAGGTG	55	
CSMD3	AGTGATGTTAAAGCCACGTC	55	598
	ATGGTAACCGTTTTGTTCTG	55	
CUL1	TTGTTTTTAGGATGGAGAAGA	55	588
	AACTGTGATGGTAGGGTCAG	55	
DBT	CAGGAGAACCATTACACCAT	55	482
	TGTACAGGAAGAAGGGTTTG	55	
DNAH10	ACAAATATTCACCCCTTCCT	55	568
	TGGCACACCAAACTACACTA	55	
FBXO30	AATGGCGGTCATGTATAATC	55	576
	TAGATCTGATCACAGCAGCA	55	
FHOD3	TGTTAAGAGATGCCTCATCC	55	472
	TGATGAGTCACTTGCATCTC	55	
FOXK2	TTTTCTAGCAGGTGTGATAGC	55	580
	CCAGCATGACTAGATTTCCT	55	
H1FOO	CCCTACAGAAGGCAGGAT	55	511

	GCCTTGCTACCTGTAGATCA	55	
IGDCC3	GCTGGTTTTCCACATCTTTA	55	579
	GATTGGTCTGGAGTCCTGTA	55	
IGSF9B	GTGGTGGCAGGTAGGACT	55	554
	CTTGCAGCTTGTCCTCTG	55	
IQGAP1	TCCAAGGAGAATGGAAATAA	55	541
	GGGCTTTCCAGATAAATTCT	55	
OAS3	CCAAGGTCACACAGTAGGTT	55	473
	CTGGTTGTGTACTGTGCAAC	55	
OR4Q3	CCCGATAAAACTTGCTAAAA	55	556
	AAACATCTCACTGGCTCCTA	55	
PARD3B	GCTCAGAGAAGTCAAACCAC	55	572
	GTTGAGCACACATTCTTGC	55	
SHMT1	TCCCTCAACTGAATCTGAAC	55	504
	CTTTAGCACAGAGGTGCTTT	55	
SLC13A1	GGGCGTGAATTACTTATTTG	55	576
	ATTAGCCCCAATAGAAAAGC	55	
SPATA20	CTGTTCTGACACTGCTTCC	55	578
	ATGGGCAGAATGGTTCTT	55	
VGLL3	TTATTGCAAAATGGCTAGGT	55	513
	GGCCTTATCCTTTGACATCT	55	

Appendix H Primers designed for HRM and Sanger sequencing validation

Primer	Gene	Exon	Forward sequence	Reverse sequence	Tm (°C)	Product (bp)
HRM	POT1	4	TGCAATGTAATTAGAGAATAAAAGCTG	ATTATACGTATTTTGGTGATTGATTCA	60	192
Sequencing	POT1	4	AAGTGCAATATCTGCCAAGT	TCCAAACAATGACAAAATCA	55	478
HRM	POT1	5	CACATGTATCTATGTGTGTGGCATA	AGCATGTAATCACATTGGAGGTT	60	181
Sequencing	POT1	5	TCAGCAGATATTCCAGACAA	AGCTTAGACAACTTTGCACAT	55	452
HRM	POT1	6	AAACTCCACCAGTTTTAATACCTACC	TACATGGATTTGCTGCTAATATGAT	60	229
Sequencing	POT1	6	AGCCAAAGAATATGCATCAG	CCATTTATAAACAAAGTTCTAAGGA	55	457
HRM	POT1	7	TTCTCTTCAAATAAATATAAGTTCTAGAC	GGTTTGGTGTTTTGAAGTAAGCA	60	281
Sequencing	POT1	7	GCAGTGGTTTGTTCAAATG	TTGCAGTGTGTATTGAAAGC	55	493
HRM	POT1	8	TGGTGCTAACTTATAATTCCCAGTATT	CCTTACGTGTTTGGGCATCT	60	183
HRM	POT1	8	AGATGCCCAAACACGTAAGG	CTGTTTTCTACTTTGCCCTACTTTC	60	197
Sequencing	POT1	8	CCACACAAATCTCATGTCAA	TCACCCAGTAAATCTCTTTAGC	55	474
HRM	POT1	9	TCAGAGATCTTGCCACATGAA	TTATGGCAGGTATGGGATGG	60	163
Sequencing	POT1	9	CATTTTACAACTTAAAAATCAAAGA	TTCCACATTACCCATATTTCA	55	487
HRM	POT1	10	TCGGCTTAATCGATACCTTATTTAC	TTTTTCCCCACTTTCTAAATAACAA	60	257
Sequencing	POT1	10	ATTTGTTTCATTTGGCTCAT	CCATGCAGCTGATATTCAA	55	483
HRM	POT1	18	TCAAGTAAAAGAAGTGTGGGATTG	AAGGACAAATTCTTCCAGATTCC	60	152
Sequencing	POT1	18	TTTGACTGCAGGAATTATGA	GATTTTGGAGTTGAGACCAG	60	475
HRM	BIRC3	7	TTCCATATAGTTATCCATTTTGAACCT	ACATACTTGATTCTTTTTCCTCAGTTG	60	259
Sequencing	BIRC3	7	TGCCTATACATTTTGTTGGTT	AAAAACCTGACTGGATTGAG	55	490
HRM	BIRC3	10	TGAAGAAGCAAACTGCCTTTTATT	AAAGTTTAGACGATGTTTTGGTTCT	60	268
Sequencing	BIRC3	10	CCACAGAAGATGTTTCAGGT	GTGCTACCTCTTTTTCGTTC	55	546
HRM	MYD88	3	TCTGACCACCACCCTTGTG	GGCCTTCTAGCCAACCTCTT	60	227
Sequencing	MYD88	3	GGCACTTTCTCTGAGGAGTA	GACAGTGCACAGCTAGGAG	55	407
HRM	MYD88	4	GCTGAACTAAGTTGCCACAGG	CCAGAGCAGGGTTGAGCTT	60	191

Sequencing	MYD88	4	CAGGGGATATGCTGAACTAA	GATCTTCAGCAGTTCTTTGG	55	266
HRM	MYD88	5	CAGGTGCCCATCAGAAGC	GGTTGGTGTAGTCGCAGACA	60	102
Sequencing	MYD88	5	GCAGAAGTACATGGACAGGCAGACAGATAC	GTTGTTAACCCTGGGGTTGAAG	55	297
HRM	U2AF1	Codon 34	TGTCGTCATGGAGACAGGTG	AGGCAAACAAACCTGGCTAA	61	66
Sequencing	U2AF1	34	TAAAAACAAGGAGTGGTGGT	GGGAAGTTGAAGTTACGTTG	55	480
HRM	U2AF1	157	CCGTGACGGACTTCAGAGA	CCTCACTCACCCCATCTCAT	60	56
HRM	SRSF2	95	GAGCTGCGGGTGCAAAT	CGGCTGTGGTGTGAGTCC	61	56

Appendix I .Illumina TruSeq Index PCR primers. Illumina technology uses combinations of 12 i7 and 8 i5 index PCR primers to multiplex samples for NGS. The 8 base sequence for each primer is show in the tables below. A green LED is used to sequence G/T and a red LED to sequence A/C. At each cycle at least one of two nucleotides for each colour channel need to be read to ensure proper registration. It is important to maintain colour balance for each base of the index read being sequenced, otherwise registration failure could occur during index read sequencing.

TruSeq Indices								
i7 Index PCR Primer	Index Sequence							
A701	A	T	C	A	C	G	A	C
A702	A	C	A	G	T	G	G	T
A703	C	A	G	A	T	C	C	A
A704	A	C	A	A	A	C	G	G
A705	A	C	C	C	A	G	C	A
A706	A	A	C	C	C	C	T	C
A707	C	C	C	A	A	C	C	T
A708	C	A	C	C	A	C	A	C
A709	G	A	A	A	C	C	C	A
A710	T	G	T	G	A	C	C	A
A711	A	G	G	G	T	C	A	A
A712	A	G	G	A	G	T	G	G
i5 Index PCR Primer	Index Sequence							
A501	T	G	A	A	C	C	T	T
A502	T	G	C	T	A	A	G	T
A503	T	G	T	T	C	T	C	T
A504	T	A	A	G	A	C	A	C
A505	C	T	A	A	T	C	G	A
A506	C	T	A	G	A	A	C	A
A507	T	A	A	G	T	T	C	C
A508	T	A	G	A	C	C	T	A

Appendix J Illumina recommends processing at least 16 samples at a time. While many index combinations will perform well in most situations, it is recommended to include the minimal set of colour-balanced indexes. The following index combinations, selected to maintain colour balance, are arranged in sets of 8. For more than 8 samples but less than a multiple of 8, the relevant number of filler samples were chosen Fillers are used sequentially.

	Index Combination Sets of 8		Index Combination Fillers
Set 1	501/701	1	501/710
	502/702	2	502/701
	503/703	3	503/702
	504/704	4	504/709
	505/705	5	505/711
	506/706	6	506/708
	507/707	7	507/709
	508/709	8	508/708
Set 2	501/702	9	501/711
	502/703	10	502/709
	503/704	11	503/708
	504/705	12	504/710
	505/706	13	505/708
	506/707	14	506/710
	507/708	15	507/710
	508/710	16	501/712
Set 3	501/703	17	508/711
	502/704	18	508/708
	503/705		
	504/706	19	501/710
	505/707	20	502/701
	506/709	21	503/702
	507/711	22	504/709
	508/712	23	505/711
		24	506/708
Set 4	501/704	25	**507**/709
	502/705	26	508/708
	503/706		
	504/707	27	501/711
	505/710	28	502/709
	506/711	29	503/708
	507/712	30	504/710
	508/701	31	505/708
		32	506/710
Set 5	501/705	33	507/710
	502/706	34	501/712
	503/707		
	504/708	35	508/711
	505/709	36	508/708
	506/712	37	501/710
	507/701	38	502/701
	508/702	39	503/702
		40	504/709
Set 6	501/706	41	505/711
	502/707	42	506/708
	503/710	43	507/709
	504/711	44	508/708
	505/712		
	506/701	45	501/711
	507/702	46	502/709
	508/703	47	503/708
		48	504/710
Set 7	501/707	49	505/708
	502/708	50	506/710
	503/709	51	507/710
	504/712	52	501/712
	505/701		
	506/702	53	508/711
	507/703	54	508/708
	508/704	55	501/710
		56	502/701
Set 8	501/708	57	503/702
	502/710	58	504/709
	503/711	59	505/711
	504/701	60	506/708
	505/702	61	507/709
	506/703	62	508/708
	507/704		
	508/705	63	501/711
		64	502/709
Set 9	501/709	65	503/708
	502/711	66	504/710
	503/712	67	505/708
	504/702	68	506/710
	505/703	69	507/710
	506/704	70	501/712
	507/705		
	508/706	71	508/711
		72	508/708

Appendix K Read depth coverage data from Run 1

Sample ID	1115	1133	1144	1146	1147	1150	1156	1169	1179	1181	1184	1191	1210	1213	1215	1219	428
total_reads	1631 239	1638 405	1515 880	2250 243	1985 508	1497 131	1534 962	1631 212	2128 291	1933 075	2119 793	1703 035	2106 098	1558 600	2020 966	1293 103	1808 275
mapped_to_target_reads	1593 448	1604 318	1486 301	2200 957	1945 269	1463 231	1501 029	1593 468	2080 547	1894 357	2058 132	1666 055	2061 095	1523 064	1965 041	1264 567	1767 939
percentage	97.68	97.92	98.05	97.81	97.97	97.74	97.79	97.69	97.76	98	97.09	97.83	97.86	97.72	97.23	97.79	97.77
mapped_to_target_reads_plus_150bp	1622 507	1634 289	1514 103	2238 037	1980 532	1490 129	1527 146	1620 760	2115 610	1927 503	2088 352	1693 485	2096 709	1549 835	1998 267	1286 653	1799 011
percentage	99.46	99.75	99.88	99.46	99.75	99.53	99.49	99.36	99.4	99.71	98.52	99.44	99.55	99.44	98.88	99.5	99.49
mean_coverage	1485. 36	1501. 98	1393. 03	2057. 61	1820. 05	1369. 69	1404. 03	1492. 43	1940. 15	1771. 88	1924. 96	1557. 84	1922. 58	1420. 5	1832. 78	1187. 73	1646. 97
accessible_target_bases	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87	1294 87
accessible_target_bases_1x	1290 79	1290 80	1290 80	1292 81	1292 80	1292 80	1292 80	1292 80	1290 81	1292 80	1292 80	1292 79	1293 76	1290 80	1292 80	1290 79	1290 80
percentage	99.68	99.68	99.68	99.84	99.84	99.84	99.84	99.84	99.68	99.84	99.84	99.84	99.91	99.68	99.84	99.68	99.68
accessible_target_bases_5x	1289 92	1290 78	1285 29	1289 40	1283 32	1290 79	1288 80	1289 18	1285 05	1290 79	1290 79	1287 20	1289 27	1285 29	1292 80	1282 38	1281 64
percentage	99.61	99.68	99.26	99.57	99.1	99.68	99.53	99.56	99.24	99.68	99.68	99.4	99.56	99.26	99.84	99.03	98.97
accessible_target_bases_10x	1279 77	1280 99	1281 44	1284 07	1276 24	1287 51	1283 30	1287 89	1275 88	1283 55	1283 60	1277 15	1278 75	1279 70	1285 21	1280 15	1274 48
percentage	98.83	98.92	98.96	99.16	98.56	99.43	99.1	99.46	98.53	99.12	99.12	98.63	98.75	98.83	99.25	98.86	98.42
target_bases_20x	1270 92	1274 10	1274 75	1276 10	1274 97	1274 22	1273 48	1272 70	1274 48	1274 47	1272 70	1272 88	1272 40	1275 25	1277 88	1274 87	1272 68
percentage	98.15	98.39	98.44	98.54	98.46	98.4	98.34	98.28	98.42	98.42	98.28	98.3	98.26	98.48	98.68	98.45	98.28
target_bases_30x	1270 91	1271 47	1271 47	1272 70	1269 87	1272 97	1270 33	1265 55	1269 90	1270 44	1272 70	1269 87	1271 48	1272 69	1272 70	1272 42	1268 54
percentage	98.15	98.19	98.19	98.28	98.06	98.31	98.1	97.73	98.07	98.11	98.28	98.07	98.19	98.28	98.28	98.26	97.96
target_bases_50x	1264 33	1268 64	1264 52	1266 50	1267 21	1262 67	1264 91	1261 07	1265 73	1264 81	1271 48	1265 16	1267 61	1266 38	1269 87	1265 55	1260 26
percentage	97.64	97.97	97.65	97.8	97.86	97.51	97.68	97.39	97.74	97.67	98.19	97.7	97.89	97.8	98.06	97.73	97.32
target_bases_100x	1255 51	1252 44	1251 14	1254 79	1255 18	1254 13	1258 33	1246 61	1259 96	1257 69	1258 27	1248 86	1255 61	1257 56	1260 15	1249 94	1250 02
percentage	96.96	96.72	96.62	96.9	96.93	96.85	97.17	96.27	97.3	97.12	97.17	96.44	96.96	97.12	97.31	96.53	96.53
target_bases_500x	1092 43	1105 88	1076 68	1152 39	1142 69	1071 07	1085 74	1077 16	1131 53	1133 44	1140 34	1078 74	1151 05	1095 27	1141 70	1041 97	1099 31
percentage	84.36	85.4	83.15	88.99	88.24	82.71	83.85	83.18	87.38	87.53	88.06	83.3	88.89	84.58	88.17	80.47	84.89

target_bases_1000x	125551	125244	125114	125479	125518	125413	125833	124661	125996	125769	125827	124886	125561	125756	126015	124994	125002
percentage	64.63	65.09	60.15	75.59	72.61	60.2	61.29	62.43	73.45	73.1	71.99	65.12	73.38	62.41	71.86	51.7	66.87

Appendix L Read depth data from run 2

Sample ID	343G	346B	348	489E	495	498	501	504A	504B	528B	530C	531D	542	551C	558B	573C	577
total_reads	1897659	1936946	2085652	1918497	1949986	2230491	2049971	2450329	2235135	1908725	2019385	1891152	1969963	2406514	2028315	2357089	1991205
mapped_to_target_reads	14695453	14478408	16218286	14794643	15139281	17249630	16095037	19104117	17664577	14278842	15523228	14871589	15239951	18455006	16052975	18760703	15336054
percentage	774.4	747.49	777.61	771.16	776.38	773.36	785.13	779.66	790.31	748.08	768.71	786.38	773.62	766.88	791.44	795.93	770.19
mapped_to_target_reads_plus_150bp	15202516	15005469	16789059	15350165	15659497	17885972	16636291	19768440	18265299	14818108	16122103	15379958	15791426	19107972	16627517	19414197	15887279
percentage	801.12	774.7	804.98	800.11	803.06	801.88	811.54	806.77	817.19	776.34	798.37	813.26	801.61	794.01	819.77	823.65	797.87
mean_coverage	2910.98	2859.34	3215.51	2929.9	2998.87	3417.78	3196.99	3786.07	3512.29	2821.35	3074.43	2951.16	3011.98	3655.7	3187.14	3727.29	3036.05
accessible_target_bases	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180
accessible_target_bases_1x	670600	669993	670367	669941	670165	669986	670280	669948	669919	671366	671791	671790	669917	669949	669914	669914	670225
percentage	96.46	96.38	96.43	96.37	96.4	96.37	96.42	96.37	96.36	96.57	96.63	96.63	96.36	96.37	96.36	96.36	96.41
accessible_target_bases_5x	669796	669873	669888	669873	669897	669873	669886	669873	669873	669873	669897	669787	669772	669874	669792	669893	669873
percentage	96.35	96.36	96.36	96.36	96.36	96.36	96.36	96.36	96.36	96.36	96.36	96.35	96.34	96.36	96.35	96.36	96.36
accessible_target_bases_10x	669491	669682	668960	669772	669688	669357	669702	669772	669661	669481	669362	668669	669697	669873	669274	669339	669408
percentage	96.3	96.33	96.23	96.34	96.33	96.28	96.33	96.34	96.33	96.3	96.28	96.18	96.33	96.36	96.27	96.28	96.29
target_bases_20x	667426	667953	666997	668024	667542	667708	668004	669488	667789	668168	668223	666433	668625	669540	667953	667080	666871
percentage	96.01	96.08	95.94	96.09	96.02	96.05	96.09	96.3	96.06	96.11	96.12	95.86	96.18	96.31	96.08	95.96	95.93
target_bases_30x	666037	666029	666512	666317	667151	666148	667685	667138	665975	665733	666748	665360	667828	668371	666427	666231	665400
percentage	95.8	95.8	95.8	95.8	95.9	95.8	96.0	95.9	95.8	95.7	95.9	95.7	96.0	96.1	95.8	95.8	95.7

	1	1	7	5	7	2	4	6		6	1	1	6	4	6	3	2
target_bases_50x	664248	665324	665360	666108	664397	663403	665286	665442	663709	665234	663400	664384	665391	667779	664286	665388	664429
percentage	95.55	95.7	95.71	95.82	95.57	95.43	95.7	95.72	95.47	95.69	95.43	95.57	95.71	96.06	95.55	95.71	95.58
target_bases_100x	662280	663033	662522	662495	662878	662329	662458	663281	662054	661858	661605	662093	662372	663226	662041	663198	662769
percentage	95.27	95.37	95.3	95.3	95.35	95.27	95.29	95.41	95.23	95.21	95.17	95.24	95.28	95.4	95.23	95.4	95.34
target_bases_500x	632345	636498	640867	634829	638642	636708	638975	645341	639251	633593	635148	633505	631319	645522	636025	642126	634356
percentage	90.96	91.56	92.19	91.32	91.87	91.59	91.91	92.83	91.95	91.14	91.36	91.13	90.81	92.86	91.49	92.37	91.25
target_bases_1000x	662280	663033	662522	662495	662878	662329	662458	663281	662054	661858	661605	662093	662372	663226	662041	663198	662769
percentage	82.15	82.85	84.57	82.46	83.98	84.52	83.96	87.23	85.02	82.63	83.09	81.99	82.21	87.6	83.52	86.25	82.69

Appendix M . Run 3 total read depth coverage

Sample ID	575	580	585	588	595	597	598	601B	604	607	610C	619	621	632	644	662	664
total_reads	2215392	2449418	2205888	2125463	2306753	1954889	2467422	3146646	1761733	1734451	2160428	1756356	1778568	2168993	2017204	2178756	1813864
mapped_to_target_reads	17385799	18927290	16645145	16357753	17553965	15085984	19727790	25006420	14048317	13689548	16835229	13175033	13816495	17729481	15546446	17318812	14606763
percentage	784.77	772.73	754.58	769.61	760.98	771.71	799.53	794.7	797.41	789.27	779.25	750.13	776.83	817.41	770.69	794.89	805.28
mapped_to_target_reads_plus_150bp	17969220	19641590	17233463	16955852	18164301	15660026	20418707	25854115	14564809	14251449	17466294	13658794	14294007	18338435	16082973	17899277	15153819
percentage	811.11	801.89	781.25	797.75	787.44	801.07	827.53	821.64	826.73	821.67	808.46	777.68	803.68	845.48	797.29	821.54	835.44
mean_coverage	3457.52	3742.98	3294.69	3237.44	3473.53	2988.25	3915.65	4976.51	2783.52	2705.67	3332.98	2597.14	2740.94	3526.49	3076.48	3443.9	2899.36
accessible_target_bases	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180
accessible_target_bases_1x	669982	669955	669966	670439	669962	672234	669984	672008	669973	669943	669952	669934	670424	670501	670000	669966	669923
percentage	96.37	96.37	96.37	96.44	96.37	96.7	96.37	96.66	96.37	96.37	96.37	96.37	96.44	96.45	96.38	96.37	96.37
accessible_target_bases_5x	669876	669873	669899	669897	669873	669647	669815	669775	669633	669733	669873	669853	669873	669873	669873	669905	669873
percentage	96.36	96.36	96.36	96.36	96.36	96.33	96.35	96.34	96.32	96.34	96.36	96.36	96.36	96.36	96.36	96.36	96.36
accessible_target_bases_10x	669873	669803	669342	669706	669702	669067	669555	669159	668387	669064	669804	669031	669059	669873	669512	669493	669600
percentage	96.36	96.35	96.28	96.33	96.33	96.24	96.31	96.25	96.14	96.24	96.35	96.24	96.24	96.36	96.31	96.3	96.32
target_bases_20x	668997	668825	668424	668870	669188	667527	667811	666528	666458	666786	668780	667147	667356	667971	667519	667094	668326
percentage	96.23	96.21	96.15	96.21	96.26	96.02	96.06	95.88	95.87	95.91	96.2	95.97	96	96.08	96.02	95.96	96.14
target_bases_30x	668024	667555	667554	667784	667551	667243	665323	665755	665422	664421	668021	666399	667130	666099	665317	665661	667525

percentage	96.09	96.02	96.02	96.06	96.02	95.98	95.7	95.77	95.72	95.57	96.09	95.86	95.96	95.81	95.7	95.75	96.02
target_bases_50x	665630	665363	665360	664384	666396	664851	665286	663763	664768	662982	664652	663211	662709	665422	665250	664384	665596
percentage	95.75	95.71	95.71	95.57	95.86	95.64	95.7	95.48	95.62	95.37	95.61	95.4	95.33	95.72	95.69	95.57	95.74
target_bases_100x	662701	663094	663293	662154	662533	662500	663051	662042	661145	659928	661770	661989	661027	663069	663123	662306	662440
percentage	95.33	95.38	95.41	95.25	95.3	95.3	95.38	95.23	95.1	94.93	95.19	95.22	95.09	95.38	95.39	95.27	95.29
target_bases_500x	642042	642551	641066	635163	645971	635332	638742	645921	615268	613313	638746	624622	627447	640548	637299	639278	632644
percentage	92.35	92.43	92.21	91.37	92.92	91.39	91.88	92.91	88.5	88.22	91.88	89.85	90.26	92.14	91.67	91.96	91
target_bases_1000x	662701	663094	663293	662154	662533	662500	663051	662042	661145	659928	661770	661989	661027	663069	663123	662306	662440
percentage	85.45	86.65	85.22	83.18	86.24	82.58	85.06	88.35	76.14	77.13	84.88	77.63	80.79	84.68	82.74	84.86	81.93

Appendix N Total read depth coverage data from TruSeq run 4

Sample ID	523D	629B	635	637	654	661	665A	668	670	674	675A	684	695	697	704	709	727
total_reads	1957533	1726578	2722421	2563729	1823915	2197376	1838187	2093148	1755091	1961161	2629759	2217494	2358212	1783984	829411	2047803	2507271
mapped_to_target_reads	15339207	13181046	20865654	19672184	14171464	16802440	14235989	16422539	13675429	15069760	20508364	17403710	18800156	13946567	6517073	15669296	19805858
percentage	783.6	763.42	766.44	767.33	776.98	764.66	774.46	784.59	779.19	768.41	779.86	784.84	797.22	781.77	785.75	765.18	789.94
mapped_to_target_reads_plus_150bp	15860020	13648391	21606595	20354561	14679719	17376858	14736193	16975739	14195740	15618086	21210118	18009137	19446156	14467470	6738316	16213862	20488195
percentage	810.2	790.49	793.65	793.94	804.85	790.8	801.67	811.01	808.83	796.37	806.54	812.14	824.61	810.96	812.42	791.77	817.15
mean_coverage	3044.31	2613.99	4143.38	3903.08	2813.74	3335.55	2823.59	3263.58	2715.73	2989.95	4063.87	3463.97	3731.9	2773.41	1296.08	3108.29	3928.85
accessible_target_bases	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180	695180
accessible_target_bases_1x	669942	670437	669953	672000	669976	670393	669980	670415	669920	669972	669965	671951	669963	670083	669967	669910	669980
percentage	96.37	96.44	96.37	96.66	96.37	96.43	96.37	96.44	96.37	96.37	96.37	96.66	96.37	96.39	96.37	96.36	96.37
accessible_target_bases_5x	669772	669767	669873	669772	669806	669873	669796	669876	669873	669873	669902	669876	669873	669768	669726	669873	669873
percentage	96.34	96.34	96.36	96.34	96.35	96.36	96.35	96.36	96.36	96.36	96.36	96.36	96.36	96.34	96.34	96.36	96.36
accessible_target_bases_10x	669488	669641	669772	669697	669702	669563	669721	669513	669702	669772	669714	669702	669505	669538	667886	669596	669713
percentage	96.3	96.32	96.34	96.33	96.33	96.31	96.34	96.31	96.33	96.34	96.33	96.33	96.31	96.31	96.07	96.32	96.34
target_bases_20x	667218	667932	668674	669267	669112	667218	668310	667906	669273	668594	668996	668708	667065	667596	665268	667606	668976
percentage	95.98	96.08	96.19	96.27	96.25	95.98	96.13	96.08	96.27	96.17	96.23	96.19	95.95	96.03	95.7	96.03	96.23

target_bases_30x	665740	666800	666829	667696	667592	666103	667910	666541	667439	668013	667821	666929	665630	666495	664465	666765	666717
percentage	95.76	95.92	95.92	96.04	96.03	95.82	96.08	95.88	96.01	96.09	96.06	95.93	95.75	95.87	95.58	95.91	95.9
target_bases_50x	665232	665325	665352	666467	665346	665423	665388	664823	664714	664528	665642	666467	665360	665363	662698	665328	666396
percentage	95.69	95.7	95.71	95.87	95.71	95.72	95.71	95.63	95.62	95.59	95.75	95.87	95.71	95.71	95.33	95.7	95.86
target_bases_100x	662929	662874	663966	664172	663389	663078	662378	663150	663264	662167	662915	663503	662488	662878	655321	662981	663033
percentage	95.36	95.35	95.51	95.54	95.43	95.38	95.28	95.39	95.41	95.25	95.36	95.44	95.3	95.35	94.27	95.37	95.37
target_bases_500x	635098	630449	648318	646386	633529	637353	637744	638407	629424	636486	643247	638689	635059	628713	545330	637097	644572
percentage	91.36	90.69	93.26	92.98	91.13	91.68	91.74	91.83	90.54	91.56	92.53	91.87	91.35	90.44	78.44	91.64	92.72
target_bases_1000x	662929	662874	663966	664172	663389	663078	662378	663150	663264	662167	662915	663503	662488	662878	655321	662981	663033
percentage	82.53	79.97	87.86	87.46	81.99	84.46	83.01	84.73	80.58	83.73	86.83	84.7	84	80.64	54.16	83.45	87.41

Appendix O Total read coverage for the 48 sample TruSeq run

Sample ID	1	2	3	4	5	6	7	8	9	10
total_reads	650616	610095	27474	681537	791617	612697	562280	699651	644154	693329
mapped_to_target_reads	1506899	1370710	61446	1647905	1841039	1364077	1312749	1655257	1516716	1651221
percentage	231.61	224.67	223.65	241.79	232.57	222.63	233.47	236.58	235.46	238.16
mapped_to_target_reads_plus_150bp	1614364	1475202	66165	1769466	1978464	1468726	1408848	1782532	1634576	1779492
percentage	248.13	241.8	240.83	259.63	249.93	239.71	250.56	254.77	253.76	256.66
mean_coverage	2225.18	1997.29	89.21	2481.59	2736.46	1996.36	1945.99	2463.25	2258.62	2462.35
accessible_target_bases	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
accessible_target_bases_1x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
accessible_target_bases_5x	80404	80404	79399	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	98.75	100	100	100	100	100	100	100
accessible_target_bases_10x	80404	80404	78282	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	97.36	100	100	100	100	100	100	100
target_bases_20x	80404	80404	72644	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	90.35	100	100	100	100	100	100	100
target_bases_30x	80390	80404	67178	80386	80404	80404	80404	80404	80404	80404
percentage	99.98	100	83.55	99.98	100	100	100	100	100	100

target_bases_50x	80386	80068	53777	80386	80386	80386	80386	80386	80404	80386
percentage	99.98	99.58	66.88	99.98	99.98	99.98	99.98	99.98	100	99.98
target_bases_100x	79680	79680	27058	79680	79680	79680	79615	79680	79770	79680
percentage	99.1	99.1	33.65	99.1	99.1	99.1	99.02	99.1	99.21	99.1
target_bases_500x	73071	71296		74706	75293	70876	72386	74136	74469	74942
percentage	90.88	88.67	0	92.91	93.64	88.15	90.03	92.2	92.62	93.21
target_bases_1000x	79680	79680	27058	79680	79680	79680	79615	79680	79770	79680
percentage	77.61	72.25	0	83.37	84.36	70.99	74.16	81.48	81.23	82.95
Sample ID	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
total_reads	565492	714486	873712	708168	795554	742955	598469	894291	739434	814817
mapped_to_target_reads	1320005	1678844	2024338	1698191	1861591	1747244	1388337	2077205	1702548	1884641
percentage	233.43	234.97	231.69	239.8	234	235.17	231.98	232.27	230.25	231.3
mapped_to_target_reads_plus_150bp	1420097	1810839	2176213	1820912	2004020	1883721	1495096	2232797	1835321	2031829
percentage	251.13	253.45	249.08	257.13	251.9	253.54	249.82	249.67	248.21	249.36
mean_coverage	1957.9	2489.01	2991.28	2549.54	2761.28	2592.28	2056.5	3077.77	2512.62	2780.94
accessible_target_bases	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
accessible_target_bases_1x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
accessible_target_bases_5x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100

accessible_target_bases_10x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
target_bases_20x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
target_bases_30x	80404	80404	80404	80404	80404	80386	80404	80404	80404	80404
percentage	100	100	100	100	100	99.98	100	100	100	100
target_bases_50x	80064	80404	80404	80386	80386	80064	80064	80064	80404	80082
percentage	99.58	100	100	99.98	99.98	99.58	99.58	99.58	100	99.6
target_bases_100x	79680	79770	79770	79680	79680	79680	79264	79680	79784	79770
percentage	99.1	99.21	99.21	99.1	99.1	99.1	98.58	99.1	99.23	99.21
target_bases_500x	72940	74262	76551	74162	75544	74799	72785	75456	73888	74251
percentage	90.72	92.36	95.21	92.24	93.96	93.03	90.52	93.85	91.9	92.35
target_bases_1000x	79680	79770	79770	79680	79680	79680	79264	79680	79784	79770
percentage	78.18	82.09	86.46	82.31	83.99	83.31	77.33	86.9	81.68	84.13
Sample ID	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
total_reads	721982	648144	717334	681790	631327	839920	795560	949840	877588	669713
mapped_to_target_reads	1685439	1503743	1688149	1618306	1459645	1915851	1850538	2226759	2032301	1547874
percentage	233.45	232.01	235.34	237.36	231.2	228.1	232.61	234.44	231.58	231.12
mapped_to_target_reads_plus_150bp	1808138	1616947	1816034	1741661	1568642	2063035	1995784	2397665	2190791	1665521
percentage	250.44	249.47	253.16	255.45	248.47	245.62	250.87	252.43	249.64	248.69

mean_coverage	2496.37	2230.71	2505.86	2411.84	2157.27	2816.07	2736.65	3295.86	3005.56	2289.09
accessible_target_bases	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
accessible_target_bases_1x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
accessible_target_bases_5x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
accessible_target_bases_10x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
target_bases_20x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
target_bases_30x	80404	80404	80404	80404	80386	80404	80404	80404	80404	80404
percentage	100	100	100	100	99.98	100	100	100	100	100
target_bases_50x	80082	80064	80404	79788	80296	80404	80404	80404	80404	80404
percentage	99.6	99.58	100	99.23	99.87	100	100	100	100	100
target_bases_100x	79770	79309	79770	79399	79634	79770	79770	79770	79770	79770
percentage	99.21	98.64	99.21	98.75	99.04	99.21	99.21	99.21	99.21	99.21
target_bases_500x	74949	72808	75532	74691	72927	74201	75300	76093	75199	73363
percentage	93.22	90.55	93.94	92.89	90.7	92.29	93.65	94.64	93.53	91.24
target_bases_1000x	79770	79309	79770	79399	79634	79770	79770	79770	79770	79770
percentage	82.9	79.24	83.43	80.88	77.85	83	84.85	87.09	86.35	79.8

Sample ID	31	32	33	34	35	36	37	38	39	40
total_reads	822208	689122	679507	718358	965077	1447636	1190183	741502	781658	762135
mapped_to_target_reads	1905416	1547257	1558470	1631151	2338503	3208892	2805705	1804342	1826329	1680567
percentage	231.74	224.53	229.35	227.07	242.31	221.66	235.74	243.34	233.65	220.51
mapped_to_target_reads_plus_150bp	2049732	1659604	1674300	1755369	2507786	3448196	3017657	1935035	1965022	1810224
percentage	249.3	240.83	246.4	244.36	259.85	238.19	253.55	260.96	251.39	237.52
mean_coverage	2820.35	2267.52	2308.24	2388.76	3529.95	4679.94	4177.96	2738.18	2712.04	2455.58
accessible_target_bases	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
accessible_target_bases_1x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
accessible_target_bases_5x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
accessible_target_bases_10x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
target_bases_20x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
target_bases_30x	80404	80404	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100	100	100
target_bases_50x	80404	79770	80386	80390	80404	80404	80404	80386	80386	80404
percentage	100	99.21	99.98	99.98	100	100	100	99.98	99.98	100

target_bases_100x	79770	79309	79680	79680	80386	80386	80386	80296	79680	79680
percentage	99.21	98.64	99.1	99.1	99.98	99.98	99.98	99.87	99.1	99.1
target_bases_500x	76085	72142	74109	72758	77517	77205	78170	77161	75093	72228
percentage	94.63	89.72	92.17	90.49	96.41	96.02	97.22	95.97	93.39	89.83
target_bases_1000x	79770	79309	79680	79680	80386	80386	80386	80296	79680	79680
percentage	85.3	77.15	80.92	78	90.16	90.33	90.37	86.65	83.38	74.87

Sample ID	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**
total_reads	707192	688854	893657	971544	804952	962895	690277	849596
mapped_to_target_reads	1633213	1648783	2090001	2294415	1910766	2261903	1604961	1991961
percentage	230.94	239.35	233.87	236.16	237.38	234.91	232.51	234.46
mapped_to_target_reads_plus_150bp	1755474	1772065	2248332	2466922	2055900	2427282	1726491	2142751
percentage	248.23	257.25	251.59	253.92	255.41	252.08	250.12	252.21
mean_coverage	2421.85	2459.1	3102.44	3414.56	2847.5	3362.63	2381.63	2958.5
accessible_target_bases	80404	80404	80404	80404	80404	80404	80404	80404
accessible_target_bases_1x	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100
accessible_target_bases_5x	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100
accessible_target_bases_10x	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100

target_bases_20x	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100
target_bases_30x	80404	80404	80404	80404	80404	80404	80404	80404
percentage	100	100	100	100	100	100	100	100
target_bases_50x	80386	80386	80404	80404	80404	80404	80404	80404
percentage	99.98	99.98	100	100	100	100	100	100
target_bases_100x	79680	79680	79770	80064	80386	80386	79705	79770
percentage	99.1	99.1	99.21	99.58	99.98	99.98	99.13	99.21
target_bases_500x	72743	73266	74942	76461	76224	76766	73021	74807
percentage	90.47	91.12	93.21	95.1	94.8	95.48	90.82	93.04
target_bases_1000x	79680	79680	79770	80064	80386	80386	79705	79770
percentage	78.62	79.34	85.68	86.88	85.25	87.77	79.32	84.29

Appendix P Full list of Southampton patients sequenced using the 25 gene TruSeq custom panel

Sample Id	sample	diagnosis	IGHV
104	17/12/1998	CLL	M
108	09/02/1999	CLL	M
110	03/02/1999	CLL	U
122	12/02/1999	CLL	M
130	04/03/1999	CLL	M
135	09/03/1999	CLL	0
273	14/04/2001	CLL	M
276	25/08/2001	CLL	M
281	28/06/2001	CLL	M
322	22/11/2001		0
351	24/05/2002	CLL	M
353	11/05/2002	CLL	M
368	31/10/2002	CLL	M
391	09/10/2003	CLL	M
424	30/11/2004	CLL	U
428	02/03/2005		U
452	01/06/2006	CLL	M
453	01/06/2006	CLL	M
455	13/06/2006	CLL	U
456	15/06/2006	CLL	M
458	22/06/2006	CLL	M
459	22/06/2006	CLL	U
460	29/06/2006	CLL	U
461	14/07/2006	CLL	M(IgG)
483	21/12/2006	CLL	M
484	04/01/2007	CLL	0
493	08/02/2007	CLL	M
498	12/04/2007	CLL	M
500	26/04/2007	CLL	U
507	31/5/11		0
520	09/08/2007	CLL	U
525	30/08/2007	CLL	M
529	20/09/2007	CLL	M
541	01/11/2007	CLL	M
542	08/11/2007	CLL	M
548	20/12/2007		0
549	03/01/2008	CLL	M
559	21/02/2008	CLL	M
561	06/03/2008	CLL	M

564	27/03/2008	CLL	M
565	16/04/2008	CLL	U
568	24/04/2008	CLL	M
570	08/05/2008	CLL	M
571	15/05/2008	CLL	M
578	12/06/2008	CLL	M
581	19/06/2008	CLL	M
583	04/07/2008	CLL	U
585	10/07/2008	CLL	U
586	24/07/2008	NHL/other	M
587	25/09/2008	CLL	M
590	09/09/2008	CLL	M
592	09/10/2008	CLL	U
594	23/10/2008	CLL	U
597	21/11/12	CLL	M
600	11/05/2008	CLL	M
602	15/01/2009	CLL	U
607	14/03/13	CLL	U
635	24/09/2009	CLL	U
637	22/10/2009	CLL	M
654	26/11/2009	CLL	0
661	14/01/2010	CLL	M
668	09/02/2010	CLL	U
670	16/03/2010	CLL	U
684	15/06/2010	CLL	M
695	03/08/2010	CLL	U
697	10/08/2010	CLL	M
704	24/09/2010	CLL	U
709	19/10/2010	CLL	M
621A	#N/A	#N/A	#N/A
662	#N/A	#N/A	#N/A
163B	08/05/2008	CLL	U
343G	22/03/2011	CLL	U
346B	01/06/2010	CLL	U
348D	04/01/2011	CLL	0
389A	29/11/2007	CLL	0
465A	06/12/2007	CLL	M
468B	28/11/12	CLL	0
482A	24/07/2008	CLL	M
488A	06/03/2008	CLL	M
489E	04/10/2011	CLL	0
495D	21/09/2010		0
501B	12/11/2009	CLL	U
504A	03/09/2008	CLL	U
504B	28/05/2009	CLL	0
523D	12/07/2011	CLL	0

528B	06/11/2008	CLL	0
530C	01/06/2010	CLL	0
531A	23/10/2008	CLL	0
531D	05/10/2010	CLL	0
551C	11/05/2010	CLL	0
558B	22/06/2010	CLL	0
573C	12/03/2011	CLL	0
575B	16/03/2010	CLL	0
577A	16/02/2010	CLL	0
580A	01/10/2009	CLL	0
588A	29/10/2009	CLL	0
595D	30/11/2010	CLL	0
598A	10/11/2009	CLL	M
601B	01/06/2010	CLL	U
604A	16/10/2009	CLL	0
610C	01/02/2011	CLL	0
619A	11/05/2010	CLL	0
629B	07/06/2011	CLL	0
632D	21/09/2010	CLL	0
644B	27/04/2010	CLL	0
664A	13/04/2010	CLL	0
665A	07/06/2011	CLL	0
674C	08/03/2011	CLL	0
675A	07/05/2010		0
727A	04/05/2011	CLL	0

Appendix Q The average (ERIC gene panel) amplicons coverage, sorted from poorest amplicon coverage to the best.

Amplicon	Average coverage per amplicon
EGR2_Cds_16702979_UserDefined (36114810)_62975106.1	96
ATM_Cds_16735823_UserDefined (31325648)_62975037.1	98
FBXW7_Cds_17001128_UserDefined (31327697)_62975103.1	177
BIRC3 + BIRC3_UserDefined (36114777)_62975237.2	229
NFKBIE_Cds_17033258_UserDefined (35731633)_62975097.1	378
EGR2_Cds_16702979_UserDefined (36114810)_62975107.1	397
NFKBIE_Cds_17033258_UserDefined (35731633)_62975094.1	427
XPO1 + XPO1_UserDefined (35731628)_62975212.1	452
NFKBIE_Cds_17033258_UserDefined (35731633)_62975093.1	473
NOTCH1_Cds_17106080_UserDefined (36114819)_62975076.1	486
BIRC3_Cds_16737250_UserDefined (36114808)_62975128.1	574
MYD88 + MYD88 + MYD88_UserDefined (36114778)_62975211.1	596
FBXW7_Cds_17000028_UserDefined (31327696)_62975101.1	637
TP53_Cds_16839835_UserDefined (36114836)_62975055.1	645
ATM_Cds_16740128_UserDefined (35731666)_62975135.1	652
EGR2_Cds_16702979_UserDefined (36114810)_62975109.1	665
NFKBIE_Cds_17036882_UserDefined (35731635)_62975073.1	709
ATM_Cds_16735661_UserDefined (33620582)_62975159.1	844
ATM_Cds_16739330_UserDefined (35731663)_62975130.1	873

ATM_Cds_16740104_UserDefined (36114787)_62975180.1	874
POT1_Cds_17065330_UserDefined (36114823)_62975065.1	885
POT1_Cds_17064340_UserDefined (36114829)_62975056.1	1012
NOTCH1_Cds_17106080_UserDefined (36114819)_62975079.1	1016
EGR2_Cds_16702979_UserDefined (36114810)_62975110.1	1027
MYD88_Cds_16961557_UserDefined (35427239)_62975090.1	1040
ATM_Cds_16740159_UserDefined (36114805)_62975137.1	1042
NOTCH1_Cds_17106080_UserDefined (36114819)_62975078.1	1189
TP53 + TP53 + TP53_UserDefined (36114772)_62975045.1	1252
NOTCH1_Cds_17106080_UserDefined (36114819)_62975075.1	1276
POT1_Cds_17064340_UserDefined (36114829)_62975057.1	1361
NOTCH1_Cds_17106080_UserDefined (36114819)_62975082.1	1400
ATM_Cds_16734636_UserDefined (31325605)_62975190.1	1403
EGR2_Cds_16702979_UserDefined (36114810)_62975111.1	1506
ATM_Cds_16735639_UserDefined (36114792)_62975177.1	1579
TP53 + TP53 + TP53_UserDefined (36114772)_62975044.1	1603
ATM_Cds_16736827_UserDefined (36114797)_62975152.1	1604
FBXW7_Cds_16997296_UserDefined (36114811)_62975116.1	1624
ATM_Cds_16739664_UserDefined (36114786)_62975195.1	1629
BIRC3 + BIRC3_UserDefined (35731626)_62975206.1	1655
ATM_Cds_16734800_UserDefined (36114781)_62975197.1	1714
NFKBIE_Cds_17033258_UserDefined (35731633)_62975096.1	1747
NFKBIE + NFKBIE_UserDefined (36114779)_62975215.1	1786
EGR2_Cds_16706074_UserDefined (35731592)_62975113.1	1878

ATM + ATM_UserDefined (36114775)_62975225.1	1893
ATM_Cds_16736017_UserDefined (31325635)_62975187.1	1949
ATM_Cds_16738611_UserDefined (36114801)_62975143.1	1949
ATM_Cds_16740128_UserDefined (35731666)_62975136.1	1975
ATM_Cds_16740292_UserDefined (35731669)_62975124.1	1983
NFKBIE + NFKBIE_UserDefined (36114779)_62975217.1	1992
ATM_Cds_16736093_UserDefined (36114782)_62975200.1	2038
SF3B1 + SF3B1_UserDefined (36114774)_62975050.1	2063
BIRC3 + BIRC3_UserDefined (36114777)_62975238.1	2090
ATM_Cds_16736794_UserDefined (36114796)_62975149.1	2091
ATM_Cds_16738115_UserDefined (36114798)_62975153.1	2119
SF3B1_Cds_16918020_UserDefined (36114834)_62975064.1	2170
XPO1 + XPO1_UserDefined (35731628)_62975214.1	2214
ATM_Cds_16740190_UserDefined (32871199)_62975121.1	2215
ATM_Cds_16736794_UserDefined (36114796)_62975150.1	2236
POT1_Cds_17067181_UserDefined (36114826)_62975068.1	2255
MYD88_Cds_16964799_UserDefined (36114818)_62975091.1	2333
ATM_Cds_16738115_UserDefined (36114798)_62975153.2	2354
POT1_Cds_17064991_UserDefined (36114830)_62975058.1	2449
NOTCH1_Cds_17106080_UserDefined (36114819)_62975074.1	2471
BIRC3_Cds_16737142_UserDefined (36114807)_62975127.1	2474
NFKBIE_Cds_17033258_UserDefined (35731633)_62975095.1	2496
FBXW7_Cds_16998341_UserDefined (36114813)_62975119.1	2498
ATM_Cds_16738542_UserDefined (32871568)_62975141.1	2518

XPO1 + XPO1_UserDefined (35731628)_62975213.1	2684
NOTCH1_Cds_17106080_UserDefined (36114819)_62975081.1	2689
TP53 + TP53_UserDefined (33879790)_62975040.1	2704
EGR2_Cds_16702979_UserDefined (36114810)_62975108.1	2739
FBXW7_Cds_16998753_UserDefined (31327863)_62975098.1	2774
ATM_Cds_16735823_UserDefined (31325648)_62975036.1	2807
ATM_Cds_16735262_UserDefined (31325602)_62975174.1	2851
ATM_Cds_16738329_UserDefined (35731658)_62975138.1	2910
ATM_Cds_16739133_UserDefined (36114802)_62975147.1	2912
ATM_Cds_16738733_UserDefined (35731661)_62975146.1	2987
FBXW7_Cds_17001128_UserDefined (31327697)_62975104.1	2999
ATM + ATM_UserDefined (33677912)_62975233.1	3000
TP53 + TP53 + TP53_UserDefined (36114772)_62975043.1	3019
TP53 + TP53_UserDefined (36114773)_62975046.1	3083
ATM + ATM_UserDefined (36114775)_62975227.1	3087
ATM_Cds_16738210_UserDefined (33677890)_62975155.1	3100
MYD88 + MYD88 + MYD88_UserDefined (36114778)_62975210.1	3143
ATM_Cds_16734834_UserDefined (31325621)_62975193.1	3169
MYD88 + MYD88 + MYD88_UserDefined (36114778)_62975208.1	3179
TP53 + TP53_UserDefined (36114773)_62975047.1	3183
ATM_Cds_16734899_UserDefined (31325657)_62975199.1	3186
ATM_Cds_16736426_UserDefined (33879770)_62975168.1	3190
MYD88_Cds_16961557_UserDefined (35427239)_62975089.1	3213

ATM_Cds_16740139_UserDefined (31325650)_62975181.1	3250
ATM_Cds_16736162_UserDefined (36114784)_62975188.1	3267
MYD88 + MYD88 + MYD88_UserDefined (36114778)_62975209.1	3269
POT1 + POT1_UserDefined (36114780)_62975222.1	3293
ATM_Cds_16738733_UserDefined (35731661)_62975145.1	3337
SF3B1_Cds_16921391_UserDefined (31327238)_62975052.1	3361
MYD88 + MYD88 + MYD88_UserDefined (36114778)_62975207.1	3379
SF3B1_Cds_16921391_UserDefined (31327238)_62975051.1	3409
TP53_Cds_16835833_UserDefined (31325106)_62975054.1	3447
TP53_Cds_16837290_UserDefined (36114835)_62975053.1	3450
ATM_Cds_16735661_UserDefined (33620582)_62975160.1	3454
ATM_Cds_16736213_UserDefined (31325628)_62975202.1	3456
ATM_Cds_16735076_UserDefined (36114789)_62975185.1	3483
ATM_Cds_16736426_UserDefined (33879770)_62975167.1	3496
FBXW7 + FBXW7_UserDefined (32455541)_62975218.1	3506
NFKBIE_Cds_17036882_UserDefined (35731635)_62975072.1	3556
ATM_Cds_16735702_UserDefined (35731651)_62975165.1	3560
FBXW7_Cds_16997179_UserDefined (31327698)_62975115.1	3573
ATM_Cds_16738115_UserDefined (36114798)_62975154.1	3578
NFKBIE_Cds_17033316_UserDefined (35731632)_62975092.1	3580
ATM_Cds_16736501_UserDefined (36114795)_62975169.1	3593
POT1_Cds_17063851_UserDefined (36114828)_62975070.1	3597
NOTCH1_Cds_17106080_UserDefined (36114819)_62975080.1	3680

ATM_Cds_16734737_UserDefined (36114785)_62975192.1	3707
ATM_Cds_16738542_UserDefined (32871568)_62975142.1	3735
POT1_Cds_17062880_UserDefined (36114827)_62975069.1	3759
FBXW7 + FBXW7_UserDefined (32455541)_62975219.1	3762
ATM_Cds_16734800_UserDefined (36114781)_62975198.1	3788
ATM_Cds_16738329_UserDefined (35731658)_62975139.1	3842
POT1_Cds_17062453_UserDefined (36114820)_62975084.1	3856
FBXW7_Cds_16998341_UserDefined (36114813)_62975120.1	3862
POT1 + POT1_UserDefined (36114780)_62975221.1	3909
POT1_Cds_17065861_UserDefined (36114832)_62975060.1	3914
FBXW7_Cds_16997876_UserDefined (36114812)_62975118.1	3916
ATM_Cds_16740292_UserDefined (35731669)_62975125.1	3924
NOTCH1_Cds_17106080_UserDefined (36114819)_62975077.1	3937
BIRC3_Cds_16740300_UserDefined (36114771)_62975038.1	3949
ATM_Cds_16736017_UserDefined (31325635)_62975186.1	3963
TP53 + TP53_UserDefined (33879790)_62975041.1	3965
ATM + ATM_UserDefined (33619696)_62975223.1	3971
ATM_Cds_16735192_UserDefined (36114790)_62975172.1	3991
EGR2_Cds_16702979_UserDefined (36114810)_62975112.1	4007
ATM_Cds_16739903_UserDefined (31325626)_62975179.1	4083
ATM + ATM_UserDefined (33677912)_62975234.1	4087
ATM_Cds_16735270_UserDefined (36114791)_62975175.1	4087
ATM_Cds_16740238_UserDefined (36114806)_62975123.1	4090
ATM_Cds_16734834_UserDefined (31325621)_62975194.1	4154

BIRC3 + BIRC3_UserDefined (36114777)_62975241.1	4216
FBXW7 + FBXW7_UserDefined (32455541)_62975220.1	4275
FBXW7_Cds_17002044_UserDefined (36114817)_62975088.1	4277
ATM_Cds_16739903_UserDefined (31325626)_62975178.1	4389
BIRC3_Cds_16737142_UserDefined (36114807)_62975126.1	4430
ATM_Cds_16738242_UserDefined (36114799)_62975158.1	4443
TP53 + TP53_UserDefined (33879790)_62975042.1	4473
FBXW7_Cds_16999053_UserDefined (36114814)_62975099.1	4475
ATM_Cds_16734636_UserDefined (31325605)_62975189.1	4490
BIRC3_Cds_16739950_UserDefined (36114809)_62975129.1	4499
NOTCH1_Cds_17106080_UserDefined (36114819)_62975083.1	4617
POT1_Cds_17066741_UserDefined (36114833)_62975062.1	4635
ATM_Cds_16735666_UserDefined (36114793)_62975162.1	4654
ATM_Cds_16735407_UserDefined (33620581)_62975176.1	4663
NFKBIE + NFKBIE_UserDefined (36114779)_62975216.1	4671
ATM_Cds_16734642_UserDefined (31325649)_62975196.1	4689
ATM_Cds_16739649_UserDefined (36114803)_62975131.1	4694
ATM_Cds_16738611_UserDefined (36114801)_62975144.1	4745
FBXW7_Cds_17000856_UserDefined (36114815)_62975102.1	4823
SF3B1 + SF3B1_UserDefined (36114774)_62975049.1	4824
ATM + ATM_UserDefined (33619698)_62975228.1	4878
BIRC3 + BIRC3_UserDefined (35731626)_62975205.1	4890
FBXW7_Cds_17000028_UserDefined (31327696)_62975100.1	4946
ATM_Cds_16735262_UserDefined (31325602)_62975173.1	4992

FBXW7_Cds_16997876_UserDefined (36114812)_62975117.1	5011
ATM + ATM_UserDefined (36114775)_62975226.1	5030
NFKBIE_Cds_17034921_UserDefined (35731634)_62975071.1	5141
ATM_Cds_16740190_UserDefined (32871199)_62975122.1	5150
ATM_Cds_16735191_UserDefined (31325604)_62975171.1	5161
POT1_Cds_17066709_UserDefined (36114825)_62975067.1	5256
ATM_Cds_16737258_UserDefined (36114783)_62975203.1	5273
ATM + ATM_UserDefined (33619696)_62975224.1	5417
BIRC3 + BIRC3_UserDefined (36114777)_62975236.1	5430
ATM_Cds_16735702_UserDefined (35731651)_62975164.1	5487
ATM_Cds_16734737_UserDefined (36114785)_62975191.1	5495
FBXW7_Cds_17001547_UserDefined (31327699)_62975087.1	5527
BIRC3 + BIRC3_UserDefined (36114777)_62975235.1	5606
FBXW7_Cds_17001254_UserDefined (36114816)_62975105.1	5663
BIRC3 + BIRC3_UserDefined (36114777)_62975239.1	5704
BIRC3 + BIRC3_UserDefined (36114777)_62975237.1	5713
ATM_Cds_16735065_UserDefined (31325629)_62975184.1	5719
POT1_Cds_17064725_UserDefined (36114822)_62975086.1	5886
BIRC3 + BIRC3_UserDefined (36114777)_62975240.1	5918
ATM_Cds_16740176_UserDefined (36114788)_62975182.1	5989
ATM + ATM_UserDefined (36114776)_62975231.1	6203
SF3B1 + SF3B1_UserDefined (36114774)_62975048.1	6247
ATM_Cds_16738210_UserDefined (33677890)_62975156.1	6333
ATM + ATM_UserDefined (33619698)_62975229.1	6338

ATM_Cds_16736786_UserDefined (33879760)_62975148.1	6338
ATM_Cds_16736386_UserDefined (36114794)_62975166.1	6356
ATM_Cds_16740176_UserDefined (36114788)_62975183.1	6368
BIRC3_Cds_16740300_UserDefined (36114771)_62975039.1	6381
ATM + ATM_UserDefined (33677912)_62975232.1	6471
POT1_Cds_17065345_UserDefined (36114831)_62975059.1	6522
ATM_Cds_16735702_UserDefined (35731651)_62975163.1	6525
ATM + ATM_UserDefined (36114776)_62975230.1	6624
ATM_Cds_16739870_UserDefined (32875552)_62975132.1	6630
ATM_Cds_16736093_UserDefined (36114782)_62975201.1	6667
ATM_Cds_16739910_UserDefined (36114804)_62975133.1	6755
ATM_Cds_16738508_UserDefined (36114800)_62975140.1	6844
POT1_Cds_17066741_UserDefined (36114833)_62975061.1	7010
ATM_Cds_16735666_UserDefined (36114793)_62975161.1	7146
ATM_Cds_16739910_UserDefined (36114804)_62975134.1	7257
ATM_Cds_16736827_UserDefined (36114797)_62975151.1	7278
POT1_Cds_17065436_UserDefined (36114824)_62975066.1	7341
BIRC3 + BIRC3_UserDefined (35731626)_62975204.1	8047
POT1_Cds_17063140_UserDefined (36114821)_62975085.1	8505
SF3B1_Cds_16918020_UserDefined (36114834)_62975063.1	10168

Appendix R Variants chosen with a VAF more than 20% for Sanger sequencing validation of the TruSeq data

PID	Chr	location	WT	Alt	Variant	exon	protein	variant	gene	VAF
585	10	64573248	G	T	nonsynonymous SNV	3	p.H334N,EGR2	exonic	*EGR2*	31%
590	11	108186599	A	G	nonsynonymous SNV	exon41	p.Y2019C,	exonic	*ATM*	11.16%
598	11	108199938	T	C	nonsynonymous SNV	49	p.L2427P,	exonic	*ATM*	47%
619	10	64573332	C	T	nonsynonymous SNV	3	p.E306K,EGR2	exonic	*EGR2*	56%
635	11	108170471	G	T	nonsynonymous SNV	34	p.G1679V,	exonic	*ATM*	53%
644	11	108201132	T	G	nonsynonymous SNV	50	p.V2500G,	exonic	*ATM*	55%
670	19	1440421	T	G	nonsynonymous SNV	4	p.I133S,	exonic	*RPS15*	43%
674	20	35547857	C	T	nonsynonymous SNV	7	p.M254I,	exonic	*SAMHD1*	73%
695	11	108186598	T	A	nonsynonymous SNV	41	p.Y2019N,	exonic	*ATM*	47%
704	2	61719186	T	C	nonsynonymous SNV	16	p.D624G,	exonic	*XPO1*	49%
727	17	7578271	T	G	nonsynonymous SNV	2	p.H61P,TP53	exonic	*TP53*	75%
343G	11	108199962	A	T	nonsynonymous SNV	49	p.N2435I,	exonic	*ATM*	43%
346B	11	102207659	-	T	frameshift insertion	10	p.Q547fs,BIRC3	exonic	*BIRC3*	38%
531A	19	1440415	C	G	nonsynonymous SNV	4	p.P131R,	exonic	*RPS15*	52%
675A	17	7578271	T	A	nonsynonymous SNV	2	p.H61L,TP53	exonic	*TP53*	87%

Appendix S Pearsons correlations of expression data and telomere length. *POT1* and *BIRC3* expresion show a positive significant correlation, whilst *TP53* and *ATM* expression are also positively, significantly correlated.

		TL	ATMexp	BIRC3exp	POT1exp	TP53exp
TL	Pearson	1	0.026	0.234	0.006	0.008
	Sig.		0.854	0.197	0.967	0.956
	N	65	51	32	44	51
ATMexp	Pearson	0.026	1	0.086	0.001	0.861
	Sig.		0.854	0.622	0.994	0
	N	51	54	35	46	54
BIRC3exp	Pearson	0.234	0.086	1	0.839	0.08
	Sig.		0.197	0.622	0	0.649
	N	32	35	35	33	35
POT1exp	Pearson	0.006	0.001	0.839	1	0.031
	Sig.		0.967	0.994	0	0.838
	N	44	46	33	46	46
TP53exp	Pearson	0.008	0.861	0.08	0.031	1
	Sig.		0.956	0	0.649	0.838
	N	51	54	35	46	54

Appendix T Filtered variants identified from the bioinformatical pipeline. Many of these were identified as false positives

PID	chr	genomic	Ref	Alt	variant	Gene	exon	coding	protein	QDEPTH	VAF
1	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1157	7.2
2	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1384	6.9
3	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	69	10.1
4	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1311	6.9
5	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	887	6.4
6	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1157	5.8
7	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	908	6.7
9	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1220	7.1
10	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1355	6.0
11	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1122	6.1
12	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1469	5.6
13	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1545	5.5
14	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	856	5.7
15	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1423	6.3
16	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1451	6.8
17	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	995	7.5
18	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1436	5.3
19	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1381	6.2
20	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1594	5.5
23	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1081	5.4
24	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	838	8.6
25	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	948	8.3
26	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1573	5.5
27	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1579	5.3
28	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1724	6.5
29	11	1081000	T	A	nonsynonym	*ATM*	4	c.T314A	p.I105N,	1740	6.4

		33			ous SNV						
30	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1090	5.5
31	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1470	8.0
32	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	889	6.9
33	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1069	6.2
34	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1244	5.3
35	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1176	6.1
36	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	2643	6.1
38	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	977	7.6
39	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1550	6.8
40	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1170	7.3
41	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1094	5.9
42	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1387	5.2
43	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1624	7.1
44	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1826	6.1
45	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1405	7.7
46	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1787	8.6
47	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1040	6.2
48	11	108100033	T	A	nonsynonymous SNV	*ATM*	4	c.T314A	p.I105N,	1829	7.1
2	7	124487022	T	C	nonsynonymous SNV	*POT1*	11	c.A587G	p.E196G,POT1	3867	7.2
3	11	108124603	A	G	nonsynonymous SNV	*ATM*	13	c.A1961G	p.Q654R,	63	6.4
3	11	108124615	A	G	nonsynonymous SNV	*ATM*	13	c.A1973G	p.D658G,	62	6.5
3	11	108164197	T	C	nonsynonymous SNV	*ATM*	31	c.T4769C	p.L1590P,	38	5.3
3	11	108173662	A	G	nonsynonymous SNV	*ATM*	36	c.A5402G	p.N1801S,	64	6.3
3	11	108173715	A	T	nonsynonymous SNV	*ATM*	36	c.A5455T	p.T1819S,	32	6.3
3	11	108192098	A	T	nonsynonymous SNV	*ATM*	45	c.A6523T	p.R2175W,	86	7.0
3	11	108202266	T	A	nonsynonymous SNV	*ATM*	51	c.T7611A	p.F2537L,	32	6.3
3	11	108204686	G	A	nonsynonymous SNV	*ATM*	54	c.G8001A	p.M2667I,	105	5.7
3	11	102195317	T	A	stopgain SNV	*BIRC3*	3	c.T77A	p.L26X,BIRC3	36	5.6
3	11	102195329	T	C	nonsynonymous SNV	*BIRC3*	3	c.T89C	p.L30P,BIRC3	35	5.7
3	3	38180419	C	A	nonsynonymous SNV	*MYD88*	1	c.C267A	p.D89E,MYD88	48	6.3
3	9	1393905	A	G	nonsynonym	*NOTC*	34	c.T7661C	p.F2554S,	29	6.9

		30			ous SNV	*H1*					
3	9	139390696	T	C	nonsynonymous SNV	*NOTCH1*	34	c.A7495G	p.S2499G,	75	5.3
3	9	139391004	T	C	nonsynonymous SNV	*NOTCH1*	34	c.A7187G	p.N2396S,	30	13.3
3	9	139391013	T	C	nonsynonymous SNV	*NOTCH1*	34	c.A7178G	p.Q2393R,	32	6.3
3	9	139391016	A	G	nonsynonymous SNV	*NOTCH1*	34	c.T7175C	p.M2392T,	32	6.3
3	9	139391050	G	A	stopgain SNV	*NOTCH1*	34	c.C7141T	p.Q2381X,	36	5.6
3	9	139391321	G	T	nonsynonymous SNV	*NOTCH1*	34	c.C6870A	p.S2290R,	16	12.5
3	9	139391364	A	T	nonsynonymous SNV	*NOTCH1*	34	c.T6827A	p.L2276Q,	16	12.5
3	9	139391481	A	G	nonsynonymous SNV	*NOTCH1*	34	c.T6710C	p.M2237T,	28	7.1
3	7	124499027	A	G	nonsynonymous SNV	*POT1*	8	c.T293C	p.V98A,POT1	26	7.7
5	11	108225537	G	T	splicing	*ATM*				1738	42.8
3	7	124499154	T	C	nonsynonymous SNV	*POT1*	8	c.A166G	p.T56A,POT1	10	20.0
5	11	108235865	T	G	stopgain SNV	*ATM*	62	c.T8907G	p.Y2969X,	1887	9.3
7	11	108202619	T	C	nonsynonymous SNV	*ATM*	52	c.T7643C	p.I2548T,	77	5.2
7	7	124482937	G	A	stopgain SNV	*POT1*	13	c.C1087T	p.R363X,POT1	2095	33.5
11	11	108202642	A	C	nonsynonymous SNV	*ATM*	52	c.A7666C	p.T2556P,	31	6.5
13	11	108141989	C	T	nonsynonymous SNV	*ATM*	20	c.C2933T	p.S978F,	4629	10.8
13	11	108236073	-	A	frameshift insertion	*ATM*	63	c.9010dupA	p.N3003fs,	7487	15.2
35	4	153247168	T	C	nonsynonymous SNV	*FBXW7*	9	c.A1280G	p.Y427C,FBXW7	2767	50.9
14	9	139391076	C	T	nonsynonymous SNV	*NOTCH1*	34	c.G7115A	p.R2372Q,	715	49.9
26	4	153249384	C	A	nonsynonymous SNV	*FBXW7*	8	c.G1040T	p.R347L,FBXW7	1910	42.1
14	2	198267480	T	C	nonsynonymous SNV	*SF3B1*	14	c.A1877G	p.N626S,	7599	11.5
17	11	108175516	A	C	nonsynonymous SNV	*ATM*	37	c.A5611C	p.T1871P,	349	5.1
29	3	38182025	G	T	nonsynonymous SNV	*MYD88*	2	c.G514T	p.V172F,MYD88	3463	12.3
9	3	38182641	T	C	stoploss SNV	*MYD88*	3	c.T478C	p.X160R,MYD88	1529	37.4
12	3	38182641	T	C	stoploss SNV	*MYD88*	3	c.T478C	p.X160R,MYD88	1403	39.4
17	3	38182641	T	C	stoploss SNV	*MYD88*	3	c.T478C	p.X160R,MYD88	1370	46.8
27	3	38182641	T	C	stoploss SNV	*MYD88*	3	c.T478C	p.X160R,MYD88	1598	50.6
28	3	38182641	T	C	stoploss SNV	*MYD88*	3	c.T478C	p.X160R,MYD88	1921	43.3
44	3	38182641	T	C	stoploss SNV	*MYD88*	3	c.T478C	p.X160R,MYD88	2136	48.9
21	11	102207526	-	T	frameshift insertion	*BIRC3*	9	c.1616dupT	p.V539fs,BIRC3	2114	25.6
22	11	1022019	-	G	frameshift	*BIRC3*	7	c.1288dup	p.E429fs,BIR	3442	10.9

		35			insertion			G	C3		
22	11	1022075 24	-	TGT T	frameshift insertion	*BIRC3*	9	c.1613_161 4insTGTT	p.D538fs,BI RC3	1941	10.7
23	11	1081418 74	G	T	Splicing	*ATM*	19	c.2921+1G>T		4291	50.3
23	11	1082245 97	G	T	nonsynonym ous SNV	*ATM*	60	c.G8776T	p.V2926F,	1452	49.2
24	11	1081551 87	-	A	frameshift insertion	*ATM*	26	c.3981dup A	p.L1327fs,	1313	48.5
26	11	1082026 24	A	G	nonsynonym ous SNV	*ATM*	52	c.A7648G	p.M2550V,	106	5.7
26	4	1532473 70	A	T	nonsynonym ous SNV	*FBXW 7*	9	c.T1078A	p.S360T,FBX W7	206	5.3
30	7	1245323 40	G	A	nonsynonym ous SNV	*POT1*	6	c.C104T	p.P35L,	1950	50.6
31	7	1244930 78	G	A	nonsynonym ous SNV	*POT1*	10	c.C817T	p.R273W,P OT1	3356	50.3
32	11	1081755 16	A	C	nonsynonym ous SNV	*ATM*	37	c.A5611C	p.T1871P,	419	5.3
32	7	1244640 66	G	A	stopgain SNV	*POT1*	18	c.C1462T	p.Q488X,PO T1	5725	19.4
33	11	1022019 39	-	A	frameshift insertion	*BIRC3*	7	c.1292dup A	p.E431fs,BIR C3	3192	11.3
34	3	3818016 8	G	C	nonsynonym ous SNV	*MYD8 8*	1	c.G16C	p.A6P,MYD8 8	433	41.8
35	11	1081984 86	G	T	.	.				1203	13.6
36	11	1081960 65	-	T	frameshift insertion	*ATM*	46	c.6602dupT	p.V2201fs,	5440	21.7
38	11	1022077 90	G	A	nonsynonym ous SNV	*BIRC3*	10	c.G1772A	p.C591Y,BIR C3	2564	76.9
40	17	7573009	C	G	.	.				154	5.2
16	2	1982666 11	C	T	nonsynonym ous SNV	*SF3B1*	16	c.G2225A	p.G742D,	3805	23.7
37	2	1982666 11	C	T	nonsynonym ous SNV	*SF3B1*	16	c.G2225A	p.G742D,	6107	52.2
46	2	1982666 11	C	T	nonsynonym ous SNV	*SF3B1*	16	c.G2225A	p.G742D,	5505	25.5
13	2	1982667 13	C	T	nonsynonym ous SNV	*SF3B1*	15	c.G2219A	p.G740E,	6762	7.2
41	2	1982668 22	T	A	nonsynonym ous SNV	*SF3B1*	15	c.A2110T	p.I704F,	3676	13.4
4	2	1982668 34	T	C	nonsynonym ous SNV	*SF3B1*	15	c.A2098G	p.K700E,	4276	11.5
19	2	1982668 34	T	C	nonsynonym ous SNV	*SF3B1*	15	c.A2098G	p.K700E,	3731	5.2
25	2	1982668 34	T	C	nonsynonym ous SNV	*SF3B1*	15	c.A2098G	p.K700E,	3303	11.3
19	2	1982673 59	C	A	nonsynonym ous SNV	*SF3B1*	14	c.G1998T	p.K666N,	1559	23.1
26	2	1982673 59	C	A	nonsynonym ous SNV	*SF3B1*	14	c.G1998T	p.K666N,	1532	43.9
48	2	1982673 61	T	G	nonsynonym ous SNV	*SF3B1*	14	c.A1996C	p.K666Q,	2295	27.7
2	2	1982673 69	G	A	nonsynonym ous SNV	*SF3B1*	14	c.C1988T	p.T663I,	810	44.6
40	7	1245036 64	T	A	nonsynonym ous SNV	*POT1*	8	c.A286T	p.I96F,	3296	49.6
36	2	1982674 81	T	C	nonsynonym ous SNV	*SF3B1*	14	c.A1876G	p.N626D,	7702	46.6
1	17	7572980	T	G	nonsynonym ous SNV	*TP53*	11	c.A1129C	p.T377P,TP5 3	48	85.4

2	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	21	76.2
4	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	50	82.0
5	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	50	66.0
6	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	32	84.4
7	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	38	86.8
8	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	38	81.6
9	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	43	72.1
10	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	37	83.8
11	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	32	59.4
12	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	45	84.4
13	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	95	86.3
14	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	41	78.1
15	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	62	83.9
16	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	39	84.6
17	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	34	70.6
18	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	68	75.0
19	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	34	79.4
20	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	51	90.2
21	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	60	71.7
22	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	49	79.6
23	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	46	73.9
24	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	42	88.1
25	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	35	68.6
26	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	33	72.7
27	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	65	81.5
28	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	46	84.8
29	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	66	72.7
30	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	49	79.6
31	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	58	82.8
32	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	38	79.0
33	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	37	75.7

34	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	67	73.1
35	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	84	71.4
36	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	124	78.2
37	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	103	82.5
38	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	68	63.2
39	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	53	79.3
40	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	18	94.4
41	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	48	77.1
42	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	21	90.5
43	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	32	75.0
44	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	33	81.8
45	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	28	71.4
46	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P,TP53	37	86.5
47	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	41	78.1
48	17	7572980	T	G	nonsynonymous SNV	*TP53*	11	c.A1129C	p.T377P,TP53	45	75.6
45	17	7577095	G	C	nonsynonymous SNV	*TP53*	4	c.C447G	p.D149E,TP53	908	45.3
45	17	7577508	T	C	nonsynonymous SNV	*TP53*	3	c.A377G	p.E126G,TP53	1066	34.4
43	17	7578224	T	A	stopgain SNV	*TP53*	2	c.A229T	p.R77X,TP53	2677	12.3
42	17	7578262	C	G	nonsynonymous SNV	*TP53*	2	c.G191C	p.R64P,TP53	2047	99.8
47	17	7578394	T	C	nonsynonymous SNV	*TP53*	1	c.A140G	p.H47R,TP53	1255	43.4
45	17	7578400	G	C	nonsynonymous SNV	*TP53*	1	c.C134G	p.P45R,TP53	1933	20.8
26	17	7578403	C	T	nonsynonymous SNV	*TP53*	1	c.G131A	p.C44Y,TP53	1234	66.1
39	17	7578550	G	A	nonsynonymous SNV	*TP53*	5	c.C380T	p.S127F,TP53	2293	82.2
46	17	7579335	-	C	frameshift insertion	*TP53*	4	c.351dupG	p.T118fs,TP53	607	91.1
16	17	7579350	A	C	nonsynonymous SNV	*TP53*	4	c.T337G	p.F113V,TP53	1070	28.6
43	17	7576928	T	C	Splicing	*TP53*				2266	26.7
41	11	108202633	C	T	nonsynonymous SNV	*ATM*	52	c.C7657T	p.P2553S,	79	5.1
47	17	7579699	C	A	Splicing	*TP53*				1343	43.0
43	17	7577610	T	A	Splicing	*TP53*				1108	21.2
10	2	61719471	T	C	nonsynonymous SNV	*XPO1*	15	c.A1712G	p.E571G,	2230	29.9
19	2	61719471	T	C	nonsynonymous SNV	*XPO1*	15	c.A1712G	p.E571G,	2410	48.0
48	2	61719471	T	A	nonsynonymous SNV	*XPO1*	15	c.A1712T	p.E571V,	2749	51.5

11	2	61719472	C	T	nonsynonymous SNV	*XPO1*	15	c.G1711A	p.E571K,	2205	54.3
26	2	61719472	C	T	nonsynonymous SNV	*XPO1*	15	c.G1711A	p.E571K,	2760	40.5
37	2	61719472	C	T	nonsynonymous SNV	*XPO1*	15	c.G1711A	p.E571K,	4414	44.6
8	17	7572980	T	G	nonsynonymous SNV	*TP53*	7	c.A733C	p.T245P	38	81.6

Appendix U Full filtered variant list from the Southampton cases before IGV confirmation

PID	chr	genomic	ref	alt	variant	Exon	coding	Protein	GENE	depth	VAF
598	11	108199938	T	C	nonsynonymous	49	c.T7280C	p.L2427P,	*ATM*	1495	47.09
695	11	108186598	T	A	nonsynonymous	41	c.T6055A	p.Y2019N,	*ATM*	1177	47.11
590	11	108186599	A	G	nonsynonymous	41	c.A6056G	p.Y2019C,	*ATM*	744	11.16
504-A-1	11	108200961	G	A	nonsynonymous	50	c.G7328A	p.R2443Q,	*ATM*	1437	15.03
504-B	11	108200961	G	A	nonsynonymous	50	c.G7328A	p.R2443Q,	*ATM*	1394	19.37
619	7	140434574	-	A	.	19	c.2128-4->T)		*BRAF*	2746	5.34
637	7	140434574	-	A	.	19	c.2128-4->T)		*BRAF*	2766	5.60
661	7	140434574	-	A	.	19	c.2128-4->T)		*BRAF*	2028	5.59
668	7	140434574	-	A	.	19	c.2128-4->T)		*BRAF*	1607	11.46
704	7	140434574	-	A	.	19	c.2128-4->T)		*BRAF*	869	5.01
709	7	140434574	-	A	.	19	c.2128-4->T)		*BRAF*	1599	11.14
346B	7	140434574	-	A	.	19	c.2128-4->T)		*BRAF*	2538	12.55
348-1	7	140434574	-	A	.	19	c.2128-4->T)		*BRAF*	2474	17.95
504-B	7	140434574	-	A	.	19	c.2128-4->T)		*BRAF*	2821	5.66
542-1	7	140434574	-	A	.	19	c.2128-4->T)		*BRAF*	1823	31.45
558B-II	7	140434574	-	A	.	19	c.2128-4->T)		*BRAF*	2688	5.48
585	10	64573248	G	T	nonsynonymous	3	c.C1000A	p.H334N,EGR2	*EGR2*	1160	31.12
619	10	64573332	C	T	nonsynonymous	3	c.G916A	p.E306K,EGR2	*EGR2*	1204	56.31
588	3	38182641	T	C	stoploss	3	c.T478C	p.X160R,MYD88	*MYD88*	1333	43.06
662	3	38182641	T	C	stoploss	3	c.T478C	p.X160R,MYD88	*MYD88*	1243	41.22
542-1	3	38182641	T	C	stoploss	3	c.T478C	p.X160R,MYD88	*MYD88*	1037	44.94
580	3	47040917	T	C	nonsynonymous	25	c.T3656C	p.V1219A,	*NBEAL2*	20	10
1215	3	47036838	G	A	nonsynonymous	13	c.G1613A	p.R538H,	*NBEAL2*	1743	49
504-B	3	47040917	T	C	nonsynonymous	25	c.T3656C	p.V1219A,	*NBEAL2*	16	12.50
607	9	139390930	C	T	nonsynony	34	c.G7261A	p.V24	*NOTCH1*	32	6.25

					mous			21M,			
674	9	139390816	G	A	stopgain	34	c.C7375T	p.Q2459X,	*NOTCH1*	2244	19.74
665A	9	139391020	G	A	stopgain	34	c.C7171T	p.Q2391X,	*NOTCH1*	57	5.26
674	20	35547857	C	T	nonsynonymous	7	c.G762A	p.M254I,	*SAMHD1*	3927	73.26
428	2	198266834	T	C	nonsynonymous	15	c.A2098G	p.K700E,	*SF3B1*	1943	48.64
632	2	198267483	C	T	nonsynonymous	14	c.G1874A	p.R625H,	*SF3B1*	1341	48.25
664	2	198266834	T	C	nonsynonymous	15	c.A2098G	p.K700E,	*SF3B1*	1988	24.55
704	2	198267371	G	T	nonsynonymous	14	c.C1986A	p.H662Q,	*SF3B1*	730	31.92
1115	2	198266611	C	T	nonsynonymous	16	c.G2225A	p.G742D,	*SF3B1*	518	35.71
1133	2	198266834	T	C	nonsynonymous	15	c.A2098G	p.K700E,	*SF3B1*	1726	47.91
1144	2	198266834	T	C	nonsynonymous	15	c.A2098G	p.K700E,	*SF3B1*	1500	26.67
1147	2	198267483	C	T	nonsynonymous	14	c.G1874A	p.R625H,	*SF3B1*	1157	40.88
590	2	198266834	T	C	nonsynonymous	15	c.A2098G	p.K700E,	*SF3B1*	1642	43.48
1156	2	198267359	C	G	nonsynonymous	14	c.G1998C	p.K666N,	*SF3B1*	1627	26.31
1169	2	198267484	G	A	nonsynonymous	14	c.C1873T	p.R625C,	*SF3B1*	981	43.73
1179	2	198266834	T	C	nonsynonymous	15	c.A2098G	p.K700E,	*SF3B1*	1640	46.83
1213	2	198266834	T	C	nonsynonymous	15	c.A2098G	p.K700E,	*SF3B1*	1912	33.89
495-1	2	198266834	T	C	nonsynonymous	15	c.A2098G	p.K700E,	*SF3B1*	1751	6.57
704	17	7578370	C	T	.	6	c.559+1G>A		*TP53*	1461	14.17
727	17	7578271	T	G	nonsynonymous	2	c.A182C	p.H61P,TP53	*TP53*	3451	74.62
551C-I	17	7574035	T	A	.	11	c.994-2A>T,NM_001276760		*TP53*	30	6.67
629B	17	7578545	C	T	nonsynonymous	5	c.G385A	p.A129T,TP53	*TP53*	100	6
675A	17	7578271	T	A	nonsynonymous	2	c.A182T	p.H61L,TP53	*TP53*	3649	87.04
644	2	61719472	C	T	nonsynonymous	15	c.G1711A	p.E571K,	*XPO1*	1650	6.97
704	2	61719186	T	C	nonsynonymous	16	c.A1871G	p.D624G,	*XPO1*	283	49.12
661	8	106814031	G	A	nonsynonymous	8	c.G1721A	p.R574Q,	*ZFPM2*	3361	16.52
1115	3	47103697	-	G	frameshift insertion	14	c.6248dupC	p.P2083fs,	*SETD2*	2807	5.43
1133	2	61722612	T	C	nonsynonymous	11	c.A1025G	p.N342S,	*XPO1*	51	5.88
1146	3	47103697	-	G	frameshift insertion	14	c.6248dupC	p.P2083fs,	*SETD2*	3432	5.12
590	3	47103697	-	G	frameshift insertion	14	c.6248dupC	p.P2083fs,	*SETD2*	1946	9.07

1156	3	47103697	-	G	frameshift insertion	14	c.6248dupC	p.P2083fs,	*SETD2*	2829	5.07
1184	3	47103697	-	G	frameshift insertion	14	c.6248dupC	p.P2083fs,	*SETD2*	2643	5.63
1191	19	1440415	C	G	nonsynonymous	4	c.C392G	p.P131R,	*RPS15*	1222	24.88
1191	3	47103697	-	G	frameshift insertion	14	c.6248dupC	p.P2083fs,	*SETD2*	2910	5.31
1215	3	47103697	-	G	frameshift insertion	14	c.6248dupC	p.P2083fs,	*SETD2*	3614	7.73
428	3	47103697	-	G	frameshift insertion	14	c.6248dupC	p.P2083fs,	*SETD2*	2855	5.02
428	7	124475324	-	CAAAG	splicing	15			*POT1*	52	5.77
346B	11	102207659	-	T	frameshift insertion	10	c.1642dupT	p.Q547fs,BIRC3	*BIRC3*	2910	38.27
661	11	102199670	T	C	nonsynonymous	6	c.T1075C	p.S359P,BIRC3	*BIRC3*	48	6.25
709	11	102199624	G	T	.	5	c.1033-4G>T,NM_182962			29	6.90
348-1	10	64573542	T	C	nonsynonymous	3	c.A706G	p.S236G,EGR2	*EGR2*	74	6.76
523D	10	64573557	T	G	nonsynonymous	3	c.A691C	p.T231P,EGR2	*EGR2*	78	6.41
530C	10	64573540	A	C	nonsynonymous	3	c.T708G	p.S236R,EGR2	*EGR2*	141	5.67
530C	10	64573542	T	C	nonsynonymous	3	c.A706G	p.S236G,EGR2	*EGR2*	85	5.81
558B-II	10	64573542	T	C	nonsynonymous	3	c.A706G	p.S236G,EGR2	*EGR2*	136	5.84
573C	10	64573557	T	G	nonsynonymous	3	c.A691C	p.T231P,EGR2	*EGR2*	135	6.67
577	10	64573557	T	G	nonsynonymous	3	c.A691C	p.T231P,EGR2	*EGR2*	42	9.52
607	10	64573540	A	C	nonsynonymous	3	c.T708G	p.S236R,EGR2	*EGR2*	99	5.05
621	10	64573557	T	G	nonsynonymous	3	c.A691C	p.T231P,EGR2	*EGR2*	36	5.56
629B	10	64573542	T	C	nonsynonymous	3	c.A706G	p.S236G,EGR2	*EGR2*	120	6.67
637	10	64573542	T	C	nonsynonymous	3	c.A706G	p.S236G,EGR2	*EGR2*	103	5.77
644	10	64573167	C	G	nonsynonymous	3	c.G1081C	p.D361H,EGR2	*EGR2*	2879	5.90
674	10	64573557	T	G	nonsynonymous	3	c.A691C	p.T231P,EGR	*EGR2*	29	6.90

								2			
675A	10	64573540	A	C	nonsynonymous	3	c.T708G	p.S236R,EGR2	*EGR2*	171	5.26
675A	10	64573557	T	G	nonsynonymous	3	c.A691C	p.T231P,EGR2	*EGR2*	117	5.98
684	10	64573557	T	G	nonsynonymous	3	c.A691C	p.T231P,EGR2	*EGR2*	114	5.22
695	10	64573540	A	C	nonsynonymous	3	c.T708G	p.S236R,EGR2	*EGR2*	56	5.36
684	4	153251918	G	A	nonsynonymous	6	c.C734T	p.T245I,FBXW7	*FBXW7*	47	6.38
504-B	20	35555658	T	C	.				*SAMHD1*	5	40
501-1	7	124503424	C	T	nonsynonymous	7	c.G133A	p.G45R,POT1	*POT1*	1319	48.75
558B-II	7	124532370	A	T	nonsynonymous	6	c.T74A	p.V25D,	*POT1*	1616	5.75
632	7	124482918	T	C	nonsynonymous	12	c.A713G	p.Y238C,POT1	*POT1*	1415	47.92
674	7	124532325	C	T	nonsynonymous	6	c.G119A	p.G40E,	*POT1*	1531	7.71
489E	7	124537218	C	T	.	6	c.9+1G>A		*POT1*	2253	48.87
531D	19	1440415	C	G	nonsynonymous	4	c.C392G	p.P131R,	*RPS15*	2321	27.88
598	19	1440059	G	A	nonsynonymous	3	c.G131A	p.R44Q,	*RPS15*	85	5.88
670	19	1440421	T	G	nonsynonymous	4	c.T398G	p.I133S,	*RPS15*	1837	25.93
530C	20	35555638	G	A	nonsynonymous	6	c.C643T	p.H215Y,	*SAMHD1*	5	60
577	3	47205404	A	G	nonsynonymous	1	c.T11C	p.L4P,	*SETD2*	25	8
585	3	47205401	T	C	nonsynonymous	1	c.A14G	p.Q5R,	*SETD2*	24	8.33
588	3	47162951	T	C	nonsynonymous	3	c.A3175G	p.S1059G,	*SETD2*	4347	26.28
588	3	47205404	A	G	nonsynonymous	1	c.T11C	p.L4P,	*SETD2*	13	15.38
604	3	47125283	T	C	nonsynonymous	12	c.A5987G	p.Q1996R,	*SETD2*	2912	50.96
684	3	47205404	A	G	nonsynonymous	1	c.T11C	p.L4P,	*SETD2*	28	7.14
697	2	61722681	A	G	nonsynonymous	11	c.T956C	p.I319T,	*XPO1*	118	5.08
343G	11	108199962	A	T	nonsynonymous	49	c.A7304T	p.N2435I,	*ATM*	2228	43.49
635	11	108170471	G	T	nonsynonymous	34	c.G5036T	p.G1679V,	*ATM*	819	53.48
644	11	108201132	T	G	nonsynonymous	50	c.T7499G	p.V2500G,	*ATM*	1773	55.22
HRM		POT1	4	TGCAATGTAATTAGAGAATAAAAGCTG				*ATTATACGTATTTTGGTGATTGATTCA*		60	192
Sequencing		POT1	4	AAGTGCAATATCTGCCAAGT				*TCCAAACAATGACAAAATCA*		55	47

						8
HRM	POT1	5	CACATGTATCTATGTGTGTGGCATA	*AGCATGTAATCACATTGGAGGTT*	60	181
Sequencing	POT1	5	TCAGCAGATATTCCAGACAA	*AGCTTAGACAACTTTGCACAT*	55	452
HRM	POT1	6	AAACTCCACCAGTTTTAATACCTACC	*TACATGGATTTGCTGCTAATATGAT*	60	229
Sequencing	POT1	6	AGCCAAAGAATATGCATCAG	*CCATTTATAAACAAAGTTCTAAGGA*	55	457
HRM	POT1	7	TTCTCTTCAAATAAATATAAGTTCTAGAC	*GGTTTGGTGTTTTGAAGTAAGCA*	60	281
Sequencing	POT1	7	GCAGTGGTTTGTTCAAATG	*TTGCAGTGTGTATTGAAAGC*	55	493
HRM	POT1	8	TGGTGCTAACTTATAATTCCCAGTATT	*CCTTACGTGTTTGGGCATCT*	60	183
HRM	POT1	8	AGATGCCCAAACACGTAAGG	*CTGTTTTCTACTTTGCCCTACTTTC*	60	197
Sequencing	POT1	8	CCACACAAATCTCATGTCAA	*TCACCCAGTAAATCTCTTTAGC*	55	474
HRM	POT1	9	TCAGAGATCTTGCCACATGAA	*TTATGGCAGGTATGGGATGG*	60	163
Sequencing	POT1	9	CATTTTACAACTTAAAAATCAAAGA	*TTCCACATTACCCATATTTCA*	55	487
HRM	POT1	10	TCGGCTTAATCGATACCTTATTTAC	*TTTTTCCCCACTTTCTAAATAACAA*	60	257
Sequencing	POT1	10	ATTTGTTTCATTTGGCTCAT	*CCATGCAGCTGATATTCAA*	55	483
HRM	POT1	18	TCAAGTAAAAGAAGTGTGGGATTG	*AAGGACAAATTCTTCCAGATTCC*	60	152
Sequencing	POT1	18	TTTGACTGCAGGAATTATGA	*GATTTTGGAGTTGAGACCAG*	60	475
HRM	BIRC3	7	TTCCATATAGTTATCCATTTTGAACCT	*ACATACTTGATTCTTTTTCCTCAGTTG*	60	259
Sequencing	BIRC3	7	TGCCTATACATTTTGTTGGTT	*AAAAACCTGACTGGATTGAG*	55	490
HRM	BIRC3	10	TGAAGAAGCAAACTGCCTTTTATT	*AAAGTTTAGACGATGTTTTGGTTCT*	60	268
Sequencing	BIRC3	10	CCACAGAAGATGTTTCAGGT	*GTGCTACCTCTTTTTCGTTC*	55	546
HRM	MYD88	3	TCTGACCACCACCCTTGTG	*GGCCTTCTAGCCAACCTCTT*	60	227
Sequencing	MYD88	3	GGCACTTTCTCTGAGGAGTA	*GACAGTGCACAGCTAGGAG*	55	40

						7
HRM	MYD88	4	GCTGAACTAAGTTGCCACAGG	*CCAGAGCAGGGTTGAGCTT*	60	191
Sequencing	MYD88	4	CAGGGGATATGCTGAACTAA	*GATCTTCAGCAGTTCTTTGG*	55	266
HRM	MYD88	5	CAGGTGCCCATCAGAAGC	*GGTTGGTGTAGTCGCAGACA*	60	102
Sequencing	MYD88	5	GCAGAAGTACATGGACAGGCAGACAGATAC	*GTTGTTAACCCTGGGGTTGAAG*	55	297
HRM	U2AF1	Codon 34	TGTCGTCATGGAGACAGGTG	*AGGCAAACAAACCTGGCTAA*	61	66
Sequencing	U2AF1	34	TAAAAACAAGGAGTGGTGGT	*GGGAAGTTGAAGTTACGTTG*	55	480
HRM	U2AF1	157	CCGTGACGGACTTCAGAGA	*CCTCACTCACCCCATCTCAT*	60	56
HRM	SRSF2	95	GAGCTGCGGGTGCAAAT	*CGGCTGTGGTGTGAGTCC*	61	56

Appendix V Nextera BTK-PLCG2 and DNAH9-SPEN indexing QC, percent of reads matching to identification index. Some samples have very low reads this may have been dure to pipetting errors.

INDEX NUMBER	SAMPLE ID	INDEX 1 (I7)	INDEX 2 (I5)	% READS IDENTIFIED (PF)
1	343G	TAAGGCGA	TCTTACGC	0.0001
2	601B	GCTACGCT	TCTTACGC	0.0001
3	573C	GGACTCCT	TCTTACGC	0.0001
4	558C	TAAGGCGA	CTCTCTAT	0.7194
5	530D	GCTACGCT	CTCTCTAT	1.5432
6	632M	GGACTCCT	CTCTCTAT	0.8036
7	558P	TAAGGCGA	AGAGTAGA	1.1696
8	573N	GCTACGCT	AGAGTAGA	0.315
9	1215	GGACTCCT	AGAGTAGA	1.2797
10	1150	TAAGGCGA	AAGGAGTA	0.5004
11	dna585	GCTACGCT	AAGGAGTA	1.1539
12	dna343G	GGACTCCT	AAGGAGTA	0.8928
13	346B	CGTACTAG	CTCTCTAT	0.9521
14	632D	CGAGGCTG	CTCTCTAT	1.2642
15	343H	TAGGCATG	CTCTCTAT	0.8159
16	601D	CGTACTAG	TATCCTCT	0.7412
17	343P	CGAGGCTG	TATCCTCT	0.4194
18	727I	TAGGCATG	TATCCTCT	0.0003
19	632AB	CGTACTAG	GTAAGGAG	0.9828
20	Control1	CGAGGCTG	GTAAGGAG	0.8391
21	1191	TAGGCATG	GTAAGGAG	0.2213
22	1184	CGTACTAG	CTAAGCCT	0.1658
23	dna598	CGAGGCTG	CTAAGCCT	0.315
24	dna531D	TAGGCATG	CTAAGCCT	0.2679
25	409F	AGGCAGAA	TATCCTCT	0.772
26	644B	AAGAGGCA	TATCCTCT	0.9779
27	346D	CTCTCTAC	TATCCTCT	0.0002
28	632E	AGGCAGAA	AGAGTAGA	0.0003
29	346N	AAGAGGCA	AGAGTAGA	0
30	573L	CTCTCTAC	AGAGTAGA	0.0002
31	644L	AGGCAGAA	ACTGCATA	0.3696
32	Control2	AAGAGGCA	ACTGCATA	0.0002
33	1169	CTCTCTAC	ACTGCATA	0.0008
34	1147	AGGCAGAA	TCTTACGC	0

35	dna619A	AAGAGGCA	TCTTACGC	0
36	dna495D	CTCTCTAC	TCTTACGC	0.0001
37	489E	TCCTGAGC	AGAGTAGA	1.0672
38	727A	GTAGAGGA	AGAGTAGA	1.0066
39	409G	CAGAGAGG	AGAGTAGA	0.0003
40	644C	TCCTGAGC	GTAAGGAG	0.6814
41	495M	GTAGAGGA	GTAAGGAG	1.6262
42	346Q	CAGAGAGG	GTAAGGAG	0.9493
43	727L	TCCTGAGC	AAGGAGTA	0.5416
44	1219	GTAGAGGA	AAGGAGTA	0.0014
45	1156	CAGAGAGG	AAGGAGTA	0.0012
46	dna695	TCCTGAGC	CTCTCTAT	1.0083
47	dna635	GTAGAGGA	CTCTCTAT	0.5501
48	dna551C	CAGAGAGG	CTCTCTAT	0.0047
49	495D	GGACTCCT	GTAAGGAG	0.9564
50	222E	TAAGGCGA	GTAAGGAG	0.7411
51	489F	GCTACGCT	GTAAGGAG	0.1322
52	727B	GGACTCCT	ACTGCATA	1.0072
53	531N	TAAGGCGA	ACTGCATA	1.4563
54	409N	GCTACGCT	ACTGCATA	1.9586
55	222L	GGACTCCT	CTAAGCCT	1.0983
56	1210	TAAGGCGA	CTAAGCCT	0.0003
57	1213	GCTACGCT	CTAAGCCT	0.0002
58	dna727A	GGACTCCT	TATCCTCT	0.2874
59	dna644B	TAAGGCGA	TATCCTCT	0.8033
60	dna604	GCTACGCT	TATCCTCT	0.2511
61	531D	TAGGCATG	ACTGCATA	0.8439
62	665A	CGTACTAG	ACTGCATA	1.0346
63	495E	CGAGGCTG	ACTGCATA	0.5071
64	222F	TAGGCATG	AAGGAGTA	0.8559
65	551N	CGTACTAG	AAGGAGTA	2.8336
66	489P	CGAGGCTG	AAGGAGTA	0.9556
67	665L	TAGGCATG	TCTTACGC	0.0002
68	1179	CGTACTAG	TCTTACGC	0.0001
69	1181	CGAGGCTG	TCTTACGC	0
70	dna346B	TAGGCATG	AGAGTAGA	1.2611
71	dna674	CGTACTAG	AGAGTAGA	0.4448
72	dna665A	CGAGGCTG	AGAGTAGA	0.4431
73	551C	CTCTCTAC	AAGGAGTA	1.0464
74	619A	AGGCAGAA	AAGGAGTA	0.6455
75	531E	AAGAGGCA	AAGGAGTA	0.1501
76	665B	CTCTCTAC	CTAAGCCT	0.1731
77	558L	AGGCAGAA	CTAAGCCT	0.2913
78	531Q	AAGAGGCA	CTAAGCCT	1.0048
79	619H	CTCTCTAC	CTCTCTAT	0.9599
80	1146	AGGCAGAA	CTCTCTAT	0.0002

81	1133	AAGAGGCA	CTCTCTAT	0.0002
82	dna501	CTCTCTAC	GTAAGGAG	0.2296
83	dna664	AGGCAGAA	GTAAGGAG	1.0644
84	dna222E	AAGAGGCA	GTAAGGAG	0.8337
85	558B	CAGAGAGG	CTAAGCCT	1.2487
86	530C	TCCTGAGC	CTAAGCCT	0.7829
87	551E	GTAGAGGA	CTAAGCCT	0.2492
88	619B	CAGAGAGG	TCTTACGC	0.0001
89	601L	TCCTGAGC	TCTTACGC	0.0002
90	551P	GTAGAGGA	TATCCTCT	0.8689
91	530K	CAGAGAGG	TATCCTCT	0.9818
92	1115	TCCTGAGC	TATCCTCT	0.0053
93	1144	GTAGAGGA	ACTGCATA	0.6574
94	dna675A	CAGAGAGG	ACTGCATA	0.7462
95	dna704	TCCTGAGC	ACTGCATA	0.9403
96	dna409F	GTAGAGGA	TCTTACGC	0.0002

Appendix W - This work contributed to published data:

Longitudinal copy number, whole exome and targeted deep sequencing of 'good risk' IGHV-mutated CLL patients with progressive disease.
Rose-Zerilli MJ, Gibson J, Wang J, Tapper W, Davis Z, Parker H, Larrayoz M, McCarthy H, Walewska R, **Forster J**, Gardiner A, Steele AJ, Chelala C, Ennis S, Collins A, Oakes CC, Oscier DG, Strefford JC.
Leukemia. Epub 2016 Feb 5. PMID: 26847028

Telomere length predicts progression and overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial.
Strefford JC, Kadalayil L, **Forster J**, Rose-Zerilli MJ, Parker A, Lin TT, Heppel N, Norris K, Gardiner A, Davies Z, Gonzalez de Castro D, Else M, Steele AJ, Parker H, Stankovic T, Pepper C, Fegan C, Baird D, Collins A, Catovsky D, Oscier DG.
Leukemia. 2015 Dec;29(12):2411-4. PMID: 26256637

Low frequency mutations independently predict poor treatment-free survival in early stage chronic lymphocytic leukemia and monoclonal B-cell lymphocytosis.
Winkelmann N, Rose-Zerilli M, **Forster J**, Parry M, Parker A, Gardiner A, Davies Z, Steele AJ, Parker H, Cross NC, Oscier DG, Strefford JC.
Haematologica. 2015 Jun;100(6). PMID: 25710457

ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q deleted chronic lymphocytic leukemia: data from the UK LRF CLL4 trial.
Rose-Zerilli MJ, **Forster J**, Parker H, Parker A, Rodríguez AE, Chaplin T, Gardiner A, Steele AJ, Collins A, Young BD, Skowronska A, Catovsky D, Stankovic T, Oscier DG, Strefford JC.
Haematologica. 2014 Apr;99(4):736-42. PMID: 24584352

The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial.
Oscier DG, Rose-Zerilli MJ, Winkelmann N, Gonzalez de Castro D, Gomez B, **Forster J**, Parker H, Parker A, Gardiner A, Collins A, Else M, Cross NC, Catovsky D, Strefford JC.
Blood. 2013 Jan 17;121(3). PMID: 23086750

Whole exome sequencing identifies novel recurrently mutated genes in patients with splenic marginal zone lymphoma.
Parry M, Rose-Zerilli MJ, Gibson J, Ennis S, Walewska R, **Forster J**, Parker H, Davis Z, Gardiner A, Collins A, Oscier DG, Strefford JC.
PLoS One. 2013 Dec 13;8(12).
PMID: 24349473

Non-coding NOTCH1 mutations in chronic lymphocytic leukemia; their clinical impact in the UK CLL4 trial.
Larrayoz M, Rose-Zerilli MJJ, Kadalayil L, Parker H, Blakemore S, **Forster J**, Davis Z, Steele AJ, Collins A, Else M, Catovsky D, Oscier DG, Strefford JC.
Leukemia. 2016. October:1–10.
PMID: 27773930

OPEN

Leukemia (2016) 30, 1301–1310

www.nature.com/leu

ORIGINAL ARTICLE

Longitudinal copy number, whole exome and targeted deep sequencing of 'good risk' IGHV-mutated CLL patients with progressive disease

MJJ Rose-Zerilli[1,7], J Gibson[2,7], J Wang[3], W Tapper[4], Z Davis[5], H Parker[1], M Larrayoz[1], H McCarthy[5], R Walewska[5], J Forster[1], A Gardiner[5], AJ Steele[1], C Chelala[3], S Ennis[4], A Collins[4], CC Oakes[6], DG Oscier[1,5] and JC Strefford[1]

The biological features of *IGHV-M* chronic lymphocytic leukemia responsible for disease progression are still poorly understood. We undertook a longitudinal study close to diagnosis, pre-treatment and post relapse in 13 patients presenting with cMBL or Stage A disease and good-risk biomarkers (*IGHV-M* genes, no del(17p) or del(11q) and low CD38 expression) who nevertheless developed progressive disease, of whom 10 have required therapy. Using cytogenetics, fluorescence *in situ* hybridisation, genome-wide DNA methylation and copy number analysis together with whole exome, targeted deep- and Sanger sequencing at diagnosis, we identified mutations in established chronic lymphocytic leukemia driver genes in nine patients (69%), non-coding mutations (*PAX5* enhancer region) in three patients and genomic complexity in two patients. Branching evolutionary trajectories predominated ($n = 9/13$), revealing intra-tumoural epi- and genetic heterogeneity and sub-clonal competition before therapy. Of the patients subsequently requiring treatment, two had sub-clonal *TP53* mutations that would not be detected by standard methodologies, three qualified for the very-low-risk category defined by integrated mutational and cytogenetic analysis and yet had established or putative driver mutations and one patient developed progressive, therapy-refractory disease associated with the emergence of an *IGHV-U* clone. These data suggest that extended genomic and immunogenetic screening may have clinical utility in patients with apparent good-risk disease.

Leukemia (2016) **30,** 1301–1310; doi:10.1038/leu.2016.10

INTRODUCTION

Clinical heterogeneity within chronic lymphocytic leukemia (CLL), especially in the majority of patients presenting with a low-tumour burden, provides a continuing impetus for the discovery of prognostic biomarkers.

Immunogenetic features such as *IGHV* mutation status and stereotypy, immunophenotypic markers, genomic abnormalities and serum markers have prognostic significance. A recently described prognostic index incorporating gender, age, performance status, *IGHV* mutation status, deletions of 11q and 17p, serum B2 microglobulin and thymidine kinase distinguished four risk categories with differing 5-year overall and progression-free survivals.[1]

Candidate gene approaches and next generation sequencing have led to the discovery of mutations in many genes, including *TP53, ATM, NOTCH1, SF3B1, BIRC3, SAMHD1 and EGR2,* with prognostic and/or predictive significance, even when first detected as small sub-clones in the case of *TP53* mutation.[2–10] A recent whole-genome study demonstrated the adverse prognostic significance of multiple driver mutations and implicated novel non-coding mutation.[11] Mutations in an intergenic region on 9p13 correlated with reduced *PAX5* expression and three prime untranslated region (3′UTR) *NOTCH1* mutations associated with a poor outcome comparable to cases with an exon 34 *NOTCH1* mutation. Retrospective analyses of non-trial cohorts show that integration of a restricted set of mutations with copy number data refines and enhances the prognostic significance of the latter and suggest that mutations may be incorporated into future prognostic indices.[12] Furthermore, copy number array and next generation sequencing data inferred from a single time-point or from sequential studies have demonstrated intra-clonal heterogeneity in CLL, the prognostic significance of sub-clonal mutations and the selective pressure of therapy in determining clonal evolution.[13,14] Recent epigenetic data has identified three CLL subtypes that correlate with B-cell maturity and possess distinct patterns of somatic instability, degree of *IGHV* mutation, mutation risk profiles and clinical outcomes.[15–18] Despite this progress there remain patients who would be classified as 'low-risk' based on biomarkers who nevertheless have progressive disease.

To obtain more information about the genomic and epigenomic landscape and clinical significance of abnormalities in *IGHV-M* CLL (M-CLL), we performed a longitudinal study at or close to diagnosis, pre-treatment and post relapse in 13 patients. These patients presented with Binet Stage A disease ($n = 10$) or clinical monoclonal B-cell lymphocytosis (cMBL) with good-risk biomarkers (*IGHV-M*

[1]Academic Unit of Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK; [2]Centre for Biological Sciences, Faculty of Natural and Environmental Studies, University of Southampton, Southampton, UK; [3]Bioinformatics Unit, Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK; [4]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK; [5]Department of Haematology, Royal Bournemouth Hospital, Bournemouth, UK and [6]Division of Hematology, Department of Internal Medicine, The Ohio State University, Columbus, USA. Correspondence: Professor JC Strefford, Academic Unit of Cancer Sciences, Faculty of Medicine, Somers Cancer research Building, Southampton General Hospital, University of Southampton, Tremona Road, Southampton SO16 6YD, UK.
E-mail: JCS@soton.ac.uk
[7]These authors share first authorship.
Received 14 September 2015; revised 21 December 2015; accepted 4 January 2016; accepted article preview online 5 February 2016; advance online publication, 26 February 2016

genes, no del(17p) or del(11q) and low CD38 expression), 10 of whom subsequently required treatment. Using a combination of DNA methylation, copy number analysis, whole-exomic sequencing (WES), targeted deep sequencing (TDR) of recurrently mutated CLL driver genes, screening of non-coding mutation and immunogenetic analysis we identified the presence or acquisition of clonal or sub-clonal driver mutations and DNA methylation changes in eight cases and the emergence of a new immunogenetic clone in one case.

METHODS

Patient data, copy number and methylation analysis

We studied 13 patients diagnosed at the Royal Bournemouth Hospital between 1992–2007 as cMBL or Binet Stage A, Rai stage 0 CLL according to the 2008 IWCLL/NCI guidelines.[19] This study was approved by the local Research Ethics Committee and informed consent obtained according to the Declaration of Helsinki. *IGHV* sequencing, CD38, cytogenetic and fluorescence *in situ* hybridisation (FISH) analyses were performed as described[20–23] and only cases with mutated *IGHV* genes (excluding major stereotypes), low CD38 expression and lacking 11q or 17p deletion and the availability of stored material were included. Germline DNA (GL) was obtained from saliva (DNAgenotek). CD19+ B-cells were taken at a median of 1 year (0–7.3 years) from diagnosis when patients had cMBL or Stage A disease (time-point 1 (TP1)). The three patients (pts 1–3) who did not require treatment remained as Stage A with a rising ($n=2$) or stable lymphocyte count. All three were sampled again (TP2) at a median of 7 years[6–10] from TP1 and one was sampled at further TP (TP3) 3 years after TP2. In the 10 patients requiring treatment, a further sample was taken at a median of 4 months, (range: 0–42) pre-treatment (TP2). In all, 6/10 patients who relapsed after first-line treatment had a sample taken post relapse (TP3) and 2/6 patients were also sampled at relapse following subsequent treatments (TP4 and 5). For 13 sample-trios (GL, TP1 and TP2), DNA regions of copy number alteration and differential methylation were identified using SNP6 arrays (Affymetrix, Santa Clara, CA USA) and 450 K arrays (Illumina Inc., San Diego, CA, USA), respectively, as described.[7,24]

Sequencing

WES libraries were prepared from 13 sample-trios (GL, TP1 and TP2) as described.[25] TDR used Haloplex (Agilent Technologies, Santa Clara, CA, USA) as described[26] to capture single nucleotide variants (SNVs) identified by WES and 22 genes (exons and 5′ & 3′-UTRs) that are frequently mutated in CLL (Supplementary Table S1) in all tumour samples. TDR libraries were sequenced at high depth (average x4000) to detect mutation down to the 1% level. For each mutation detected by TDR, variant allele frequencies (VAFs) were adjusted for tumour purity estimated as %CD19+ cells. Clonal or sub-clonal mutations were further classified according to ref. 26. We subjected the TDR data to SciClone analysis[27] to define the clonal dynamics of mutation clusters into three types: (1) static: clusters remain the same over time. (2) Expanding: all mutations in a cluster increase over time. (3) Evolving: new mutations in later samples or one or more mutations in a cluster increase over time. Phylosub was used for tumour phylogeny analysis to predict the most likely order of mutation events and classify either linear or branching evolution patterns.[28] *PAX5* enhancer region was screened as described in ref. 11. We defined mutated genes into those recurrently mutated from previous CLL studies, non-coding mutation and genes mutated in other haematological malignancies (All excluding copy number changes). Supplementary Methods are available on-line.

RESULTS

Genomic landscape of progressive M-CLL

Clinical features, treatment regimens and the genomic landscape at multiple TPs are summarised in Table 1 and Figure 2.

We employed WES and TDR to identify somatically acquired mutation in tumour samples from 13 cases with mean coverage of 77x (min–max: 43–127) and 3681x (2142–5268), with $>86\%$ of all bases covered at >20x and >200x, respectively. (Figure 1, Supplementary Tables S2 and S3). Of the filtered WES variants (Supplementary Table S4), TDR confirmed the presence of 224/312 (72%) SNVs and 7/9 (78%) indels (when present at both TP1 and TP2), respectively (Supplementary Table S5 and Supplementary Methods). We used the TDR variants to study temporal clonal evolution and demonstrated that our TP1 and TP2 samples harboured a similar mutation burden, with on average 17 (min–max: 9–26) and 19,[8–29] respectively. After adjusting for tumour purity, we observed no difference in the mean number of clonal (6 vs 6) or sub-clonal mutations (10 vs 12) in either untreated TP (TP1 vs TP2). All reported VAFs (%VAF) are adjusted for tumour purity.

Focusing on genes previously shown to be recurrently mutated in CLL; at TP2, clonal mutations were detected in *MYD88* (p.L265P; pt-2), and *CHD2* (pt-1) among the three untreated patients and in *ATM* (pt-4), *DDX3X* (pt-13), *NOTCH1* (pts-6,13), *SF3B1* (pts-6, 8), *TP53* (pts-8, 9), *NFKBIE* (pt-5), *SPEN* (pt-9), *ZMYM3* (pt-6), *KLHL6* (pt-10), *BIRC3* (pt-13) and *IRF4* (pt-13) among the 10 patients who received treatment. Only five of these mutations were clonal (Figure 2). In addition, three patients (pts-3, 11, 12) exhibited missense mutations (damaging by Polyphen-2) in genes known to have a role in other haematological tumours, *LTF* (pt-3), *ITGA6* (pt-11) and a frame-shift in *TNFAIP3* (pt-12) and all were present at sub-clonal levels (11–42% VAF). Only one case (pt-7) lacked any recurrently mutated driver mutation documented in CLL or other haematological malignancies. However, WES did identify 10 mutated genes from which two candidates emerged, namely missense mutations in *ZBTB7C* a kidney cancer-related gene that interacts with p53 (ref. 29 and *S1PR4* a receptor expressed in hematopoietic cells that interacts with *MAPK3* (ERK1), placing it in the B-cell receptor pathway.[30]

The CLL driver mutations (with the exception of *BIRC3* in pt-13) were detectable (by the presence of one or more mutated reads) at TP1 supporting the hypothesis that identification of mutations at diagnosis may identify individuals later requiring therapy. The CLL driver gene mutations with VAFs $<1\%$ (*IRF4*, *NOTCH1, SF3B1 and TP53;* in pts-6, 8, 9, 13) at TP1 were ascertained by manual curation of the TDR sequencing reads (Supplementary Table S6) after being originally detected in later tumour TPs with a higher VAF, suggesting a larger sub-clonal population at progression. Pileup of reads across all samples provided statistical confidence for calling *TP53* mutation in patient 8 and 9 below a 1% VAF ($P=0.013$ & $P<0.001$; Supplementary Table S6). Droplet-digital PCR analysis of patient 13 confirmed the presence of the *NOTCH1* mutation at TP1 (Supplementary Figure S1). Together, this would suggest sequencing depths much $>$x4000 will be required to robustly identify all sub-clonal mutations, for example, patient 8 had a *TP53* mutation (29%VAF) at TP2, detectable at TP1 in 9/15581 reads (0.1%VAF), equating to the presence of one mutant cell in ~1000 CLL B-cells. In patient 13, the *BIRC3* mutation at TP2 was not identified at TP1 (0/3624 reads) and conversely a clonal *DDX3X* mutation at TP1 was detected as a small sub-clone at TP2. This lead to a re-appraisal of this case which is discussed in detail later. At relapse, we identified mutations in *SF3B1* (pt-6) and *TP53* (pt-9) with VAF's of 17 and 3.3%, respectively, which had VAF's of $<1\%$ pre-treatment.

We screened for non-coding mutations.[11] *PAX5* enhancer region mutations were detected in three patients, estimated at TP1 to be clonal in one case (pt-4) and sub-clonal in the other cases (pts-2, 6; Figure 2 and Supplementary Figure S2). These mutations co-occurred with other mutations: *MYD88* (pt-2), *ATM* (pt-4) and *NOTCH1, SF3B1* and *ZMYM3* (pt-6). We also detected at TP2, the presence of a sub-clonal mutation (6% VAF; chr17:56408615:T$>$C) in the mature sequence of hsa-mir-142, this co-occurred with a *ITGA6* mutation in patient 11 (Supplementary Figure S2). The *NOTCH1* 3′UTR mutations, previously observed solely in cases of U-CLL,[11] were absent.

Combining karyotypic, FISH and SNP6 data, at TP1, only two patients (pts-8, 10) had no copy number abnormality or translocation, while the remainder had mono ($n=5$) or mono +biallelic loss of 13q14. Two patients (pts-12, 13) with del13q also had trisomy 12. Two patients (pts-6, 9) had a complex genome ($\geqslant 3$ copy number alterations), defined as previously reported.[31] SNP6 confirmed the absence of 11q or 17p deletion (Supplementary Table S7). At TP2, additional abnormalities were detected in three

Table 1. Overview of patient biomarker and clinical data

Patient ID	*Age at diagnosis*	*Tumour time-point 1*				*Tumour time-point 2*				*LDT*	*First treatment (response)*	*Treatment at relapse (response)*	*Current status at last follow-up (03/8/15)*
		IGHV (% identity)	*FISH/karyo*	*%CD38 expression*	*Lymphocyte count (x10^9/L)*	*IGHV (% identity)*	*FISH/karyo*	*%CD38 expression*	*Lymphocyte count (x10^9/L)*				
1	52	IGHV4–61 (93)	del(13q) +/− (54%)	2	6	IGHV4–61 (93)	del(13q) +/−, −/− (10, 85%)	1	142	> 1year	–	–	Stable CLL
2	72	IGHV3–73 (91)	del(13q) 46, XY, der4(4) t(4;12)(q35; q13)	1	5	IGHV3–73 (91)	No change	1	25	> 1year	–	–	Stable CLL
3	57	IGHV2–70 (93)	del(13q) +/− (90%)	1	92	IGHV2–70 (93)	No change	1	88	> 1year	–	–	Stable CLL
4	61	IGHV4–59 (89)	del(13q) +/− (6%) 45, X-Y, t(7,13) (a11.2;q14)	15	34	IGHV4–59 (89)	del(13q) +/−, −/− (76, 8%), 45, X –Y, t(7;13) (q11.2;q14)	50	136	> 1year	BR (CR)	–	High-risk MDS. Died.
5	79	IGHV4–61 (92)	del(13q) +/−, −/− (9/86%) 46, XY, t(6;13) (q26;q14)	1	107	IGHV4–61 (92)	del(13q) +/−, −/− (14/86%) 46, XY, t(6;13)(q26;q14)	–	198	> 1year	Chlor (CR)	BR (PR)	Stable CLL
6	70	IGHV3–48 (97)	del(13q) +/− (88%)	1	20	IGHV3–48 (97)	del(13q) +/−, −/− (53, 10%)	2	127	6–12 months	Chlor (GR)	BR (CR)	In remission
7	47	IGHV4–34 (92)	del(13q) +/−, −/− (19, 72%)	6	59	IGHV4–34 (92)	No change	1	81	> 1year	Chlor R (PR)	Alemtuz (CR, MRD +ve)	In remission
8	59	IGHV3–23 (96)	Normal	3	21	IGHV3–23 (96)	del(13q) +/− (66%) 46, XY, del(9)(q21), t(12;15)(p11; q15)	1	48	> 1year	Chlor Of (CR)	–	In remission
9	74	IGHV3–7 (89)	del(13q) +/− (54%)	1	17	IGHV3–7 (89)	del(13q) +/− (91%) (+ del17p at TP3)	1	185	6–12 months	Chlor (PR)	BR (CR)	On Ibrutinib
10	56	IGHV3–48 (93)	Normal	5	181	IGHV3–48 (93)	No change	2	139	> 1year	Chlor (PR)	Continuum	Stable CLL
11	64	IGHV3–23 (91)	46, XY. No 13q FISH	1	58	IGHV3–23 (91)	NT	1	145	> 1year	B Of (CR)	–	In remission
12	63	IGHV4–34 (96)	47, XY,+12. No 13q FISH	1	42	IGHV4–34 (96)	NT	1	77	6–12 months	Chlor R (CR)	BR (CR, MRD +ve) Continuum	In remission
13	67	IGHV3–48 (92)	Tri 12 (2%) del(13q) −/− (55%)	9	18	IGHV3–48 (92) & IGHV5-10-1*01 (100)	Tri 12 (73%) (Tri 12 (75%) at TP3)	21	158	> 1year	Chlor (NR)	BR (PR), Of (PR)	Richters syndrome, NR to CHOP of. Died.

Abbreviations: BR, bendamustine plus rituximab; FISH, fluorescence *in situ* hybridisation; LDT, lymphocyte doubling time (from diagnosis for the first year of follow-up); NT, not tested.

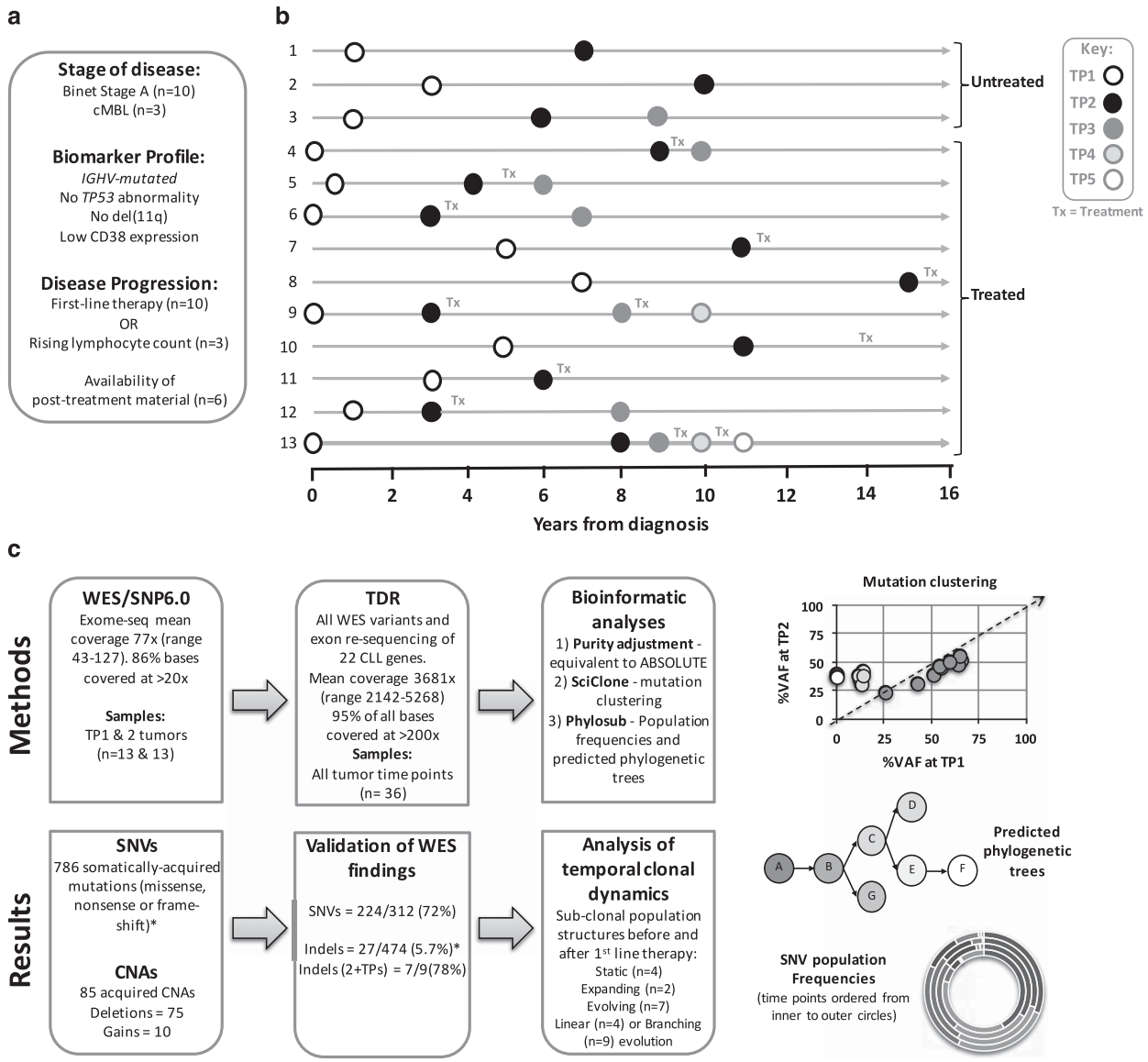


Figure 1. Study overview. (**a**) Inclusion criteria for study and definition of disease progression. (**b**) Tumour TP sampling time line for the 13 patients. Tx, treatment. (**c**) Flow diagram describing genomic analyses and result summaries. Example data plots for SciClone mutation clustering, Phylosub phylogenetic trees and concentric pie charts (each layer, inner to outer, is a sampling TP) displaying imputed SNV population frequencies at each phylogenetic node. *For indel filtering we accepted a high-false positive WES rate to ensure we could capture all of the 'true' somatically acquired indel variants by TDR (Supplementary Methods). When considering indels present in two or more tumour TPs (2+TPs) our indel TDR validation rate (78%, 7/9%) was in line with the SNV rate (72%).

patients (pts-4, 5, 8). Interestingly, the complex karyotypic abnormality in patient 8 was associated with expansion of a sub-clonal *TP53* mutation (from 0.1% at TP1 to 29% at TP2) without *TP53* loss. Three patients (pts-2, 4, 5) had an unbalanced translocation. Patient 13 had a remarkable change in copy number and is discussed later.

Intra-clonal heterogeneity in progressive M-CLL

Our longitudinal approach provided an opportunity to evaluate intra-clonal heterogeneity both before and following therapy.

SciClone analysis of TP1 and TP2 data from patients before therapy enabled us to make the following observations: (1) Four cases (pts-5, 7, 10, 12) had a static sub-clonal structure with mutation clusters present at similar VAFs at both TPs (pt-12 in Supplementary Figure S3A). (2) Two cases (pts-2, 4) had an expanding population where all mutations in a cluster were more dominant at TP2 (pt-4 in Supplementary Figure S3A). (3) Seven patients (pts-1, 3, 6, 8, 9, 11, 13) had an evolving genome where new mutations appeared ($n=8$ in four patients (pts-6, 8, 9, 13) and/or one or more mutations in a cluster increased, at TP2 (pt-6 in Supplementary Figure S3A). Six of these new mutations had low read depths ($<\times 4000$; ranging: 138–3642) in TP1 samples, suggesting there may be a lack of detection sensitivity. The three remaining mutations (*DSG4*, *SIM1* and *SLC8A2*) had adequate depth (4218–9043) at TP1, suggesting these mutations are either very rare (in <0.5–1% of cells) or represent acquired mutations at disease progression (TP2). For the six patients with post therapy TPs, there were no new mutations and we observed two patients (pts-5, 12) exhibiting a static structure with same distribution of sub-clones pre- and post-treatment (pt-12 in Supplementary Figure S3B), while four cases (pts-4, 6, 9, 13) showed expansion (pt-4 in Supplementary Figure S3B) or evolution (pt-6 in Supplementary Figure S3B) following an apparent therapy-related sweep selecting

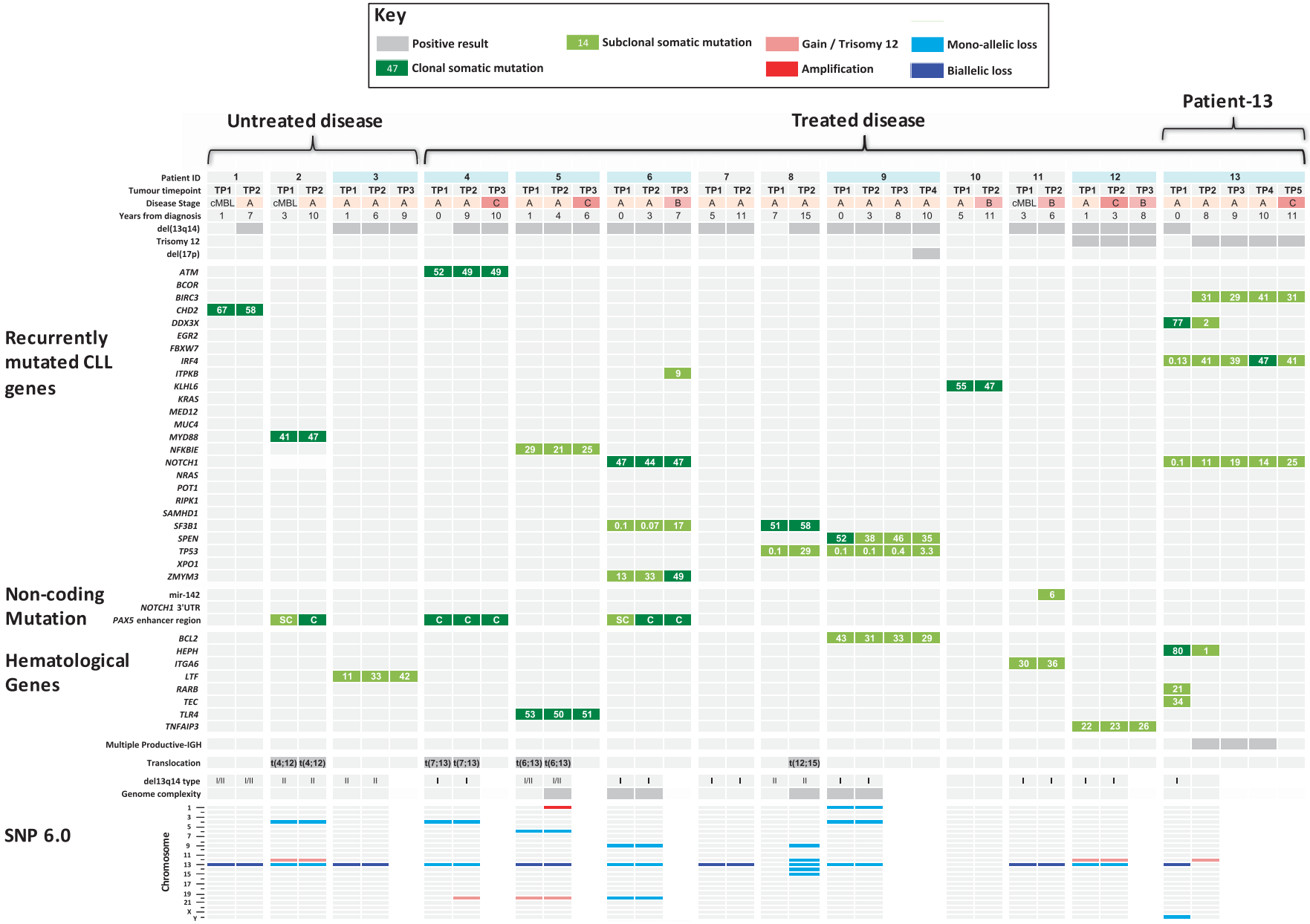


Figure 2. Heat-map representation of tumour TPs analysed by WES and targeted deep re-sequencing. From top to bottom: key to heat-map cell shading. Patient characteristics, light blue cell shading indicates patients with follow-on tumour samples (that is, TP3, TP4 and TP5); dark-grey cells indicated a positive result. Sub-clonal (light-green cells) and clonal mutations (dark-green cells) in each patient, grouped into recurrently mutated CLL driver genes, non-coding mutation described in Puente *et al*.[6] and genes mutated in haematological malignancies. Numbers in cells denote tumour purity-adjusted %VAFs from TDR. SC, sub-clonal; C, clonal from Sanger-seq traces. Presence of multiple productive-*IGH* relating to patient 13, chromosomal translocation or genome complexity is denoted by dark-grey cells. SNP6.0 data for TP1 and TP2 samples.

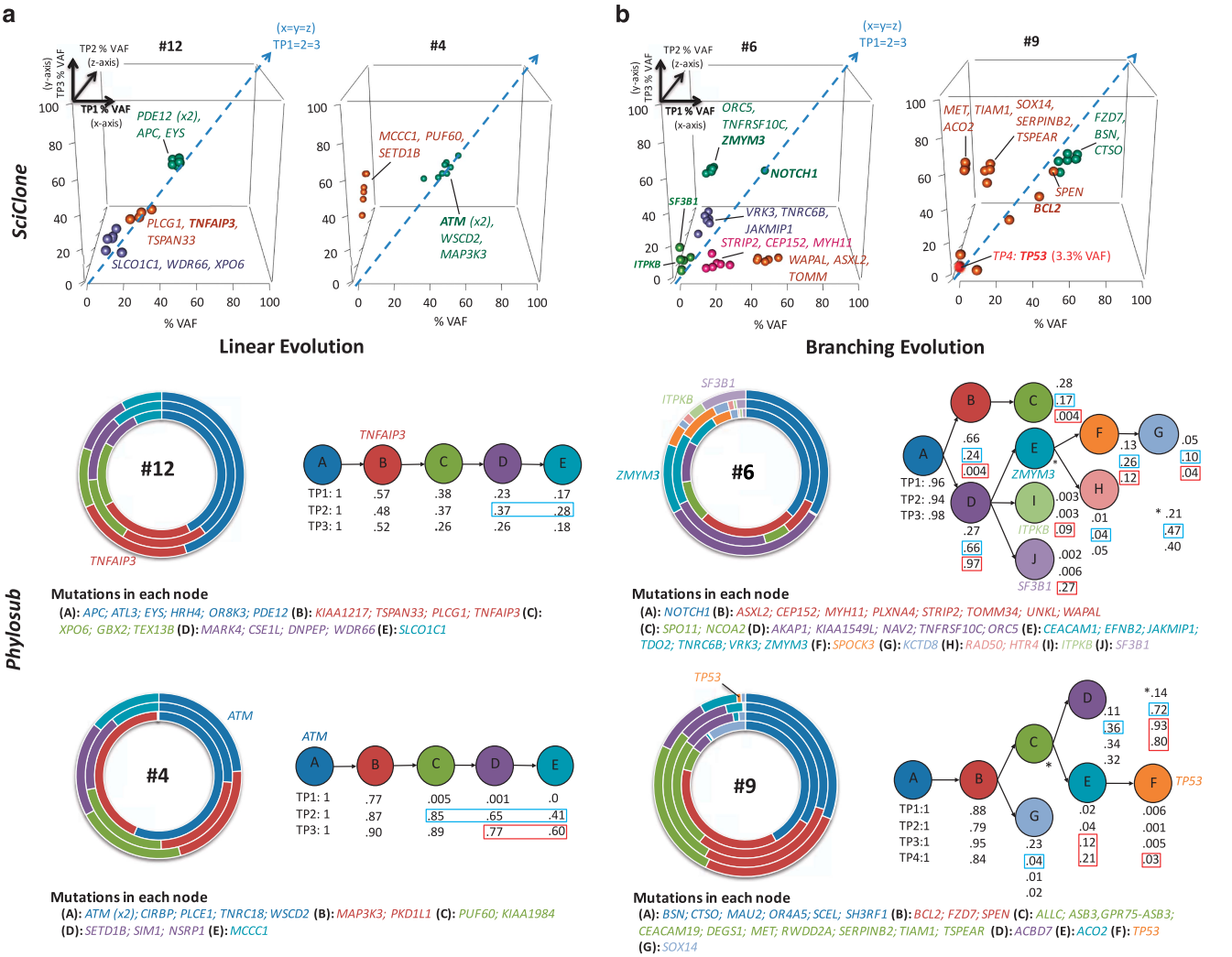


Figure 3. PhyloSub analysis of TDR results in predicted linear and branching clonal evolutionary pattern. By columns (**a**) two patient examples (pts-12 and 4) of linear evolution path, (**b**) two patient examples (pts-6 and 9) of complex branching trajectories. (**a** and **b**) Top panel: XYZ scatter-graphs displaying the SciClone mutation clustering analysis on TDR data sets from sequential tumour TPs TP1 (*x* axis; first tumour sample), 2 (*z* axis; progression) and 3 (*y* axis; post-treatment). Data point symbols denote a distinct mutation cluster and the $x=y=z$ line is displayed as a dashed blue arrow and denotes no change in the tumour purity-adjusted %VAF of mutation clusters between TPs (clonal equilibrium). Selected gene symbols are displayed adjacent to its corresponding mutation cluster. (**a** and **b**) Bottom panel: concentric pie charts (each layer, inner to outer, is an early to later sampling TP) displaying imputed SNV population frequencies at each phylogenetic node. Predicted phylogenetic tree structure (best model shown), with population frequencies for each node from Phylosub analysis. Blue and red boxes denote large changes SNV population frequencies before and after first-line treatment, suggesting ongoing clonal dynamics and selection by therapy, respectively. Selected gene symbols are displayed adjacent to the corresponding segment of the pie chart or phylogenetic node.

resistant/fitter sub-clones. Interestingly, these four cases had either expanding or evolving mutation clusters before therapy. Results for the remaining patients are provided in Supplementary Figure S4.

Phylosub analysis predicted a linear evolutionary path, where progeny replaced ancestral clones, in a minority of patients ($n=4$) in whom SciClone analysis had identified either static (pts-7, 12) or expanding (pts-2, 4) mutation clusters. Phylosub also predicted that the *ATM*, *MYD88, S1PR4, TNFAIP3* and *ZBTB7C* mutations in these patients were early evolutionary events (placed in the first or second nodes of each tree). Complex branching trajectories were predicted in the remaining nine patients, including two of the three patients (pts-1, 3) with no indication for therapy, and provided the following insights: (1) The *BCL2*, *CHD2*, *NOTCH1*, *SF3B1* and *TLR4* mutations were all predicted to be early events. (2) Generally located at branch points, the sub-clones that appear to have good fitness, or are selected for at later tumour TPs, contained CLL drivers (*ITPKB*, *NFKBIE*, *SF3B1*, *TP53* and *ZMYM3*) or genes mutated in other haematoloical malignancies (*ITGA6 and LTF*) supporting the role of genes in the latter two categories as candidate drivers of progression in those patients. (3) Convergent evolution was only found in two patients (pts-4, 8) who exhibited two mutations in a single gene (*ATM*:p.I2606M/p.Q2733K and *IGLL5*:p.C31Y/p.P50S). All four mutations were clonal and only the *IGLL5* mutations were close enough to be detected on separate overlapping reads, but as the VAFs were similar (54–57%) they could have arisen in the same cell.

Phylosub and SciClone analysis of patients-4, 6, 9 and 12 is displayed in Figure 3. These four patients had three (pts-4, 6, 12) or four (pt-9) tumour TPs presenting static, expanding or evolving mutation clusters with predicted linear (pts-4, 12) or branching (pts-6, 9) evolution trajectories before and after therapy. At TP2 in patient 4 we observe expanding populations (nodes C–E)

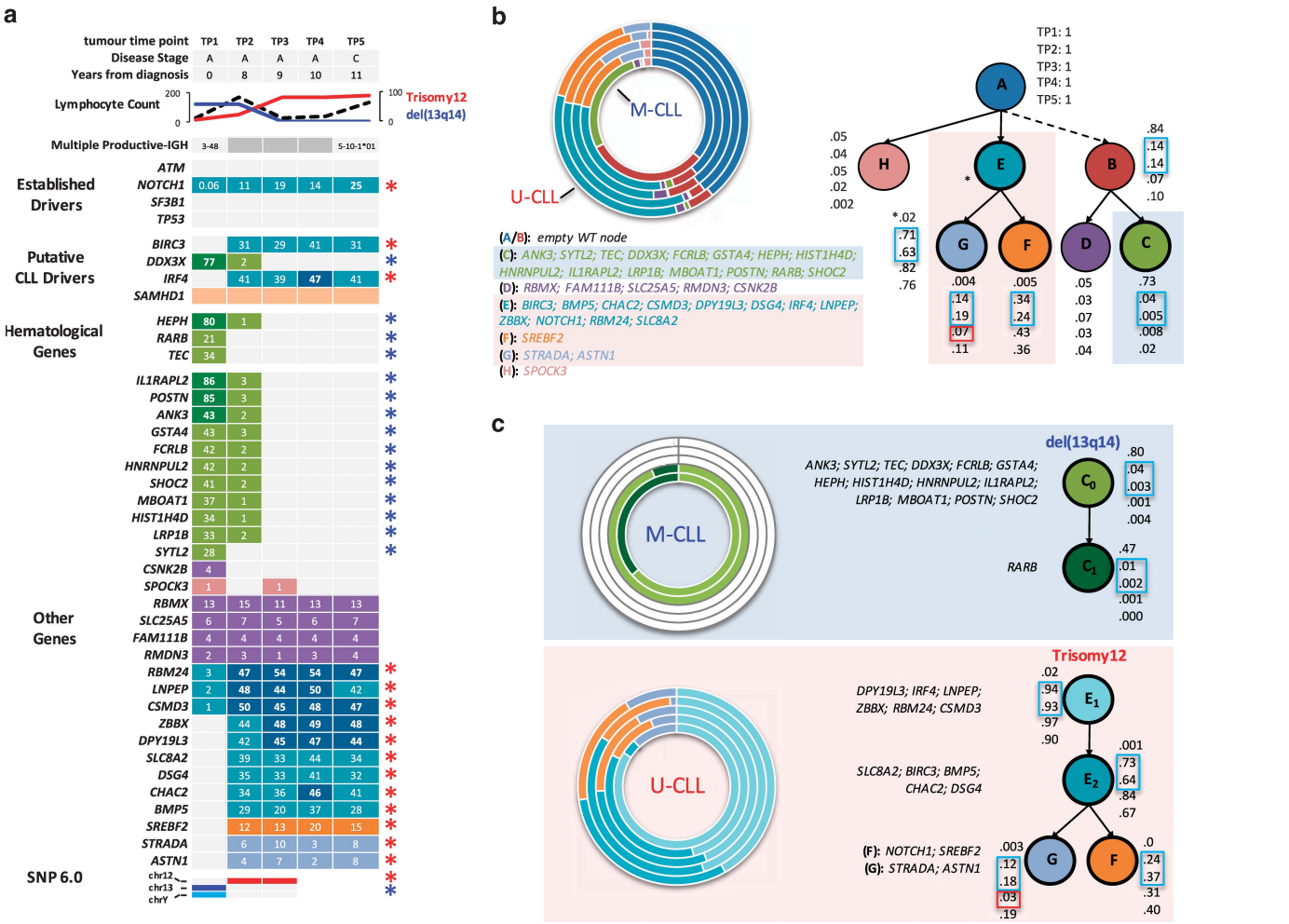


Figure 4. Evolution of multiple productive-*IGH* in CLL patient 13. (**a**) From top to bottom: five tumour TPs with corresponding clinical, cytogenetic and immunogenetic data. Mutation heat-map representation of five tumour TPs analysed by targeted deep re-sequencing. Numbers in cells denote tumour purity-adjusted %VAFs from TDR. Cell colours are linked to the SNV population nodes/frequencies displayed in part **b**. Lighter shading indicates a sub-clonal mutation. Blue asterisks/del13q14 = M-CLL clone (IGHV3–48; 92% identity to germ-line; del (13q14)) and red asterisks/trisomy 12 = U-CLL clone (IGHV5–10*01; 100% identity to germ-line; trisomy 12). (**b** and **c**) Filled light blue and red boxes denote mutations and cytogenetic abnormalities inferred into the M-CLL and U-CLL clone, respectively. From left to right: concentric pie charts (each layer, inner to outer, is an early to later sampling TP) displaying imputed SNV population frequencies at each phylogenetic node. Predicted phylogenetic tree structure, with population frequencies for each node from Phylosub analysis. Best models are displayed for analyses using all mutations (**b**) or only mutations associated with either the M-CLL or U-CLL clone providing insights into the probable order of mutation (**c**). Note from TP3 onwards the mutations associated with the M-CLL clone are not detectable by sequencing. Open blue and red boxes denote large changes SNV population frequencies before and after first-line treatment, suggesting ongoing clonal dynamics and selection by therapy, respectively.

replacing the ancestral population, suggesting that these later mutations are associated with disease progression. Following treatment with bendamustine plus rituximab (TP3), we observed no reduction in population frequencies. Conversely, in patient 12, we observed a reduction back to baseline (TP1) for nodes C–E, suggesting a similar sensitivity of the descendant clones to chlorambucil–rituximab therapy; this patient remains in remission. Interestingly, in patient 6 the *ITPKB*, *SF3B1* and *ZMYM3* mutations were predicted to be in distinct populations (nodes I, J and E, respectively) and following chlorambucil treatment at TP3 the population frequencies of both the I (*ITPKB*) and J (*SF3B1*) nodes increased in comparison to other nodes. In patient 9 following two rounds of therapy (chlorambucil and bendamustine plus rituximab) we observed 75% del(17p) loss by FISH and a sub-clonal *TP53* mutation (p.Y234C, 3.3%VAF), at TP4. A complex karyotype was also observed at TP3. The difference between del(17p) FISH clone size and %VAF of the *TP53* mutation would suggest evolutionary independent events, with a rare *TP53* mutated clone detectable at presentation and later acquisition of del(17p) loss in another *TP53*-wildtype clone. Following exposure to chemo-immunotherapy, the resistant *TP53* aberrant clones accumulate and dominate the tumour. Phylosub results for the remaining patients are provided in Supplementary Figure S4.

The emergence of an *IGHV-U* immunogenetic clone can drive progression

Substantial differences in the clone size of copy number alterations and mutations between TP1 (diagnosis) and TP2 (+8 years) in patient 13, and the detection of a *BIRC3* mutation only at TP2, led to a review of karyotypic, FISH, SNP6 and mutational data and targeted re-sequencing analysis of samples taken after treatment with chlorambucil, bendamustine plus rituximab and ofatumumab, all of which were ineffective. This

showed a remarkable temporal shift in genomic aberrations supporting a dominant population at diagnosis containing a deletion of 13q, loss of chromosome Y and a unique set of mutations, including *DDX3X, HEPH, RARB* and *TEC*. These were gradually replaced by a 47, XY, trisomy 12, population with mutations in 18 genes including *BIRC3*, *NOTCH1* and *IRF4* (Figure 4 and Supplementary Table S8). As the dominant mutations present at TP2 are more frequently or exclusively associated with *IGHV-U* genes, we reanalyzed the *IGHV* status at TP2, and identified a dominant *IGHV5-10-1*01* (100% identity to germ-line) clone in addition to the *IGHV3–48* clone with 92% germ-line identity present at diagnosis. *IGHV* analysis of six intermediate samples between TP1 and 2 detected the *IGHV-U* clone as far back as 4 years post diagnosis (Supplementary Table S9). Importantly, both the *NOTCH1* mutated and trisomy 12 sub-clones were detectable at diagnosis using TDR and FISH, respectively, demonstrating the presence of the *IGHV5-10-1*01* clone at diagnosis but at a level which was undetectable using standard immunogenetic assays. This patient's tumour eventually transformed to a diffuse large B-cell lymphoma, using the *IGHV5-10-1*01* clone. Unfortunately, TDR of the Richter's node biopsy was not successful.

Additional *IGHV* sequencing on the other cases failed to identify any additional patients with evolution of an *IGHV-U* clone (Supplementary Table S9).

DNA methylation subtyping and co-evolution of epigenetic changes

We performed clustering analysis of the TP1 methylation data together with a reference sample set where the three epigenetic subtypes were defined previously (Oakes, in press)[32] and determined that 12/13 patients belonged to the high-programmed CLL DNA methylation subtype, consistent with the selection of our patients based on the presence of mutated *IGHV* (Supplementary Figure S5A). In all except a single patient, adjacent clustering of the TP2 data confirmed the clonal relationship between tumour TPs. The exception was patient 13 which clustered in the low-programmed CLL subtype at TP2, consistent with the emergence of the *IGHV-U* clone. Patients with limited genetic evolution had relatively few differences in overall methylation, where as those that exhibited either expansion or evolution of genetic sub-clones showed higher proportions of altered CpG methylation (Supplementary Figure S5B).

DISCUSSION

Patients with *IGHV-M* genes, defined as $< 98\%$ identity to the germ-line sequence, have a better outcome than those with *IGHV-U* genes.[33,34] While stable cMBL and Stage A CLL are strongly enriched for cases with M-CLL, 37–39% of patients entered into the UKCLL4 and CLL8 trials of first-line therapy had M-CLL[8,35] and the key biological features responsible for progression are still poorly understood.[11] M-CLL is biologically heterogeneous and studies have shown that CD38, CD49d and ZAP70 expression, serum markers, stereotypic subset-2,[36] telomere length, del(11q) and del (17p) and genomic abnormalities influence time to first treatment or outcome following therapy.[37–42] In this study, we performed longitudinal genomic and epigenomic characterisation before and after therapy in a cohort of M-CLL cases presenting with Stage A disease or cMBL. All cases lacked established biomarkers associated with progression: namely, high CD38 expression, del(11q), del(17p). Nevertheless, 10/13 cases subsequently required treatment.

Before treatment we found mutations in genes that are recurrently mutated in CLL, (*ATM*, *BIRC3*, *CHD2*, *DDX3X*, *IRF4*, *ITPKB*, *KLHL6*, *MYD88, NOTCH1*, *NFKBIE*, *SF3B1*, *SPEN*, *TP53 and ZMYM3*) in nine patients (69%) with a mean of one mutation per case (min–max: 0–4). This mutation incidence is consistent with a recent whole genomic and exomic study, where 83% of M-CLL cases had a driver mutation.[11] Nine patients had one or more clonal mutations detected pre-treatment and the majority of the sub-clonal mutations were detectable at the earliest TP at or soon after diagnosis. Although many of the above genes are associated with disease progression and/or resistance to treatment, the clinical significance of others (*ITPKB, KLHL6 and SPEN*) is less certain. A further three patients had mutations in other genes (*LTF*, *ITGA6* and *TNFAIP3*) implicated in other haematological malignancies (COSMIC v 73;[43]).

We also screened for non-coding mutation in the *PAX5* enhancer region. These mutations co-existed with *MYD88*, *ATM, NOTCH1*, *SF3B1* and *ZMYM3* mutations in contrast to the original description where they were either the sole recurrent mutation or occurred in conjunction with 13q loss.[11]

SciClone and Phylosub analysis provided novel insight into the extent of intra-clonal heterogeneity. Clonal expansion or evolution, predominantly in a branching pattern, was found in nine cases pre-treatment, indicating that sub-clonal competition occurs in the absence of selective pressure through therapy, with the resistant/fitter sub-clones dominating the tumour at post therapy TPs. A recent WES study comparing matched pre-treatment and relapse samples demonstrated that clonal evolution was the rule after therapy and the resistant clone could be detected before treatment in ~ 30% of cases.[14] Six cases developed isolated splenomegaly, two had splenomegaly and lymphadenopathy and two had lymphadenopathy before treatment. Further spatial-temporal studies will be required to determine the site(s) of clonal evolution. Phylosub analysis also demonstrated the selection of sub-clones containing either *SF3B1*, *TP53, ITGA6*, *ITPKB*, *LTF*, *NFKBIE*, or *ZMYM3* mutations, supporting their role as candidate drivers of progression.

One unexpected finding was the emergence of an *IGHV-U* clone in one patient who after 6 years of stable disease, developed progressive, therapy-refractory disease and culminating in a clonally related lymphomatous transformation. Each clone had a unique spectrum of gene mutations and epigenetic profiles consistent with two distinct and competing leukaemic clones originating from a different pool of lymphocyte progenitors. Bi, or more rarely multi-clonal *IGHV* rearrangements have been documented. From a cohort of 1147 cases, Plevova *et al*. identified seven cases with both mutated and unmutated clones in which serial studies showed diminution of an *IGHV-M* clone with persistence of a co-existing *IGHV-U* clone, resulting in re-classification to U-CLL.[44] Clinically, this was associated with progressive lymphocytosis, disease progression and in some cases, the selection of a *TP53*-defect post therapy. Our case is unusual in that the unmutated clone was not detectable until 4 years after diagnosis using standard methodologies for *IGHV* sequencing, even though a sub-clonal trisomy 12 and *NOTCH1* mutation, associated with the unmutated clone were detected at diagnosis using more sensitive techniques.

Previous whole-genome longitudinal studies of copy number[45–47] and/or genomic mutations[48–51] in CLL have included a higher percentage of cases with U-CLL than M-CLL and with progressive rather than stable disease. However, a picture has emerged of clonal evolution which is usually branched rather than linear and is more frequent in cases with progressive disease who have required treatment for which the majority of driver mutations can be detected at initial testing, often as small sub-clones. Our study confirms many of these findings but also highlights the high frequency of CLL driver mutations in a progressive cohort miss-labelled as 'good-risk' and the extent of clonal evolution before therapy. DNA methylation analysis revealed that co-evolution of genetic and epigenetic changes is a prominent feature and that this exists regardless of *IGHV* subtype and mutational risk assessment, supporting the perspective that evolution is an important predictor of disease progression.[13,14] A recent study[52] also found the highest number of differentially methylated CpGs were in cases with genetically evolving and expanding sub-clones.

From a clinical perspective, a key question is whether the additional information that genomic and epigenomic screening provides in this group is likely to improve patient outcome. Although it would be unwise to draw general conclusions from this small study, it does offer three examples where screening could have clinical utility. Firstly, 2/13 cases had small *TP53* mutated clones early in the disease with evidence of clonal selection post therapy. These cases had no *TP53* loss detectable by FISH when the mutant clone was initially detected. Secondly, 6/11 cases with 13q loss fell into the 'very-low' risk category defined by Rossi *et al.* in which predominantly *IGHV-M* cases with isolated 13q loss, lacking mutations in *TP53, BIRC3, NOTCH1, SF3B1* and *MYD88* had a prolonged time to first treatment and an expected overall survival similar to the matched general population.[12] Two of our six cases with isolated 13q loss had a progressive lymphocytosis with no indication for treatment and a mutation in *CHD2* or *LTF* while three of the four who required treatment had mutations in *NFKBIE, TNFAIP3* or *ITGA6* suggesting that more extensive screening may aid in the differentiation of cases destined to have stable or progressive disease. Finally, the emergence of an *IGHV-U* clone, not evident at diagnosis, which eventually lead to the patient's death is a rare event but supports either repeat immunogenetic analysis in cases with unexplained progressive or therapy-refractory disease, or the use of more sensitive assays capable of detecting small clones. In summary this study does not evaluate the role of other factors such as cell signalling as an explanation for disease progression, but does support the role for sequential genomic/epigenomic screening as a means of identifying potential driver mutations and predicting progressive disease.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was funded by The Kay Kendall Leukaemia Fund (KKL 584) and Wessex Medical Research (Innovation fund 2011 R06). We thank the patients for supplying tissue and the infrastructure support from a CR-UK centre grant (C34999/A18087), Leukaemia and Lymphoma Research grant (12021) and ECMC grant (C24563/A15581). This work was funded by grants awarded to JCS and MJJRZ.

AUTHOR CONTRIBUTIONS

MJJRZ, JF, HP, ML and HP performed the experimental work; AG and HP performed the molecular diagnostic assays. MJJRZ, JG and AC conducted the exome analyses; CO analysed the methylation data; DGO contributed patient samples and data. DGO, MJJRZ and JCS initiated and designed the study; MJJRZ, JG, DGO and JCS wrote the manuscript with contributions from AC and CO; and all authors critically reviewed the final manuscript.

REFERENCES

1 Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T *et al.* Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood* 2014; **124**: 49–62.
2 Skowronska A, Austen B, Powell JE, Weston V, Oscier DG, Dyer MJ *et al.* ATM germline heterozygosity does not play a role in chronic lymphocytic leukemia initiation but influences rapid disease progression through loss of the remaining ATM allele. *Haematologica* 2012; **97**: 142–146.
3 Clifford R, Louis T, Robbe P, Ackroyd S, Burns A, Timbs AT *et al.* SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood* 2014; **123**: 1021–1031.
4 Rossi D, Khiabanian H, Spina V, Ciardullo C, Bruscaggin A, Famà R *et al.* Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood* 2014; **123**: 2139–2147.
5 Lionetti M, Fabris S, Cutrona G, Agnelli L, Ciardullo C, Matis S *et al.* High-throughput sequencing for the identification of NOTCH1 mutations in early stage chronic lymphocytic leukaemia: biological and clinical implications. *Br J Haematol* 2014; **165**: 629–639.
6 Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med* 2011; **365**: 2497–2506.
7 Jeromin S, Weissmann S, Haferlach C, Dicker F, Bayer K, Grossmann V *et al.* SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia* 2014; **28**: 108–117.
8 Stilgenbauer S, Schnaiter A, Paschka P, Zenz T, Rossi M, Döhner K *et al.* Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood* 2014; **123**: 3247–3254.
9 Baliakas P, Hadzidimitriou A, Sutton LA, Rossi D, Minga E, Villamor N *et al.* Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia* 2015; **29**: 329–336.
10 Guièze R, Robbe P, Clifford R, de Guibert S, Pereira B, Timbs A *et al.* Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. *Blood* 2015; **126**: 2110–2117.
11 Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI *et al.* Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* 2015; **526**: 519–524.
12 Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C *et al.* Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* 2013; **121**: 1403–1412.
13 Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS *et al.* *Cell* 2013; **152**: 714–726.
14 Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J *et al.* Mutations driving CLL and their evolution in progression and relapse. *Nature* 2015; **526**: 525–530.
15 Kulis M, Heath S, Bibikova M, Queirós AC, Navarro A, Clot G *et al.* Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet* 2012; **44**: 1236–1242.
16 Queirós AC, Villamor N, Clot G, Martinez-Trillos A, Kulis M, Navarro A *et al.* A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia* 2015; **29**: 598–605.
17 Oakes CC, Claus R, Gu L, Assenov Y, Hüllein J, Zucknick M *et al.* Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer Discov* 2014; **4**: 348–361.
18 Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H *et al.* Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* 2014; **26**: 813–825.
19 Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H *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**: 5446–5456.
20 van Dongen JJ, Langerak AW, Brüggemann M, Evans PA, Hummel M, Lavender FL *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003; **17**: 2257–2317.
21 Best OG, Ibbotson RE, Parker AE, Davis ZA, Orchard JA, Oscier DG. ZAP-70 by flow cytometry: a comparison of different antibodies, anticoagulants, and methods of analysis. *Cytometry B Clin Cytom* 2006; **70**: 235–241.
22 Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK, Unmutated Ig V(H. genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* 1999; **94**: 1848–1854.
23 Smoley SA, Van Dyke DL, Kay NE, Heerema NA, Dell' Aquila ML, Dal Cin P *et al.* Standardization of fluorescence in situ hybridization studies on chronic lymphocytic leukemia (CLL) blood and marrow cells by the CLL Research Consortium. *Cancer Genet Cytogenet* 2010; **203**: 141–148.
24 Parker H, Rose-Zerilli M, Parker A, Chaplin T, Chen X, Wade R *et al.* 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. *Leukemia* 2011; **25**: 489–497.
25 Parry M, Rose-Zerilli MJ, Gibson J, Ennis S, Walewska R, Forster J *et al.* Whole exome sequencing identifies novel recurrently mutated genes in patients with splenic marginal zone lymphoma. *PLoS One* 2013; **8**: e83244.
26 Parry M, Rose-Zerilli MJJ, Ljungström V, Gibson J, Wang J, Walewska R *et al.* Genetics and Prognostication in Splenic Marginal Zone Lymphoma: Revelations from Deep Sequencing. *Clin Cancer Res.* 2015; **21**: 4174–4183.
27 Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL *et al.* SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput Biol* 2014; **10**: e1003665.
28 Jiao W, Vembu S, Deshwar AG, Stein L, Morris Q. Inferring clonal evolution of tumors from single nucleotide somatic mutations. *BMC Bioinformatics* 2014; **15**: 35.
29 Jeon BN1 Kim MK, Choi WI, Koh DI, Hong SY, Kim KS *et al.* KR-POK interacts with p53 and represses its ability to activate transcription of p21WAF1/CDKN1A. *Cancer Res* 2012; **72**: 1137–1148.

30 Seitz G, Yildirim S, Boehmler AM, Kanz L, Möhle R. Sphingosine 1-Phosphate (S1P) Induces Migration and ERK/MAP-Kinase-Dependent Proliferation in Chronic Lymphocytic Leukemia (B-CLL) Due to Expression of the G Protein-Coupled Receptors S1P1/4. *Blood (ASH Annual Meeting Abstracts)* 2005; **106**, Abstract 4996.
31 Ouillette P, Fossum S, Parkin B, Ding L, Bockenstedt P, Al-Zoubi A *et al.* Aggressive chronic lymphocytic leukemia with elevated genomic complexity is associated with multiple gene defects in the response to DNA double-strand breaks. *Clin Cancer Res* 2010; **16**: 835–847.
32 Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowitz M, Ruppert AS *et al.* DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat Genet* 2015; e-pub ahead of print 18 January 2016; doi:10.1038/ng.3488.
33 Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL *et al.* Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* 1999; **94**: 1840–1847.
34 Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* 1999; **94**: 1848–1854.
35 Catovsky D, Richards S, Matutes E, Oscier D, Dyer MJ, Bezares RF *et al.* Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. *Lancet* 2007; **370**: 230–239.
36 Strefford JC, Sutton LA, Baliakas P, Agathangelidis A, Malčíková J *et al.* Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. *Leukemia* 2013; **27**: 2196–2199.
37 Del Giudice I, Morilla A, Osuji N, Matutes E, Morilla R, Burford A *et al.* Zeta-chain associated protein 70 and CD38 combined predict the time to first treatment in patients with chronic lymphocytic leukemia. *Cancer* 2005; **104**: 2124–2132.
38 Bomben R, Dal-Bo M, Benedetti D, Capello D, Forconi F, Marconi D *et al.* Expression of mutated IGHV3-23 genes in chronic lymphocytic leukemia identifies a disease subset with peculiar clinical and biological features. *Clin Cancer Res* 2010; **16**: 620–628.
39 Lin TT, Norris K, Heppel NH, Pratt G, Allan JM, Allsup DJ *et al.* Telomere dysfunction accurately predicts clinical outcome in chronic lymphocytic leukaemia, even in patients with early stage disease. *Br J Haematol* 2014; **167**: 214–223.
40 Strefford JC, Kadalayil L, Forster J, Rose-Zerilli MJJ, Parker A, Lin TT *et al.* Telomere length predicts progression and overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. *Leukemia* 2015; **29**: 2411–2414.
41 Skowronska A, Parker A, Ahmed G, Oldreive C, Davis Z, Richards S *et al.* Biallelic ATM inactivation significantly reduces survival in patients treated on the United Kingdom Leukemia Research Fund Chronic Lymphocytic Leukemia 4 trial. *J Clin Oncol.* 2012; **30**: 4524–4532.
42 Gonzalez D, Martinez P, Wade R, Hockley S, Oscier D, Matutes E *et al.* Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. *J Clin Oncol* 2011; **29**: 2223–2229.
43 Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* 2004; **91**: 355–358.
44 Plevova K, Francova HS, Burckova K, Brychtova Y, Doubek M, Pavlova S *et al.* Multiple productive immunoglobulin heavy chain gene rearrangements in chronic lymphocytic leukemia are mostly derived from independent clones. *Haematologica* 2014; **99**: 329–338.
45 Gunnarsson R, Mansouri L, Isaksson A, Göransson H, Cahill N, Jansson M *et al.* Array-based genomic screening at diagnosis and during follow-up in chronic lymphocytic leukemia. *Haematologica* 2011; **96**: 1161–1169.
46 Knight SJ, Yau C, Clifford R, Timbs AT, Sadighi Akha E, Dréau HM *et al.* Quantification of subclonal distributions of recurrent genomic aberrations in paired pre-treatment and relapse samples from patients with B-cell chronic lymphocytic leukemia. *Leukemia* 2012; **26**: 1564–1575.
47 Braggio E, Kay NE, VanWier S, Tschumper RC, Smoley S, Eckel-Passow JE *et al.* Longitudinal genome-wide analysis of patients with chronic lymphocytic leukemia reveals complex evolution of clonal architecture at disease progression and at the time of relapse. *Leukemia* 2012; **26**: 1698–1701.
48 Ouillette P, Saiya-Cork K, Seymour E, Li C, Shedden K, Malek SN. Clonal evolution, genomic drivers, and effects of therapy in chronic lymphocytic leukemia. *Clin Cancer Res* 2013; **19**: 2893–2904.
49 Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R *et al.* Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* 2012; **120**: 4191–4196.
50 Ojha J, Secreto C, Rabe K, Ayres-Silva J, Tschumper R, Dyke DV *et al.* Monoclonal B-cell lymphocytosis is characterized by mutations in CLL putative driver genes and clonal heterogeneity many years before disease progression. *Leukemia* 2014; **28**: 2395–2398.
51 Ojha J, Ayres J, Secreto C, Tschumper R, Rabe K, Van Dyke D *et al.* Deep sequencing identifies genetic heterogeneity and recurrent convergent evolution in chronic lymphocytic leukemia. *Blood* 2015; **125**: 492–498.
52 Smith EN, Ghia EM, DeBoever CM, Rassenti LZ, Jepsen K, Yoon KA *et al.* Genetic and epigenetic profiling of CLL disease progression reveals limited somatic evolution and suggests a relationship to memory-cell development. *Blood Cancer J* 2015; **5**: e303.

Supplementary Information accompanies this paper on the *Leukemia* website (http://www.nature.com/leu)

L Eder-Azanza[1,2], P Evans[3], C Wickham[4], S Akiki[5], JL Vizmanos[2], A Chase[1,6] and NCP Cross[1,6]
[1]*Wessex Regional Genetics Laboratory, Salisbury NHS Foundation Trust, Salisbury, UK;*
[2]*Department of Biochemistry & Genetics, School of Science, University of Navarra, Pamplona, Spain;*
[3]*Haematological Malignancy Diagnostic Service, St James's Institute of Oncology, Bexley Wing, St James's University Hospital, Leeds, UK;*
[4]*Molecular Genetics Department, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK;*
[5]*West Midlands Regional Genetics Laboratory, Birmingham Women's NHS Foundation Trust, Birmingham, UK and*
[6]*Faculty of Medicine, University of Southampton, Southampton, UK*
E-mail: ncpc@soton.ac.uk

REFERENCES

1 Tefferi A, Pardanani A. Genetics: CALR mutations and a new diagnostic algorithm for MPN. *Nat Rev Clin Oncol* 2014; **11**: 125–126.
2 Jones AV, Cross NC. Inherited predisposition to myeloproliferative neoplasms. *Ther Adv Hematol* 2013; **4**: 237–253.
3 Jones AV, Chase A, Silver RT, Oscier D, Zoi K, Wang YL *et al.* JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet* 2009; **41**: 446–449.
4 Olcaydu D, Harutyunyan A, Jager R, Berg T, Gisslinger B, Pabinger I *et al.* A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat Genet* 2009; **41**: 450–454.
5 Kilpivaara O, Mukherjee S, Schram AM, Wadleigh M, Mullally A, Ebert BL *et al.* A germline JAK2 SNP is associated with predisposition to the development of JAK2(V617F)-positive myeloproliferative neoplasms. *Nat Genet* 2009; **41**: 455–459.
6 Tapper W, Jones AV, Kralovics R, Harutyunyan AS, Zoi K, Leung W *et al.* Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. *Nat Commun* 2015; **6**: 6691.
7 Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD *et al.* Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med* 2013; **369**: 2379–2390.
8 Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC *et al.* Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med* 2013; **369**: 2391–2405.
9 Eder-Azanza L, Navarro D, Aranaz P, Novo FJ, Cross NC, Vizmanos JL. Bioinformatic analyses of CALR mutations in myeloproliferative neoplasms support a role in signaling. *Leukemia* 2014; **28**: 2106–2109.
10 Tefferi A, Lasho TL, Finke C, Belachew AA, Wassie EA, Ketterling RP *et al.* Type 1 vs type 2 calreticulin mutations in primary myelofibrosis: differences in phenotype and prognostic impact. *Leukemia* 2014; **28**: 1568–1570.
11 Tefferi A, Wassie EA, Guglielmelli P, Gangat N, Belachew AA, Lasho TL *et al.* Type 1 versus Type 2 calreticulin mutations in essential thrombocythemia: a collaborative study of 1027 patients. *Am J Hematol* 2014; **89**: E121–E124.
12 Harutyunyan AS, Jager R, Chen D, Berg T, Rumi E, Gisslinger B *et al.* Allelic imbalance in CALR somatic mutagenesis. *Leukemia* 2015; **29**: 1431–1435.
13 Jones AV, Ward D, Lyon M, Leung W, Callaway A, Chase A *et al.* Evaluation of methods to detect CALR mutations in myeloproliferative neoplasms. *Leuk Res* 2015; **39**: 82–87.

Telomere length predicts progression and overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial

Leukemia (2015) **29,** 2411–2414; doi:10.1038/leu.2015.217

Telomere erosion and fusion have an important role in the pathology of many common human malignancies including chronic lymphocytic leukemia (CLL).[1,2] Previous studies in CLL have shown that short telomeres defined on the basis of the median value or receiver operating characteristic analysis are associated with unmutated *IGHV* genes, poor-risk genomic abnormalities, genomic complexity and high expression of CD38, CD49d and ZAP70 whereas long telomeres are associated with increasing *IGHV* mutational load, isolated deletion of 13q and low CD49d expression. In addition, in predominantly diagnostic or mixed patient cohorts, telomere length (TL) predicts time to first treatment and/or overall survival (OS) in multivariate analyses of models incorporating established biomarkers.[3–7] However uncertainties about the most clinically relevant measure of TL, the optimal choice of assay, the need for assay standardisation and the lack of published data on the prognostic value of TL in patients entered into randomised trials have hindered the implementation of TL measurement into routine clinical practice. We have attempted to address these issues by measuring TL using monochrome multiplex Q-PCR (MMQ-PCR) in 384 patients at randomisation into the UK LRF CLL4 phase 3 chemotherapy trial (Supplementary Table S1), of whom 111 samples were also screened by single telomere length analysis (STELA). TL and established biomarkers were measured as previously described.[8–13]

The mean TL assessed by MMQ-PCR in 384 cases was 3.57 TLU (TL units, range 0.61–19.05). For 111 patients randomised to the fludarabine plus cyclophosphamide arm analysed by both MMQ-PCR and STELA, we showed an excellent correlation between these two TL measurements (Spearman correlation 0.80; Supplementary Fig. S1), permitting the calibration of the mean TLU to a mean absolute TL of 3.39 kb (range 1.93–11.06 kb, median 2.92 kb (inter quartile range: 2.58, 3.60)) for our cohort. All subsequent analyses were based on data derived from MMQ-PCR expressed in kb rather than TLU. TL was found to be significantly associated with *IGHV* mutation status, ZAP70 and CD38 expression, serum beta-2 microglobulin, *TP53* abnormality, 11q deletion, genomic complexity, *ATM* and *SF3B1* mutation but not trisomy 12 or *NOTCH1* mutation (Supplementary Table S2).

Increasing TL, entered as a continuous variable, was associated with a significant reduction in risk of a progression free survival (PFS) event (hazard ratio (HR) = 0.89, 95% confidence interval (CI): 0.85–0.93, $P < 0.001$) and longer OS (HR = 0.84, 95% CI: 0.80–0.89, $P < 0.001$; Supplementary Table S3). To determine a single TL cut-off value with maximum prognostic power, we employed recursive partitioning and identified the 75th and 80th percentile for PFS (HR: 2.42, $P < 0.001$) and OS (HR: 3.17, $P < 0.001$), respectively (Supplementary Fig. S2), demonstrating that identifying cases with the longest telomeres is key to maximising the prognostic value of these data. We then performed Kaplan–Meier analysis and generated a categorical variable with three groups for TL—short (< 50 percentile), intermediate (50–75 percentile) and long (> 75 percentile; Figure 1 and Supplementary Fig. S3).

Accepted article preview online 10 August 2015; advance online publication, 4 September 2015

The range for the mean TL in the short group was 1.93–2.92 kb (median: 2.58 kb), in the intermediate group, 2.91–3.57 kb (median: 3.14 kb), and in the long group, 3.64–11.06 kb (median: 4.81 kb). Interestingly, we have previously shown that telomere fusions were never detected when the mean TL exceeded 3.81 kb (as defined by STELA), and the range of TLs in our long and short telomere groups would suggest that telomere fusions would be expected to occur predominantly in the latter group.[14] The risk of progression was increased twofold for the intermediate group (HR: 2.07, 95% CI: 1.52–2.82, $P < 0.001$) and by 2.7 times for the short group (HR: 2.67, 95% CI: 2.03–3.53, $P < 0.001$) when compared with the long group while the risk for OS was increased 2.3 and 3.5 times for the intermediate and short groups, respectively (Supplementary Table S3 and Supplementary Fig. S4). The median PFS and OS for the 96 patients in the long TL group was 4.0 and 9.9 years, respectively, and these patients were 63% (HR: 0.37, 95% CI: 0.28–0.49, $P < 0.001$) and 72% (HR: 0.28, 95% CI: 0.21–0.39, $P < 0.001$) less likely to progress or die compared with patients within the short TL group (Supplementary Table S4).

Further investigations into the genomic and immunogenetic context of the three TL groups in 170 CLL4 patients with complete *TP53* and *ATM* deletion and mutational data showed that poor-risk features, such as unmutated *IGHV* genes, *TP53* abnormalities, biallelic *ATM* inactivation, and genomic complexity were found at a higher frequencies in the intermediate and especially in the short TL groups (Figure 1, Supplementary Table S5). The short TL group also included occasional patients that lacked poor-risk biomarkers, thus identifying additional patients with poor outcome after first-line chemotherapy. Conversely, the long TL group captured patients with extended survival despite having poor-risk features. Interestingly in the long TL group, patients with *TP53* abnormalities or biallelic *ATM* inactivation predominantly had mutated *IGHV* genes whereas they were unmutated in the short TL group.

Next, we estimated the adjusted impact of the TL categories on PFS and OS after controlling for confounding variables using multivariate Cox regression. The short TL category, treatment allocation (chlorambucil, fludarabine and cyclophosphamide plus fludarabine), unmutated *IGHV* genes, the presence of an 11q deletion and *TP53* abnormality emerged as strong predictors of shorter PFS when the stepwise backward and forward selection method was employed to arrive at a final model (Table 1). The highest increase in risk of progression was for patients with a *TP53* abnormality (HR: 2.51, 95% CI: 1.66–3.81, $P < 0.001$) followed by the

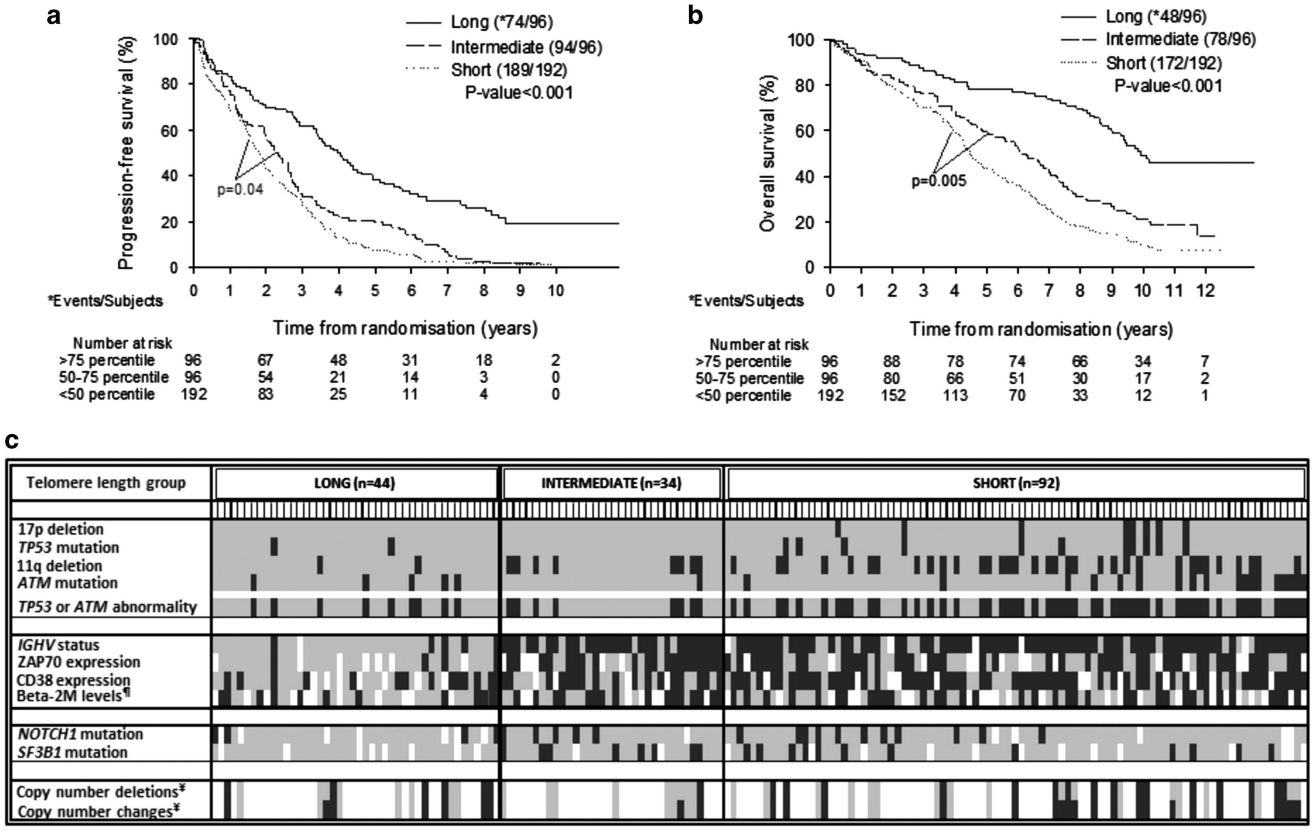


Figure 1. Kaplan–Meier plots of mean TL (from MMQ-PCR) divided into three groups for PFS (**a**) and OS (**b**), and distribution of CLL biomarkers within the TL groups for 170 patients with complete data for *TP53* and *ATM* status (**c**). For (**a** and **b**), log-rank *P*-value for each of the lower quartiles when compared with the long group is < 0.001. Long: >75 percentile; intermediate: 50–75 percentile; short: <50 percentile. The median PFS and OS for the 96 patients in the long TL group was 4.0 and 9.9 years, respectively. This longer median survival was sustained when the analyses were performed with similarly categorised TL groups using measurements from both MMQ-PCR and STELA in the 111 patients with data for both (PFS: 8.1 and 5.1 years; OS: not reached and 9.8 years, Supplementary Figs S5 and 6). For (**c**) data are presented in decreasing TL as measured by MMQ-PCR and divided into the three groups as described in the text. Long: >75 percentile, intermediate: 50–75 percentile, short: <50 percentile. Each short vertical line (below the TL group name) corresponds to a patient. The presence and absence of each of the biomarker status listed on the left are represented by black and grey boxes, respectively, whereas white boxes indicate missing data on the biomarker status. *ZAP70 is expressed if $>10\%$-positive cells by flow cytometry, #CD38 is expressed if $>7\%$-positive cells by flow cytometry, ¶Beta- M: beta-2 microglobulin (present: >4 mg/l), ¥Present: if: >3 deletions/changes per patient.

Table 1. Predictors of overall and progression-free survival

Factor	*HR (95% CI)*	P-*value*
PFS		
Male vs female	1.34 (0.999–1.80)	0.05
Treatment		< 0.001
Fludarabine vs chlorambucil	0.72 (0.52–0.996)	0.047
(Fludarabine+cyclophosphamide) vs chlorambucil	0.38 (0.28–0.50)	< 0.001
TL		< 0.001
Intermediate vs long	1.31 (0.83–2.05)	0.246
Short vs long	2.10 (1.37–3.21)	0.001
IGHV unmutated vs mutated	1.59 (1.12–2.25)	0.01
TP53 abnormal vs normal	2.51 (1.66–3.81)	< 0.001
11q deleted vs not deleted	1.46 (1.07–1.98)	0.02
OS		
TL		0.003
Intermediate vs long	1.34 (0.72–2.46)	0.355
Short vs long	2.21 (1.27–3.87)	0.005
IGHV unmutated vs mutated	2.08 (1.22–3.57	0.01
TP53 abnormal vs normal	2.11 (1.26–3.53)	0.004
ZAP70 expressed vs not expressed[a]	0.65 (0.45–0.96)	0.03
13q deleted vs not deleted	0.67 (0.47–0.95)	0.02

Abbreviations: CI, confidence interval; HR, hazard ratio; OS, overall survival; PFS, progression free survival; TL, telomere length. TL groups: long: > 75 percentile; intermediate: 50–75 percentile; short: < 50 percentile. Candidates entered in the iterative backward-forward selection method were factors with *P*-values ⩽0.05 in the univariable analysis (see Fig. 2). Age was entered as a continuous variable for the multivariable analysis. There was a small but significant negative association between age and shorter TL (data not included). The final models were based on 292 subjects and 269 events for PFS and 201 subjects and 153 deaths for OS. [a]Expressed if expression levels > 10%

short group for TL (short vs long HR: 2.10, 95% CI: 1.37–3.21). It is interesting to note that short TL had a more detrimental effect on progression than unmutated *IGHV* genes or 11q deletion. Shorter TL, unmutated *IGHV* genes and a *TP53* abnormality remained as strong predictors of shorter OS as well.

We then used sensitivity–specificity analysis to derive likelihood ratios (LR^+ and LR^- and the LR^+/LR^- ratio) to judge the relative discriminatory power of each of the four biomarkers (TL, *IGHV*, *TP53* and 11q deletion) that were strong predictors of PFS and/or OS in our whole cohort, to predict both the presence and absence of PFS and OS events at last follow-up in the 292 patients with complete TL, *IGHV*, 11q deletion and *TP53* abnormality data (Supplementary Tables S6-S9). One caveat was that LR^+/LR^- for PFS events could not be estimated for *TP53* alone as the false negative rates were zero.

The best predictor of PFS events was the combined short and intermediate TL groups (cut-off < 75th percentile), alone (LR^+/LR^-: 15.54) or in combination with a *TP53* abnormality (LR^+/LR^-: 16.35; Supplementary Table S6). In all, 81% of patients who progressed or died had short or intermediate length telomeres. Consistent with the recursive partitioning data, the best predictor of long PFS (absence of PFS events) was the long TL group alone (LR^+/LR^-: 15.26) or in combination with wild type *TP53* (LR^+/LR^-: 17.20; Supplementary Table S7). Long TL correctly predicted the 18/23 patients who did not have a PFS event (sensitivity of 78.3%), while 218/269 patients without long TL did have a PFS event during follow-up (specificity of 81%). As with the presence of PFS events, the inclusion of IGHV mutational status alone or combined with TL and *TP53* data did not increase the LR^+/LR^- ratio further (Supplementary Table S7). For OS, the LR^+/LR^- for the short TL sub-group was 4.70, with the highest ratio for *TP53* abnormality (LR^+/LR^-: 8.83), with sensitivity, specificity and accuracy rates of 84.5%, 51.5% and 77.1%, respectively (Supplementary Table S8). The long TL category provided a LR^+/LR^- ratio of 5.82 to predict the absence of an OS event, second to *TP53* (LR^+/LR^-: 8.61) but with much higher specificity (Supplementary Table S9). In each case LR ratios were improved for TL in combination with IGHV mutational status, *TP53* data, or both.

In summary, our data confirm that both MMQ-PCR and STELA can provide clinically relevant information in the range of TLs found in CLL. Importantly we demonstrate for the first time that TL is superior both to established and to recently discovered genomic biomarkers for predicting prolonged PFS following chemotherapy. A study of TL using MMQ-PCR in 620 patients entered into the CLL8 trial currently presented in abstract form only, identified TL as an independent marker of PFS but not OS in models, which incorporated a similar range of biomarkers to those employed in our study, suggesting that TL also has prognostic significance in patients treated with chemo-immunotherapy.[15] It remains to be seen whether TL will retain prognostic value in patients treated solely with non-genotoxic therapies. Although these novel therapies show very significant promise, long-term data on side effects and drug resistance are awaited and they may be unsuitable or unaffordable for some patients. In contrast, immunochemotherapy can offer good long-term outcomes with acceptable toxicity to subgroups of patients with CLL.[14] In conclusion, TL, which usually remains stable on sequential analyses[5] warrants further evaluation in CLL patients receiving therapy. Factors such as cost, reproducibility and availability will determine the assay method of choice.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

JC Strefford[1,7], L Kadalayil[2,7], J Forster[1], MJJ Rose-Zerilli[1], A Parker[3], TT Lin[4], N Heppel[4], K Norris[4], A Gardiner[3], Z Davies[3], D Gonzalez de Castro[5], M Else[5], AJ Steele[1], H Parker[1], T Stankovic[6], C Pepper[4], C Fegan[4], D Baird[4], A Collins[2], D Catovsky[5] and DG Oscier[1,3]

[1]*Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK;*
[2]*Genetic Epidemiology and Bioinformatics, Faculty of Medicine, University of Southampton, Southampton, UK,;*
[3]*Department of Pathology, Royal Bournemouth Hospital, Bournemouth, UK;*
[4]*CLL Research Group, Institute of Cancer & Genetics, School of Medicine, Cardiff University, Cardiff, UK;*
[5]*Haemato-oncology Research Unit, Division of Molecular Pathology, The Institute of Cancer Research, London, UK and*
[6]*Division of Cancer Sciences, School of Medicine, University of Birmingham, Birmingham, UK*
E-mail: JCS@soton.ac.uk or David.Oscier@sky.com
[7]*These authors contributed equally to this study and share first authorship.*

REFERENCES

1 Lin TT, Letsolo BT, Jones RE, Rowson J, Pratt G, Hewamana S *et al.* Telomere dysfunction and fusion during the progression of chronic lymphocytic leukemia: evidence for a telomere crisis. *Blood* 2010; **116**: 1899–1907.
2 Jones CH, Pepper C, Baird DM. Telomere dysfunction and its role in haematological cancer. *Br J Haematol* 2012; **156**: 573–587.
3 Rossi D, Lobetti Bodoni C, Genuardi E, Monitillo L, Drandi D, Cerri M *et al.* Telomere length is an independent predictor of survival, treatment requirement and Richter's syndrome transformation in chronic lymphocytic leukemia. *Leukemia* 2009; **23**: 1062–1072.
4 Roos G, Kröber A, Grabowski P, Kienle D, Bühler A, Döhner H *et al.* Short telomeres are associated with genetic complexity, high-risk genomic aberrations, and short survival in chronic lymphocytic leukemia. *Blood* 2008; **111**: 2246–2252.
5 Mansouri L, Grabowski P, Degerman S, Svenson U, Gunnarsson R, Cahill N *et al.* Short telomere length is associated with NOTCH1/SF3B1/TP53 aberrations and poor outcome in newly diagnosed chronic lymphocytic leukemia patients. *Am J Hematol* 2013; **88**: 647–651.
6 Grabowski P, Hultdin M, Karlsson K, Tobin G, Aleskog A, Thunberg U *et al.* Telomere length as a prognostic parameter in chronic lymphocytic leukemia with special reference to VH gene mutation status. *Blood* 2005; **105**: 4807–4812.

7 Sellmann L, de Beer D, Bartels M, Opalka B, Nückel H, Dührsen U *et al.* Telomeres and prognosis in patients with chronic lymphocytic leukaemia. *Int J Hematol* 2011; **93**: 74–82.
8 Skowronska A, Parker A, Ahmed G, Oldreive C, Davis Z, Richards S *et al.* Biallelic ATM inactivation significantly reduces survival in patients treated on UK CLL4 trial. *J Clin Oncol* 2012; **30**: 4524–4532.
9 Rose-Zerilli M, Forster J, Parker H, Parker A, Rodriguez A, Chaplin T *et al.* ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q-deleted chronic lymphocytic leukemia, data from the UK LRF CLL4 trial. *Haematologica* 2014; **99**: 736–742.
10 Oscier DG, Rose-Zerilli MJJ, Winkelmann N, Gonzalez de Castro D, Gomez B, Forster J *et al.* The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. *Blood* 2013; **120**: 4441–4443.
11 Oscier D, Wade R, Davis Z, Morilla A, Best G, Richards S *et al.* Prognostic factors identified three risk groups in the LRF CLL4 trial, independent of treatment allocation. *Haematologica* 2010; **95**: 1705–1712.
12 Gonzalez D, Martinez P, Wade R, Hockley S, Oscier D, Matutes E *et al.* Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. *J Clin Oncol* 2011; **29**: 2223–2229.
13 Catovsky D, Richards S, Matutes E, Oscier D, Dyer MJ, Bezares RF *et al.* Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. *Lancet* 2007; **370**: 230–239.
14 Lin TT, Norris K, Heppel NH, Pratt G, Allan JM, Allsup DJ *et al.* Telomere dysfunction accurately predicts clinical outcome in chronic lymphocytic leukaemia, even in patients with early stage disease. *Br J Haematol* 2014; **167**: 214–223.
15 Jebaraj BM, Busch R, Zenz T, Bühler A, Winkler D, Schnaiter A *et al.* Telomere length and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood* 2013; **122**: 671.

Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu)

Long-term complete clinical and hematological responses of the TEMPI syndrome after autologous stem cell transplantation

Leukemia (2015) **29,** 2414–2416; doi:10.1038/leu.2015.298

The TEMPI syndrome is a recently described entity characterized by the pentad of Telangiectasias, Erythrocytosis with elevated erythropoietin, Monoclonal gammopathy, Perinephric fluid collections and Intrapulmonary shunting.[1] The prevalence of the disease is not known and few cases have been reported in the literature so far.[1–5] The exact pathophysiology is not clear either. Erythrocytosis seems to be the initial manifestation of the disease and the condition has been therefore misdiagnosed as Polycythemia Vera in few patients. Erythrocytosis is associated with extreme elevation of erythropoietin level that seems to be correlating with response to therapy. The clinical course is typically followed by severe hypoxemia,[1,4] which has been attributed to the combination of intrapulmonary shunting and iatrogenic iron deficiency anemia. Perinephric fluids are thought to be related to malformation of lymphatic vessels.[6] In most of the reported cases, the monoclonal protein is found to be IgG kappa.[1,4]

Treatment with bortezomib-based regimens has been used in few of these patients and variable responses have been reported.[5] Some of these responses were remarkable, suggesting a central role for the monoclonal gammopathy in the pathophysiology of this syndrome.[5] This also points to the hypothesis that TEMPI might represent a paraneoplastic syndrome related to the monoclonal gammopathy, in a similar way to POEMS syndrome (Polyneuropathy, Organomegaly, Endocrinopathy, M-protein and Skin abnormalities), scleromyxoedema or Schnitzler syndrome.[7–9]

In this letter, we report a patient with the TEMPI syndrome that had a dramatic and durable complete clinical and hematological response after high dose chemotherapy and autologous stem cell transplantation (ASCT).

A 49-year-old female was referred to our institution for further evaluation of progressive hypoxia requiring the continuous use of high concentrations of oxygen. She had a 15-year history of erythrocytosis, thought to be related to Polycythemia Vera, and has been treated with intermittent phlebotomy since then. The phlebotomy procedures have resulted in long-standing iatrogenic iron deficiency anemia. Physical examination was significant for remarkable telangiectasia on the face and upper chest and oxygen saturation was 77% on room air. Hemoglobin at presentation was 10.7 g/dl with indices consistent with iron deficiency anemia. However, the erythrocyte count was 7.1 × 10(12)/l (reference: 3.9–5.03) and the erythropoietin level was 8144 ml U/ml (reference: 2.6–18.5). Serum protein

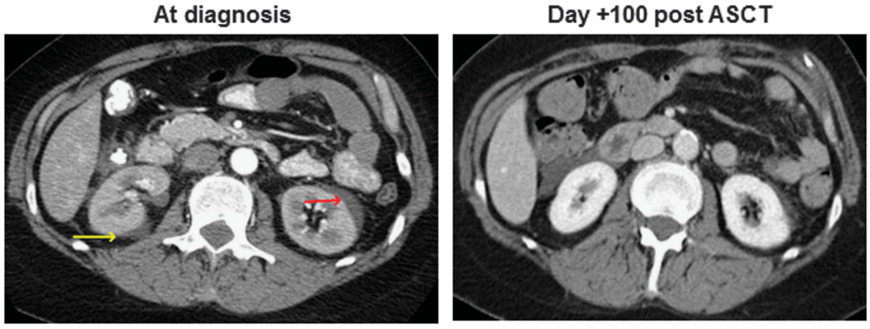


Figure 1. Perinephric fluids at diagnosis and on Day +100 post autologous stem cell transplantation (ASCT).

Accepted article preview online 26 October 2015; advance online publication, 17 November 2015

EUROPEAN HEMATOLOGY ASSOCIATION
haematologica
Journal of the European Hematology Association

Low frequency mutations independently predict poor treatment free survival in early stage chronic lymphocytic leukemia and monoclonal B-cell lymphocytosis

by Nils Winkelmann, Matthew Rose-Zerilli, Jade Forster, Marina Parry, Anton Parker, Anne Gardiner, Zadie Davis, Andrew J. Steele, Helen Parker, Nicholas C.P. Cross, David G. Oscier, and Jonathan C. Strefford

Haematologica 2015 [Epub ahead of print]

Citation: Winkelmann N, Rose-Zerilli M, Forster J, Parry M, Parker A, Gardiner A, Davis Z, Steele AJ, Parker H, Cross NC, Oscier DG, and Strefford JC. Low frequency mutations independently predict poor treatment free survival in early stage chronic lymphocytic leukemia and monoclonal B-cell lymphocytosis. Haematologica. 2015; 100:xxx
doi:10.3324/haematol.2014.120238

Low frequency mutations independently predict poor treatment free survival in early stage chronic lymphocytic leukemia and monoclonal B-cell lymphocytosis

Running title: Acquired mutations in CLL and cMBL

Nils Winkelmann [1, 2, *], Matthew Rose-Zerilli [3, *], Jade Forster [3], Marina Parry [3], Anton Parker [3,5], Anne Gardiner [5], Zadie Davies [5], Andrew J Steele [3], Helen Parker [3], Nicholas C P Cross [1,4], David G Oscier [3,5] and Jonathan C Strefford [3].

[1] Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, UK

[2] Klinik für Innere Medizin II, Universitätsklinikum Jena, Jena, Germany

[3] Academic Unit of Cancer Sciences, Faculty of Medicine, University of Southampton, UK

[4] Human Development and Health, Faculty of Medicine, University of Southampton, UK

[5] Department of Molecular Pathology, Royal Bournemouth Hospital, Bournemouth, UK

[*] NW and MRZ contributed equally to this manuscript

Correspondence to: Professor Jonathan C Strefford, Cancer Genomics, Cancer Sciences Unit, Somers Cancer research Building, Southampton General Hospital, Tremona Road, Southampton SO16 6YD. Tel: 44 23 8079 5246. E-mail: JCS@soton.ac.uk

Recent studies employing next generation sequencing (NGS) technologies have identified novel recurring mutations in monoclonal B-cell lymphocytosis (cMBL) and chronic lymphocytic leukemia (CLL) (1). *NOTCH1* and *SF3B1* mutations are the most prevalent and are associated with reduced survival independent of established clinical and biological variables. However, we have limited understanding of the impact on clinical outcome of less prevalent mutations, especially when identified at diagnosis in patients with cMBL or Binet stage A CLL. To address this question, we have investigated the clinical significance of *SF3B1, NOTCH1* and four 'low frequency' mutations: *POT1, XPO1, MYD88, BIRC3* in a single centre cohort of well characterized patients. A fifth gene (*FBXW7*) was screened only in cases with trisomy 12, in view of the known strong association between the two abnormalities (2).

This study included 206 previously untreated patients with a diagnosis of either Binet stage A CLL [n=116] or cMBL [n=-90] (**Supplementary Table 1**) according to the NCI-WG criteria updated in 2008 (3). Follow up ranged from 1 to 35 years with a median of 10 years. One hundred and ten patients (53%) remained in stable Binet Stage A or cMBL while 29/90 (32%) of patients with cMBL evolved to Binet stage A during the observational period. Sixty-seven (33%) patients progressed to Binet stage B or C disease and required treatment. Of these, 21 (23%) progressed from cMBL and 46 (40%) from Binet stage A CLL. The indications for treatment and the treatment regimens were based on current guidelines at the time of treatment, with most patients receiving an alkylating agent or purine analogue. Biomarker studies and mutational analysis (described in **Supplementary methods**) were performed on samples stored at, or within 6 months of diagnosis. Mutational analysis was performed at a second time point in 84 patients with a median of 72 months (range: 24 to 145 months) between initial and subsequent testing (**Supplementary Table 2**). This study was implemented in accordance with the Declaration of Helsinki and has been ethically approved by the Regional Ethics Committee (REC).

At diagnosis, mutations were detected in *SF3B1* (16/ 199, 8%), *NOTCH1* (11/ 203, 5%), *POT1* (8/ 198, 4%), *XPO1* (2/ 172, 1%) *MYD88* (3/ 198, 1.5 %), *BIRC3* (1/ 197, 0.5%) and *FBXW7* (2/31 trisomy 12 cases, 6.5%). The majority of mutations have been previously observed in CLL, are annotated in COSMIC and predicted to have deleterious functional consequences based on PolyPhen and SIFT scores (**Supplementary Table 3**). These figures are broadly comparable to those in a recent large study of 1160 previously untreated patients of all stages of whom 82% were screened at diagnosis: *NOTCH1, SF3B1, XPO1, FBWX7* and *MYD88* mutations were found 12.3, 9.0, 3.4, 2.5 and 1.5 % of cases, respectively (2). The clinical and biological features of patients with *NOTCH1, SF3B1* and low frequency mutations are shown in **Supplementary Table 4** and **Figure 1A**. The expected associations were apparent, such as both *NOTCH1* and *SF3B1* mutations being significantly associated with *IGHV* unmutated genes, high CD38 and ZAP70 expression (4). Three of the eight *POT1* mutated cases had mutated *IGHV* genes in contrast to previous data (5), in which *POT1* mutations have occurred exclusively in cases with unmutated *IGHV* genes. In view of the reported increased incidence of

telomere-containing chromosome fusions in patients with *POT1* mutations, we examined the incidence of chromosomal complexity and instability in our 8 sequential cases with *POT1* mutations. Based on cytogenetic and FISH data at diagnosis and later time points we detected a del(13q14) at the second time point only in 2 cases, but found no evidence of genomic complexity **(Supplementary table 5)**. Larger studies in patients pre- and post therapy, using more sensitive methods to detect complexity and instability, will be required to determine the biological significance of *POT1* mutations. The low incidence of the other mutations precluded any meaningful analysis of their associations with other biomarkers. We found no difference in the frequency of genomic mutations or any other biomarker, apart from a borderline higher incidence of CD38 expression in cMBL, between cases presenting with cMBL or stage A CLL. This is consistent with other recent data confirming the biological similarity between cMBL and Rai O CLL, including a similar incidence of *NOTCH1* and *SF3B1* mutations (6) and provides justification for analyzing the outcome of the combined cMBL and stage A cohorts.

We then assessed the prognostic significance of mutations in our cohort. In view of the low frequency of all mutations other than *NOTCH1* and *SF3B1*, and the known favorable outcome of cases with an *MYD88* mutation (7), we hypothesized that the collective analysis of the other low frequency mutations might provide insight into their biological importance and clinical utility in early stage disease. Accordingly, we grouped mutations of *POT1, XPO1, BIRC3* and *FBXW7* together and compared the outcome of cases with any of these 'low frequency' mutations with those that were wild type for *NOTCH1*, *SF3B1* and the four 'low frequency' genes. In support of this hypothesis, we identified an enrichment of *NOTCH1* and *SF3B1* mutations in Stage A CLL and cMBL patients who ultimately developed progressive disease, and also showed that the collective presence of a low frequency mutation was significantly associated with subsequent disease progression (OR: 17.4, p=0.002) and need for treatment (OR: 18.0 and P=0.002) (**Supplementary Table 4**).

Univariate analysis confirmed recent data (8) showing that the presence of *NOTCH1* (median 58 vs. 105 months, HR 3.9, P=<0.001), and *SF3B1* mutations (median 44 vs. 103 months, HR 2.8, P= 0.001) were significantly associated with reduced treatment free survival (TFS) but not overall survival (OS) (**Table 1**). However, the presence of a low frequency mutation was also associated with reduced TFS (median 42 vs. 113 months, HR 8.6, P<0.001) (**Figure 1B**).

We then estimated the impact of these 'low frequency' mutations on TFS and OS after controlling for confounding variables in multivariate Cox proportional hazard analysis depicted in **Table 2**. Along with the 'low frequency' mutations, other variables included in the analysis were *SF3B1* and *NOTCH1* status, trisomy 12, del(11q) and the presence or absence of mutated *IGHV* genes. Loss of chromosome 17p was omitted due to the very low frequency of del(17p) cases, all of which exhibited mutated IGHV genes (9). As expected, unmutated *IGHV* genes remained the strongest predictor of poor treatment free (HR: 5.7, 95% CI: 2.7-11.9, P<0.001) and overall survival (HR: 3.3, 95%: CI 1.9-5.7, P<0.001). Loss of chromosome 11q was confirmed as an adverse prognostic factor in treatment free

survival (HR: 3.7, 95% CI: 1.5-8.6, P=0.003) and *SF3B1* mutations showed borderline significance as an independent predictor of reduced overall survival (HR: 1.3, 95% CI: 1.02-5.3, P=0.045). In addition, our 'low frequency mutations' variable retained significance for reduced treatment free survival (HR: 3.7, 95% CI: 1.3-10.5, P=0.016). While we recognize that the apparent poorer outcome of cases with low frequency mutations might reflect the presence of other undetected mutations or genomic instability rather than the mutations we detected, none had del(17p) or cytogenetic evidence of genomic complexity and only one patient showed a del(11q).

Finally, eighty-four patients were screened sequentially for mutations compared to those screened only at diagnosis in order to document clonal evolution in CLL. The sequential and single time point cases are shown in **Supplementary table 2** and no significant differences were observed between the two groups, apart from a borderline higher frequency of trisomy 12 in the cases tested only at diagnosis. Among the sequential cases, forty-seven patients (56%) remained stable during observation time, but 37 (44%) progressed and 36 patients received treatment between diagnosis and second testing. We detected mutations of *SF3B1* in four cases and *XPO1* in one case, not found on screening at diagnosis. Of these, three presented with cMBL, and two with Stage A CLL and 3/5 had progressive disease. Additional characteristics of these patients are depicted in **Supplementary Table 6**. Although the number of patients screened was small, these results were consistent with those of a recent large multinational study in which the incidence of *SF3B1*, but not *NOTCH1*, mutations rose with increasing time from diagnosis to the date of sampling (8).

In summary, our study suggests that screening for these low frequency mutations may have utility in the clinical management of early stage CLL and cMBL, and future larger studies should evaluate the incidence and clinical significance of low frequency potential driver mutations in early disease to assess their relevance in new molecular prognostication systems.

Authorship and Disclosures

JCS, DO and NC coordinated the research and helped write the paper; DO recruited the patients, provided the clinical data; NW. MRZ, JF, AP, AG, MP and HP performed laboratory work; NW and MRZ analyzed the data and wrote the paper. The authors report no potential conflicts of interest.

Acknowledgements

NW acknowledges the support of the Dr. Mildred Scheel Stiftung for Cancer Research (Bonn, Germany). This work was funded by Leukaemia and Lymphoma Research, the Kay Kendall Leukaemia Fund, Cancer Research UK and Wessex Medical Research. The authors gratefully acknowledge all patients who participated in the trial.

References

1. Strefford JC. The genomic landscape of chronic lymphocytic leukaemia: biological and clinical implications. Br J Haematol. 2014:[Epub ahead of print].

2. Jeromin S, Weissmann S, Haferlach C, 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):108-117.

3. Hallek M, Cheson BD, Catovsky D, 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):5446-5456.

4. Oscier DG, Rose-Zerilli MJJ, Winkelmann N, et al. The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. Blood. 2013;120(22):4441-4443.

5. Ramsay AJ, Quesada V, Foronda M, et al. POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. Nat Genet. 2013;45(5):526-530.

6. Ojha J, Secreto C, Rabe K, et al. Monoclonal B-cell lymphocytosis is characterized by mutations in CLL putative driver genes and clonal heterogeneity many years before disease progression. Leukemia. 2014; 28(12):2395-2398.

7. Martínez-Trillos A, Pinyol M, Navarro A, et al. Mutations in TLR/MYD88 pathway identify a subset of young chronic lymphocytic leukemia patients with favorable outcome. Blood. 2014;123(24):3790-3796.

8. Baliakas P, Hadzidimitriou A, Sutton LA, et al. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. Leukemia. 2015; 29(2):329-336.

9. Best OG, Gardiner AC, Davis ZA, et al. A subset of Binet stage A CLL patients with TP53 abnormalities and mutated IGHV genes have stable disease. Leukemia. 2009;23(1):212-214.

Table 1: Univariate Cox proportional hazard analysis of treatment-free and overall survival.

		Treatment Free Survival / TFS (months)							Overall Survival/ OS (months)						
Mutation/Biomarkers	status	total	events	median TFS	95% CI	HR	95% CI	P-Value	total	events	median OS	95% CI	HR	95%CI	P-Value
NOTCH1	wild type	137	52	105	88-128				190	91	117	107-133			
	mutated	8	8	58	26-126	3.9	1.8-8.2	**<0.001**	11	7	116	50-173	1.6	0.8-3.5	0.14
SF3B1	wild type	127	50	103	86-120				181	89	120	108-138			
	mutated	15	11	44	22-113	2.8	1.4-5.4	**0.001**	16	12	97	49-149	2.0	0.95-4.1	0.06
Low frequency mutations	wild type	106	34	113	91-127				154	74	122	112-143			
	mutated	7	6	42	11-67	8.6	3.3-22	**<0.001**	7	3	85	77-156	2.3	0.7-7.4	0.2
Gender	female	57	20	142	73-181				88	41	142	120-162			
	male	90	42	91	72-107	1.9	1.1-3.2	**0.04**	116	59	106	94-119	1.6	1.1-2.4	**0.017**
Disease	cMBL	65	20	107	74-124				89	40	118	107-143			
	CLL stage A	82	42	101	53-121	1.7	1.1-3.1	**0.02**	115	60	119	97-138	1.25	0.8-2.1	0.4
Disease progression	absent	85	1	121	102-152				137	56	119	103-138			
	present	62	61	44	30-101	144	20-1039	**<0.001**	67	44	116	102-150	1.7	1.1-2.5	**0.009**
IGHV-status	mutated	100	22	122	113-152				141	55	139	119-155			
	unmutated	47	40	34	22-53	10.7	6.1-19	**<0.001**	63	45	91	70-110	3.7	2.4-5.5	**<0.001**
CD38-status (30%)	negative	92	23	121	106-149				127	52	142	121-162			
	positive	43	30	68	42-84	4.8	2.7-8.5	**<0.001**	56	36	97	86-112	2.9	1.9-4.5	**<0.001**
ZAP70-Status	negative	83	21	122	109-152				119	53	141	120-162			
	positive	40	29	52	33-104	4.8	2.7-8.6	**<0.001**	50	31	107	85-121	2.4	1.5-3.8	**<0.001**
del(11q)	absent	128	44	112	89-123				175	79	121	112-140			
	present	12	11	19	4-86	7.2	3.6-14.3	**<0.001**	15	11	91	48-114	2.6	1.4-4.9	**0.002**
Trisomy 12	absent	89	37	118	103-143				124	64	131	116-162			
	present	26	16	81	49-132	1.9	1.1-3.5	**0.026**	40	24	108	78-142	1.8	1.1-2.9	**0.02**

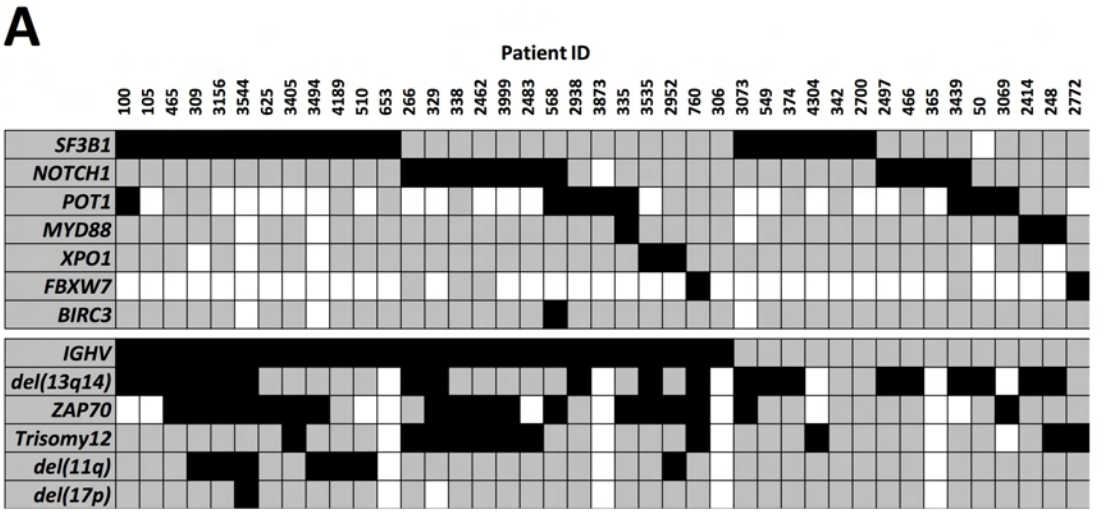
Table 2: Multivariate Cox proportional hazard analysis of treatment-free and overall survival.

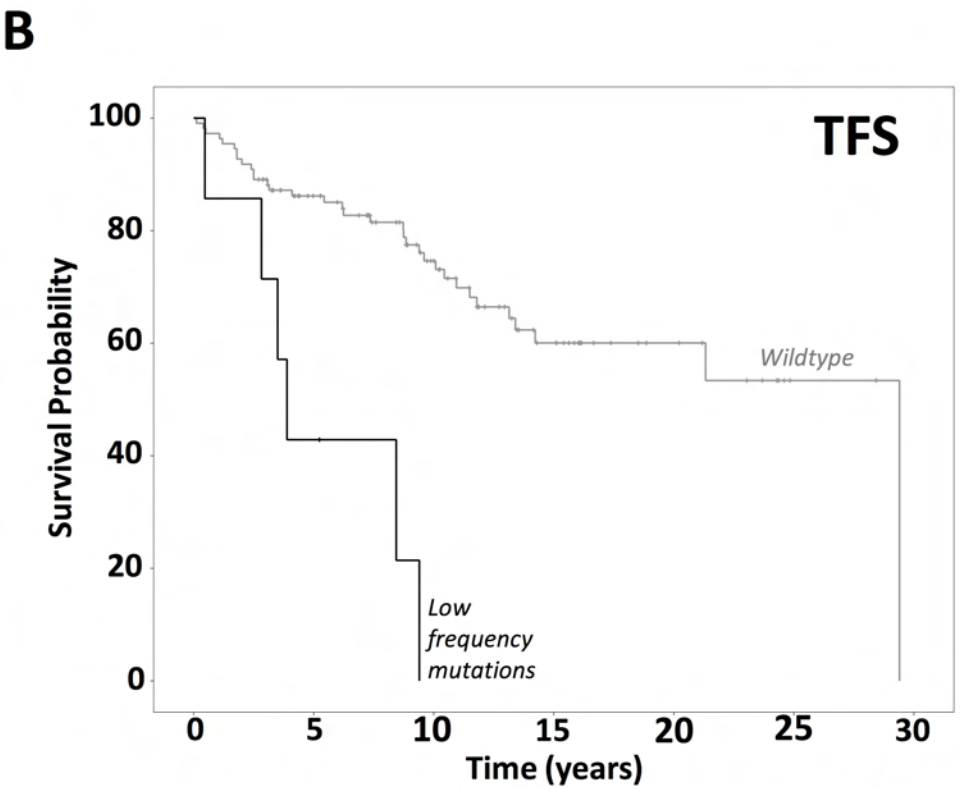
Variable	Treatment Free Survival			Overall Survival		
	HR	95% CI	P-Value	HR	95% CI	P-Value
Low frequency mutations	**3.7**	**1.3-10.5**	**0.016**	1.9	0.8-4.7	0.14
NOTCH1 mutations	1.4	0.5-4.0	0.5	1.1	0.5-2.7	0.8
SF3B1 mutations	1.9	0.9-4.3	0.1	**1.3**	**1.02-5.3**	**0.045**
Trisomy 12	1.5	0.7-3.2	0.3	1.2	0.7-2.1	0.6
del(11q)	**3.7**	**1.5-8.6**	**0.003**	1.3	0.6-2.7	0.5
IGHV-unmutated	**5.7**	**2.7-11.9**	**<0.001**	**3.3**	**1.9-5.7**	**<0.001**

Footnote: TFS multivariate: 109 cases with 51 events, 97 cases with missing data; OS multivariate: 154 cases with 82 events, 52 cases with missing data

Figure 1. The associations between gene mutations, established biomarkers and time to first treatment in our series of cMBL and Stage A CLL patients. (A) shows the mutual relationship between gene mutations and other genetic lesions and biomarkers in CLL, sorted by IGHV mutational status. Rows correspond to specific lesions/biomarkers, and columns represent individual patients (only patients with mutations in the genes tested are shown). Boxes coloured black and grey show the presence and absence of a lesion/biomarkers. A white box denotes that no data is available. (B) Shows TFS for patients with 'low frequency' mutations compared to wild-type controls. The *P* value is derived from Kaplan-Meier analysis with a log-rank test and median survival times with 95% confidence intervals.

Figure 1





Low frequency mutations	No. of cases	No. of events	Median TFS	95% CI	HR	95% CI	P-Value
mutated	7	6	42	11-67	8.6	3.3-22	**<0.001**
unmutated	106	34	113	91-127			

Supplementary Material

Supplementary Methods

FISH Analysis

FISH analysis was undertaken in all cases using commercially available probes (Abbott Diagnostics, Maidenhead, UK; DakoCytomation, Glostrup, Denmark) to detect deletion of 13q, 11q, 17p and trisomy 12, according to the manufacturers' instructions as previously described (1). Chromosomal analysis was performed and described according to the International System for Human Cytogenetic Nomenclature (2). ZAP70 and CD38 expression were determined as previously described (3) (4), where 10% and 30% positive cells were classed as positive or highly expressive, respectively. IGHV genes were sequenced as previously described (5) and a cut-off of ≥98% germ-line homology was taken to define the unmutated subset.

Mutational screening and sequencing

For each of the seven genes, we aimed to capture the majority of previously reported CLL specific somatic variations. Each genomic DNA sample (gDNA) was subjected to whole genome amplification (WGA) using the Illustra GenomiPhi V2 Amplification Kit® (GE Healthcare) prior to mutational screening. With the exception of the *FBXW7* (direct Sanger Sequencing for all samples), all genes were screened using high-resolution melt (HRM) analysis and subsequent Sanger sequencing as previously described (6, 7). Primer sequences are shown in **Supplementary Table 7.** PCR products showing altered melt patterns were sequenced. All detected changes were sequence validated on genomic DNA (gDNA) from the archival sample. A randomly selected subset of 100 genomic DNA samples that exhibited normal HRM melt profiles on WGA material was directly sequenced for *SF3B1* and *XPO1*. In doing so, we found no additional mutations. For *NOTCH1* we screened a 399bp section of exon 34 (amino acids 2405-2525), where all of the PEST domain mutations have been found (8-11). Further analysis of the c.7544_7545delCT, referred to as 'delCT' variant was performed using PCR-based fragment analysis and KASPar genotyping and is detailed below. We screened exons 14, 15 and 16 of *SF3B1* and exons 15 and 16 of *XPO1* capturing the majority of previously reported disease specific variations (8, 12-14). For *MYD88*, we screened for Exons 3-5 including the known L265P somatic variant by HRM and subsequent sequencing (8). Exons 4 to 10, and exon 18 of *POT1* and Exons 7 and 10 of *BIRC3* were screened using HRM and subsequent Sanger sequencing. For sequentially acquired mutations sample accuracy was confirmed by powerplex testing (Powerplex® 16 System, Promega, Madison, USA). A subset of 31 patients with confirmed trisomy 12 including

sequential samples for 10 patients were screened for mutations in all coding exons of *FBXW7* by Sanger sequencing.

Further analysis of the NOTCH delCT variant

To further assess the sensitivity of the HRM approach for *NOTCH1*, we identified cases with the P2515Rfs*4 (c.7544_7545delCT, referred to as 'delCT') variant using PCR-based fragment analysis [PCR-FA, n= 372] and allele-specific PCR [n= 213]as previously described for the CLL4 cohort (7). For the PCR-FA, the delCT variant was identified using PEST2 domain-specific primers (Forward: GTGACCGCAGCCCAGTTC, reverse: AAAGGAAGCCGGGGTCTC) as previously described (15). PCR products were sized using the LT3500 (Life Technologies) and the 271bp wild-type and 269bp mutant fragments were identified using GeneMapper (v4.1) software (Life Technologies). Allele-specific PCR for the deflect variant was performed by KBiosciences (http://www.kbiosciences.co.uk) using their fluorescence-based competitive allele-specific PCR (KASPar). 106 duplicates were included and the concordance between duplicates was >99%. Sample processing with PCR-FA and KASPar technologies were performed blindly in independent laboratories. Using these two approaches, we confirm the sensitivity of our HRM approach which showed a 100% concordance with our PCR-FA and KASPar analysis.

Supplementary Tables

Supplementary table 1. Clinical, cytogenetic and molecular characteristics of the complete cohort (n=206) and cMBL (n=90) vs. CLL stage A cases (n=116).

	complete cohort	cMBL	CLL Stage A	cMBL vs. CLL Stage A
variable	**N (%)**	**N (%)**	**N (%)**	**P-Value**
Screened Cases	206(100)	90(44)	116(56)	
Female	89(43)	37(41)	52(45)	
Male	117(57)	53(59)	64(55)	0.3
Age (median)	70	70	70	
IGHV unmutated	64(31)	28(31)	36(31)	
mutated	142(69)	62(69)	80(69)	0.6
CD38 negative (30%)	128(70)	52(63)	76(75)	
CD38 positive (30%)	56(30)	31(37)	25(25)	0.05
ZAP 70 negative	120(71)	54(69)	66(72)	
ZAP 70 positive	50(29)	24(31)	26(28)	0.4
del(11q) negative	176(92)	80(94)	96(91)	
del(11q) positive	15(8)	5(6)	10(9)	0.3
del(13q14) negative	69(38.5)	35(45)	34(34)	
del(13q14)positive (monoallelic)	110(61.5)	44(55)	66(66)	0.1
Trisomy 12 negative	124(76)	47(69)	77(80)	
Trisomy 12 positive	40(24)	21(31)	19(20)	0.08
del(17p) negative	185(97)	81(96)	104(98)	
del(17p) positive [1]	5(3)	3(4)	2(2)	0.4
***NOTCH1* wild type**	192(95)	84(94)	108(95)	
***NOTCH1* mutation**	11(5)	5(6)	6(5)	0.6
***SF3B1* wild type**	183(92)	82(94)	101(90)	
***SF3B1* mutation**	16(8)	5(6)	11(10)	0.2
***POT1* wild type**	190(96)	83(95)	107(96)	
***POT1* mutation**	8(4)	4(5)	4(4)	0.5
***XPO1* wild type**	170(99)	72(99)	98(99)	
***XPO1* mutation**	2(1)	1(1)	1(1)	0.7
***FBXW7* wild type (tri 12 cases only)**	29(93.5)	15(94)	14(93)	
***FBXW7* mutation (tri 12 cases only)**	2(6.5)	1(6)	1(7)	0.7
***MYD88* wild type (exon 3,4,5)**	195(98.5)	86(99)	109(98)	
***MYD88* mutation (exons 3,4,5)**	3(1.5)	1(1)	2(2)	0.6
***BIRC3* wild type (exons 7,10)**	196(99.5)	86(99)	110(100)	
***BIRC3* mutation (exons 7,10)**	1(0.5)	1(1)	0(0)	0.4
total number of mutations	43(21)	18(20)	27(23)	0.3
total number of low frequency mutations (excluding *SF3B1, Notch1* and *MYD88*)	13(6)	7(8)	6(5)	0.4

[1] all the del(17p) cases have mutated *IGHV* genes

Supplementary table 2. Clinical, cytogenetic and molecular characteristics of sequential (n=84) vs. single time point cases (n=122).

	sequential cases	single time point cases	
variable	**N (%)**	**N (%)**	**P-Value**
Screened cases	84(41)	122(59)	
Female	34(41)	55(45)	
Male	50(60)	67(55)	0.3
Age (median)	66	72	
Stage A CLL	49(58)	67(55)	
cMBL	35(42	55(45)	0.4
IGHV unmutated	28(33)	36(30)	
IGHV mutated	56(67)	86(71)	0.3
CD38 negative (30%)	54(70)	74(69)	
CD38 positive (30%)	23(30)	33(31)	0.5
ZAP70 negative	53(69)	67(72)	
ZAP70 positive	24(33)	26(28)	0.4
del(11q) negative	70(91)	106(93)	
del(11q) positive	7(9)	8(7)	0.6
del(13q14) negative	31(40)	38(37)	
del(13q14) positive (monoallelic)	46(60)	64(63)	0.4
Trisomy 12 negative	63(83)	61(69)	
Trisomy 12 positive	13(17)	27(33)	**0.05**
del(17p) negative	76(99)	109(96)	
del(17p) positive	1(1)	4(4)	0.3
***NOTCH1* wild type**	79(95)	113(94)	
***NOTCH1* mutation**	4(5)	7(6)	0.5
***SF3B1* wild type**	69(91)	113(93)	
***SF3B1* mutation**	7(9)	9(7)	0.4
***POT1* wild type**	80(96)	110(96)	
***POT1* mutation**	3(4)	5(4)	0.6
***XPO1* wild type**	67(100)	103(98)	
***XPO1* mutation**	0(0)		0.5
***FBXW7* wild type (tri 12 cases only)**	11(100)	18(90)	
***FBXW7* mutation (tri 12 cases only)**	0(0)	2(10)	0.4
***MYD88* wild type (exon 3,4,5)**	82(99)	113(98)	
***MYD88* mutation (exons 3,4,5)**	1(1)	2(2)	0.6
***BIRC3* wild type (exons 7,10)**	81(99)	115(100)	
***BIRC3* mutation (exons 7,10)**	1(1)	0(0)	0.4

Supplementary Table 3. Mutation characteristics and mutation prediction

No.	Chromosome (forward strand)	Gene Symbol (Hugo)	Mutation	Nature of Mutation	COSMIC variation	Somatic in CLL	conservation data(USCSC)	dbsnp/ 1000 genomes	SIFT	SIFT_INT	Polyphen	Polyphen_INT
1	2:198267490	*SF3B1*	32327A>AG:623Y>Y/C	missense variant	COSM1651682	yes	conserved to zebrafish	no known snp	0	deleterious	0.98	probably damaging
2	2:198266611	*SF3B1*	33205G>GA:742G>G/D	missense variant	COSM145923	yes	conserved to zebrafish	no known snp	0.12	tolerated	0.77	possibly damaging
3	2:198266834	*SF3B1*	32982A>AG:700K>K/E	missense variant	COSM84677	yes	conserved to zebrafish	no known snp	0.01	deleterious	0.96	probably damaging
4	2:198267361	*SF3B1*	32455A>AG:666K>K/E	missense variant	COSM110694	yes	conserved to zebrafish	no known snp	0	deleterious	0.98	probably damaging
5	2:198267369	*SF3B1*	32447C>CT:663T>T/I	missense variant	COSM145921	yes	conserved to zebrafish	no known snp	0	deleterious	1	probably damaging
6	2:198266709	*SF3B1*	33107G>GC:741K>K/N	missense variant	COSM572730	yes	conserved to zebrafish	no known snp	0	deleterious	0.99	probably damaging
7	2:198267484	*SF3B1*	32332C>CT:625R>R/C	missense variant	COSM255276	yes	conserved to zebrafish	no known snp	0	deleterious	1	probably damaging
8	2:198267545	*SF3B1*	32269G>GA:604A>A/T	missense variant	no data	unknown	conserved to zebrafish	no known snp	no data		no data	
9	2:198266795	*SF3B1*	33021G>GA:713A>A/T	missense variant	COSM1691803	yes	conserved to zebrafish	no known snp	0	deleterious	1	probably damaging
No.	**Chromosome (forward strand)**	**Gene Symbol (Hugo)**	**Mutation**	**Nature of Mutation**	**COSMIC variation**	**Somatic in CLL**	**conservation data(USCSC)**	**dbsnp/ 1000 genomes**	**SIFT**	**SIFT_INT**	**Polyphen**	**Polyphen_INT**
10	9:139390912	*NOTCH1*	c.7275delCAGCGG:p.S2426Hfs*128-deletion contains a known snp	frameshift variant	no data	unknown	conserved to rhesus	rs370722609	0.3	tolerated	0.09	benign
11	9:139390637-139390646	*NOTCH1*	c.7544_7545delCT P2515Rfs	frameshift variant	COSM13071	yes	conserved to zebrafish	no known snp	no data		no data	
12	9:139391076	*NOTCH1*	7115G>G/A:2372R>R/Q	missense variant	no data	unknown	conserved to x-tropicalis	rs373119531	0.22	tolerated	0.76	possibly damaging

No.	Chromosome (forward strand)	Gene Symbol (Hugo)	Mutation	Nature of Mutation	COSMIC variation	Somatic in CLL	conservation data(USCSC)	dbsnp/ 1000 genomes	SIFT	SIFT_INT	Polyphen	Polyphen_INT
13	7:124892275	*POt1*	37710A>T:K39I	missense variant	COSM131380	yes for cancers	conserved to mammals	no known snp	0	deleterious	0.06	benign
14	7:124826283	*POt1*	UTR g.104749T>G	intron variant	no data	unknown	conserved to rhesus	rs113233510	no data		no data	
15	7:124853116	*POt1*	76868A>AG:242Y>Y/C	missense variant	COSM1659034	yes for cancers	conserved to mammals	no known snp	0	deleterious	0.97	probably damaging
16	7:124503424	*POt1*	526G>GA:176G>G/R (known variant: 176G>G/V)	missense variant	COSM1547942	unknown	conserved to x-tropicalis	no known snp	0.01	deleterious	0.34	benign
17	7:124537282	*POt1*	in between two polymorphic regions in intron 4, chr7:124537282 T>A	intron variant	no data	no data	conserved to rhesus	no known snp	no data		no data	
18	7:124824015-124824016	*POt1*	c.1458_1459del:p.486_487del and c.1851_1852del:p.617_618del	frameshift variant, feature truncation	TMP_ESP_7_124464069_124464070	no data	conserved to rhesus	no known snp	no data		no data	
19	7:37338304	*POt1*	UTR g.23777T>G	intron variant	no data	no data	conserved to humans	no known snp	no data		no data	
20	7:124537197	*POt1*	UTR g.32906A>G	intron variant	no data	no data	conserved to mammals	no known snp	no data		no data	
No.	**Chromosome (forward strand)**	**Gene Symbol (Hugo)**	**Mutation**	**Nature of Mutation**	**COSMIC variation**	**Somatic in CLL**	**conservation data(USCSC)**	**dbsnp/ 1000 genomes**	**SIFT**	**SIFT_INT**	**Polyphen**	**Polyphen_INT**
21	2:61492337	*XPO1*	46290G>GA:571E>E/K	missense variant	COSM96797	yes	conserved to rhesus	no known snp	0	deleterious	1	probably damaging
No.	**Chromosome (forward strand)**	**Gene Symbol (Hugo)**	**Mutation**	**Nature of Mutation**	**COSMIC variation**	**Somatic in CLL**	**conservation data(USCSC)**	**dbsnp/ 1000 genomes**	**SIFT**	**SIFT_INT**	**Polyphen**	**Polyphen_INT**
22	4:152326215	*FBXW7*	208806C>C/G:479R>R/G	missense variant	COSM3127988	yes	conserved to rhesus	no known snp	0	deleterious	1	probably damaging
23	4:152332710	*FBXW7*	202312A>AC:291Y>Y/S (291 Y>Y/N is a known snp)	missense variant	no data	no data	conserved to rhesus	rs369187069 SNP	0.18	tolerated	0.07	benign

No.	Chromosome (forward strand)	Gene Symbol (Hugo)	Mutation	Nature of Mutation	COSMIC variation	Somatic in CLL	conservation data(USCSC)	dbsnp/ 1000 genomes	SIFT	SIFT_INT	Polyphen	Polyphen_INT
24	3:38140534	*MYD88*	c.649G>T, p.V217F	missense variant	COSM85941	yes for cancers	conserved to mammals	no known snp	0.36	tolerated	0.62	possibly damaging
25	3:38181970-38181972	*MYD88*	c.594_596delAAA, p.N199del	missense variant	no data	no data	conserved to rhesus	no known snp	no data		no data	
26	3:38141150	*MYD88*	ENSP00000379625.3:p.Leu265Pro, , c.794T>C, p.L265P	missense variant	no data	yes	conserved to mammals	rs387907272	0	deleterious	1	probably damaging
No.	**Chromosome (forward strand)**	**Gene Symbol (Hugo)**	**Mutation**	**Nature of Mutation**	**COSMIC variation**	**Somatic in CLL**	**conservation data(USCSC)**	**dbsnp/ 1000 genomes**	**SIFT**	**SIFT_INT**	**Polyphen**	**Polyphen_INT**
27	11:102207657	*BIRC3*	g.19464delC, c.1639delc, p.Q547Nfs*21, comment: close to COSM1627718, L548L	missense variant	no data	no data	conserved to zebrafish	no known snp	no	data	no data	no data

Supplementary table 4: Univariate Cox proportional hazard analysis of treatment-free and overall survival.

	NOTCH1 (all PEST terminating mutations)						*SF3B1 (Exon 14-16)*					*Low Frequency Mutations*				
variable	**status**	**total(%)**	**Wild-type (%)**	**Mutated (%)**	**OR**	**P-Value**	**total (%)**	**Wild-type (%)**	**Mutated (%)**	**OR**	**P-Value**	**total (%)**	**Wild-type (%)**	**Mutated (%)**	**OR**	**P-Value**
Screened Cases		203(99)	192 (95)	11(5)			199(100)	183(92)	16(8)			163 (79)	156(96)	7(4)		
Gender	Female	87(43)	83(95)	4(5)			87(44)	81(93)	6(7)			72(44)	69(96)	3(4)		
	Male	116(57)	109(94)	7(6)	1.3	0.45	112(56)	102(91)	10(9)	1.3	0.4	91(56)	87(96)	4(4)	1.0	0.6
Median Age		70	70	72			70	70	69			70	70	74		
Disease Stage	cMBL	89(44)	84(94)	5(6)			87(44)	82(94)	5(6)			73(45)	69(95)	4(5)		
	Stage A CLL	114(56)	108(95)	6(5)	1.1	0.57	112(56)	101(90)	11(10)	0.6	0.4	90(55)	87(97)	3 (3)	1.7	0.4
IGHV	unmutated	62(30)	54(87)	8(13)			63(32)	54(86)	9(14)			42(26)	38(90.5)	4(9.5)		
	mutated	141(70)	138(98)	3(2)	**6.8**	**0.004**	136(63)	129(95)	7(5)	**3.1**	**0.03**	121(74)	118(97.5)	3(2.5)	4.1	0.07
CD38 (30%)	negative	128(70)	126(98)	2(2)			123(70)	118(96)	5(4)			111(75.5)	107(96)	4(4)		
	positive	54(30)	47(87)	7(13)	**9.4**	**0.003**	54(31)	45(83)	9(17)	**4.7**	**0.012**	36(24.5)	33(92)	3(8)	2.4	0.2
ZAP70	negative	119(71)	116(97.5)	3(2.5)			114(70)	109(96)	5(4)			101(75)	99(98)	2(2)		
	positive	49(29)	44(90)	5(10)	**4.4**	**0.047**	49(30)	42(86)	7(14)	**3.6**	**0.03**	34(25)	30(88)	4(12)	**6.6**	**0.04**
del(11q)	absent	174(92)	165(95)	9(5)			169(92)	160(95)	9(5)			143(94)	137(96)	6(4)		
	present	15(8)	15(100)	0(0)	0.95	0.47	15(8)	9(60	6(40)	**11.8**	**<0.001**	9(6)	8(89)	1(11)	2.9	0.4
del(13q14)	absent	69(39)	63(91)	6(9)			67(39)	62(92.5)	5(7.5)			54(37)	52(96)	2(4)		
	present	108(61)	104(96)	4(4)	0.4	0.31	105(61)	96(91)	9(9)	1.2	0.5	86(63)	82(95)	4(5)	1.3	0.6
Trisomy 12	absent	123(76)	120(98)	3(2)			118(75)	105(89)	13(11)			97(76)	94(96)	3(4)		
	present	39(24)	33(85)	6(15)	**7.3**	**0.007**	39(25)	37(95)	2(5)	0.4	0.3	30(24)	28(93)	2(7)	2.2	0.4
del(17p)	absent	183(97)	175(96)	8(4)			178(97)	164(92)	14(8)			148(97)	141(95)	7(5)		
	present	5(3)	5(100)	0(0)	0.9	0.8	5(3)	4(80	1(20)	3.0	0.4	4(3)	4(100)	0(0)	0.95	0.8
Progressive Disease	absent	138(68)	135(98)	3(2)			133(67)	127(95.5)	6(5.5)			117(72)	116(99)	1(1)		
	present	65(32)	57(88)	8(12)	**6.3**	**0.005**	66(33)	56(85)	10(15)	**3.8**	**0.012**	46(28)	40(87)	6(13)	**17.4**	**0.002**
Treatment	no treatment	138(68)	135(98)	3(2)			133(67)	128(96)	5(4)			118(72)	117(99)	1(1)		
	treatment	65(32)	57(88)	8(12)	**6.3**	**0.005**	66 (33)	55(83)	11(17)	**5.1**	**0.003**	45(28)	39(87)	6(13)	**18.0**	**0.002**

Frequencies of *NOTCH1, SF3B1* and flow frequency mutations with biologic and clinical characteristics in Binet Stage A CLL and cMBL (n=206): *NOTCH1*-, *SF3B1*- and low frequency mutations are all associated with progressive disease and need for therapy

Supplementary table 5. Sequential data of patients with a *POT1* mutation (n=8)

ID	diagnosis	gender	age at diagnosis	IGHV	year 1	cytogenetics at diagnosis	FISH 1	*POT1* mutation	disease stability	year 2	cytogenetics 2nd timepoint	FISH2	treatment	clonal evolution
50	stage A CLL	female	59ys	mutated	1985	normal	normal	g.104749T>G	stable	2004	normal	13q loss +/-	no	yes
335	cMBL	male	77ys	unmutated	1994	normal	normal	g.76868A>G,c.1321A>G,p.Y242C	stable	2002	normal	normal	no	no
568	cMBL	male	79ys	unmutated	1997	normal	normal	UTR g.32906A>G	stable	2004	normal	13q loss +/-	no	yes
2938	stage A CLL	male	72ys	unmutated	2005	46, XY, del (13) q14q22	13q loss +/-	g.37710A>T;p.K39I	progressive	2011	same as 2005	13q loss +/-	yes	no
3069	cMBL	female	63ys	mutated	2003	normal	normal	124503424C>T;c.526G>A;p.G176R	progressive	2007	normal	normal	yes	no
3439	stage A CLL	female	69ys	unmutated	2004	normal	13q loss +/-	c.1458_1459del:p.486_487del and c.1851_1852del:p.617_618del	progressive	2007	normal	13q loss +/-	yes	no
3873	stage A CLL	female	n/a	unmutated	2008	normal	normal	UTR g.23777T>G	progressive	2010	not done	normal	yes	no
4286	cMBL	male	59ys	mutated	2008	normal	13q loss +/-	intron between two polymorphic regions in intron 4, chr7:124537282 T>A	stable	2011	n/a	n/a	no	n/a

Supplementary table 6. Clinical data from patients with mutations in *SF3B1* and *XPO1*, detected at the second time point only (n=5).

Gene	ID	diagnosis	gender	age	IGHV	cytogenetics and molecular prognostic factor analysis at diagnosis	acquired mutation	disease stability	treatment	clonal evolution	follow up data
SF3B1	338	cMBL	male	76ys	unmutated	trisomy 12, ZAP70 positive, 7544_7545delCT NOTCH1 mutation	33205G>GA:742G>G/D	stable	no	n/a	died 36 months after diagnosis independent from disease
	656	CLL Binet A	male	82ys	mutated	normal, no initial mutations	32982A>AG:700K>K/E	progressive	yes	yes	died 7 months after detection of mutation in progressive disease
	948	cMBL	male	78ys	mutated	normal, no initial mutations	32455A>AG:666K>K/E	stable	no	no	died 17 years after diagnosis independent from disease
	3439	CLL Binet A	female	69ys	unmutated	n/a	18771A>AG:626N/S	progressive	yes	n/a	progression and treatment from 36 months after diagnosis
XPO1	306	cMBL	male	n/a	unmutated	n/a	46290G>GA:571E>E/K	progressive	yes	n/a	progression and treatment from 3 months after diagnosis

Supplementary table 7. Primer Sequences.

	Primers and sequences	
Gene name	**Primer name**	**Primer sequence (5'-3')**
SF3B1	SF3B1_Ex14F	CCAACTCATGACTGTCCTTTCTT
	SF3B1_Ex14R	GGGCAACATAGTAAGACCCTGT
	SF3B1_Ex14R_Seq*	CAAGATGGCACAGCCCATAA
	SF3B1_Ex15F	TTGGGGCATAGTTAAAACCTG
	SF3B1_Ex15R	AAATCAAAAGGTAATTGGTGGA
	SF3B1_Ex16F	TCTTCATTAAAGTTAAGGCGACA
	SF3B1_Ex16R	TGTTAGAACCATGAAACATATCCA
NOTCH1	NOTCH1_Ex34_1F	AGCAAACATCCAGCAGCAG
	NOTCH1 _Ex34_1R	GCTCTCCTGGGGCAGAATA
	NOTCH1_Ex34_2F	GAGCTTCCTGAGTGGAGAGC
	NOTCH1_Ex34_2R	GTGAGGAAGGGGTGCTCAG
	NOTCH1_Ex34_3F	CACTATTCTGCCCCAGGAGA
	NOTCH1_Ex34_3R	CAGTCGGAGACGTTGGAATG
	NOTCH1_Ex34F_Seq*	GCCACAAAACTTACAGATGC
	NOTCH1_Ex34R_Seq*	CGCCGTTTACTTGAAGG
XPO1	XPO_Ex15_16F	TTAGGAAATGTACTTGTAGTTTCTA
	XPO_Ex15_16R	GGGTCTCTAACAAGACAAAAACAT
	XPO1_Ex15F_Seq*	GCTCATTATTTGCATTTGAAACC
	XPO1_Ex15R_Seq*	TCTAATTCATACCTATCCCTTGCAT
	XPO1_Ex16F_Seq*	TGCAAGGGATAGGTATGAATTAGA
	XPO1_Ex16R_Seq*	TTTTGTCCTGGACTCCATCAT
POT1	POT1_Ex4F_HRM**	TGCAATGTAATTAGAGAATAAAAGCTG
	POT1_Ex4R_HRM**	ATTATACGTATTTTGGTGATTGATTCA
	POT1_EX4F_Seq*	AAGTGCAATATCTGCCAAGT
	POT1_Ex4R_Sseq*	TCCAAACAATGACAAAATCA
	POT1_Ex5F_HRM**	CACATGTATCTATGTGTGTGGCATA
	POT1_Ex5R_HRM**	AGCATGTAATCACATTGGAGGTT
	POT1_Ex5F_Seq*	TCAGCAGATATTCCAGACAA
	POT1_Ex5R_Seq*	AGCTTAGACAACTTTGCACAT
	POT1_Ex6F_HRM**	AAACTCCACCAGTTTTAATACCTACC
	POT1_Ex6R_HRM**	TACATGGATTTGCTGCTAATATGAT
	POT1_Ex6F_Seq*	AGCCAAAGAATATGCATCAG
	POT1_Ex6R_Seq*	CCATTTATAAACAAAGTTCTAAGGA

POT1	POT1_Ex7F_HRM**	TTCTCTTCAAATAAATATAAGTTCTAGAC
(continued)	POT1_Ex7R_HRM**	GGTTTGGTGTTTTGAAGTAAGCA
	POT1_Ex7F_Seq*	GCAGTGGTTTGTTCAAATG
	POT1_Ex7R_Seq*	TTGCAGTGTGTATTGAAAGC
	POT1_Ex8F_HRM1**	TGGTGCTAACTTATAATTCCCAGTATT
	POT1_Ex8R_HRM1**	CCTTACGTGTTTGGGCATCT
	POT1_Ex8F_HRM2**	AGATGCCCAAACACGTAAGG
	POT1_Ex8R_HRM2**	CTGTTTTCTACTTTGCCCTACTTTC
	POT1_Ex8F_Seq*	CCACACAAATCTCATGTCAA
	POT1_Ex8R_Seq*	TCACCCAGTAAATCTCTTTAGC
	POT_Ex9F_HRM**	TCAGAGATCTTGCCACATGAA
	POT1_Ex9R_HRM**	TTATGGCAGGTATGGGATGG
	POT1_Ex9F_Seq*	CATTTTACAACTTAAAAATCAAAGA
	POT1_Ex9R_Seq*	TTCCACATTACCCATATTTCA
	POT1_Ex10F_HRM**	TCGGCTTAATCGATACCTTATTTAC
	POT1_Ex10R_HRM**	TTTTTCCCCACTTTCTAAATAACAA
	POT1_Ex10F_Seq*	ATTTGTTTCATTTGGCTCAT
	POT1_Ex10R_Seq*	CCATGCAGCTGATATTCAA
	POT1_Ex18F_HRM**	TCAAGTAAAAGAAGTGTGGGATTG
	POT1_Ex18R_HRM**	AAGGACAAATTCTTCCAGATTCC
	POT1_Ex18F_Seq*	TTTGACTGCAGGAATTATGA
	POT1_Ex18R_Seq*	GATTTTGGAGTTGAGACCAG
BIRC3	BIRC3_Ex7F_HRM**	TTCCATATAGTTATCCATTTTGAACCT
	BIRC3_Ex7R_HRM**	ACATACTTGATTCTTTTTCCTCAGTTG
	BIRC3_Ex7F_Seq*	TGCCTATACATTTTGTTGGTT
	BIRC3_Ex7R_Seq*	AAAAACCTGACTGGATTGAG
	BIRC3_Ex10F_HRM**	TGAAGAAGCAAACTGCCTTTTATT
	BIRC3_Ex10R_HRM**	AAAGTTTAGACGATGTTTTGGTTCT
	BIRC3_Ex10F_Seq*	CCACAGAAGATGTTTCAGGT
	BIRC3_Ex10R_Seq*	GTGCTACCTCTTTTTCGTTC
MYD88	MYD88_Ex3F_HRM**	TCTGACCACCACCCTTGTG
	MYD88_Ex3R_HRM**	GGCCTTCTAGCCAACCTCTT
	MYD88_Ex3F_Seq*	GGCACTTTCTCTGAGGAGTA
	MYD88_Ex3R_Seq*	GACAGTGCACAGCTAGGAG
	MYD88_Ex4F_HRM**	GCTGAACTAAGTTGCCACAGG
	MYD88_Ex4R_HRM**	CCAGAGCAGGGTTGAGCTT
	MYD88_Ex4F_Seq*	CAGGGGATATGCTGAACTAA
	MYD88_Ex4R_Seq*	GATCTTCAGCAGTTCTTTGG
	MYD88_Ex5F_HRM**	CAGGTGCCCATCAGAAGC
	MYD88_Ex5R_HRM**	GGTTGGTGTAGTCGCAGACA
	MYD88_Ex5F_Seq*	GCAGAAGTACATGGACAGGCAGACAGATAC
	MYD88_Ex5R_Seq*	GTTGTTAACCCTGGGGTTGAAG

FBXW7	FBXW7_Ex2F	ATTTTCCCCTGCAGAATGTG
	FBXW7_Ex2R	TTTAGTAATACAAAGACTGTGAGGAAA
	FBXW7_Ex3F	TGACTCAAGATTTGATAGTTAGACGA
	FBXW7_Ex3R	AAACTAAAACACTTTCAGAATCAACTC
	FBXW7_Ex4F	TCTTTGCTTTCACTTTTGTTTTT
	FBXW7_Ex4R	GCAGCAATTAAGTGAGGCATT
	FBXW7_Ex4Fseq	ACCATGTTCAGCAACACCAA
	FBXW7_Ex5F	GCCTGTAATTTGGGACATCTG
	FBXW7_Ex5R	CAAACTGACAATACCGAATACCA
	FBXW7_Ex6F	TCAAGTATCTCATCCTGTGGAGAA
	FBXW7_Ex6R	TTCGGCTCATCTGAATGTGT
	FBXW7_Ex7F	TGGTTTTGAGCAGAGAGATGG
	FBXW7_Ex7R	TTTCTTTCTACAGAAGAGGAGTGTCA
	FBXW7_Ex8F	TGTTCCTGTTTATGCCTTCATTT
	FBXW7_Ex8R	CCAGTTGCTACTTGCAATGAT
	FBXW7_Ex9F	TCACTTTTCCTTTCTACCCAAAA
	FBXW7_Ex9R	CTACACAGAAAGGGCCCAAA
	FBXW7_Ex10F	AAAAATTCTAAACGTGGGTTTTT
	FBXW7_Ex10R	TGGATCAGCAATTTGACAGTG
	FBXW7_Ex11F	TCCTCTTCCCCCTTTCCTAC
	FBXW7_Ex11R	TTTTGTGATGCTAAGGCTCCAT
	FBXW7_Ex12F	TTTCAAATGTTGCATTTATTGTATG
	FBXW7_Ex12R	CAACATCCTGCACCACTGAG

* Theses primers were used as a sequencing primers only

** These primers were used as High Resolution Melt primers only

References (Supplement)

1. Parker H, Rose-Zerilli MJ, Parker A, et al. 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. Leukemia. 2011;25(3):489-497.

2. Shaffer MLS, and L.J. Campbell. . ISCN, An International System for Human Cytogenetic Nomenclature: Recommendations of the International Standing Committee on Human Cytogenetic Nomenclature Published in collaboration with 'Cytogenetic and Genome Research' Basel: SKarger. 2009.

3. Best OG, Ibbotson RE, Parker AE, Davis ZA, Orchard JA, Oscier DG. ZAP-70 by flow cytometry: a comparison of different antibodies, anticoagulants, and methods of analysis. Cytometry Part B, Clinical cytometry. 2006;70(4):235-241.

4. Oscier DG, Gardiner AC, Mould SJ, et al. Multivariate analysis of prognostic factors in CLL: clinical stage, IGVH gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors. Blood. 2002;100(4):1177-1184.

5. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood. 1999;94(6):1848-1854.

6. White HE, Hall VJ, Cross NC. Methylation-sensitive high-resolution melting-curve analysis of the SNRPN gene as a diagnostic screen for Prader-Willi and Angelman syndromes. Clin Chem. 2007;53(11):1960-1962.

7. Oscier DG, Rose-Zerilli MJJ, Winkelmann N, et al. The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. Blood. 2013;120(22):4441-4443.

8. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011;475(7354):101-105.

9. Fabbri G, Rasi S, Rossi D, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. J Exp Med. 2011;208(7):1389-1401.

10. Rossi D, Fangazio M, Rasi S, et al. Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. Blood. 2012;119(12):2854-2862.

11. Gianfelici V. Activation of the NOTCH1 pathway in chronic lymphocytic leukemia. Haematologica. 2012;97(3):328-330.

12. Wang L, Lawrence MS, Wan Y, et al. SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. N Engl J Med. 2011;365(26):2497-2506.

13. Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nature Genet. 2011;44(1):47-52.

14. Rossi D, Bruscaggin A, Spina V, et al. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. Blood. 2011;118(26):6904-6908.

15. Weng AP, Ferrando AA, Lee W, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science. 2004;306(5694):269-271.

haematologica
the hematology journal

Early Release Paper

ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q-deleted chronic lymphocytic leukemia, data from the UK LRF CLL4 trial

by Matthew J.J. Rose-Zerilli, Jade Forster, Helen Parker, Anton Parker, Ana E. Rodriguez, Tracy Chaplin, Anne Gardiner, Andrew J. Steele, Andrew Collins, Bryan D. Young, Anna Skowronska, Daniel Catovsky, Tatjana Stankovic, David G. Oscier, and Jonathan C. Strefford

Haematologica 2013 [Epub ahead of print]

Citation: Rose-Zerilli MJ, Forster J, Parker H, Parker A, Rodriguez AE, Chaplin T, Gardiner A, Steele AJ, Collins A, Young BD, Skowronska A, Catovsky D, Stankovic T, Oscier DG, and Strefford JC. ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q-deleted chronic lymphocytic leukemia, data from the UK LRF CLL4 trial. Haematologica. 2013; 98:xxx doi:10.3324/haematol.2013.098574

ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q-deleted chronic lymphocytic leukemia, data from the UK LRF CLL4 trial

Running title: 11q23 deletions, ATM and BIRC3 mutations in CLL

Matthew J.J. Rose-Zerilli [1], Jade Forster [1], Helen Parker [1], Anton Parker [2], Ana E. Rodríguez [3], Tracy Chaplin [4], Anne Gardiner[2], Andrew J. Steele[1], Andrew Collins [5], Bryan D. Young [4], Anna Skowronska [6], Daniel Catovsky [7], Tatjana Stankovic [6], David G. Oscier [1,2] and Jonathan C. Strefford [1]

[1]Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[2]Department of Haematology, Royal Bournemouth Hospital, Bournemouth, United Kingdom

[3]Cancer Research Center (IBMCC-CSIC). University of Salamanca, Salamanca, Spain

[4]Centre for Medical Oncology, Barts and The London School of Medicine and Dentistry, London, United Kingdom

[5]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[6]School of Cancer Sciences, University of Birmingham, Birmingham, United Kingdom

[7]Haemato-oncology Research Unit, Division of Molecular Pathology, The Institute for Cancer Research, Sutton, United Kingdom

Correspondence
Jonathan C. Strefford, Cancer Genomics Group, Cancer Sciences Unit, Somers Cancer research Building, Southampton General Hospital, Tremona Road, Southampton SO16 6YD.
E-mail: JCS@soton.ac.uk

Acknowledgments

The authors would like to thank Leukaemia and Lymphoma Research, the Kay Kendall Leukaemia Fund and Wessex Medical Research for funding this study.

Funding

The LRF CLL4 trial was funded by a core grant from Leukaemia and Lymphoma Research, with associated research work supported by the MRC (G8223452) and Cancer Research UK, and laboratory studies by the Arbib Foundation, Schering Healthcare UK, Schering AG - Germany and Leukaemia and Lymphoma Research. The authors gratefully acknowledge all patients and clinicians who participated in the trial. The authors would also like to thank Professor Nick Cross for help with the preparation of this manuscript.

Abstract

ATM mutation and *BIRC3* deletion and/or mutation have independently been shown to have prognostic significance in chronic lymphocytic leukemia. However the relative clinical importance of these abnormalities in patients with a deletion of 11q encompassing the *ATM* gene has not been established. We screened a cohort of 166 patients enriched for 11q-deletions for *ATM* mutations and *BIRC3* deletion and mutation and determined the overall and progression-free survival among the 133 of these cases treated within the UK LRF CLL4 trial. SNP6 profiling demonstrated that *BIRC3* deletion occurred in 83% of 11q-deleted cases and always co-existed with *ATM* deletion. For the first time we have demonstrated that 40% of *BIRC3*-deleted cases have concomitant deletion and mutation of *ATM.* While *BIRC3* mutations were rare, they exclusively occurred with *BIRC3* deletion and a wild-type residual *ATM* allele. In 11q-deleted cases we confirmed that *ATM* mutation was associated with a reduced overall and progression-free survival comparable to that seen with *TP53* abnormalities whereas *BIRC3* deletion and/or mutation had no impact on overall and progression free survival. In conclusion, in 11q-deleted patients treated with first line chemotherapy, *ATM* mutation rather than *BIRC3* deletion and/or mutation identifies a sub-group with a poorer outcome.

Introduction

Deletion of chromosome 11q (termed del11q) was first recognized as a recurrent karyotypic abnormality acquired during the course of the disease in patients with progressive chronic lymphocytic leukemia (CLL).[1] Subsequent interphase FISH analysis identified 11q deletion in approximately 20% of patients with CLL, and associations with bulky lymphadenopathy, and a poorer outcome for patients under the age of 55 years were noted.[2,3] Subsequent studies have documented associations with unmutated *IGHV* genes, del13q, genomic complexity, short telomeres, progressive disease and a poor outcome in response to alkylating agent or purine analogue treatment, which was improved by their use in combination and ameliorated by the further addition of rituximab.[4–10]

Genomic profiling studies have refined previous karyotypic and FISH studies showing that 11q deletions are mono-allelic, frequently large and include a minimally deleted region (MDR), which encompasses the *ATM* gene.[11–14] Evidence that *ATM* is a key target of 11q deletions derives from: 1) mutation of the *ATM* gene is found in 30 – 40% of patients with an 11q deletion;[15,16] 2) the presence of an *ATM* mutation results in impaired DNA damage responses;[15,17–19] and 3) patients in the UK CLL4 trial with biallelic *ATM* abnormalities (deletion and mutation) have a poorer outcome following the initial therapy with alkylating agent and/or purine analogue therapy compared to those with mono-allelic *ATM* deletion or mutation.[20]

However, there is still uncertainty as to whether the poorer outcome of patients with 11q deletion in the absence of an *ATM* mutation is simply a consequence of *ATM* haploinsufficiency. Alternative possibilities that have been considered include deletion, mutation or epigenetic silencing of other genes either within or outside the MDR or the associated genomic complexity.[8,21–23] Neither candidate gene sequencing nor whole exome sequencing studies have identified mutations within other genes located in the MDR.[24,25] However, recent data has revealed a high incidence of deletion or more rarely mutation of *BIRC3*, a negative regulator of non-canonical NF□B signaling located at 11q22. It has been reported that *BIRC3* deletion and/ or mutation (previously termed '*BIRC3* disruption') occurs in a mutually exclusive manner with *TP53* abnormalities, is associated with fludarabine-resistance,[26] and when detected at diagnosis predicts for poor overall survival independent of 11q deletion.[27]

The above data strongly suggest that there are subsets of del11q patients that exhibit differing responses to standard treatment. However, the relative frequency and clinical significance of *ATM* and *BIRC3* abnormalities is unclear. This study addresses this issue in a large cohort of 11q-deleted patients detected by SNP6 profiling and screened for *ATM* and *BIRC3* mutations. As a consequence, in the context of a phase III clinical trial of

chemotherapy, we show that *ATM* mutational status remains the most clinically informative genomic lesion in 11q-deleted CLL, identifying cases with outcome comparable to *TP53* deletion and/or mutation patients, and that the presence of *BIRC3* deletion and/or mutation is associated with outcome comparable to other 11q-deleted CLL cases.

Methods

Patients and molecular diagnostic assays

166 untreated CLL patients diagnosed based on standard morphologic and immunophenotypic criteria were included in this study (**Table 1** and **supplementary table 1**). This cohort principally included patients from the LRF UK CLL4 trial[9] [n=133] which allowed accurate clinical correlations to be made, but also included 33 additional 11q-deleted patients, to permit more significant associations between the 11q deletion and other genomic variables to be made, such as with mutational status of *ATM* and *BIRC3*. The additional del11q patients were sampled subsequent to the development of progressive disease, with a median time from diagnosis of 3.2 ± 5.2 years (±1SD; ranging from 1 month to 27 years). For the entire cohort of 166 patients, mutational data was available for *TP53* [n=125], *SF3B1* [n=140] and *NOTCH1* [n=146] [28,29]. **Details on the molecular diagnostic assays[30,31] are available in supplementary methods**. Informed consent was obtained from all patients in accordance with the Helsinki declaration and our local ethics committee approved this study.

DNA extraction, SNP6 array hybridization, data extraction and analysis

Genomic DNA was extracted from CLL B-cells [n=166] and buccal swabs [n=32], prior to being purified, amplified, labeled and hybridized to the Affymetrix SNP6 platform (Affymetrix, Santa Clara, CA) as previously described.[32] (**Supplementary methods**)

Mutational analysis of *ATM* and *BIRC3 g*enes

The presence or absence of somatically-acquired single nucleotide variants (SNVs or mutations) in *ATM* and *BIRC3* were successfully ascertained in 105 [CLL4 cases: n=79/133] and 162 [CLL4 cases: n=131/133] patients, respectively. Denaturing high-performance liquid chromatography (DHPLC), high-resolution melt (HRM) PCR analysis and Sanger sequencing was utilised.[20,26,27,29] **Extended methods and the strategy for assigning the somatic nature of each SNV are detailed in supplementary methods**.

Statistical analysis

Statistical analysis was performed using SPSS (v20). Associations with treatment response and clinical outcome were only performed on the LRF UK CLL4 cases due to the homogeneous and well-annotated nature of this cohort (**Supplementary methods**). The LRF UK CLL4 cohort included in this current study did not differ significantly from the entire cohort for an extensive panel of variables, with the enrichment exception of del11q (P=0.002) (**Supplementary Table 1**). Kaplan-Meier analysis with the Log Rank test or Cox regression were used for survival

analyses on Overall (OS) and Progression-Free Survival PFS. Chi-squared (Pearson's or Fisher's Exact test when necessary) were also employed for some comparisons and are described where relevant in the main text or table footnotes. Results were considered statistically significant at the 5% level (P=0.05). Previous reports have demonstrated that *BIRC3* disruption occurs in 50% of 11q deleted cases and is associated with survival comparable to *TP53* deleted CLL.[26,27] Therefore, with clinical follow-up of 10 years, our LRF CLL4 cohort [del11q, n=36] had 97% power with a significance level = 0.05 to detect a OS difference between cases with and without *BIRC3* deletion and/ or mutation based on the median CLL4 OS times for cases with del11q (53 months) and del17p (14 months) detected by FISH.

Results

Incidence of *ATM* and *BIRC3* deletions

We analysed all 166 patients for 11q deletions and copy number neutral-loss of heterozygosity (CNNLOH) events using the SNP 6.0 genome-wide copy number arrays. Sixty-nine patients had 11q deletions that all included a minimally deleted region (MDR) which was 416 Kb in size (chr11:107.498-107.914Mb) and contained the following genes: *ACAT1, NPAT, ATM, C11orf65, KDELC2, LOC1000127964, LOC1000128794* and *EXPH5* (**Figure 1A**). We observed that the majority of 11q-deleted cases (67/69) had deletions much larger than the MDR and 57/69 (83%) lost one copy of the *BIRC3* gene. In 56 of these patients, *ATM* and *BIRC3* genes were lost as part of a single deletion event while in one patient the genes were lost as a consequence of two separate deletions events. In four patients (6%) the centromeric deletion breakpoint was within the *BIRC3* gene body, resulting in loss of the 3' end of the gene that contains the C terminal RING domain important for the proteasomal degradation of *MAP3K14* (**Figure 1B**).

Incidence of *ATM* and *BIRC3* mutations

BIRC3 mutation screening was successfully ascertained in 162/166 patients. A *BIRC3* gene mutation was detected in only three patients all of whom had a large mono-allelelic 11q-deletion encompassing both the *ATM* gene and the *BIRC3* gene located at 11q22. The *BIRC3* gene mutations are predicted to be deleterious for BIRC3 protein function, the p.R434Kfs*10 (frameshift in/del) and p.L585* (STOP-codon) mutations both target highly conserved amino acid residues, resulting in truncation of BIRC3 protein with loss of the C terminal CARD-RING and RING domains, respectively. The p.I570del (3bp in-frame deletion) mutation targets another highly conserved amino, resulting in loss of the C-beta branched amino acid residue, isoleucine at the C terminal end of the beta-sheet structure in the RING domain (**Figure 1C**). *ATM* gene mutation screening was successfully completed in 105 patients and identified mutations in 36 patients. The mutations were classified as somatic and pathogenic based on our previously reported criteria (**See Supplementary Methods**) and occurred in exons 2-63, between amino acids 37-3047, affecting the TAN, FAT, PI3K and FATc domains of the ATM protein (**Supplementary Table 3**). *ATM* mutations occurred in the absence of an 11q-deletion in 27% (13/48) of patients screened and in 40% (23/57) of 11q-deleted patients. Interestingly, we identified that single case with an *ATM* mutation (c.8428-8450del23:p.Lys2810fs) also harbored CNNLOH of the region duplicating the variant. This case highlights that in rare instances biallelic inactivation of *ATM* may occur independently of chromosome deletion (this case was not included in our survival analyses), yet current molecular diagnostic tests do not detect these events. Importantly, we observed that within

11q-deleted patients, with loss of both *ATM* and *BIRC3* genes and successfully screened for gene mutation, 40% (19/48) with a *BIRC3* deletion also had an *ATM* mutation (**Figure 1D**).

Relative importance of *ATM* and *BIRC3* abnormalities on outcome of patients with 11q deletions

Within the CLL4 cohort [n=133], there were 36 patients with an 11q deletion, of whom 32 had loss of both *ATM* and *BIRC3* genes, two had a *BIRC3* mutation and 14 had an *ATM* mutation. Initially, we confirmed in univariate analysis that within this cohort, unmutated IGHV genes and deletions of 11q and *TP53* were associated with short PFS and OS, as found in larger studies of prognostic markers in this trial. (**Table 2 model A**).[9,20,28,29,33] Specifically, 11q deleted patients exhibited a median OS and PFS of 53 months (95%CI:36-69; P=0.004) and 20 months (95%CI:6-33; P<0.001), respectively.[9,20,28,29,33] Inclusion of *BIRC3* deletion and mutation into the model (**Table 2 model B**) showed that *BIRC3* deleted/mutated cases had a reduced survival compared to non-11q deleted cases, with a median OS of 53 months (95% CI: 39-67) and PFS of 17 (95% CI: 4-30). As expected there was no detectable difference (OS: P=0.52 & PFS: P=0.42) in outcome between all cases with del11q and those with *BIRC3* deletion/mutation, consistent with the high incidence of *BIRC3* loss in del 11q cases (**Figure 2a & b and Table 2 model B**).

Finally, we demonstrated that the most significant reduction in OS and PFS in 11q-deleted patients was observed in patients with mutations targeting the *ATM* gene. These patients exhibited a median OS and PFS of 42 (95% CI:13-71) and 10 months (95% CI:6-15), respectively. Importantly, cases with a *BIRC3* deletion/mutation that did not have an *ATM* mutation had significantly longer survival times for OS (76 vs. 42 months; P=0.05) and PFS (28 vs. 10 months; P=0.01) than cases with biallelic inactivation of *ATM* (**Table 2, model C**).

Discussion

This study extends recent observations on the incidence and clinical significance of *ATM* and *BIRC3* loss and/or mutation (SNV) in CLL patients with an 11q deletion. We had previously shown that the combination of an 11q deletion encompassing the *ATM* gene and mutation of the remaining *ATM* allele was associated with a shorter progression -free and overall survival than a del11q with wild type *ATM* in patients receiving first line chemotherapy in the UK CLL4 trial.[20] This observation is consistent with the importance of functional ATM protein in the response to DNA damage.[17,34–36] While several smaller studies have not identified an impact of *ATM* mutational status on patient survival, they did not differentiate *ATM* mutated individuals into those with and without deletion of 11q23.[37,38] Rossi and colleagues independently reported that *BIRC3* disruption resulting from complete or partial *BIRC3* loss with or without mutation of the remaining allele is common in patients who are refractory to, but not in those sensitive to fludarabine-containing regimens.[26] Specifically they noted that: 1) recurrent mutations target *BIRC3* in CLL, albeit at low frequencies; 2) *BIRC3* is recurrently deleted principally due to large genomic deletions on 11q; 3) in a single CLL case the *BIRC3* gene was deleted without concomitant loss of *ATM* and 4) the gene-body of *BIRC3* is targeted by rare proximal 11q-deletion breakpoints. This data is consistent with the tumour suppressor role of BIRC3 as a negative regulator of the non-canonical NF-KB pathway and the known importance of NK-KB pathway activation in CLL by multiple mechanisms (39-43)

By studying a relatively large cohort of patients with del 11q screened for both *ATM* and *BIRC3* abnormalities, we confirm the high incidence of *BIRC3* deletion in cases with del 11q and the lack of *BIRC3* deletion in the absence of *ATM* deletion. In support of previously published work,[26] our data also shows that *BIRC3* mutations are rare in a cohort of untreated CLL patients with progressive disease, suggesting limited clonal selection of these mutations prior to treatment. Our most important observation is the high frequency of missense or nonsense *ATM* mutations in *BIRC3* deleted cases (40%).

Investigating the clinical significance of these findings in patients entered into the UK CLL4 trial confirmed that patients with bi-allelic *ATM* abnormalities had a shorter PFS and OS than cases with del11q and wild type *ATM*. Although 11q-deleted cases with *BIRC3* deletion and/or mutation did exhibit reduced OS and PFS compared to non-11q-deleted cases, this was not significantly different from 11q-deleted cases without *BIRC3* deletion and/or mutation and much superior to cases with bi-allelic *ATM* abnormalities and *TP53* deletion/mutation. Furthermore, when we stratified the CLL4 cases that have large 11q23 deletions encompassing both *BIRC3* and *ATM* genes (n=24) by presence or absence of *ATM* mutation we were able to identify that patients with an *ATM* mutation had a shorter PFS (HR = 3.03, 95% CI: 1.22-7.49; P=0.017), but not OS, when compared to the patients with only *BIRC3*

deletion/mutation (**data not shown**). Our negative finding for CLL4 overall survival in this sub-group analysis must be interpreted with caution due to the lack of power (29%; Supplementary methods).

Several limitations to this study are noted: 1) In view of the frequency with which *ATM* and *BIRC3* deletions co-exist, a much larger study would be required to identify differences in outcome between patients with *ATM* deletion compared to those with *ATM* and *BIRC3* deletion 2) Given that the LRF CLL4 trial compared different chemotherapy regimens in previously untreated patients, it will be important to validate these observations in the context of chemo-immunotherapy trials, in studies of novel agents, such as B-cell receptor (BCR) signalling inhibitors and in patients with relapsed or refractory disease. 3) Our study has focused on *ATM* and *BIRC3* abnormalities. However deletion of 11q frequently results in the loss of many other genes involved in key regulatory pathways that could impact on CLL pathogenesis.[8] As an example, *MRE11A* and *H2AFX,* both of which are involved in DNA damage response are also deleted in 58% [n=40] and 26% [n=18], respectively. Indeed, 53% (17/31) of our *BIRC3*-deleted cases also lost *H2AFX* and/or *MRE11A*. Furthermore, although we identified recurrent deletion breakpoints within the gene body of *BIRC3*, we also identified recurrent breakpoints in the bodies of other 11q genes (**Supplementary Table 4**), including *CEP164*, which encodes a mediator protein required for the maintenance of genomic instability.[39] Similarly we have not investigated the potential clinical consequences of the elevated genomic complexity associated with del11q nor the incidence of epigenetic silencing of genes on the retained 11q allele. As some previous studies have noted an association between del11q and *SF3B1* gene mutations we considered whether the incidence of *SF3B1* mutations might differ between cases with or without an *ATM* mutation but observed no significant enrichment of *SF3B1* mutations in either the 11q-deleted, *ATM* mutated or patients with biallelic *ATM* abnormalities suggesting that this mutation does not contribute to the poor prognosis associated with 11q abnormalities (**data not shown**).

In conclusion, we confirm that in patients treated with chemotherapy as part of the LRF CLL4, the presence of an *ATM* mutation subdivided 11q-deleted CLL into a sub-group with significantly shorter overall and progression-free survival. *BIRC3* deletion/mutation did not identify an 11q-deleted patient subgroup with dismal outcome comparable to *TP53* deletion/mutation, in fact within 11q-deleted patients *BIRC3* deletion/mutation was associated with significantly longer survival times than *ATM* mutation. We would caution against interpreting the clinical impact of *BIRC3* abnormalities without knowledge of *ATM* mutational status.

Authorship

Contribution: This work was funded by grants awarded to J.C.S.; M.J.RZ, H.P, T.C. and B.D.Y. performed the microarray analysis; J.F performed the *BIRC3* mutational analysis; A.S. and T.S. performed the *ATM* mutation analysis. A.G and A.P. performed the cytogenetic and molecular diagnostic assays. M.J.RZ. and A.C. conducted statistical analyses; and D.C. and D.G.O. contributed patient samples and data; D.G.O and J.C.S. initiated and designed the study; M.J.RZ. and J.C.S wrote the manuscript with contributions from A.J.S., T.S. and D.G.O.; and all authors critically reviewed the final manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

References

1. Fegan C, Robinson H, Thompson P, Whittaker JA WD. Karyotypic evolution in CLL: identification of a new sub-group of patients with deletions of 11q and advanced or progressive disease. Leukemia. 1995;9(12):2003–8.

2. Döhner H, Stilgenbauer S, James MR, Benner a, Weilguni T, Bentz M, et al. 11q deletions identify a new subset of B-cell chronic lymphocytic leukemia characterized by extensive nodal involvement and inferior prognosis. Blood. 1997;89(7):2516–22.

3. Neilson JR, Auer R, White D, Bienz N, Waters JJ, Whittaker JA, et al. Deletions at 11q identify a subset of patients with typical CLL who show consistent Cytogenetic analysis. Leukemia. 1997; 11(11):1929–32.

4. Trbusek M, Malcikova J, Smardova J, Kuhrova V, Mentzlova D, Francova H, et al. Inactivation of p53 and deletion of ATM in B-CLL patients in relation to IgVH mutation status and previous treatment. Leukemia. 2006;20(6):1159–61.

5. Hallek M, Fischer K, Fingerle-Rowson G, Fink a M, Busch R, Mayer J, 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):1164–74.

6. Saiya-Cork K, Collins R, Parkin B, Ouillette P, Kuizon E, Kujawski L, et al. A pathobiological role of the insulin receptor in chronic lymphocytic leukemia. Clin Cancer Res. 2011;17(9):2679–92.

7. Britt-Compton B, Lin TT, Ahmed G, Weston V, Jones RE, Fegan C, et al. Extreme telomere erosion in ATM-mutated and 11q-deleted CLL patients is independent of disease stage. Leukemia. 2012;26(4):826–30.

8. Ouillette P, Fossum S, Parkin B, Ding L, Bockenstedt P, Al-Zoubi A, et al. Aggressive chronic lymphocytic leukemia with elevated genomic complexity is associated with multiple gene defects in the response to DNA double-strand breaks. Clin Cancer Res. 2010;16(3):835–47.

9. Catovsky D, Richards S, Matutes E, Oscier D, Dyer MJS, Bezares RF, et al. Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. Lancet. 2007;370(9583):230–9.

10. Tsimberidou A-M, Tam C, Abruzzo L V, O'Brien S, Wierda WG, Lerner S, et al. Chemoimmunotherapy may overcome the adverse prognostic significance of 11q deletion in previously untreated patients with chronic lymphocytic leukemia. Cancer. 2009;115(2):373–80.

11. Schwaenen C, Nessling M, Wessendorf S, Salvi T, Wrobel G, Radlwimmer B, et al. Automated array-based genomic profiling in chronic lymphocytic leukemia: development of a clinical tool and discovery of recurrent genomic alterations. Proc Natl Acad Sci U S A. 2004;101(4):1039–44.

12. Pfeifer D, Pantic M, Skatulla I, Rawluk J, Kreutz C, Martens UM, et al. Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays. Blood. 2007;109(3):1202–10.

13. Lehmann S, Ogawa S, Raynaud SD, Sanada M, Nannya Y, Ticchioni M, et al. Molecular allelokaryotyping of early-stage, untreated chronic lymphocytic leukemia. Cancer. 2008;112(6):1296–305.

14. Edelmann J, Holzmann K, Miller F, Winkler D, Bühler A, Zenz T, et al. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. Blood. 2012;120(24):4783–94.

15. Austen B, Powell JE, Alvi A, Edwards I, Hooper L, Starczynski J, et al. Mutations in the ATM gene lead to impaired overall and treatment-free survival that is independent of IGVH mutation status in patients with B-CLL. Blood. 2005;106(9):3175–82.

16. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med. 2011;365(26):2497–506.

17. Austen B, Skowronska A, Baker C, Powell JE, Gardiner A, Oscier D, et al. Mutation status of the residual ATM allele is an important determinant of the cellular response to chemotherapy and survival in patients with chronic lymphocytic leukemia containing an 11q deletion. J Clin Oncol. 2007;25(34):5448–57.

18. Pettitt a. R. p53 dysfunction in B-cell chronic lymphocytic leukemia: inactivation of ATM as an alternative to TP53 mutation. Blood. 2001;98(3):814–22.

19. Best OG, Gardiner a C, Majid a, Walewska R, Austen B, Skowronska a, et al. A novel functional assay using etoposide plus nutlin-3a detects and distinguishes between ATM and TP53 mutations in CLL. Leukemia. 2008;22(7):1456–9.

20. Skowronska A, Parker A, Ahmed G, Oldreive C, Davis Z, Richards S, et al. Biallelic ATM inactivation significantly reduces survival in patients treated on the United Kingdom Leukemia Research Fund Chronic Lymphocytic Leukemia 4 trial. J Clin Oncol. 2012;30(36):4524–32.

21. Auer RL, Starczynski J, McElwaine S, Bertoni F, Newland AC, Fegan CD, et al. Identification of a potential role for POU2AF1 and BTG4 in the deletion of 11q23 in chronic lymphocytic leukemia. Genes Chromosomes Cancer. 2005;43(1):1–10.

22. Visone R, Rassenti LZ, Veronese A, Taccioli C, Costinean S, Aguda BD, et al. Karyotype-specific microRNA signature in chronic lymphocytic leukemia. Blood. 2009;114(18):3872–9.

23. Marasca R, Maffei R, Martinelli S, Fiorcari S, Bulgarelli J, Debbia G, 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):348-55.

24. Kalla C, Scheuermann MO, Kube I, Schlotter M, Mertens D, Döhner H, et al. Analysis of 11q22-q23 deletion target genes in B-cell chronic lymphocytic leukaemia: evidence for a pathogenic role of NPAT, CUL5, and PPP2R1B. Eur J Cancer. 2007;43(8):1328–35.

25. Landau D a, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell. 2013;152(4):714–26.

26. Rossi D, Fangazio M, Rasi S, Vaisitti T, Monti S, Cresta S, et al. Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. Blood. 2012;119(12):2854–62.

27. Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. Blood. 2013;121(8):1403–12.

28. Gonzalez D, Martinez P, Wade R, Hockley S, Oscier D, Matutes E, et al. Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. J Clin Oncol. 2011;29(16):2223–9.

29. Oscier DG, Rose-Zerilli MJJ, Winkelmann N, Gonzalez de Castro D, Gomez B, Forster J, et al. The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. Blood. 2013;121(3):468–75.

30. Best OG, Ibbotson RE, Parker AE, Davis ZA, Orchard JA, Oscier DG. ZAP-70 by Flow Cytometry : A Comparison of Different Antibodies , Anticoagulants , and Methods of Analysis. Cytometry B Clin Cytom 2006;70(4):235-241.

31. Hamblin TJ, Davis Z, Gardiner a, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood. 1999;94(6):1848–54.

32. Parker H, Rose-Zerilli MJJ, Parker a, Chaplin T, Wade R, Gardiner a, et al. 13Q Deletion Anatomy and Disease Progression in Patients With Chronic Lymphocytic Leukemia. Leukemia. 2011;25(3):489–97.

33. Oscier DG, Gardiner AC, Mould SJ, Glide S, Davis ZA, Ibbotson RE, et al. Multivariate analysis of prognostic factors in CLL: clinical stage, IGVH gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors. Blood. 2002;100(4):1177–84.

34. Starostik P, Manshouri T, Brien SO, Leukemia BCL, Lerner S, Keating M. Deficiency of the ATM protein expression defines an aggressive subgroup of B-cell chronic lymphocytic leukemia. Cancer Res 1998;58(20):4552-4557

35. Grossmann V, Kohlmann A, Schnittger S, Weissmann S, Jeromin S, Kienast J, et al. Recurrent ATM and BIRC3 Mutations in Patients with Chronic Lymphocytic Leukemia (CLL) and Deletion 11q22-q23. Blood (ASH Annu Meet Abstr). 2012;115(Abstract No.):1771.

36. Stankovic T. Ataxia telangiectasia mutated-deficient B-cell chronic lymphocytic leukemia occurs in pregerminal center cells and results in defective damage response and unrepaired chromosome damage. Blood. 2002;99(1):300–9.

37. Ouillette P, Li J, Shaknovich R, Li Y, Melnick A, Shedden K, et al. Incidence and Clinical Implications of ATM Aberrations in Chronic Lymphocytic Leukemia. Genes, Chromosomes & Cancer. 2012;51(12):1125–32.

38. Lozanski G, Ruppert AS, Heerema N a, Lozanski A, Lucas DM, Gordon A, et al. Variations of the ataxia telangiectasia mutated gene in patients with chronic lymphocytic leukemia lack substantial impact on progression-free survival and overall survival: a Cancer and Leukemia Group B study. Leuk Lymphoma. 2012;53(9):1743–8.

39. Sivasubramaniam S, Sun X, Pan Y-R, Wang S, Lee EY-HP. Cep164 is a mediator protein required for the maintenance of genomic stability through modulation of MDC1, RPA, and CHK1. Genes Dev. 2008;22(5):587–600.

Table 1: Cohort Characteristics

Characteristics	Sub-groups	All Cases (%)	CLL4 Cases n (%) [2]	
			non-del(11q)	del(11q)
Number of cases		166 (100)	97 (73)	36 (27)
Gender	Male	124 (75)	71 (73)	27 (75)
	Female	42 (25)	26 (27)	9 (25)
Age at Diagnosis	Mean	63	71	62
	Range	39-89	56-86	44-80
Binet Stage at Diagnosis or Randomisation	A	48 (31)	23 (24)	7 (19)
	B	69 (44)	47 (49)	18 (50)
	C	40 (25)	27 (28)	11 (31)
Trisomy 12 by SNP6	Trisomy	13 (8)	9 (9)	3 (8)
	Normal	153 (92)	88 (91)	33 (92)
13q by SNP6	Deleted	94 (66)	51 (53)	22 (61)
	Normal	72 (44)	46 (47)	14 (39)
17p by SNP6	Deleted	12 (7)	9 (9)	1 (3)
	Normal	154 (93)	88 (91)	35 (97)
***IGHV* Mutation Status** [a]	Unmutated	98 (68)	56 (64)	30 (83)[1]
	Mutated	47 (32)	32 (36)	6 (17)
CD38 Positivity [b]	+	68 (50)	32 (42)	18 (55)
	-	69 (50)	44 (58)	15 (45)
ZAP70 Positivity [c]	+	74 (54)	42 (51)	21 (64)
	-	63 (46)	40 (49)	12 (36)
***ATM* mutation status** [d]	T-Mutated	17 (16)	5 (10)	10 (32)[3]
	NT-Mutated	19 (18)	8 (17)	4 (13)[4]
	Unmutated	69 (66)	35 (73)	17 (55)
***TP53* mutation status** [e]	Mutated	9 (7)	7 (8)	2 (6)
	Unmutated	116 (93)	76 (92)	30 (94)
***BIRC3* deletion status**	Deleted	57 (34)	0 (0)	32 (89)
	Normal	109 (66)	97 (100)	4 (11)
***BIRC3* mutation status**	Mutated	3 (2)	0 (0)	2 (6)
	Unmutated	159 (98)	95 (100)	34 (94)
***NOTCH1* mutation status** [f]	Mutated	14 (10)	10 (12)	2 (6)
	Unmutated	132 (90)	75 (88)	32 (94)
***SF3B1* mutation status** [g]	Mutated	28 (20)	19 (24)	3 (9)
	Unmutated	112 (80)	61 (76)	29 (91)

Footnote: A proportion of cases were not screened for our panel of molecular and cytogenetic biomarkers; not screened=21 [a], 29 [b], 29 [c], 61 [d], 41 [e], 20 [f], 26 [g]. [1] A significant positive association was identified between the presence of del11q and an unmutated IGHV sequences in the CLL4 cases (P=0.03, 2x2 Chi-squared test). [2] The columns for del11q and non-del11q are based on the SNP6.0 profiling data. [3] A significant positive association was identified between the frequency of truncating ATM

mutations (T-Mutated) and presence of del11q. (P=0.02, 2x2 Chi-squared test). [4] *The frequency of non-truncating ATM mutations (NT-Mutated) was similar in non-del11q and del11q cases. (P=0.76, 2x2 Chi-squared test).*

Table 2: Survival models of TP53, ATM and BIRC3 disruption in CLL4 cases

Mutation/ Biomarkers		Overall Survival (years)						Progression Free Survival (years)					
		Events/ Total	Median[1]	95% CI[1]	HR[2]	95% CI[2]	P-value[2]	Events/ Total	Median[1]	95% CI[1]	HR[2]	95% CI[2]	P-value[2]
Model A: Established prognostication model [a]	**mutated IGHV**	18/38	118	96-139	-	-	-	23/38	61	24-99	-	-	-
	un-mutated IGHV	75/86	56	49-64	3.19	1.89-5.39	**<0.001**	83/86	25	16-33	4.05	2.46-6.66	**<0.001**
	trisomy 12	8/9	79	35-123	1.65	0.76-3.61	0.208	9/9	35	19-50	2.55	1.21-5.34	0.013
	del(11q)	30/35	53	36-69	2.11	1.26-3.51	**0.004**	35/35	20	6-33	3.15	1.95-5.10	**<0.001**
	del(17p)	10/10	8	0-16	44.28	16.11-121.67	**<0.001**	10/10	4	2-5	23.33	10.16-53.56	**<0.001**
Model B: Inclusion of ***BIRC3*** deletion and /or mutation [b]	**Wild-type**	48/72	78	55-102	-	-	-	55/72	43	39-48	-	-	-
	del(11q)	3/3	42	0-86	3.00	0.93-9.73	0.066	3/3	38	0-81	18.0	0.56-5.79	0.323
	***BIRC3* deletion and/ or mutation** (with/ without ATM mutation)	24/29	53	39-67	1.91	1.16-3.15	**0.011**	29/29	17	4-30	3.26	2.03-5.22	**<0.001**
	***TP53* deletion and/ or mutation**	15/15	14	5-22	5.76	3.17-10.44	**<0.001**	15/15	5	2-7	5.20	2.89-9.38	**<0.001**
Model C: Inclusion of ***BIRC3*** and ***ATM*** mutation [c]	**Wild-type**	16/28	91	48-134	-	-	-	19/28	46	5-87	-	-	-
	del(11q) (*ATM* deletion without *BIRC3* deletion and /or mutation)	2/2	15	-	3.16	0.71-13.76	0.132	2/2	11	-	2.31	0.53-10.01	0.265
	***BIRC3* deletion and/ or mutation** (without *ATM* mutation)	7/11	76	40-111	1.41	0.57-3.45	0.456	11/11	28	16-40	2.80	1.29-6.04	**0.009**
	Biallelic inactivation of *ATM* (deletion and mutation of *ATM*)	13/13	42	13-71	3.42	1.63-7.21	**0.001**	13/13	10	6-15	6.07	2.80-13.18	**<0.001**
	***TP53* deletion and/ or mutation**	15/15	14	5-22	6.27	3.03-12.96	**<0.001**	15/15	5	2-7	6.13	3.03-12.38	**<0.001**

Footnote: Survival analysis is based on [1] Log-Rank testing and [2] univariate regression analysis. Significant P-values are shown in ***bold****. [a] All 133 CLL4 cases were used, subdivided based on the presence of established aberrations identified by SNP6 and sub-grouped based on the established Döhner model (46). [b] 119 CLL4 cases with the presence of del11q, BIRC3 and TP53 deletions and/or mutations determined by SNP6 profiling and mutational analysis. [c] 69 CLL4 cases with del11q, BIRC3 and TP53 deletions, and mutations of ATM, BIRC3 and TP53 based on SNP6 profiling data and mutational analysis.*

Figure 1: Chromosome 11 architecture, 11q23 minimally deleted region and BIRC3 deletion and/or mutation in 11q-deleted CLL

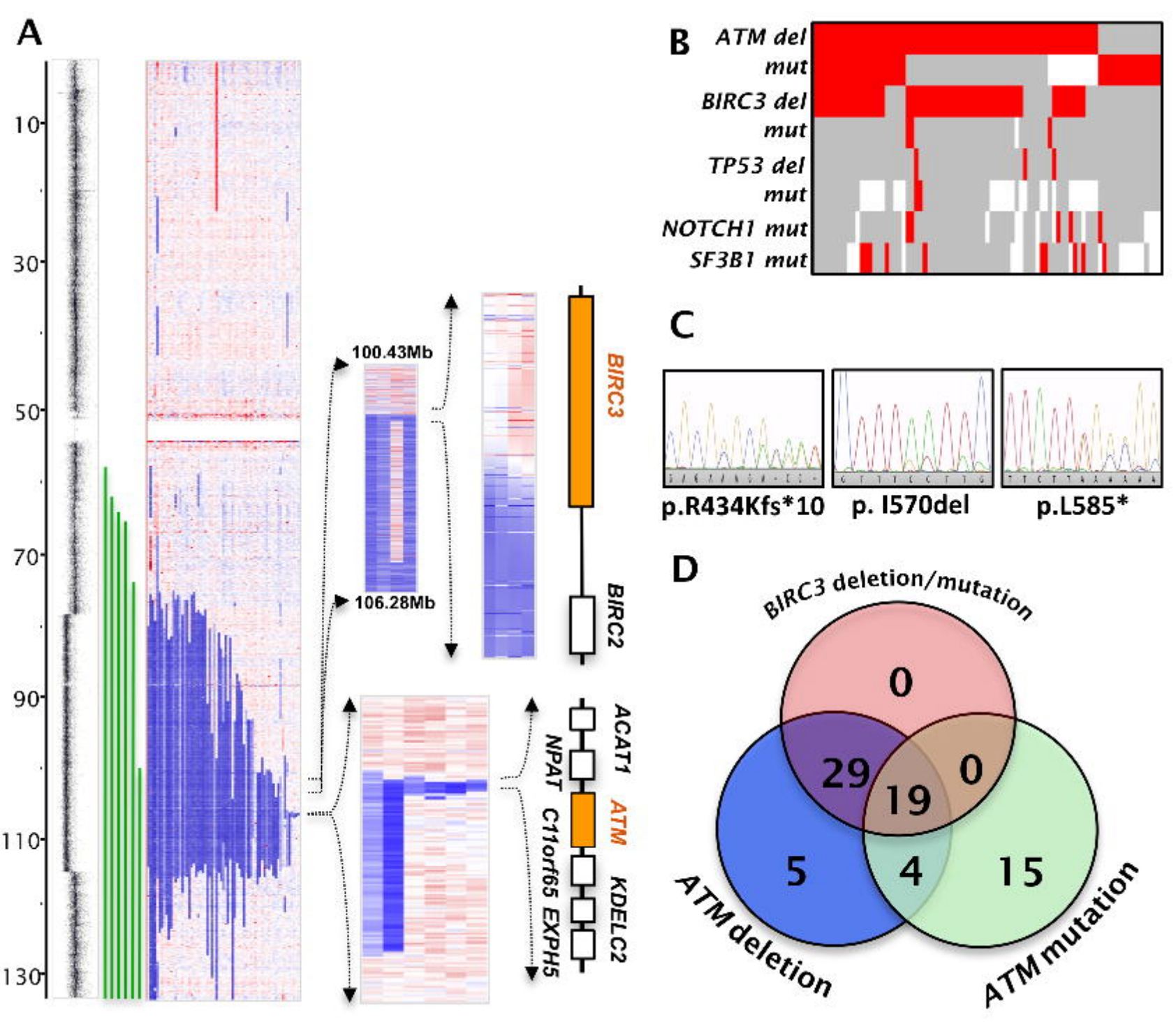
[In colour]

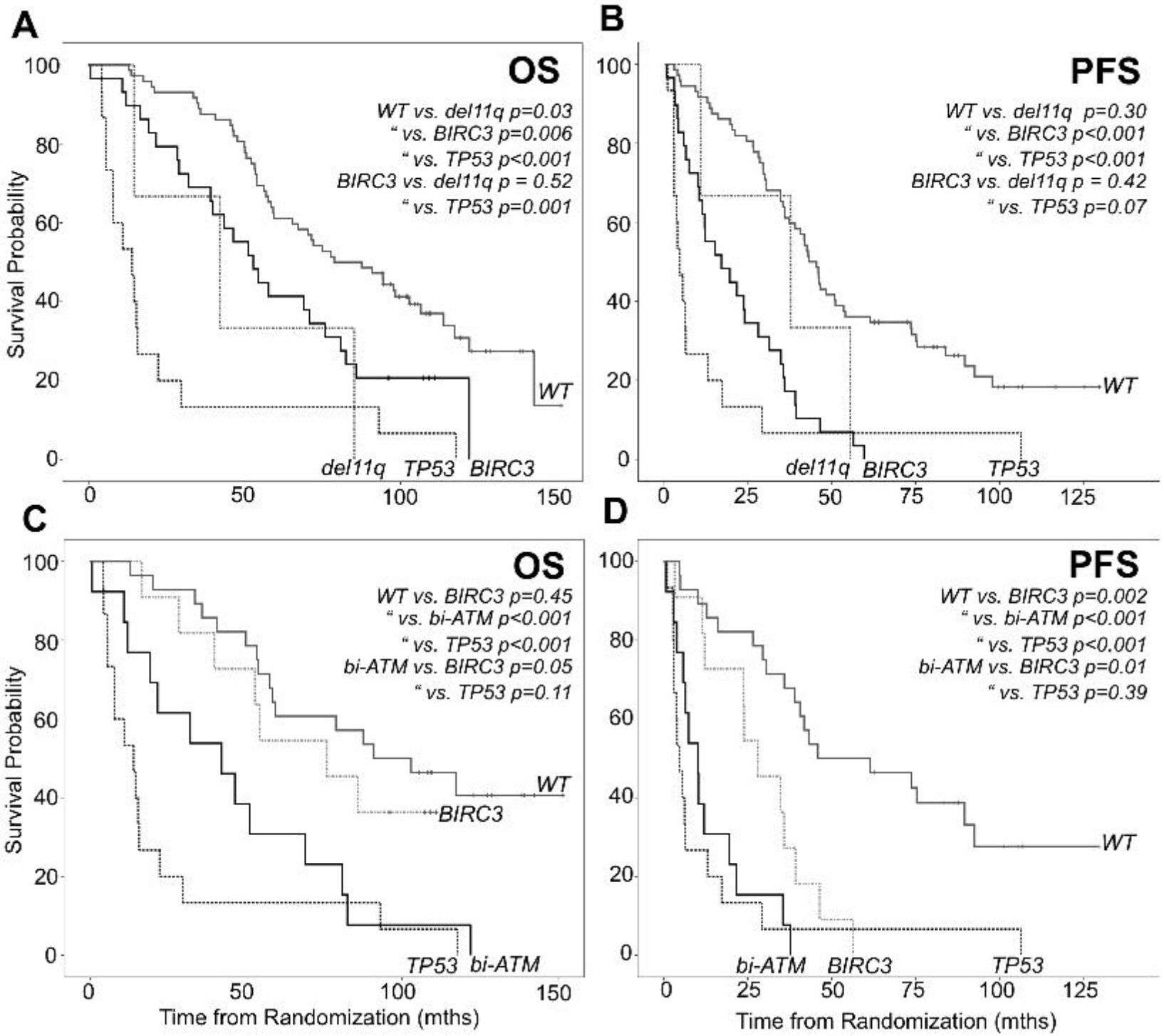
(A) Left to right: Genomic location in Mb from the telomere of 11p (top), through the centromere to the telomere of 11q (bottom); an example 11q deleted SNP6.0 probe profile showing an interstitial deletion between genomic locations 74 and 103Mb which includes the *ATM* and *BIRC3* loci; representative CNNLOH (green lines) observed in our cohort where the size and position of the line shows the genomic location and size of the CNNLOH event; heatmap of 69 11q deleted cases, where white, red and blue shown regions of no copy number change, duplications and deletions, respectively. Focussed heatmap views of *ATM* and *BIRC3* genes demonstrating the *ATM* MDR and the four cases with a telomeric deletion breakpoint in the 3' *BIRC3* gene locus are also shown (the second case from the left has deletion of both *BIRC3* and *ATM*, which is the result of two deletion events in this patient) (B) Abnormality matrix of cases with *ATM* and *BIRC3* lesions. Each row is a genetic lesion present in our cohort, and each column is a patient where red, grey and white, shows the presence, absence or no data for the lesion in question, respectively. (C) Sanger sequencing traces of the three *BIRC3* gene mutations identified in our study (D) Venn-diagram of the relationship between *BIRC3* deletion and/ or mutation and *ATM* deletion and mutation (n = number of observations).

Figure 2: Survival probability plots for Overall and Progression free survival in CLL4 cases

[In B&W]

(A) and (B) show CLL4 overall and progression-free survival, respectively; for cases sub-grouped by *TP53* deletion/SNV (dashed black line), *BIRC3* deletion and/ or mutation (black line), 11q deletion without *BIRC3* (dashed grey line) or wild-type (grey line) for *TP53*, *BIRC3* genes (no deletion or mutation) and 11q-deletions. (C) and (D) show CLL4 overall and progression-free survival, respectively; for the cases sub-divided into *TP53* deletion/mutated (dashed black line), bi-allelic inactivation of *ATM* (black line), *BIRC3* deletion and/ or mutation (dashed grey line) and wild-type (grey line) for *TP53*, *BIRC3*, *ATM* genes (no deletion or mutation) and 11q deletion. The survival curve for del11q (without an *ATM* mutation or *BIRC3* deletion and/ or mutation) was precluded due to only two survival events.


A
10
30
50
70
90
110
130
100.43Mb
106.28Mb
BIRC3
BIRC2
ACAT1
NPAT
ATM
C11orf65
KDELC2
EXPH5
B
ATM del
mut
BIRC3 del
mut
TP53 del
mut
NOTCH1 mut
SF3B1 mut
C
p.R434Kfs*10
p. I570del
p.L585*
D
BIRC3 deletion/mutation
0
29
19
0
5
4
15
ATM deletion
ATM mutation


A
OS
WT vs. del11q p=0.03
" vs. BIRC3 p=0.006
" vs. TP53 p<0.001
BIRC3 vs. del11q p = 0.52
" vs. TP53 p=0.001
Survival Probability
WT
del11q
TP53
BIRC3
B
PFS
WT vs. del11q p=0.30
" vs. BIRC3 p<0.001
" vs. TP53 p<0.001
BIRC3 vs. del11q p = 0.42
" vs. TP53 p=0.07
WT
del11q
BIRC3
TP53
C
OS
WT vs. BIRC3 p=0.45
" vs. bi-ATM p<0.001
" vs. TP53 p<0.001
bi-ATM vs. BIRC3 p=0.05
" vs. TP53 p=0.11
Survival Probability
WT
BIRC3
TP53
bi-ATM
Time from Randomization (mths)
D
PFS
WT vs. BIRC3 p=0.002
" vs. bi-ATM p<0.001
" vs. TP53 p<0.001
bi-ATM vs. BIRC3 p=0.01
" vs. TP53 p=0.39
WT
bi-ATM
BIRC3
TP53
Time from Randomization (mths)

Supplementary Methods

Molecular diagnostic assays

FISH analysis was performed for the presence of established rearrangements using a range of commercially available probes (Abbott Diagnostics, Maidenhead, UK; DakoCytomation, Glostrup, Denmark). ZAP70 and CD38 expression was determined as previously described (30) where 10% and 30% positive cells were classed as positive, respectively. *IGHV* genes were sequenced as previously described (31) and a cut-off of ≥98% germ-line identity was taken to define the unmutated sub-set.

DNA extraction, SNP6 array hybridization, data extraction and analysis

The data was aligned (Build 36.3) and analyzed by two independent researchers using Partek Genomics Suite (Partek Inc, Missouri, USA). Copy number alterations (CNAs) were defined as a deviation of 50 consecutive probes from a normal value of 2 (±0.3), within a consecutive genomic window of 50 Kilobases. As we have previously shown that aberration identification are not compromised by the absence of germ-line DNA profiling (32), the 270 HapMap Reference baseline (Affymetrix) was used as a control and germline copy number variants were excluded based on the Database of Genomic Variants (http://projects.tcag.ca/variation/). The allele ratio was calculated for each sample using the HapMap Allele Reference baseline (Affymetrix) and in the absence of paired normal DNA; copy number neutral loss of heterozygosity (CNNLOH) was defined as a region greater than 20Mb, extending to a telomere.

Mutational analysis of *ATM* and *BIRC3 g*enes

Denaturing high-performance liquid chromatography (DHPLC) was applied to high-molecular weight genomic DNA to identify SNVs in the 62 coding exons and flanking intronic sequences of the *ATM* gene as previously reported (20). Sequence changes were confirmed by direct Sanger sequencing and SNVs were categorized as previously reported (20). For *BIRC3* analysis, whole genome amplified DNA (Illustra GenomiPhi V2 Amplification Kit®, GE Healthcare) was screened using high-resolution melt (HRM) analysis as previously reported (29). *BIRC3* exons 7 and 10 (Transcript NM_182962.2) were analyzed as these exons capture all previously reported CLL-specific somatic variation (26, 27). *BIRC3* HRM-PCR primer sequences and reaction conditions are described in **Supplementary table 2**. Products showing abnormal melt patterns were sequenced. All SNVs were sequence validated on genomic DNA (gDNA) from the archival sample. Furthermore, we sequenced the aforementioned regions of *BIRC3* in gDNA samples from 35 cases that exhibited normal HRM melt profiles on WGA material. In doing so, we found no additional SNVs present in

the gDNA. The strategy for assigning the somatic nature of each SNV is detailed in the Supplementary methods section.

Due to the historical nature of this cohort, matched germ-line DNA was not available on the majority of cases, so we ascertained the somatic nature of each SNV by ensuring the variant was not annotated as a polymorphism in dbSNP132 (www.ncbi.nlm.nih.gov/projects/SNP/). For *ATM*, SNVs were defined as pathogenic if they were; a) 'truncating' due to a sequence alteration (frame-shift or nonsense-STOP-codon or short in-frame deletion or splice site defect), that was predicted to cause premature termination of the protein, or b) 'non-truncating' if they were missense, either reported in AT patients or predicted to cause a non-synonymous amino acid substitution in the translated protein within the region encoding the functional domain of the ATM protein (20). For *BIRC3*, SNVs were defined as a) resulting in a protein sequence change (i.e. a nonsynonymous amino acid change) or b) were predicted to cause premature termination of the protein or a small in-frame deletion (26).

Statistical analysis

Overall response rate (ORR) was defined as complete (CR), nodular partial (nPR), partial response (PR) and non-response/ progressive disease (NR/PD) and was available on 131 of the 133 CLL4 cases. Overall survival (OS) was defined as time from randomization to death, or to the follow up date (August 2012) for survivors. Progression free survival (PFS) was defined as time from randomization to relapse needing further therapy, progression or death, or to the follow-up date (Oct 2010; final LRF CLL4 PFS update) for those with no progression/death.

Post-hoc sub-group power calculation for OS and PFS:

With clinical follow-up of 10 years, our CLL4 sub-group (11q23 deletions encompassing *BIRC3* and *ATM*, n=24) had 29% power with a significance level = 0.05 to detect a OS difference between cases with *ATM* and *BIRC3* deletion and those with *BIRC3* deletion and biallelic loss of *ATM* via mutation of the non-deleted allele, based on the median CLL4 OS times of 76 and 42 months, respectively. For, PFS we had 81% power with a significance level = 0.05 to detect a PFS difference between cases based on the median CLL4 OS times of 28 and 10 months, respectively.

Supplementary Table 1: Comparison of the CLL4 cases to the full LRF CLL4 trial

Variable	CLL4 Cases	%	CLL4 trial	%	P-Value
Total cases	133	-	777	-	
Male	98	74	573	74	ns
Female	35	26	204	26	
Age	63	-	64	-	ns
Binet Stage					
A	30	23	191	25	ns
B	65	49	352	45	
C	38	28	234	30	
IGHV unmutated	86	69	327	61	ns
mutated	38	31	206	39	
CD38 -ve	59	54	299	56	ns
+ve	50	46	236	44	
ZAP70 -ve	52	45	234	49	ns
+ve	63	55	244	51	
TP53 normal	102	89	532	92	ns
del/mut	13	11	48	8	
del(11q) -ve	84	67	462	80	0.002
+ve	42	33	116	20	
del(13q) -ve	50	40	361	62	<0.001
+ve	75	60	217	38	
del(17p) -ve	114	91	538	94	ns
+ve	11	9	33	6	
Tri12 -ve	110	87	487	84	ns
+ve	16	13	91	16	

TP53 abnormalities defined by deletion and/ or mutation (del/mut)

Supplementary Table 2: *BIRC3* primer sequences

Primer name	Primer sequence (5'- 3')	PCR conditions
***BIRC3* exon 7F**	TTCCATATAGTTATCCATTTTGAACCT	HRM-PCR: Ta=60°C
	TGCCTATACATTTTGTTGGTT	Seq-PCR: Ta=55°C
***BIRC3* exon 7 R**	ACATACTTGATTCTTTTTCCTCAGTTG	HRM-PCR: Ta=60°C
	AAAAACCTGACTGGATTGAG	Seq-PCR: Ta=55°C
***BIRC3* exon 10 F**	TGAAGAAGCAAACTGCCTTTTATT	HRM-PCR: Ta=60°C
	CCACAGAAGATGTTTCAGGT	Seq-PCR: Ta=55°C
***BIRC3* exon 10 R**	AAAGTTTAGACGATGTTTTGGTTCT	HRM-PCR: Ta=60°C
	GTGCTACCTCTTTTTCGTTC	Seq-PCR: Ta=55°C

Footnote: HRM-PCR = High-Resolution Melting PCR. Seq-PCR = Sanger Sequencing PCR

Supplementary Table 3: *ATM* mutations in our CLL patients

Regid	BIRC3 del	ATM del	del(11q) FISH	del(11q) FISH Clone Size (%)	ATM Mutation Nomenclature	c.DNA position	Amino Acid number	ATM mutation Type	Consequence
353	N	N	N	-	c.217_218del2:p.Gln73*fs*	217	73	T	Frameshift
37	Y	Y	Y	95	c.478_482del5:p.Ser160*fs*	478	160	T	Frameshift
349	N	N	N	-	c.1006_1020del15:p.Phe336_Ala340del5	1006	340	T	Frameshift
52	N	N	N	-	c.1048G>A:p.Ala350Thr	1048	350	NT	Nonsynonymous
48	N	N	N	-	c.1066-6T>G	1066	n/a	T	Splicing defect – termination (exon11)
346	N	Y	Y	-	c.1120C>T:pGlu374*	1120	374	T	STOP codon
181	Y	Y	Y	98	c.1402_1403del2:p.Lys468*fs*	1402	468	T	Frameshift
54	N	N	N	-	c.2193delC:p.Tyr731*fs*	2193	731	T	Frameshift
348	Y	Y	Y	30	c.2417T>G:p.Leu806Trp	2417	806	NT	Nonsynonymous
42	Y	Y	Y	93	c.2720_2723del4:p.Cys907*fs*	2720	907	T	Frameshift
39	N	Y	Y	87	c.3712_3716del5:p.Leu1238*fs*	3712	1238	T	Frameshift
72	Y	Y	Y	45	c.3720_3736del17:p.Asn1240*fs*	3720	1240	T	Frameshift
352	Y	Y	Y	89	3883_3885delCTT:p.Leu1295del	3883	1295	NT	Inframe deletion
46	Y	Y	Y	83	c.5006-2A>G	5006	n/a	T	Splicing defect - termination (exon 36)
55	N	N	N	-	c.5857A>G:p.Thr1953Ala	5857	1953	NT	Nonsynonymous
43	N	N	Y	11	c.6067G>A:p.Gly2023Arg	6067	2023	NT	Nonsynonymous
342	Y	Y	Y	86	c.6106T>A:p.Tyr2036Asn	6106	2036	NT	Nonsynonymous
51	Y	Y	Y	84	c.6375insT:p.Glu2126*	6375	2126	T	STOP codon
45	Y	Y	Y	75	c.6989_6995del7:p.Leu2330*fs*	6989	2330	T	Frameshift
49	Y	Y	Y	84	c.7327C>G:p.Arg2443Gly	7327	2443	NT	Nonsynonymous
174	N	N	N	-	c.7390T>C:p.Cys2464Arg	7390	2464	NT	Nonsynonymous
56	N	N	N	-	c.7438C>T:p.His2480Tyr	7438	2480	NT	Nonsynonymous
50	Y	Y	Y	97	c.7638_7646del9:p.Arg2547_Ser2549del3	7638	2547	T	Frameshift
354	Y	Y	Y	84	c.7883_7888del5	7883		T	Frameshift
341	Y	Y	Y	-	c.8056T>C:p.Phe2686Leu	8056	2686	NT	Nonsynonymous
57	N	N	N	-	c.8095C>T:p.Pro2699Ser	8095	2699	NT	Nonsynonymous
60	N	Y	Y	23	c.8161G>A:p.Asp2721Asn	8161	2721	NT	Nonsynonymous

36	Y	Y	Y	76	c.8249T>C:p.Leu2750Ser	8249	2750	NT	Nonsynonymous
61	N	Y	Y	69	c.8249T>C:p.Leu2750Ser	8249	2750	NT	Nonsynonymous
351	N	N	N	-	c.8428_8450del23:p.Lys2810*fs*	8428	2810	T	Frameshift
58	N	N	N	-	c.8663T>C:p.Ile2888Thr	8663	2888	NT	Nonsynonymous
47	Y	Y	Y	89	c.8672G>A:p.Gly2891Asp	8672	2891	NT	Nonsynonymous
357	Y	Y	Y	57	c.8787-1G>T	8787	n/a	NT	Splicing defect - termination (exon 62)
197	Y	Y	Y	88	c.9023G>A:p.Arg3008His	9023	3008	NT	Nonsynonymous
59	N	N	N	-	c.9032T>A:p.Met3011Lys	9032	3011	NT	Nonsynonymous
38	Y	Y	Y	66	c.9139C>T:p.Arg3047*	9139	3047	T	STOP codon

Footnote: In the BIRC3 del column underlined Y indicates an 11q23 deletion breakpoint in the BIRC3 gene body. T = Truncating mutation, NT = Non-Truncating mutation.

Supplementary Table 4: Gene body disruption by 11q23 deletion breakpoints which include *ATM* (excluding the *BIRC3* gene breakpoints)

Gene Symbol	Gene location	Sample ID with deletion breakpoint within a gene body	
		centromeric	telomeric
PHCA (ACER3)	76249601-76411617	72, *350*	
RSF1	77054922-77209528	42	
RAB30	82370126-82460532	354	
DLG2	82843701-84312113	*181*, *220*	
CNTN5	98397081-99732683	247, 264, 266	
GRIA4	104986010-105358029	44, 107	
CUL5	107384618-107483698	60, 61, 244	
DDX10	108041026-108316860		62, 346
CADM1	114549555-114880451		74, 180, 352
CEP164	116703781-116789192		37, 263, *350*
RNF214	116608614-116661614		50, 64, 76, 182, 348
DSCAML1	116803699-117173186		40, *181*
ASAM	122448230-122571217		49, *220*

Footnote: Reference genome is hg18 (Mar. 2006 NCBI36).

LYMPHOID NEOPLASIA

CME article

The clinical significance of *NOTCH1* and *SF3B1* mutations in the UK LRF CLL4 trial

*David G. Oscier,[1] *Matthew J. J. Rose-Zerilli,[2] *Nils Winkelmann,[3,4] David Gonzalez de Castro,[5] Belen Gomez,[5] Jade Forster,[2] Helen Parker,[2] Anton Parker,[1] Anne Gardiner,[1] Andrew Collins,[6] Monica Else,[5] Nicholas C. P. Cross,[2,3] Daniel Catovsky,[5] and Jonathan C. Strefford[2]

[1]Department of Molecular Pathology, Royal Bournemouth Hospital, Bournemouth, United Kingdom; [2]Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom; [3]Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, United Kingdom; [4]Klinik für Innere Medizin II, Universitätsklinikum Jena, Jena, Germany; [5]Haemato-oncology Research Unit, Division of Molecular Pathology, The Institute for Cancer Research, Sutton, United Kingdom; and [6]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

Key Points

- This is the first study to validate the importance of *NOTCH1* and *SF3B1* gene mutations in the context of a randomized, prospective clinical trial.
- Mutations in both genes are independent prognostic biomarkers, and therefore have clinical utility in the accurate risk-adapted stratification of CLL patients.

***NOTCH1* and *SF3B1* mutations have been previously reported to have prognostic significance in chronic lymphocytic leukemia but to date they have not been validated in a prospective, controlled clinical trial. We have assessed the impact of these mutations in a cohort of 494 patients treated within the randomized phase 3 United Kingdom Leukaemia Research Fund Chronic Lymphocytic Leukemia 4 (UK LRF CCL4) trial that compared chlorambucil and fludarabine with and without cyclophosphamide in previously untreated patients. We investigated the relationship of mutations in *NOTCH1* (exon 34) and *SF3B1* (exon 14-16) to treatment response, survival and a panel of established biologic variables. *NOTCH1* and *SF3B1* mutations were found in 10% and17% of patients, respectively. *NOTCH1* mutations correlated with unmutated *IGHV* genes, trisomy 12, high CD38/ ZAP-70 expression and were associated with reduced overall (median 54.8 vs 74.6 months, $P = .02$) and progression-free (median 22.0 vs 26.4 months, $P = .02$) survival. *SF3B1* mutations were significantly associated with high CD38 expression and with shorter overall survival (median 54.3 vs 79.0 months, $P < .001$). Furthermore, multivariate analysis, including baseline clinical variables, treatment, and adverse prognostic factors demonstrated that although *TP53* alterations remained the most informative marker of dismal survival in this cohort, *NOTCH1* (HR 1.58, $P = .03$) and *SF3B1* (HR 1.52, $P = .01$) mutations have added independent prognostic value. (*Blood*. 2013;121(3):468-475)**

Medscape EDUCATION Continuing Medical Education online

This activity has been planned and implemented in accordance with the Essential Areas and policies of the Accreditation Council for Continuing Medical Education through the joint sponsorship of Medscape, LLC and the American Society of Hematology. Medscape, LLC is accredited by the ACCME to provide continuing medical education for physicians.

Medscape, LLC designates this Journal-based CME activity for a maximum of 1.0 **AMA PRA Category 1 Credit(s)™**. Physicians should claim only the credit commensurate with the extent of their participationin the activity.

All other clinicians completing this activity will be issued a certificate of participation. To participate in this journal CME activity: (1) review the learning objectives and author disclosures; (2) study the education content; (3) take the post-test with a 70% minimum passing score and complete the evaluation at **http://www.medscape.org/journal/blood**; and (4) view/print certificate. For CME questions, see page 566.

Disclosures

Associate Editor John G. Gribben served as an advisor or consultant for Celgene and Roche and as a speaker or a member of a speakers bureau for Roche, Jensen, and Celgene. The authors and CME questions author Laurie Barclay, freelance writer and reviewer, Medscape, LLC, declare no competing financial interests.

Submitted May 11, 2012; accepted October 4, 2012. Prepublished online as *Blood* First Edition paper, October 18, 2012; DOI 10.1182/blood-2012-05-429282.

*D.G.O., M.J.J.R.-Z., and N.W. share first authorship.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

Medscape EDUCATION **Continuing Medical Education online**

Learning objectives

Upon completion of this activity, participants will be able to:

1. Describe the frequency of *NOTCH1* and *SF3B1* mutations in patients with chronic lymphocytic leukemia (CLL), and their correlations with other genetic markers.
2. Describe survival in CLL patients with *NOTCH1* mutations, and the prognostic value of this mutation.
3. Describe survival in CLL patients with *SF3B1* mutations, and the prognostic value of this mutation.

Release date: January 17, 2013; Expiration date: January 17, 2014

Introduction

Deletions and duplications of genomic DNA are detectable in more than 80% of patients with chronic lymphocytic leukemia (CLL).[1] Together with mutations of *TP53* and *ATM* genes located within minimally deleted regions on 17p and 11q, respectively, they constitute the most powerful predictors of the natural history and response to therapy in CLL. However, neither these genomic abnormalities, nor biomarkers that reflect the ability of leukemic cells to respond to their environment, identify all cases with early disease destined to progress. Nor do they identify ~ 50% of the cases with primary or acquired resistance to chemo-immunotherapy.

The recent application of whole genome sequencing to discovery CLL cohorts followed by targeted resequencing of recurring variants in larger cohorts has identified clinically significant mutations in genes not affected by copy number alterations.[2-6] Mutations within the PEST domain of *NOTCH1* have been found in 8% to 12% of patients at diagnosis, in 21% of patients with alkylating agent or purine analog-refractory disease and in 30% of cases with the diffuse large B-cell variant of Richter syndrome (RS).[4,5,7] They have been correlated with advanced stage at diagnosis, an increased risk of RS, unmutated *IGHV* genes, trisomy 12 (especially in unmutated *IGHV* cases), but are underrepresented in patients with del13q or a *TP53* abnormality.[7,8] To date, *NOTCH1* mutations have been associated with a shorter time to first treatment and shorter overall survival independent of other prognostic factors, such as *TP53* abnormalities and *IGHV* gene mutation status.[4,5,7,8] Mutations in RNA splicing genes have been recently discovered in myeloid malignancies[9] with a strong association between *SF3B1* gene mutations and cases of the myelodysplastic syndrome with increased ringed sideroblasts.[10] Subsequently, *SF3B1* mutations have also been found in CLL but not in other chronic B-cell lymphoproliferative disorders.[2,3,6] Mutations were documented in 5% to 17% of patients and were associated with advanced stage, fludarabine-refractory disease in cases with no *TP53* abnormality, 11q23 deletions and with short time to first treatment and overall survival independent of other prognostic variables.[2,3,6]

These data suggest that these mutations represent novel independent prognostic factors that may influence clinical practice. However, mutations in neither gene have been evaluated within the context of a large clinical trial. To address this we have screened patients entered into the United Kingdom Leukaemia Research Fund Chronic Lymphocytic Leukemia 4 (UK LRF CCL4) trial for *NOTCH1* and *SF3B1* mutations and correlated their presence with a comprehensive panel of established biomarkers and with clinical outcome. With this approach, we show that although *TP53* alterations remained the most informative predictor of poor outcome in our study, *NOTCH1* and *SF3B1* mutations are independent markers of shorter overall survival.

Methods

Patients and biomarker data

Between 1999 and 2004, the UK LRF CLL4 trial randomly assigned 777 patients to first-line treatment with chlorambucil (CHL), fludarabine (FDR), or fludarabine plus cyclophosphamide (FC).[11] Up to 494 patients with DNA available were included in the current study and their characteristics did not differ significantly from the entire trial cohort, in terms of treatment allocation, cytogenetics, age, gender, disease stage, ZAP-70 and CD38 expression, or *IGHV* status (supplemental Table 1, available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article). All samples were taken at trial entry before initiation of treatment and, on average, were more than 80% tumor cells. The assessment of established biomarkers was performed as previously described,[12] and a combined *TP53* deletion and/or mutation variable was used.[13] Because of sample availability from United Kingdom DNA repositories, the number screened for *NOTCH1* and *SF3B1* varied slightly. This study was conducted in accordance with the Declaration of Helsinki and was ethically approved by the local REC.

NOTCH1 and *SF3B1* mutational screening and sequencing

NOTCH1 and *SF3B1* mutations were successfully ascertained in 466 and 437 patients, respectively. Each genomic DNA sample was subjected to whole genome amplification (WGA) using the Illustra GenomiPhi V2 Amplification Kit (GE Healthcare), before mutational screening using high-resolution melt (HRM) analysis as previously reported.[14] For each gene we targeted our screening approach to capture the majority of previously reported CLL-specific somatic variation; for *NOTCH1* we screened a 399bp section of exon 34 (amino acids 2405-2525, where all of the PEST domain mutations have been found)[4,5,7,15] and for *SF3B1* we screened exons 14-16 (capturing all the previously reported disease-specific variation[2,3,6]; see supplemental information for primer sequences). Products showing abnormal melt patterns were sequenced. All mutations were sequence validated on genomic DNA (gDNA) from the archival sample. Furthermore, we sequenced the aforementioned regions of *NOTCH1* and *SF3B1* in gDNA samples from 58 cases that exhibited normal HRM melt profiles on WGA material. In doing so, we found no additional mutations present in the gDNA.

To further assess the sensitivity of the HRM approach, we identified cases with the P2515Rfs*4 (c.7544_7545delCT, referred to as "delCT") variant using PCR-based fragment analysis [PCR-FA, n = 372] and allele-specific PCR [n = 213]. For the PCR-FA, the delCT variant was identified using PEST2 domain-specific primers (FW: GTGACCGCAGCCCAGTTC, RV: AAAGGAAGCCGGGGTCTC) as previously described.[16] PCR products were sized using the LT3500 (Life Technologies) and the 271bp wild-type and 269bp mutant fragments were identified using GeneMapper Version 4.1 software (Life Technologies). Allele-specific PCR for the delCT variant was performed by KBioscience (http://www.kbiosciences.co.uk) using their own novel fluorescence-based competitive allele-specific PCR (KASPar). Duplicates (n = 106) were included and the concordance between duplicates was > 99%. Sample processing using the PCR-FA and KASPar technologies were performed blindly in independent laboratories.

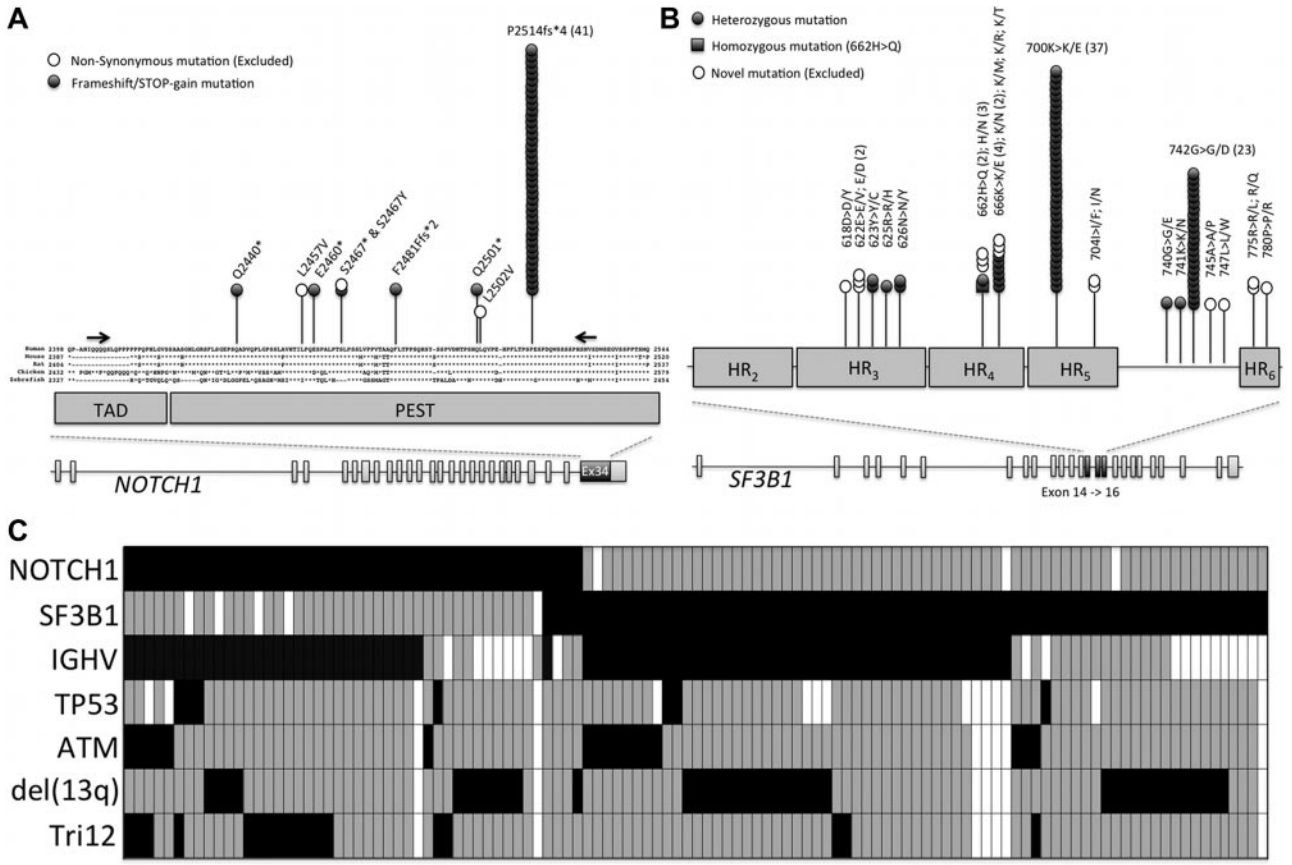


Figure 1. *NOTCH1* and *SF3B1* mutational analysis in the LRF CLL4 patients. (A) The distribution of mutations in *NOTCH1*. The *NOTCH1* gene contains 34 exons and encodes a protein with a C-terminal TAD-PEST domain, which is a hotspot for mutation in CLL. Part of exon 34 is magnified and the location of each mutation is shown, along with evolutionary conservation of amino acids. (B) The distribution of missense mutations in *SF3B1*. The *SF3B1* gene contains 25 exons and encodes a protein with a C-terminal domain consisting of 22 HEAT domains. Exons 14, 15, and 16 are magnified and the location of each mutation is shown. (C) The mutual relationship between *NOTCH1* and *SF3B1* mutations and other gene lesions in CLL. Rows correspond to specific genes and columns represent individual patients (only patients with either a *NOTCH1* or *SF3B1* mutation are shown). Boxes colored black and gray show the presence or absence of a mutation of *NOTCH1* or *SF3B1*, with deletion of *11q*, deletion or mutation of *TP53*, an unmutated *IGHV* sequence, deletion of 13q, or the presence of trisomy 12. A white box denotes that no data were available.

Using these 2 approaches, we confirm the sensitivity of our HRM approach which showed a 100% concordance with our PCR-FA and KASPar analysis.

Statistical analysis

Statistical analysis was performed using STATA Version 9.2 and SPSS Version 19. We used univariate logistic regression and dependent variables examined included the presence/absence of a *NOTCH1/SF3B1* mutation in relation to a range of biomarkers and clinical data. We included age at trial entry and gender as covariates in all models. Overall response status (ORS) was defined as complete (CR), nodular partial (nPR), and partial response (PR). Overall survival (OS) was defined as time from randomization to death, or to the follow-up date (August 2012) for survivors. Progression free survival (PFS) was defined as time from randomization to relapse needing further therapy, progression, or death; or to the follow-up date (October 2010; final LRF CLL4 PFS update) for those with no progression/death. Kaplan-Meier analysis and log-rank test or Cox regression were undertaken for survival analyses examining the impacts of mutation on OS and PFS. Fixed covariates included in the Cox regression models were: age at trial entry, gender, Binet stage (A versus B/C), *IGHV* mutation status, 11q deletion, and *TP53* deletion and/or mutation (*TP53* del/mut). Based on previous *NOTCH1* survival studies,[7] with a 10% mutational prevalence, a mean survival time of 3.5 years for mutated individuals and a follow-up of 10 years, our study had 90% power with a significance level = 0.05 to detect a minimum hazard ratio (HR) of 1.35. This HR is lower than that observed previously.[7] Results were determined to be statistically significant at the 5% level.

Results

NOTCH1 mutations are associated with the presence of unmutated *IGHV* genes, trisomy 12, CD38 expression, and with shorter progression-free and overall survival

We identified 49 patients with putative mutations within exon 34 of *NOTCH1* (Figure 1A). Mutations were most represented (41/49, 84%) by the recurrent 2-bp frameshift deletion (delCT, Figure 1A). 63% (5/8) of the remaining mutations were previously reported and/or are predicted to disrupt the *NOTCH1* PEST domain (supplemental Table 2). As patients were randomized 8 to 13 years ago, germ-line material was not available to confirm the somatic-nature of our novel mutations. Therefore, to ensure that likely pathogenic mutations were included in subsequent analysis, only PEST domain-truncating mutations were analyzed further in this study; providing a *NOTCH1* mutational frequency of 10% (46/466, Table 1) in the LRF CLL4 trial. Of the 348 deaths in patients with *NOTCH1* data, 12 (3%) were because of Richter syndrome (RS). Only 1 of 12 patients (8%) had a *NOTCH1* mutation. The incidence of death because of RS was not significantly different in the mutated and the wild-type groups.

The clinical and biologic features of *NOTCH1* mutated CLL are summarized in Tables 1 and 2. Importantly, mutations showed a

Table 1. Frequencies of NOTCH1 and SF3B1 mutations and clinical characteristics of the 494 LRF CLL4 patients

	NOTCH1 (all PEST terminating mutations)				*SF3B1* (previously shown as somatic in CLL)			
Variable	Total N (%)	Wild-type N (%)	Mutated N (%)	*P*	Total N (%)	Wild-type N (%)	Mutated N (%)	*P*
Screened cases	466 (94)	420 (90)	46 (10)		437 (88)	364 (83)	73 (17)	
Male	344 (74)	307 (89)	37 (11)	NS	319 (73)	263 (82)	56 (18)	NS
Female	122 (26)	113 (93)	9 (7)		118 (27)	101 (86)	17 (14)	
Age (range)*		63 (38-86)	66 (48-83)			63 (38-83)	64 (42-82)	
Binet stage*				NS				NS
A	106 (23)	96 (91)	10 (9)		97 (22)	85 (88)	12 (12)	
B	218 (47)	192 (88)	26 (12)		209 (48)	172 (82)	37 (18)	
C	142 (30)	132 (93)	10 (7)		131 (30)	107 (82)	24 (18)	
IGHV unmutated	240 (63)	209 (87)	31 (13)	.01	228 (64)	184 (81)	44 (19)	NS
mutated	144 (37)	137 (95)	7 (5)		131 (36)	115 (88)	16 (12)	
CD38 −ve	212 (55)	203 (96)	9 (4)	.0001	193 (55)	173 (90)	20 (10)	.007
+ve	170 (44)	139 (82)	31 (18)		160 (45)	126 (79)	34 (21)	
ZAP70 −ve	164 (47)	155 (95)	9 (5)	.008	154 (47)	132 (86)	22 (14)	NS
+ve	186 (53)	159 (85)	27 (15)		174 (53)	141 (81)	33 (19)	
TP53 normal	408 (93)	368 (90)	40 (10)	NS	388 (94)	323 (83)	65 (17)	NS
del/mut	32 (7)	28 (87)	4 (13)		24 (6)	21 (88)	3 (12)	
del(11q) −ve	352 (80)	314 (89)	38 (11)	NS	330 (80)	273 (83)	57 (7)	NS
+ve	89 (20)	83 (93)	6 (7)		83 (20)	72 (87)	11 (13)	
del(13q) −ve†	272 (62)	240 (88)	32 (12)	NS	254 (62)	214 (85)	40 (15)	NS
+ve	166 (38)	154 (93)	12 (7)		156 (38)	128 (82)	28 (18)	
Tri12 −ve	369 (84)	340 (92)	29 (8)	.002	343 (83)	278 (81)	65 (19)	.001
+ve	72 (16)	57 (79)	15 (21)		70 (17)	67 (96)	3 (4)	

*Age and disease stage were both assessed at trial entry. NS indicates nonsignificant (significance level = .05). *TP53* abnormalities defined by deletion and/or mutation (del/mut).

†The presence of a 13q deletion as a sole abnormality using a standard FISH panel. *P* values were calculated from 2 × 2 or 2 × 3 χ^2 tests (Fisher exact test was used when observations were < 5).

significant association with unmutated *IGHV* genes (OR 2.91, $P = .014$, Table 2) and positive expression of ZAP-70 (OR 2.92, $P = .007$) and CD38 (OR 5.03, $P < .001$, Table 2). In contrast, the presence of a mutation was not associated with DNA damage response biomarkers such as deletion of 11q or 17p or mutation of *TP53*. The presence of a *NOTCH1* mutation was significantly associated with the presence of trisomy 12 (OR 3.09, $P = .001$); within those cases, 13 of 15 *NOTCH1* mutated-trisomy 12 cases were *IGHV* unmutated (Figure 1C).

Next, we determined the impact of a *NOTCH1* mutation on overall response status (ORS), overall survival (OS) and progression-free survival (PFS). Mutations were not associated with ORS within any of the 3 treatment arms (data not shown). However, univariate analysis revealed that the presence of a mutation was associated with a significant reduction in median OS of 54.8 versus 74.6 months (HR 1.47, 95% CI 1.06-2.05, $P = .02$) and median PFS of 22.0 versus 26.4 months (HR 1.46, 95% CI 1.07-1.99, $P = .02$), for mutant and wild-type, respectively (Figure 2A, Table 3).

SF3B1 mutations are associated with high CD38 expression and overall survival

Of the 92 putative mutations identified in exon 14, 15, and 16 of *SF3B1*, 73 patients harbored 75 mutations that were previously shown to be somatically acquired in CLL,[2,3,6] providing a frequency of

Table 2. Association between *NOTCH1/SF3B1* gene mutation status and clinical characteristics of the LRF CLL4 patients

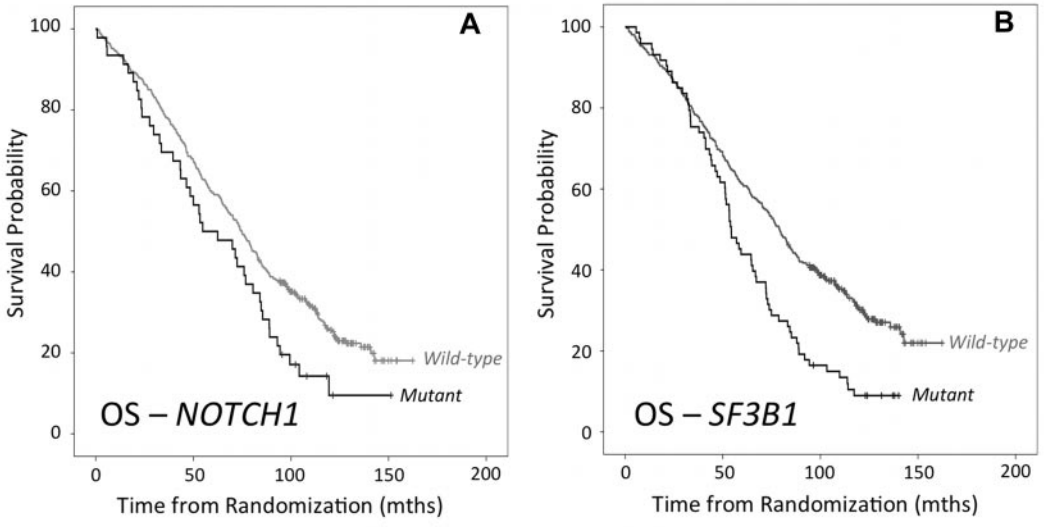
	NOTCH1 mutated cases (all PEST terminating mutations)					*SF3B1* mutated cases (previously shown as somatic in CLL)				
Variable	OR	95% CI	SE	*P*	n = *	OR	95% CI	SE	*P*	n = *
Age at diagnosis	1.04	0.99-1.07	0.02	.06	466	1.01	0.98-1.04	0.02	.46	437
Sex	0.66	0.31-1.41	0.26	.29	466	0.79	0.44-1.43	0.30	.43	437
Binet stage (A, B, or C)	0.85	0.56-1.29	0.18	.45	466	1.23	0.86-1.75	0.18	.26	437
CD38 positive	5.03	2.32-10.90	1.98	< .001	382	2.33	1.28-4.25	0.71	.005	353
ZAP70 positive	2.92	1.32-6.42	1.17	.007	350	1.40	0.78-2.53	0.42	.26	328
IGHV unmutated	2.91	1.24-6.80	0.15	.014	384	1.72	0.93-3.19	0.18	.09	359
TP53 del/mut	1.31	0.44-3.94	0.56	.63	440	0.71	0.21-2.45	0.63	.59	412
del(11q)	0.60	0.24-1.46	0.27	.26	441	0.73	0.37-1.47	0.26	.38	413
del(13q)†	0.54	0.27-1.09	0.19	.09	441	1.08	0.63-1.83	0.28	.79	413
Tri12	3.09	1.56-6.11	1.08	.001	441	0.19	0.06-0.63	0.12	.006	413
Response to treatment‡	0.88	0.57-1.35	0.19	.58	441	0.95	0.67-1.36	0.18	.79	413
LRF CLL4 treatment arm§	0.68	0.45-1.03	0.14	.07	466	1.15	0.84-1.56	0.16	.38	437

*Shows the number of observations included in each logistic regression analysis.

†The presence of a 13q deletion as a sole abnormality using a standard FISH panel.

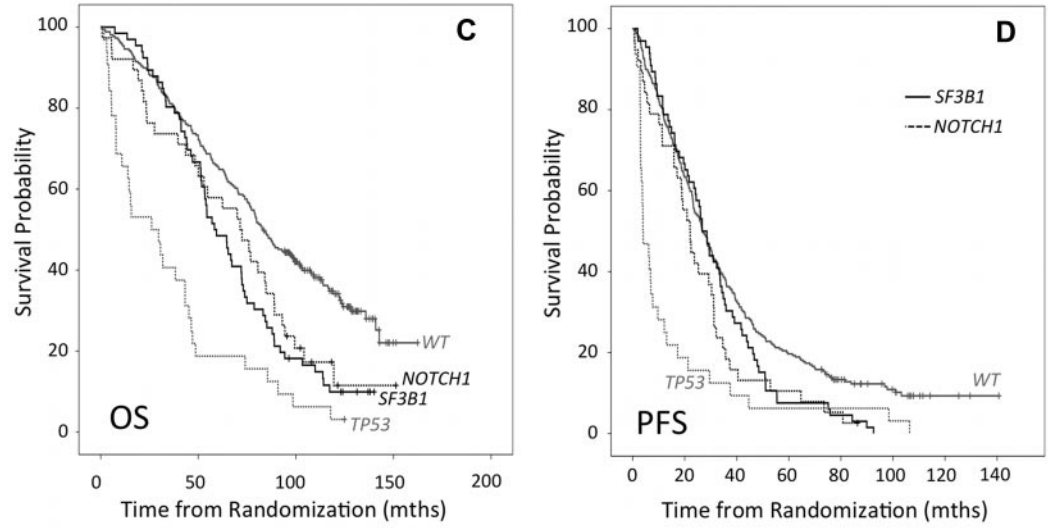
‡Defined as response (CR/nPR, PR, NR) to any treatment.

§Any treatment (CHL, FC, FDR). *TP53* abnormalities defined by deletion and/ or mutation (del/mut).



NOTCH1	Sur. (mths)	95% CI	Events	P-value
Wild-type	75	68-81	308	0.02
Mutated	55	31-79	40	

SF3B1	Sur. (mths)	95% CI	Events	P-value
Wild-type	79	72-86	250	<0.001
Mutated	54	47-61	66	



Status	Sur. (mths)	95% CI	Events	P-value
Wild-type	83	74-92	170	ref
NOTCH1	72	51-92	32	0.02
SF3B1	58	47-70	59	<0.001
TP53 del/mut	26	5-47	31	<0.001

Status	Sur. (mths)	95% CI	Events	P-value
Wild-type	27	23-31	230	ref
NOTCH1	22	17-27	37	0.02
SF3B1	27	23-31	66	0.08
TP53 del/mut	4	0.6-7.3	32	<0.001

Figure 2. Survival analysis of LRF CLL4 cases accounting for *NOTCH1*, *SF3B1* and *TP53* gene status. (A-B) Overall survival for *NOTCH1* and *SF3B1* mutational status, respectively. (C-D) Overall survival and progression free survival, respectively. Patient subgroups with mutations in *NOTCH1* (but wild-type for *SF3B1* and *TP53*), *SF3B1* (but wild-type for *NOTCH1* and *TP53*), and *TP53* del/mut cases (but wild-type for *SF3B1* and *NOTCH1*); highlighted by a solid black, dashed black, and dashed gray line, respectively. The solid-gray line identifies the WT subgroup that is wild-type for *NOTCH1*, *SF3B1*, and *TP53* (that lack either a deletion or mutation). The *P* values are derived from Kaplan-Meier analysis with a log-rank test and median survival times with 95% confidence intervals.

confirmed somatic mutations in *SF3B1* of 17% (73/437; Table 1) in the LRF CLL4 trial. In addition to these 75 mutations, we identified 17 sequence variants that have not been previously identified in CLL, 6 of which have been previously observed in myeloid disease[9,10] (supplemental Table 3) and the remaining 11 have been predicted to be functional deleterious using PolyPhen and SIFT analysis. However, as these could not be somatically confirmed in the context of CLL they were precluded from the main statistical analysis. *SF3B1* was altered by missense mutations that clustered in the HEAT3 to HEAT6 repeats of the SF3B1 protein (Figure 1B). Six codons were targeted by recurrent mutations, including codons 666, 700, and 742 in 6, 37, and 23 patients, respectively. The clinical and biologic features of *SF3B1* mutated CLL are summarized in Tables 1 and 2. Mutations occurred irrespective of the *IGHV* status, ZAP-70 expression, deletion of 11q or 17p and mutation of *TP53*. However, mutations were enriched in CD38 positive patients (OR 2.33, $P = .005$) and those without trisomy 12 (OR 0.19, $P = .006$, Table 2).

With univariate analysis we found that *SF3B1* mutations were significantly associated with a reduced median OS of 54.3 versus

Table 3. Univariate Cox proportional hazard analysis of OS and PFS in the LRF CLL4 patients

			Overall Survival (mouths)							Progression-free survival (months)						
	Mutation/Biomarkers		**Total**	**Events**	**Median**	**95% CI**	**HR**	**95% CI**	***P***	**Total**	**Events**	**Median (mean)**	**95% CI**	**HR**	**95% CI**	***P***
Gene mutation	*NOTCH1*	Wild-type	420	308	74.6	68.4-80.9				420	387	26.4	23.6-29.3			
		Mutated	46	40	54.8	31.0-78.5	1.47	1.06-2.05	.02	46	45	22.0	17.2-26.9	1.46	1.07-1.99	.02
	SF3B1	Wild-type	364	250	79.0	71.8-86.3				364	326	26.5 (39.1)	23.1-29.9			
		Mutated	73	66	54.3	47.3-61.4	1.71	1.30-2.25	< .001	73	73	26.5 (29.4)	22.4-30.7	1.32	1.02-1.71	.03
Clinical feature	Age						1.05	1.04-1.07	< .001					1.00	0.99-1.01	.87
	Sex	Male	365	280	70.3	61.4-79.2				365	340	25.2	22.3-28.2			
		Female	129	86	79.7	66.5-93.0	0.77	0.61-0.98	.04	129	115	29.4	25.5-33.3	0.79	0.64-0.98	.03
	Binet Stage (A/B/C)	A	112	76	80.6	63.4-97.7				112	104	27.2	23.8-30.7			
		B	231	174	76.7	70.2-83.3				231	213	26.5	22.5-30.6			
		C	151	116	54.3	44.8-63.8	1.21	1.04-1.39	.01	151	138	24.4	18.4-30.4	1.02	0.89-1.16	.77
	Binet Stage (A vs. B/C)	A	112	76	80.6	63.4-97.7				112	104	27.2	23.8-30.7			
		B/C	382	290	71.5	64.8-78.2	1.29	1.01-1.67	.04	382	351	26.1	23.0-29.1	1.00	0.81-1.25	.98
Biomarker	*IGHV* status	Mutated	115	91	104.2	93.3-115.1				115	128	34.8	29.3-40.4			
		Unmutated	254	215	60.6	52.5-68.7	2.18	1.70-2.79	< .001	254	249	22.7	20.2-25.2	1.86	1.49-2.31	< .001
	ZAP70 positive	Negative	177	117	84.7	69.6-99.8				177	154	30.5	25.5-35.4			
		Positive	197	162	69.0	59.7-78.2	1.47	1.16-1.86	.002	197	188	23.0	20.2-25.8	1.42	1.15-1.76	.001
	CD38 positive	Negative	223	149	84.7	73.8-95.6				223	199	28.5	24.4-32.7			
		Positive	177	145	64.5	54.2-74.9	1.54	1.22-1.94	< .001	177	169	21.6	17.6-25.7	1.39	1.13-1.70	.002
	TP53 del/mut	Normal	431	313	75.9	69.7-82.1				431	395	26.7	23.9-29.5			
		Del/mut	32	31	26.1	4.9-47.3	3.19	2.20-4.62	< .001	32	32	3.9	0.57-7.3	2.70	1.88-3.89	< .001
	del(11q)	Undeleted	372	266	75.0	67.5-82.6				372	340	28.5	26.0-31.0			
		Deleted	92	79	57.7	42.4-73.0	1.62	1.26-2.09	< .001	92	88	17.9	13.4-22.5	1.49	1.18-1.88	.001
	del(13q)sole	Undeleted	284	227	58.4	48.7-68.1				284	270	23.0	20.6-25.4			
		Deleted	177	115	88.9	73.2-104.5	0.59	0.47-0.74	< .001	177	155	32.3	28.0-36.5	0.67	0.56-84.0	< .001
	Trisomy12	Absent	388	284	73.9	67.3-80.4				388	357	26.5	23.5-29.5			
		Present	76	61	59.4	38.9-80.1	1.30	0.98-1.71	.07	76	71	19.4	11.9-26.8	1.28	0.99-1.65	.06

The double *SF3B1* mutated (700 and 742) cases (n = 2) remain in this analysis. We performed the analysis excluding these cases and obtained similar results. *TP53* abnormalities defined by deletion and/or mutation (del/mut).

Table 4. Multivariate Cox proportional hazard analysis of OS and PFS in the LRF CLL4 patients

	Overall survival			Progression-free survival		
Variable	HR	95% CI	*P*	HR	95% CI	*P*
NOTCH1	1.58	1.05-2.38	.03	1.29	0.87-1.90	.21
SF3B1	1.52	1.10-2.12	.01	1.31	0.97-1.78	.08
Age	1.05	1.03-1.07	< .001	0.99	0.98-1.01	.29
Sex	0.72	0.53-0.99	.04	0.82	0.63-1.08	.15
Binet stage (A vs B/C)	1.45	1.02-2.04	.04	1.02	0.77-1.36	.90
del(11q)	1.39	1.04-1.87	.03	1.61	1.22-2.11	.001
IGHV unmutated	1.85	1.38-2.50	< .001	1.86	1.43-2.42	< .001
TP53 del/mut	2.48	1.50-4.08	< .001	2.09	1.29-3.37	.003
LRF CLL4 arm (Chl vs FDR/FC)	1.01	0.78-1.30	.97	0.50	0.40-0.63	< .001

OS Multivariate: 330 cases with 244 events; 164 cases with missing data. PFS Multivariate: 330 cases with 306 events; 164 cases with missing data. *TP53* abnormalities defined by deletion and/or mutation (del/mut).

79.0 months for mutant and wild-type alleles, respectively (HR 1.71; 95% CI 1.30-2.25, $P < .001$), which was independent of treatment arm (Figure 2B). The presence of a *SF3B1* mutation was associated with a reduced mean PFS (29.4 versus 39.1mths) in the LRF CLL4 cohort (HR 1.32, 95% CI 1.02-1.71, $P = .03$; Table 3). However, wild-type and mutant alleles had a similar median PFS of 26.5 months. By investigating the impact of mutations in each trial arm, we show that, after controlling for *TP53* del/mut cases, a mutant *SF3B1* gene was most associated with reduced PFS for FC-treated patients (HR 2.08; 95% CI 1.29-3.34, $P = .002$), with a median survival of 46.0 and 29.4 months for a wild-type and mutant *SF3B1* allele, respectively. *SF3B1* mutations were not associated with ORS in any of the 3 treatment arms (data not shown).

NOTCH1 and *SF3B1* mutations identify additional patients with poor outcome after treatment with chemotherapy

We estimated the adjusted impact of *NOTCH1* and *SF3B1* mutations on PFS and OS after controlling for confounding variables using multivariate Cox proportional hazard analysis. Along with *NOTCH1* and *SF3B1* mutations other variables included in the analysis were age, gender, stage, deletion of 11q, *IGHV* mutation status, and deletion and/or mutation of *TP53*. We developed this model for OS and PFS as it is comparable with those that were previously used to show the prognostic independence of both genes in other studies.[6,7] We confirmed the independent prognostic significance of several established biomarkers, including *IGHV* mutational status and deletion and/or mutation of *TP53* in our patients (Table 4). Moreover, multivariate analysis selected *NOTCH1* (HR 1.58, 95% CI 1.05-2.38, $P = .03$) and *SF3B1* (HR 1.52, 95% CI 1.10-2.12, $P = .01$) mutations as independent risk factors of OS but not PFS (Table 4).

It has been suggested that mutations in *NOTCH1* and *SF3B1* identify patients with dismal outcome comparable with those with *TP53* del/mut.[6] However, in LRF CLL4, *TP53* del/mut patients exhibited a considerably shorter OS and PFS than wild-type *TP53* patients with either mutations of *NOTCH1* or *SF3B1* (Table 3, Figure 2C-D). Specifically, *TP53* del/mut patients, within the subgroup of patients analyzed for *NOTCH1* and *SF3B1,* exhibited a median OS of 26.1 months compared with 75.9 months for wild-type individuals (HR 3.19, $P < .001$; Table 3). This shows that in the context of patients receiving first-line chemotherapy, *TP53* abnormalities remain the most significant marker of reduced survival.

Discussion

Initial whole-genome sequencing studies have identified a panel of novel recurrently mutated genes in CLL including *NOTCH1* and *SF3B1*.[2-6] Although these lesions have been linked to clinical outcome, these associations have largely been defined in heterogeneous retrospective CLL cohorts. However, given the highly heterogeneous natural history of CLL and the often-serendipitous date of initial diagnosis, it is important to confirm the prognostic relevance of novel biomarkers in the context of randomized clinical trials. Herein, we present just such a study of our findings from the LRF CLL4 trial, where we show that *NOTCH1* and *SF3B1* mutations occur at a frequency of 10% and 17%, respectively, in samples taken at trial entry.

The incidence of *NOTCH1* mutations is greater than that previously seen in MBL (3.2%)[17] and comparable with the frequency found at CLL diagnosis (12%).[4-6] Our data support previously published findings, that showed a significant enrichment of *NOTCH1* mutations in *IGHV* unmutated, ZAP-70 positive patients,[4-6] and between *NOTCH1* mutations and trisomy 12 as defined by others.[3,8] The explanation for this association between *NOTCH1* mutations and other biomarkers is currently unclear. In contrast to previous findings,[7] we show that 22% of *NOTCH1* mutations (10/46) occurred concomitantly with alterations of *ATM*/*TP53*. We did not detect a difference in the incidence of death from RS between *NOTCH1* mutated and wild-type individuals. However, we did not have the necessary data to further examine the relationship between *NOTCH1* mutated CLL and subsequent Richter transformation.[4,5,7] We confirm previous observations that *NOTCH1* mutational status impacts on OS as an isolated variable in univariate analysis.[4,5,7] We also show a novel association between *NOTCH1* mutations and shorter PFS. The multivariate survival analysis identified *NOTCH1* mutation as an independent marker of poor prognosis (OS not PFS) when we controlled for a comprehensive panel of covariates, including deletion of *ATM* and *TP53*. However, our data suggest that in a homogeneous population of CLL patients requiring first-line treatment, *NOTCH1* mutations identify patients with intermediate survival, comparable with 11q loss, rather than the poor survival exhibited by *TP53* deleted or mutated individuals. Taken together, our data demonstrate that *NOTCH1* mutational status is an independent marker that identifies patients with intermediate outcome after initial therapy with DNA damaging agents.

SF3B1 is a critical component of the RNA splicing machinery that achieves successful transcription and guarantees the functional diversity of protein species using alternative spicing. The precise consequences of *SF3B1* mutations are not known but it is possible that differences in the transcriptome of lymphoid and myeloid cells would result in distinct functional consequences. This may be relevant to the observation that mutations are associated with poorer overall survival in CLL,[2,3,6] wheres in the myelodysplastic syndrome (MDS) they correlate with a more favorable disease subtype and improved survival even though the strikingly similar distribution of mutations suggest similar mechanisms of action.[10,18-20] It was previously demonstrated that chemo-refractory CLL is enriched for *SF3B1* mutations (17%) compared with 4% at diagnosis.[6] In contrast, we observed a similar high frequency (17%) in previously untreated patients and no correlation between the presence of a mutation and response to treatment. In univariate analysis, we have confirmed the previous association between mutation of *SF3B1* and shorter overall survival. We did observe a weak association between *SF3B1* mutation and PFS, which was most evident in the FC arm of the trial. Critically, our multivariate survival analysis demonstrated that *SF3B1* mutations are an independent marker for intermediate OS similar to 11q loss or *NOTCH1* mutation, but they are not an independent marker for PFS in the LRF CLL4 trial.

In conclusion, using *NOTCH1* and *SF3B1* as examples, we present the first study to validate the importance of novel mutated cancer genes identified by ongoing CLL sequencing initiatives in the context of a randomized, prospective clinical trial. Given the treatment modalities used in the LRF CLL4 trial, it will be important to validate this observation in the context of chemo-immunotherapy trials and in studies of novel agents, such as B-cell receptor (BCR) signaling inhibitors. Importantly, we show the independent prognostic importance of *NOTCH1* and *SF3B1* mutations, thereby providing valuable validation of the clinical utility of these gene mutations in the accurate risk-adapted stratification of CLL patients. As important, is our observation that alterations of the *TP53* gene remains the most informative marker of dismal survival in this trial, supporting strategies for interrogating the P53 pathway for the identification of poor prognosis patients. However, our data add to a growing body of evidence demonstrating the value of mutational analysis of other key genes. This will ultimately establish a stratified and individualized approach to care, including the potential for targeted therapy.

Acknowledgments

The authors thank all patients and clinicians who participated in the trial.

This study was funded by The Kay Kendall Leukemia Fund, Leukemia and Lymphoma Research, Cancer Research UK, and Wessex Medical Research. The LRF CLL4 trial was funded by a core grant from Leukemia and Lymphoma Research. D.G. and D.C. acknowledge the support by The Royal Marsden Hospital and The Institute of Cancer Research National Institute of Health Research Biomedical Research Center. N.W. and M.E. acknowledge the support of Dr Mildred Scheel Stiftung for Cancer Research (Bonn, Germany) and the Arbib Foundation, respectively.

Authorship

Contribution: N.W., M.J.J.R.-Z., D.G.d.C., J.F., H.P., and B.G. performed the experimental work; A.G., D.G.d.C., and A.P. performed the molecular diagnostic assays; M.J.J.R.-Z., A.P., M.E., and A.C. conducted the statistical analyses; D.G.O. and D.C. contributed patient samples and data; D.G.O., N.C.P.C., and J.C.S. initiated and designed the study; M.J.J.R.-Z., D.G.O., and J.C.S. wrote the paper with contributions from N.C.P.C., D.G.d.C., and N.W; and all authors critically reviewed the final paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Jonathan C. Strefford, Cancer Genomics Group, Cancer Sciences Unit, Somers Cancer research Bldg, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, United Kingdom; e-mail: JCS@soton.ac.uk.

References

1. Döhner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med.* 2000;343(26):1910-1916.
2. Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genet.* 2012;44(1):47-52.
3. Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med.* 2011;365(26):2497-2506.
4. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature.* 2011;475(7354):101-105.
5. Fabbri G, Rasi S, Rossi D, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J Exp Med.* 2011;208(7):1389-1401.
6. Rossi D, Bruscaggin A, Spina V, et al. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood.* 2011;118(26):6904-6908.
7. Rossi D, Rasi S, Fabbri G, et al. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood.* 2012;119(2):521-529.
8. Balatti V, Bottoni A, Palamarchuk A, et al. NOTCH1 mutations in CLL associated with trisomy 12. *Blood.* 2012;119(2):329-331.
9. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature.* 2011;478(7367):64-69.
10. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med.* 2011;365(15):1384-1395.
11. Catovsky D, Richards S, Matutes E, et al. Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. *Lancet.* 2007;370(9583):230-239.
12. Oscier D, Wade R, Davis Z, et al. Prognostic factors identified three risk groups in the LRF CLL4 trial, independent of treatment allocation. *Haematologica.* 2010;95(10):1705-1712.
13. Gonzalez D, Martinez P, Wade R, et al. Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. *J Clin Oncol.* 2011;29(16):2223-2229.
14. White HE, Hall VJ, Cross NC. Methylation-sensitive high-resolution melting-curve analysis of the SNRPN gene as a diagnostic screen for Prader-Willi and Angelman syndromes. *Clin Chem.* 2007;53(11):1960-1962.
15. Gianfelici V. Activation of the NOTCH1 pathway in chronic lymphocytic leukemia. *Haematologica.* 2012;97(3):328-330.
16. Weng AP, Ferrando AA, Lee W, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science.* 2004;306(5694):269-271.
17. Rasi S, Monti S, Spina V, Foa' R, Gaidano G, Rossi D. Analysis of NOTCH1 mutations in monoclonal B cell lymphocytosis. *Haematologica.* 2012;97(1):153-154.
18. Patnaik MM, Lasho TL, Hodnefield JM, et al. SF3B1 mutations are prevalent in myelodysplastic syndromes with ring sideroblasts but do not hold independent prognostic value. *Blood.* 2012;119(2):569-572.
19. Malcovati L, Papaemmanuil E, Bowen DT, et al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood.* 2011;118(24):6239-6246.
20. Damm F, Thol F, Kosmider O, et al. SF3B1 mutations in myelodysplastic syndromes: clinical associations and prognostic implications. *Leukemia.* 2012;26(5):1137-1140.

2013 121: 468-475
doi:10.1182/blood-2012-05-429282 originally published online October 18, 2012

The clinical significance of *NOTCH1* and *SF3B1* mutations in the UK LRF CLL4 trial

David G. Oscier, Matthew J. J. Rose-Zerilli, Nils Winkelmann, David Gonzalez de Castro, Belen Gomez, Jade Forster, Helen Parker, Anton Parker, Anne Gardiner, Andrew Collins, Monica Else, Nicholas C. P. Cross, Daniel Catovsky and Jonathan C. Strefford

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.

Whole Exome Sequencing Identifies Novel Recurrently Mutated Genes in Patients with Splenic Marginal Zone Lymphoma

Marina Parry[1☯], Matthew J. J. Rose-Zerilli[1☯], Jane Gibson[2☯], Sarah Ennis[2], Renata Walewska[3], Jade Forster[1], Helen Parker[1], Zadie Davis[3], Anne Gardiner[3], Andrew Collins[2], David G. Oscier[1,3], Jonathan C. Strefford[1*]

1 Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom, **2** Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom, **3** Department of Pathology, Royal Bournemouth Hospital, Bournemouth, United Kingdom

Abstract

The pathogenesis of splenic marginal zone lymphoma (SMZL) remains largely unknown. Recent high-throughput sequencing studies have identified recurrent mutations in key pathways, most notably *NOTCH2* mutations in >25% of patients. These studies are based on small, heterogeneous discovery cohorts, and therefore only captured a fraction of the lesions present in the SMZL genome. To identify further novel pathogenic mutations within related biochemical pathways, we applied whole exome sequencing (WES) and copy number (CN) analysis to a biologically and clinically homogeneous cohort of seven SMZL patients with 7q abnormalities and *IGHV*1-2*04 gene usage. We identified 173 somatic non-silent variants, affecting 160 distinct genes. In additional to providing independent validation of the presence of mutation in several previously reported genes (*NOTCH2, TNFAIP3, MAP3K14, MLL2* and *SPEN*), our study defined eight additional recurrently mutated genes in SMZL; these genes are *CREBBP, CBFA2T3, AMOTL1, FAT4, FBXO11, PLA2G4D, TRRAP* and *USH2A*. By integrating our WES and CN data we identified three mutated putative candidate genes targeted by 7q deletions (*CUL1, EZH2* and *FLNC*), with *FLNC* positioned within the well-characterized 7q minimally deleted region. Taken together, this work expands the reported directory of recurrently mutated cancer genes in this disease, thereby expanding our understanding of SMZL pathogenesis. Ultimately, this work will help to establish a stratified approach to care including the possibility of targeted therapy.

Citation: Parry M, Rose-Zerilli MJJ, Gibson J, Ennis S, Walewska R, et al. (2013) Whole Exome Sequencing Identifies Novel Recurrently Mutated Genes in Patients with Splenic Marginal Zone Lymphoma. PLoS ONE 8(12): e83244. doi:10.1371/journal.pone.0083244

Editor: Jose Angel Martinez Climent, University of Navarra, Spain

Received July 9, 2013; **Accepted** November 1, 2013; **Published** December 13, 2013

Funding: The manuscript was supported by: Leukaemia and Lymphoma Research, Cancer Research UK, Kay Kendall Leukaemia Fund, Wessex Medical Research, Bournemouth Leukaemia Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: JCS@soton.ac.uk

☯ These authors contributed equally to this work.

Introduction

Splenic Marginal Zone Lymphoma (SMZL) is a low grade chronic B cell lymphoproliferative disorder that predominantly affects elderly patients and involves the spleen, bone marrow, and peripheral blood [1]. Although the median survival is around 10 years, approximately 70% of SMZL patients require treatment, of whom 25% experience progressive disease, leading to early death [1].

Our understanding of the molecular pathogenesis of SMZL remains limited. Early cytogenetic studies identified recurrent deletions of 7q31-q32 and duplications of 3q in approx. 30% and 20% of cases, respectively [2], but subsequent molecular investigations have failed to identify causative genes within these regions [3]. Candidate gene studies are limited to mutations in *TP53*, which is disrupted in 10-15% of cases [2], and to genes within the NF-ĸB pathway, which are mutated in a third of all cases [4,5]. The presence of a highly restricted immunoglobulin gene repertoire, in particular the selective usage of the immunoglobulin heavy chain variable (*IGHV*) 1-2*04 allele in 20-30% of patients, suggests that antigenic stimulation may be important in the pathogenesis of this disease [6].

The recent application of whole exome sequencing to frozen splenic tissue from 14 patients with SMZL followed by targeted resequencing of recurrent variants in larger cohorts has identified further biologically relevant genes [7,8]. Mutations in *NOTCH2*, which eliminate the C-terminal PEST domain and

result in compromised protein degradation, were identified in 20 - 25% of cases although there was no consensus as to the clinical significance of these mutations between studies [7,8]. Gene mutations in modulators or other members of the Notch signalling pathway and in other pathways, such as chromatin remodelling and transcriptional regulation were also implicated [8].

In view of the relatively small number of patients investigated so far and the biological heterogeneity of SMZL, it is vital to perform additional gene discovery experiments to fully catalogue the molecular lesions that contribute to disease pathogenesis. To this aim, we performed whole exome sequencing and copy number analysis of tumour and germ-line DNA extracted from a clinically homogeneous cohort of SMZL patients. In doing so, we expand the reported directory of recurrently mutated cancer genes in this disease, thereby expanding our understanding of SMZL pathogenesis that will ultimately facilitate improvements in disease management and the promise of novel therapies.

Materials and Methods

Patients and biomarker analysis

Seven patients were included in this current study, all met established diagnostic criteria [1], and 5/7 underwent a splenectomy with histology typical of SMZL in each case and no evidence of transformation to a high-grade lymphoma. Each patients harboured chromosomal aberrations targeting 7q and *IGHV1-2*04* usage (Table **S1**), ensuring the exclusion of other types of splenic lymphoma from our analysis and maximizing the likelihood of identifying pathogenic mutations within related biochemical pathways. Informed patient consent was obtained according to the declaration of Helsinki, and the study was ethically approved by the local REC.

Chromosomal analysis was performed and described according to the International System for Human Cytogenetic Nomenclature [9]. Immunoglobulin variable region genes were sequenced from either cDNA or gDNA as previously described [6]. cDNA was synthesised by reverse transcription according to the manufacturers protocol (Promega). gDNA was extracted using the Qiagen Blood Mini Kit and amplified using the BIOMED 2 protocol [10]. PCR products were sequenced directly using an ABI 310 genetic analyser and sequences were aligned to the IMGT-V-Quest database.

High-throughput sequencing, variant calling and Sanger validation

Using targeted exome capture (SureSelect Human All Exon 51Mb V4, 50Mb V3, Agilent) we prepared sequencing libraries from high-molecular weight genomic DNA from CD19 positive-purified tumour cells (five cases extracted from the spleen and two from peripheral blood) and matched saliva cells (Oragene DNA kit, DNA Genotek) prior to high-throughput sequencing with the Illumina HiSeq system. The paired-end sequencing data were aligned against the human genome reference sequence (hg19/GRCh37) using the Novoalign software (novoalignMPI V2.08.02, Novocraft Technologies, Selangor, Malaysia). Duplicate reads, resulting from PCR clonality or optical duplicates, and reads mapping to multiple locations were excluded from downstream analysis. Depth and breadth of sequence coverage was calculated with custom scripts and the BedTools package (v2.13.2) [11] and is included in table **S2.**

Germ-line-Tumour paired datasets were analysed to identify single nucleotide variations (SNVs) and small insertion and deletions using Varscan 2.3.3 [12] (http://varscan.sourceforge.net). The minimum variant allele frequency threshold was set to 10% with a minimum read depth of 4. Variants were filtered using the 'somaticFilter' command to remove clusters of false positives and SNV calls near indels with the same frequency and depth thresholds.

Variants were annotated with respect to genes and transcripts and filtered using the Annovar software tool (v2012Jun21) [13]. Variants were cross referenced with databases of known variation were downloaded from the Annovar website (June 2012); data from the 1000 Genomes Project (2012 April release)[14], dbSNP135 (and a version with SNPs flagged as rare <1% frequency or clinically associated by NCBI) and data from 4300 European American samples from The National Heart Lung and Blood Institute Exome Sequencing Project Exome Variant Server (http://evs.gs.washington.edu/EVS/), (ESP6500 release). Using conventional Sanger sequencing, we confirmed the presence of 38/45 somatic variants (84.4%) and those non-concordant cases were due to low exome read-depth in the tumour sample.

SNP6.0 array hybridization, data extraction and analysis

Tumour and germ-line DNA was purified, amplified, labelled and hybridized to the Affymetrix SNP6.0 platform (Affymetrix, Santa Clara, CA) as previously described [15]. For copy number analysis, two independent researchers visually inspected parallel copy number profiles (aligned to hg19/GRCh37) from tumour and germ-line samples using Partek Genomics Suite (Partek Inc, Missouri, USA), and lesions were classified as somatic if they were present and absent in the tumour and germ-line material, respectively. Copy number alterations (CNAs) were defined as a deviation of 50 consecutive array features (probes) from a normal value of 2 (±0.3), within a consecutive genomic window of 50 Kilobases. The allele ratio was calculated for each sample using the HapMap Allele Reference baseline (Affymetrix) and copy number neutral loss of heterozygosity (CNNLOH) event were defined as somatic if they were present and absent in the tumour and germ-line material, respectively.

Results and Discussion

Exome-capture and high-throughput sequencing allowed us to align approx. 41.9 million reads per sample at a mean depth of 69x (range, 43-109x). In total, an average of 82.2% (range, 70-95%) of target sequences captured at 20x. Our analytical pipeline identified 176 somatic non-silent variants, affecting 165 distinct genes (Table **S2**). These variants were base-pair transitions (34%), transversions (28%), insertions (6%) and

Table 1. Summary of recurrently mutated genes in our cases and a comparison with previously published studies.

	Genes	Accession number	Variant nomenclature		SIFT score prediction	Polyphen-2 score prediction	Case no[6]. 1	2	3	4	5	6	7	Published Study
			Nucleotide change[4]	Amino acid change	consequences	consequences								[7]
ESTABLISHED:	***NOTCH2***	NM_024408	c.C7081T°	p.Q2361X	Damaging	Truncating					✓			[7,8]
Genes			c.6836delA§	p.H2279fs	Truncating	Truncating	✓							
recurrently	***MAP3K14***	NM_003954	c.C200G*	p. A67G	Damaging	Probably damaging					✓			[8]
mutated	***TNFAIP3***	NM_006290	c.T1343A*	p.M448K	Tolerated	Truncating						✓		[4,5,8]
in SMZL[1]			c.C1681T*	p. P561S	Tolerated	Truncating						✓		
			c.A328T*	p. T110S	Tolerated	Truncating				✓				
	MLL2	NM_003482	c.2507_2508insC	p.Q836fs	Truncating	Truncating						✓		[8]
	SPEN	NM_015001	c.C5179T°	p.Q1727X	Tolerated	Truncating						✓		[8]
			c.10286_10289del§	p.3429_3430del	Truncating	Truncating						✓		
NOVEL:	***AMOTL1***	NM_130847	c.G1270A*	p.A424T	Tolerated	Benign						✓		[8]
Recurrent	***FAT4***	NM_024582	c.G6628A*	p.A2210T	Tolerated	Benign	✓							[8]
genes across	***FBXO11***	NM_001190274	c.G1587C*	p.W529C	Damaging	Truncating				✓				[8]
studies[2]	***PLA2G4D***	NM_178034	c.23delG§	p.G8fs	Truncating	Truncating			✓					[8]
	TRRAP	NM_003496	c.367-10T>-^	Splicing[5]	Truncating	Truncating					✓			[8]
	USH2A	NM_206933	c.G7553C*	p.S2518T	Tolerated	Benign				✓				[8]
NOVEL:	***CBFA2T3***	NM_175931	c.C464T*	p.P155L	Damaging	Probably damaging					✓			Novel
recurrent			c.G1445A	p. S482N	Damaging	Probably damaging							✓	
genes in our study[3]	***CREBBP***	NM_001079846	c.A4349G*	p.Y1450C	Damaging	Truncating					✓	✓		Novel

[1] Identifies those genes that have previously been shown to be targeted by recurrent mutations in SMZL. [2] Shows those genes that were mutated in single SMZL cases in both our current study and in previously published work. [3] Shows the novel genes targeted by recurrent mutations in our study.

[4] Identified non-synonymous (*), splice-site (^), frameshift (§) and stopgain (°) mutations

[5] The *TRRAP* mutation in case 5 occurred within a splice-site and is predicted to resulted in aberrant splicing

[6] Showed the presence (✓) and absence (white box) of each mutation in the patients in our series

[7] Highlights the published studies that identified the mutations in each of the genes listed

doi: 10.1371/journal.pone.0083244.t001

deletions (31%). Copy number analysis identified 28 somatically-acquired copy number deletions (66%) and duplications (33%), (Table **S2**). Considering the mutation and copy number data together, our patients exhibited an average of 25 somatic mutations (range, 9-40) and four copy number alterations (range, 2-9) per tumour sample.

We initially investigated our exome sequencing data for the presence of somatic variants in genes known to be recurrently mutated in SMZL. In doing so, we identified mutations in *NOTCH2* [exon 34, n=2], *TNFAIP3* [n=3], *MAP3K14* [n=2], *MLL2* [n=1] and *SPEN* [n=1] (Table **1**). As the exome capture efficiency of *NOTCH2* can compromise variant identification, we also performed Sanger sequencing of exon 34 as previously reported [8]. In doing so, we found no additional mutations. Furthermore, we identified mutations in six genes that have previously been shown to harbour mutations in single SMZL cases [8] (Table **1**). This observation implicates these genes as recurrent mutational targets in SMZL. Mutations in several of these genes have been identified in other tumour types, for example *FBXO11*, which is recurrently mutated in diffuse large B-cell lymphoma (DLBCL) and promote leukaemogenesis by stabilization of BCL6 [16] (Table **1**).

Next we investigated our SMZL cases for recurrent mutations in genes that have not been previously identified in SMZL (Table **1**). This analysis identified two genes, *CREBBP* and *CBFA2T3*, both mutated in two patients, which in the context of the published literature provides a potential prevalence of approx. 10% in SMZL. Both of the *CREBBP* mutations were the Y1412C variant previously identified in DLBCL [17]. *CREBBP* is involved in chromatin remodelling and transcription factor recognition, and this mutation has been shown to compromise the protein's ability to acetylate BCL6 and p53 [17]. The *CBFA2T3* gene, a core binding factor from the myeloid translocation gene family, is targeted by recurrent chromosomal rearrangements in both lymphoid and myeloid malignancies. Whilst non-synonymous in nature, our mutations were not located within the key ETO, MTG16 or TAFH functional domains of the protein. In pediatric B-cell lymphoma, *CBFA2T3* has been implicated as a cellular proto-oncogene as in rare cases the gene is juxtaposed to the immunoglobulin locus [18]. In AML chromosomal inversions involving *CBFA2T3* can directly increase the self-renewal capacity of hematopoietic progenitors [19]. Mutations in both these genes were present in approx. 50% of reads, suggesting they are heterozygous mutations present in the dominant tumour clone.

To further assess the potential biological impact of the mutations observed in our cases, pathway analysis was performed using the Database for Annotation, Visualisation and Integrated Discovery (DAVID) (Table **2**). In addition to identifying pathways already implicated in SMZL pathogenesis, such as notch signalling (*NOTCH2*, *NOTCH4*), we also show that genes within MAPK signalling pathway are targeted by

Table 2. Summary of the pathways in which mutated genes in our SMZL cohort can be found and their predicted functional consequences.

DAVID Pathway	Genes	Accession numbers	Variant nomenclature		SIFT score prediction	Polyphen-2 score prediction	Case no[2].						
			Nucleotide change	Amino acid change	consequences	consequences	1	2	3	4	5	6	7
MAP kinase	*CACNA1E*	NM_001205293	c.G1069C	p.E357Q	Damaging	Damaging				✓			
	CACNA1H	NM_021098	c.391delG	p.E131fs	Truncating	Truncating			✓				
	CACNA2D2	NM_001174051	c.2837delC	p.P946fs	Truncating	Truncating		✓					
	FLNC	NM_001458	c.C3179T	p.P1060L	Damaging	Probably damaging					✓		
	MAP3K14	NM_003954	c.C200G	p. A67G	Truncating	Truncating				✓	✓		
	MAPK8IP3	NM_001040439	c.743delA	p.Q248fs	Truncating	Truncating		✓					
	RASA1	NM_002890	c.C142A	p. P48T	Damaging	Truncating						✓	
	TAOK3	NM_016281	c.438-7-T)	Splicing[1]	Truncating	Truncating		✓					
Notch	*NOTCH2*	NM_024408	c.C7081T	p.Q2361X	Truncating	Truncating					✓		
			c.6836delA	p.H2279fs	Truncating	Truncating	✓						
	PIWIL3	NM_001008496	c.2242delA	p.T748fs	Truncating	Truncating		✓					
	NOTCH4	NM_004557	c.C5877G	p.C1959W	Truncating	Damaging						✓	
	MAML3	NM_018717	c.1513_1514del	p. 505_505del	Truncating	Truncating						✓	
Cell cycle	*CUL1*	NM_003592	c.T469G	p.Y157D	Damaging	Probably damaging				✓			
	CREBBP	NM_001079846	c.A4349G	p.Y1450C	Damaging	Truncating				✓		✓	
	CDC27	NM_001114091	c.A701C	p. Y234S	Tolerated	Benign						✓	
Cytokine-cytokine receptor interaction	*FLT1*	NM_002019	c.2594_splice	splicing	Truncating	Truncating	✓						
	CRLF2	NM_022148	c.G340C	p.V114L	Tolerated	Probably damaging				✓			

[1] The *TRRAP* mutation in case 5 occurred within a splice-site and is predicted to resulted in aberrant splicing

[2] Showed the presence (✓) and absence (white box) of each mutation in the patients in our series

doi: 10.1371/journal.pone.0083244.t002

somatic non-synonymous mutations in the majority of our cases (5/7, 71%). Whilst a biological role of these genes in SMZL required functional confirmation, our data does suggest that the MAPK signalling pathway is a major target for somatic mutations in this sub-group of SMZL.

Finally, we identified somatically acquired mutations in genes also targeted by 7q deletions in our patients. In doing so, we found *CUL1*, *FLNC* and *EZH2* mutations in individual cases (Table **2**). Of these gene mutations, only *FLNC* was located within the published 7q MDR [3,20]. *FLNC* mutations have not been previously identified in a series eight del(7q) cases [20], suggesting that the prevalence of *FLNC* mutation is low in this sub-type of SMZL. However, further research will be required to establish if rare mutations represent only one mechanism of gene deregulation, as repression of *FLNC* transcription by promoter methylation in prevalent in several other human cancer types [21-24]. The somatic variant we identified in *EZH2* (p.K199N), which is located outside the SET protein domain, is not the activating mutation prevalent in follicular lymphoma and DLBCL [25] nor has it been previously reported in AML or MDS [26].

Herein, for the first time, we report the analysis of a homogeneous cohort of SMZL cases using whole exome sequencing and copy number analysis. In doing so, we validate the presence of recurrent mutations in several genes with established importance in SMZL. Furthermore, we expand the reported directory of recurrently mutation cancer genes in this disease, with the most significant observation being the identification of recurrent mutations in *CREBBP* and *CBFA2T3*. The importance of *CREBBP* is further strengthen by the presence of a single SMZL case in the literature with a small deletion that juxtaposes 16 exons of *CREBBP* with the *ZNF434* gene, resulting in loss of the acetyltransferase domain of the CREBBP protein [8]. Furthermore, we show the majority of cases in our series carried mutations within MAPK signalling genes, suggesting that mutations in these genes are strongly associated with 7q-rearranged SMZL with *IGHV1-2*04* usage. Whilst our analysis identifies a series of novel genes mutated in SMZL, a larger study is required to determine the frequency of these events and any utility in the risk-adapted stratification of SZML patients. To this aim, we are currently coordinating a pan-European study into the presence of somatic mutations in approx. 750 genes with a known or postulated role in SMZL pathophysiology in a cohort of more than 300 SMZL cases. This will ultimately establish the frequency and clinical importance of gene mutations in SMZL and help to establish a stratified approach to care including the possibility of targeted therapy.

Supporting Information

Table S1. Clinical characteristics of each patient included in the study.
(DOCX)

Table S2. Shows the 176 non-silent somatic mutations identified in our cases.
(XLSX)

Author Contributions

Conceived and designed the experiments: JS DO. Performed the experiments: MP MRZ JF HP ZD AG. Analyzed the data: JG SE AC. Contributed reagents/materials/analysis tools: JG SE AC DO RW. Wrote the manuscript: MP DO JS.

References

1. Matutes E, Oscier D, Montalban C, Berger F, Callet-Bauchu E et al. (2008) Splenic marginal zone lymphoma proposals for a revision of diagnostic, staging and therapeutic criteria. Leukemia 22: 487-495. doi: 10.1038/sj.leu.2405068. PubMed: 18094718.
2. Salido M, Baró C, Oscier D, Stamatopoulos K, Dierlamm J et al. (2010) Cytogenetic aberrations and their prognostic value in a series of 330 splenic marginal zone B-cell lymphomas: a multicenter study of the Splenic B-Cell Lymphoma Group. Blood 116: 1479-1488. doi:10.1182/blood-2010-02-267476. PubMed: 20479288.
3. Watkins AJ, Hamoudi RA, Zeng N, Yan Q, Huang Y et al. (2012) An integrated genomic and expression analysis of 7q deletion in splenic marginal zone lymphoma. PLOS ONE 7: e44997. doi:10.1371/journal.pone.0044997. PubMed: 23028731.
4. Rossi D, Deaglio S, Dominguez-Sola D, Rasi S, Vaisitti T et al. (2011) Alteration of BIRC3 and multiple other NF-κB pathway genes in splenic marginal zone lymphoma. Blood 118: 4930-4934. doi:10.1182/blood-2011-06-359166. PubMed: 21881048.
5. Novak U, Rinaldi A, Kwee I, Nandula SV, Rancoita PM et al. (2009) The NF-{kappa}B negative regulator TNFAIP3 (A20) is inactivated by somatic mutations and genomic deletions in marginal zone lymphomas. Blood 113: 4918-4921. doi:10.1182/blood-2008-08-174110. PubMed: 19258598.
6. Bikos V, Darzentas N, Hadzidimitriou A, Davis Z, Hockley S et al. (2012) Over 30% of patients with splenic marginal zone lymphoma express the same immunoglobulin heavy variable gene: ontogenetic implications. Leukemia 26: 1638-1646. doi:10.1038/leu.2012.3. PubMed: 22222599.
7. Kiel MJ, Velusamy T, Betz BL, Zhao L, Weigelin HG et al. (2012) Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. J Exp Med 209: 1553-1565. doi:10.1084/jem.20120910. PubMed: 22891276.
8. Rossi D, Trifonov V, Fangazio M, Bruscaggin A, Rasi S et al. (2012) The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. J Exp Med 209: 1537-1551. doi:10.1084/jem.20120904. PubMed: 22891273.
9. ISCN (2009) An International System for Human Cytogenetic Nomenclature: Recommendations of the International Standing Committee on Human Cytogenetic Nomenclature Shaffer LG, Slovak ML, Campbell LJ, editors. Basel: S.Karger.
10. Matthews C, Catherwood M, Morris TC, Alexander HD (2004) Routine analysis of IgVH mutational status in CLL patients using BIOMED-2 standardized primers and protocols. Leuk Lymphoma 45: 1899-1904. doi:10.1080/10428190410001710812. PubMed: 15223652.
11. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842. doi:10.1093/bioinformatics/btq033. PubMed: 20110278.
12. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD et al. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22: 568-576. doi: 10.1101/gr.129684.111. PubMed: 22300766.
13. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38: e164. doi:10.1093/nar/gkq603. PubMed: 20601685.
14. Consortium. GP, Durbin RM, Abecasis GR, Altshuler DL, Auton A et al. (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061-1073. doi:10.1038/nature09534. PubMed: 20981092. Available online at: doi:10.1038/nature09534 Available online at: PubMed: 20981092
15. Parker H, Rose-Zerilli MJ, Parker A, Chaplin T, Chen X et al. (2011) 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. Leukemia 25: 489-497. doi:10.1038/leu.2010.288. PubMed: 21151023.
16. Duan S, Cermak L, Pagan JK, Rossi M, Martinengo C et al. (2012) FBXO11 targets BCL6 for degradation and is inactivated in diffuse large B-cell lymphomas. Nature 481: 90-93. PubMed: 22113614.
17. Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A et al. (2011) Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature 471: 189-195. doi:10.1038/nature09730. PubMed: 21390126.
18. Salaverria I, Akasaka T, Gesk S, Szczepanowski M, Burkhardt B et al. (2012) The CBFA2T3/ACSF3 locus is recurrently involved in IGH chromosomal translocation t(14;16)(q32;q24) in pediatric B-cell lymphoma with germinal center phenotype. Genes Chromosomes Cancer 51: 338-343. doi:10.1002/gcc.21919. PubMed: 22420028.
19. Gruber TA, Larson Gedman A, Zhang J, Koss CS, Marada S, et al. (2012) An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. Cancer Cell 22: 683-697
20. Fresquet V, Robles EF, Parker A, Martinez-Useros J, Mena M et al. (2012) High-throughput sequencing analysis of the chromosome 7q32 deletion reveals IRF5 as a potential tumour suppressor in splenic marginal-zone lymphoma. Br J Haematol 158: 712-726. doi:10.1111/j.1365-2141.2012.09226.x. PubMed: 22816737.
21. Yi JM, Dhir M, Van Neste L, Downing SR, Jeschke J et al. (2011) Genomic and epigenomic integration identifies a prognostic signature in colon cancer. Clin Cancer Res 17: 1535-1545. doi: 10.1158/1078-0432.CCR-10-2509. PubMed: 21278247.
22. Vanaja DK, Ehrich M, Van den Boom D, Cheville JC, Karnes RJ et al. (2009) Hypermethylation of genes for diagnosis and risk stratification of prostate cancer. Cancer Invest 27: 549-560. doi: 10.1080/07357900802620794. PubMed: 19229700.
23. Kim JH, Jung EJ, Lee HS, Kim MA, Kim WH (2009) Comparative analysis of DNA methylation between primary and metastatic gastric carcinoma. Oncol Rep 21: 1251-1259. PubMed: 19360301.
24. Imura M, Yamashita S, Cai LY, Furuta J, Wakabayashi M et al. (2006) Methylation and expression analysis of 15 genes and three normally-methylated genes in 13 Ovarian cancer cell lines. Cancer Lett 241: 213-220. doi:10.1016/j.canlet.2005.10.010. PubMed: 16303245.
25. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J et al. (2010) Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet 42: 181-185. doi:10.1038/ng.518. PubMed: 20081860.
26. Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C et al. (2010) Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet 42: 722-726. doi:10.1038/ng.621. PubMed: 20601953.

participated in data analysis and manuscript preparation; JSC designed the study, analyzed and interpreted data, and wrote the manuscript. All authors approved the manuscript before submission.

V Visconte[1], B Przychodzen[1], Y Han[1], ST Nawrocki[2], S Thota[1], KR Kelly[3], BJ Patel[1], C Hirsch[1], AS Advani[4], HE Carraway[4], MA Sekeres[1,4], JP Maciejewski[1] and JS Carew[1,2]
[1]*Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA;*
[2]*Department of Medicine, Division of Translational and Regenerative Medicine, University of Arizona Cancer Center, Tucson, AZ, USA;*
[3]*Department of Medicine, USC Norris Comprehensive Cancer Center, Los Angeles, CA, USA and*
[4]*Leukemia Program, Department of Hematology/Oncology, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA*
E-mail: jcarew@email.arizona.edu

REFERENCES

1 Kroemer G. Autophagy: a druggable process that is deregulated in aging and human disease. *J Clin Invest* 2015; **125**: 1–4.
2 Thorburn A, Thamm DH, Gustafson DL. Autophagy and cancer therapy. *Mol Pharmacol* 2014; **85**: 830–838.
3 White E. The role for autophagy in cancer. *J Clin Invest* 2015; **125**: 42–46.
4 Park SM, Ou J, Chamberlain L, Simone TM, Yang H, Virbasius CM *et al.* U2AF35 (S34F) promotes transformation by directing aberrant ATG7 pre-mRNA 3' end formation. *Mol Cell* 2016; **62**: 479–490.
5 Visconte V, Tabarroki A, Zhang L, Parker Y, Hasrouni E, Mahfouz R *et al.* Splicing factor 3b subunit 1 (Sf3b1) haploinsufficient mice display features of low risk Myelodysplastic syndromes with ring sideroblasts. *J Hematol Oncol* 2014; **7**: 89.
6 Harada H, Harada Y. Recent advances in myelodysplastic syndromes: molecular pathogenesis and its implications for targeted therapies. *Cancer Sci* 2015; **106**: 329–336.
7 Liew E, Owen C. Familial myelodysplastic syndromes: a review of the literature. *Haematologica* 2011; **96**: 1536–1542.
8 Churpek JE, Pyrtel K, Kanchi KL, Shao J, Koboldt D, Miller CA *et al.* Genomic analysis of germ line and somatic variants in familial myelodysplasia/acute myeloid leukemia. *Blood* 2015; **126**: 2484–2490.
9 Gong C, Song E, Codogno P, Mehrpour M. The roles of BECN1 and autophagy in cancer are context dependent. *Autophagy* 2012; **8**: 1853–1855.
10 Liang C, Jung JU. Autophagy genes as tumor suppressors. *Curr Opin Cell Biol* 2010; **22**: 226–233.
11 Kim HJ, Zhong Q, Sheng ZH, Yoshimori T, Liang C, Jung JU. Beclin-1-interacting autophagy protein Atg14L targets the SNARE-associated protein Snapin to coordinate endocytic trafficking. *J Cell Sci* 2012; **125**: 4740–4750.

Supplementary Information accompanies this paper on the *Leukemia* website (http://www.nature.com/leu)

OPEN

Non-coding *NOTCH1* mutations in chronic lymphocytic leukemia; their clinical impact in the UK CLL4 trial

Leukemia (2017) **31,** 510–514; doi:10.1038/leu.2016.298

In chronic lymphocytic leukemia (CLL), 'coding' *NOTCH1* mutations were initially detected in exon 34 where they result in truncation of the C-PEST regulatory protein sequence with consequent impaired degradation of the Notch1 intracellular domain (NCID), constitutive activation of Notch signalling and increased cell survival and resistance to apoptosis.[1–3] Mutations occur in 6–10% of cases at diagnosis, with increasing prevalence in advanced disease stages, treatment-refractory disease and after transformation to Richter syndrome.[4,5] In diagnostic and clinical trial cohorts, patients with *NOTCH1* mutations exhibited reduced survival.[5,6] In 2015, Puente and colleagues identified recurrent 'non-coding' mutations clustered to the 3′-UTR of *NOTCH1* in 2% (11/506) previously untreated patients with CLL or monoclonal B-cell lymphocytosis.[7] The presence of these 3′-UTR mutations cause a novel splicing event, preferentially between a cryptic donor site located in the last exon and a newly created acceptor site in the 3′-UTR of exon 34, resulting in the removal of the PEST sequence and constitutive activation of downstream signaling.[7] Patients with non-coding *NOTCH1* mutations had similar outcomes to those with coding mutations, with shorter time to first treatment and shorter overall survival than wild-type cases.[7,8]

Given the highly variable natural history of CLL and the often-serendipitous date of initial diagnosis, we aimed to establish the clinical significance of non-coding *NOTCH1* mutations in DNA samples available from 489 patients at enrolment to the United Kingdom Leukemia Research Fund Chronic Lymphocytic Leukemia 4 (UK LRF CLL4) chemotherapy trial.[9] *NOTCH1* 3′-UTR mutations were identified by High Resolution Melt (HRM) analysis in whole genome amplified DNA (F: TGCTCGTTCAACTTCC CTTC; R: CAAGCAAGTTCTGAGAGCCA) and confirmed by Sanger sequencing of genomic DNA (F: CCTAACAGGCAGGTGATGCT; R: ATCTGGCCCCAGGTAGAAAC) The results were combined with the data pertaining to coding *NOTCH1* mutations in the same patient cohort from our previous publication.[5] Fifty-three patients with wild-type HRM traces were sequenced, and no additional non-coding mutations were identified. It was not possible to differentiate between clonal and subclonal *NOTCH1* mutations using our HRM/Sanger approach. We defined associations between the presence of *NOTCH1* coding and non-coding mutation and a comprehensive panel of clinical and biological features reported in previous CLL4 papers,[10–13] by univariate logistic regression. Kaplan–Meier, log-rank test and Cox regression analysis were used to assess the impact of *NOTCH1* status on survival using Stata, where overall (OS) and progression-free (PFS) survival were defined as time from randomization to death from any cause and to relapse needing treatment, progression or death from any cause at last follow-up, respectively.

In addition to exon 34 coding mutations observed in 47/489 (9.6%) CLL4 patients, we detected an additional 11/489 (2.2%) patients harbouring the non-coding mutations 139390152A > G ($n = 7$) and 139390145A > G ($n = 4$; Figure 1a), both previously reported to result in aberrant *NOTCH1* splicing.[7] Importantly, the non-coding variants were mutually exclusive to coding variants, constituting 19% of the total *NOTCH1* mutational burden of CLL4 cases, with 11.8% of the patients carrying either type of *NOTCH1* mutation. *NOTCH1* non-coding mutations were not identified in cases with mutations of *TP53, BIRC3, BRAF* (V660E), *MYD88* (L265P), *NFKBIE* and *RPS15* mutations, but did co-occur with *SF3B1* ($n = 2$) and *ATM* ($n = 2$) mutations (Figure 1b). Next, we evaluated the association between the *NOTCH1* mutations and the main clinico-biological characteristics in CLL (Supplementary Table S1).

Accepted article preview online 26 October 2016; advance online publication, 11 November 2016

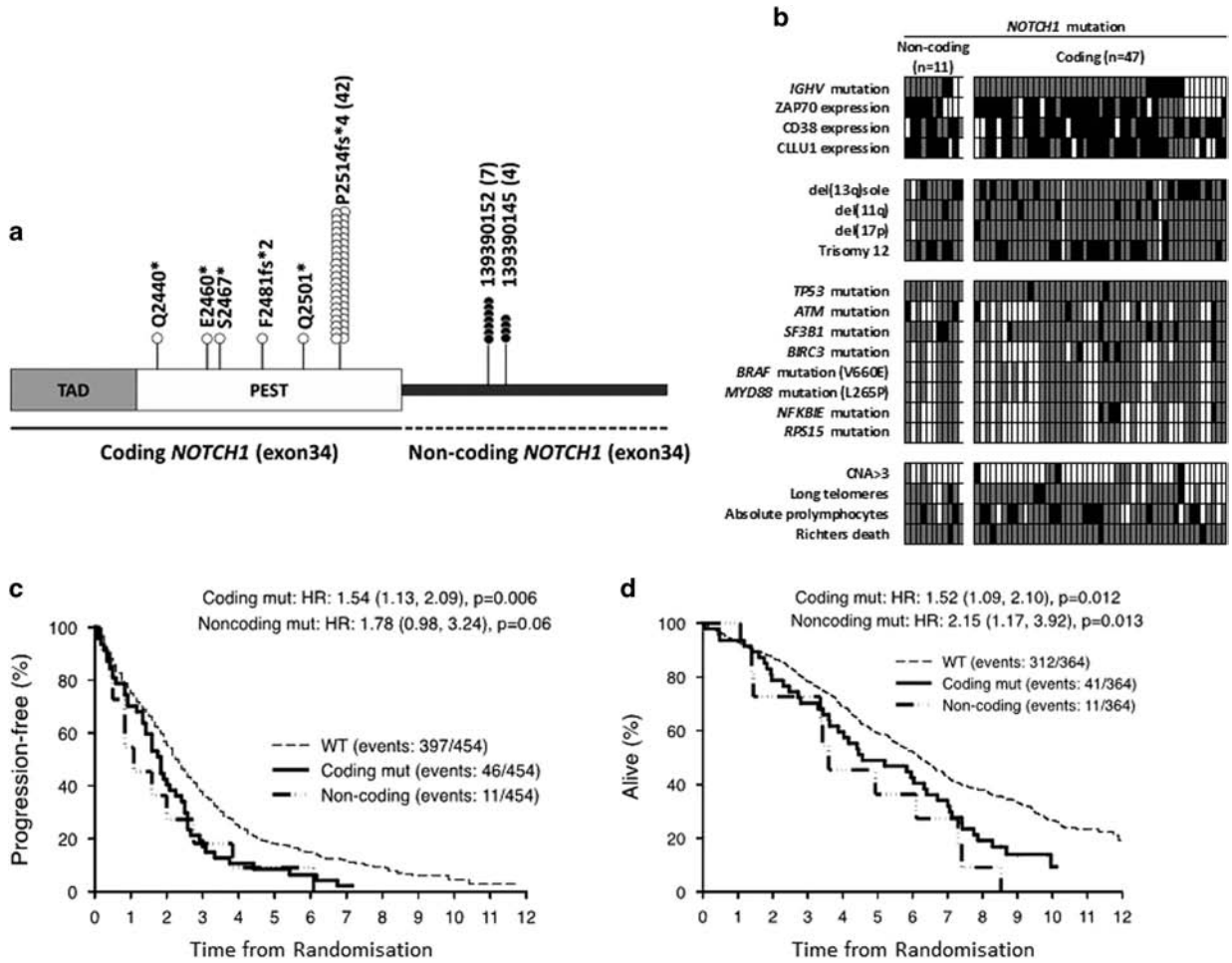


Figure 1. The genomic and clinical characteristics of *NOTCH1* non-coding and coding mutations in the LRF CLL4 trial. (**a**) The distribution of mutations in *NOTCH1*. The *NOTCH1* gene contains 34 exons and encodes a protein with a C-terminal TAD-PEST domain, which is a hotspot for mutation in CLL. Part of exon 34 and the 3′-UTR are magnified and the location of each mutation is shown; coding (white) and non-coding mutations (black) are indicated. Each dot represent a single mutation. (**b**) The mutual relationship between coding and non-coding *NOTCH1* mutations and other clinico-biological characteristics in CLL. Rows correspond to specific clinical and biological features and columns represent individual patients (only patients with a *NOTCH1* mutation are shown). Boxes colored black and grey show the presence or absence of a parameter. A white box denotes that no data were available. (**c**) and (**d**) Kaplan–Meir plots showing progression-free survival and overall survival, respectively.

As expected, when all 58 mutations were considered together, *NOTCH1* mutations were significantly more prevalent in CLL4 cases with unmutated *IGHV* genes (OR: 2.9, 95% CI: 1.4–6.2, $P=0.005$), CD38 (OR: 4.5, 95% CI: 2.3–8.7, $P<0.001$) and ZAP70 positivity (OR: 3.1, 95% CI: 1.5–6.4, $P=0.002$), high expression of CLLU1 (OR: 2.33, 95% CI: 1.2–4.4, $P=0.01$), trisomy 12 (OR: 4.0, 95% CI: 2.2–7.4, $P<0.001$) and $\geqslant 15\times10^9$/l absolute pro-lymphocytes (OR: 3.12, 95% CI: 2.0–7.9, $P<0.001$). However, for non-coding mutations on its own only the association with Trisomy 12 remained significant (OR: 5.6, 95% CI: 1.6–18.8, $P=0.006$), in spite of the limited number of cases with these mutations. Of the 364 deaths in CLL4 patients with the *NOTCH1* data, 14 (4%) were due to Richter's syndrome (RS). With non-coding *NOTCH1* mutations included, 4 of 14 (29%) Richter's deaths occurred in patients with *NOTCH1* mutation, an association that was non-significant ($P=0.062$).

In our previous CLL4 study, we confirmed the independent prognostic significance of a number of biomarkers, including coding *NOTCH1* mutations.[5] In our current study, we determined the impact of coding and non-coding mutations on overall response rate (ORR), OS and PFS. Coding and non-coding mutations, inspected together or separately, were not associated with ORR in any of the three treatment arms (data not shown). Considered separately, univariate Cox regression analysis showed that patients with *NOTCH1* non-coding or coding mutations exhibited a significantly shorter OS (median survival times: 43.2 and 54.8 months, respectively) than patients with wild-type *NOTCH1* (median: 74.6 months). Non-coding and coding *NOTCH1* mutations were also associated with reduced PFS (median survival times: 22.0 and 13.0 months respectively) compared with the wild-type *NOTCH1* (28 months, Figure 1c and d). In further support of their clinical importance, cases with non-coding *NOTCH1* mutations showed a two-fold increase in the risk of mortality when compared with wild type (HR: 2.15, 95% CI: 1.17–3.92, $P=0.013$) and an 80% increase in the risk of progression or death (HR: 1.78, 95% CI: 0.98–3.24, $P=0.05$). The impact of coding and non-coding *NOTCH1* mutations together on OS was sustained in a multivariable model where *NOTCH1* status was controlled for gender, age, stage, *IGHV* and *SF3B1* mutational status, 11q deletion, and *TP53* mutation/ deletion (adjusted HR: 1.5, 95% CI: 1.0–2.1, $P=0.04$, Table 1). On the contrary, the association between *NOTCH1* mutational status and PFS was not significant when adjusted for the other variables listed above (adjusted HR: 1.3, 95% CI: 0.9–1.9, $P=0.108$). Taken together, we show that *NOTCH1* status, based on

Table 1. Univariate and multivariate Cox proportional hazard analysis of OS and PFS in CLL4 patients

Variable	*Overall survival*										*Progression-free survival*									
	Univariate							*Multivariate*			*Univariate*							*Multivariate*		
	Total	*Events*	*Median*	*95% CI*	*HR*	*95% CI*	*P*-value	*HR*	*95% CI*	*P*-value	*Total*	*Events*	*Median*	*95% CI*	*HR*	*95% CI*	*P*-value	*HR*	*95% CI*	*P*-value
NOTCH1																				
Wild type	431	312	74.6	67.8–81.5							431	394	27.6	24.9–30.4						
Mutated	58	52	53.4	35.9–70.9	1.6	1.2–2.2	0.001	1.5	1.0–2.1	0.04	58	57	19.3	15.0–23.5	1.6	1.2–2.1	0.001	1.3	0.9–1.9	0.108
SF3B1																				
Wild type	364	250	79.1	71.8–86.3							364	326	26.5	23.1–29.9						
Mutated	73	66	54.3	47.3–61.4	1.7	1.3–2.2	< 0.001	1.5	1.1–2.1	0.014	73	73	26.5	22.4–30.7	1.3	1.0–1.7	0.033	1.3	0.9–1.8	0.071
Age					1.1	1.0–1.1	< 0.001	1.1	1.0–1.1	< 0.001					1	0.9–1.1	0.663	0.9	0.9–1.0	0.387
Sex																				
Male	366	281	70.1	61.4–78.9							366	341	25.0	21.9–28.0						
Female	129	86	79.6	66.5–93.0	0.8	0.6–1.0	0.056	0.8	0.6–1.1	0.121	129	115	29.4	25.5–33.3	0.8	0.7–1.0	0.055	0.9	0.7–1.1	0.338
Binet stage																				
A	112	76	80.6	63.4–97.7							112	104	27.2	23.8–30.7						
B/C	383	291	71.5	64.6–78.3	1.3	1.0–1.7	0.049	1.5	1.1–2.1	0.013	383	352	26.1	23.0–29.1	0.9	0.8–1.3	0.995	1.2	0.9–1.5	0.433
Del(11q)																				
Undeleted	373	267	75	67.5–82.6							373	267	75	67.4–82.6						
Deleted	92	79	57.7	42.4–73.0	1.6	1.3–2.1	< 0.001	1.4	1.1–1.9	0.023	92	79	57.7	42.4–73.0	1.5	1.2–1.9	0.001	1.7	1.3–2.2	< 0.001
IGHV *status*																				
Mutated	155	91	104.2	93.3–115.1							155	91	104.2	93.3–115.1						
Unmutated	255	216	60.6	52–8–68.4	2.2	1.7–2.8	< 0.001	1.9	1.4–2.5	< 0.001	255	216	60.6	52.8–68.4	1.9	1.6–2.4	< 0.001	1.8	1.4–2.4	< 0.001
TP53 *status*																				
Normal	431	313	75.9	69.3–82.1							431	313	75.9	69.7–82.1						
Del/Mut	32	31	26.1	4.9–47.4	3.1	2.2–4.6	< 0.001	2.5	1.5–4.1	< 0.001	32	31	26.1	4.9–47.4	2.7	1.9–3.9	< 0.001	2.2	1.3–3.5	0.002
Treatment arm																				
Chl	238	178	76.8	70.1–83.4							238	178	76.8	70.1–83.4						
FDR/FC	257	189	68	57.9–78.1	1.1	0.9–1.3	0.426	0.9	0.8–1.3	0.854	257	189	68	57.9–78.1	0.6	0.5–0.7	< 0.001	0.5	0.4–0.6	< 0.001

Abbreviations: Chl, chlorambucil; FC, fludarabine plus cyclophosphamide; FDR: fludarabine. OS multivariate, 342 cases with 252 events; 153 missing data. PFS multivariate, 342 cases with 315 events, 153 missing data.

the presence of either mutational type, is an independent risk factor for OS but not for PFS. The association between OS or PFS and the occurrence of non-coding mutations could not be estimated reliably in a multivariable analysis because of the small number of cases with such mutations in our series.

Finally, we attempted to quantify the improved discriminatory power of including non-coding *NOTCH1* mutations to coding mutations as a test to predict both the presence and absence of PFS and OS events at last follow-up using sensitivity-specificity analysis. The analysis was carried out on all 489 cases. *NOTCH1* coding mutations correctly predicted 46/454 PFS (sensitivity of 10.1%) and 43/393 (sensitivity of 10.9%) OS events (Supplementary Table S2A and S3A). As expected, the sensitivity for OS and PFS was higher when both mutational types were considered than when coding mutation alone was analysed: 13.7 versus 10.9% for OS and 12.6 versus 10.1% for PFS events (Supplementary Table S2A and S3A). This increase reflected the fact that all 11 patients with non-coding *NOTCH1* mutations exhibited an adverse OS and PFS event, resulting in 100% specificity for non-coding *NOTCH1* mutation as a test. Accuracy assesses the capability of a given biomarker to correctly predict both the presence and absence of a survival event. Coding *NOTCH1* mutations displayed 16.4 and 27.6% accuracy for correctly predicting the presence or absence of a PFS and OS, respectively. Accuracy was increased to 18.6 and 29.9% for PFS and OS, respectively, when non-coding mutations were included in this analysis. The likelihood ratio, LR+, which adjusts sensitivity for false positives and LR−, which adjusts specificity for false negatives are prevalence-independent and their ratio, LR+/LR− (diagnostic odds ratio), is an indicator of the predictive power of the biomarker. A biomarker with a higher LR+/LR− value is a better predictor of the disease outcomes. Consistent with the increased sensitivity and higher accuracy, we observe increased LR +/LR− ratios for both PFS (3.81 versus 4.88) and OS (2.43 versus 3.66) when both coding and non-coding mutations were considered together (Supplementary Table S2A and S3A). In addition, the positive predictive value (PPV), which is a measure of the proportion of true positives out of all the outcomes predicted by the biomarker, is higher when non-coding mutation was included in the test than when coding-mutation alone was used as the test biomarker (98.3 versus 97.9% for PFS and 93.1 versus 91,5% for OS, Supplementary Table S2B and S3B).

In summary, our data confirm the prognostic importance of non-coding *NOTCH1* mutations in patients requiring first-line treatment with chemotherapy as part of the UK CLL4 trial. Importantly, restricted analysis of exon 34 neglected to identify 19% of patients with pathogenic *NOTCH1* mutations in its 3′-UTR region. In addition, we show that the discriminatory power of *NOTCH1* mutation status to predict outcomes is improved with the inclusion of non-coding mutations. Taken together, our study supports the analysis of the 3′-UTR region of the *NOTCH1* gene to identify additional patients with reduced survival. Several recent studies have provided conflicting data on the clinical significance of clonal and subclonal *NOTCH1* mutations.[8,14,15] Most recently, Nadeu and colleagues demonstrated that the clonal mutations predicted for short OS, while subclonal mutations predicted for short time to first treatment.[9] It will be important to employ these same deep sequencing approaches to ascertain the clinical significance of subclonal *NOTCH1* mutations in the clinical trials setting. The UK CLL4 trial benefits from long-term clinical follow-up and the expansive-associated clinico-biological data but only assessed the utility of traditional chemotherapy. Therefore, it will be necessary to establish the impact of non-coding *NOTCH1* mutations in patients treated with chemo-immunotherapy, where they are likely to identify a significant number of additional patients destined to respond poorly to rituximab-containing treatment regimens.[6] Mutant *NOTCH1* currently represents a therapeutic target in T-ALL, with several mechanistic approaches under clinical development, including γ-secretase and metalloproteinases inhibitors, antibodies directed against the extracellular domain of Notch1 and antagonists that act by directly targeting the Notch transactivation domain. Screening for non-coding *NOTCH1* mutations identifies additional CLL patients with Notch1 activation, offering motivation for clinical trials development. Assuming these approaches are ultimately approved for the treatment of CLL, it will be critical to identify all patients that will benefit from these treatments, as there will be important clinical and cost implications. These studies will help establish a stratified and individualized approach to clinical management, including the more accurate selection of patients for targeted therapy.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all patients and clinicians who participated in the trial. This work was funded by Bloodwise (11052, 12036), the Kay Kendall Leukaemia Fund (873), Cancer Research UK (C34999/A18087, ECMC C24563/A15581), Wessex Medical Research and the Bournemouth Leukaemia Fund. The LRF CLL4 trial was funded by a core grant from Leukemia and Lymphoma Research. DC acknowledge the support by The Royal Marsden Hospital and The Institute of Cancer Research National Institute of Health Research Biomedical Research Center.

AUTHOR CONTRIBUTIONS

ML, MJJR-Z, HP, SB, JF and ZD performed the experimental work; ML, MJJR-Z, LK, AC and ME conducted the statistical analysis; DGO, ME and DC contributed patient samples and data; JCS designed the study; ML, LK, DGO and JCS wrote the paper; all the authors critically reviewed the final paper.

M Larrayoz[1], MJJ Rose-Zerilli[1], L Kadalayil[2], H Parker[1], S Blakemore[1], J Forster[1], Z Davis[3], AJ Steele[1], A Collins[2], M Else[4], D Catovsky[4], DG Oscier[3] and JC Strefford[1]

[1]Caner Genomics, Academic Unit of Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK;
[2]Genetic Epidemiology and Bioinformatics, Faculty of Medicine, University of Southampton, Southampton, UK;
[3]Department of Molecular Pathology, Royal Bournemouth Hospital, Bournemouth, UK and
[4]Division of Molecular Pathology, The Institute of Cancer Research, London, UK
E-mail: JCS@soton.ac.uk

REFERENCES

1 Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukemia. *Nature* 2011; **475**: 101–105.
2 Arruga F, Gizdic B, Serra S, Vaisitti T, Ciardullo C, Coscia M *et al.* Functional impact of NOTCH1 mutations in chronic lymphocytic leukemia. *Leukemia* 2014; **28**: 1060–1070.
3 Rosati E, Sabatini R, Rampino G, Tabilio A, Di Ianni M, Fettucciari K *et al.* Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. *Blood* 2009; **113**: 856–865.
4 Baliakas P, Hadzidimitriou A, Sutton L-A, Rossi D, Minga E, Villamor N *et al.* Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia* 2015; **29**: 329–336.
5 Oscier DG, Rose-Zerilli MJJ, Winkelmann N, Gonzalez de Castro D, Gomez B, Forster J *et al.* The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. *Blood* 2013; **120**: 4441–4443.
6 Stilgenbauer S, Schnaiter A, Paschka P, Zenz T, Rossi M, Döhner K *et al.* Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood* 2014; **123**: 3247–3254.

7 Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI *et al.* Non-coding recurrent mutations in chronic lymphocytic leukemia. *Nature* 2015; **526**: 519–524.
8 Nadeu F, Delgado J, Royo C, Baumann T, Stankovic T, Pinyol M *et al.* Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. *Blood* 2016; **127**: 2122–2130.
9 Catovsky D, Richards S, Matutes E, Oscier D, Dyer MJ, Bezares RF *et al.* Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukemia (the LRF CLL4 Trial): a randomised controlled trial. *Lancet* 2007; **370**: 230–239.
10 Oscier D, Else M, Matutes E, Morilla R, Strefford JC, Catovsky D. The morphology of CLL revisited: the clinical significance of prolymphocytes and correlations with prognostic/molecular markers in the LRF CLL4 trial. *Br J Haematol* 2016; **174**: 767–775.
11 Rose-Zerilli M, Forster J, Parker H, Parker A, Rodriguez A, Chaplin T *et al.* ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q-deleted chronic lymphocytic leukemia, data from the UK LRF CLL4 trial. *Haematologica* 2014; **99**: 736–742.
12 Gonzalez D, Martinez P, Wade R, Hockley S, Oscier D, Matutes E *et al.* Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. *J Clin Oncol* 2011; **29**: 2223–2229.
13 Strefford JC, Kadalayil L, Forster J, Rose-Zerilli MJ, Parker A, Lin TT *et al.* Telomere length predicts progression and overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. *Leukemia* 2015; **29**: 2411–2414.
14 Lionetti M, Fabris S, Cutrona G, Agnelli L, Ciardullo C, Matis S *et al.* High-throughput sequencing for the identification of NOTCH1 mutations in early stage chronic lymphocytic leukaemia: biological and clinical implications. *Br J Haematol* 2014; **165**: 629–639.
15 Rasi S, Khiabanian H, Ciardullo C, Terzi-di-Bergamo L, Monti S, Spina V *et al.* Clinical impact of small subclones harboring NOTCH1, SF3B1 or BIRC3 mutations in chronic lymphocytic leukemia. *Haematologica* 2016; **101**: e135–e138.

Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu)

OPEN

A novel t(3;13)(q13;q12) translocation fusing FLT3 with GOLGB1: toward myeloid/lymphoid neoplasms with eosinophilia and rearrangement of FLT3?

Leukemia (2017) **31,** 514–517; doi:10.1038/leu.2016.304

According to the 2016 World Health Organization classification, myeloid neoplasms with eosinophilia (MPN-Eo) are associated with genetic abnormalities of genes coding for type III tyrosine kinase (TK) receptors, mainly PDGFRA, PDGFRB and FGFR1, but also JAK2.[1] Beside these translocations, very rare FLT3 gene rearrangements have been reported, which raises the double question of its association with myeloid neoplasms and of its specific targeted therapy.[2–7]

A new t(3;13)(q13;q12) was found from a case of atypical mixed lymphoid/myeloid neoplasm. This case, diagnosed MPN-Eo, was characterized by the coexistence of bone marrow myeloproliferation with circulating hypereosinophilia and T-cell lymphoblastic lymphoma in lymph node (Supplementary Results for detailed description). The patient could not benefit from new tyrosine kinase inhibitors. Evolution was fatal in 3 months despite conventional CHOP chemotherapy (Cyclophosphamide, Hydroxydaunorubicin, Oncovin and Prednisolone).

Karyotype of tumor cells from lymph nodes and bone marrow revealed a single clonal t(3;13)(q13;q12) translocation (Figure 1a, left panel). Absence of *FGFR1* gene rearrangement was checked by fluorescence *in situ* hybridization (FISH) and RT-PCR according to methods described by others.[8] BCR-ABL gene translocation, FLT3-ITD and D835 mutation were also absent. FISH walking on both chromosomes 3 and 13 with BAC and fosmid probes showed that the breakpoint was located in a 58.6 kb region encompassing *HCLS1* and *GOLGB1* on chromosome 3 and in a 65.5 kb region containing the *FLT3* locus on chromosome 13 (Figure 1a, right panel).

FLT3 maps to band q12 of chromosome 13 and *GOLGB1* to chromosome band 3q13. We hypothesized that this translocation would lead to a fusion transcript. Since the breakpoint region covered 15 out of the 23 exons of the *GOLGB1* gene, we hypothesized that *GOLGB1* gene could be a fusion partner. *FLT3* gene was the only candidate on chromosome 13. A multiplex PCR amplified a specific product located between exons 13 and 15 of *GOLGB1* and *FLT3* respectively (Figure 1b). Direct sequencing showed that this 2000 bp PCR product was specific. The rearrangement fused exons 14 of both *GOLGB1* and *FLT3* genes. Moreover, 36 bp of intron 14 of *GOLGB1* were inserted between the two exons 14 of *GOLGB1* and *FLT3* (Figure 1c). The genomic fragment corresponding to the der(3) contains the 5′sequence of *GOLGB1* fused in frame to the 3′ sequence of *FLT3* at nucleotide 8841 which corresponds to the beginning of exon 14. Genomic DNA sequencing showed that breakpoints were within *GOLGB1* intron 14 and *FLT3* exon 14 (not shown).

This t(3;13)(q13;q12) translocation identifies GOLGB1 as a new partner of FLT3. GOLGB1 encodes for giantin, a golgin subfamily B member 1 and the largest golgi complex-associated protein (372 kD), with numerous coiled-coil regions. GOLGB1-FLT3 protein fused together the three coiled-coil GOLGB1 domains with the split kinase TK domain of FLT3, that could lead to a constitutively multimerized active protein. Alternatively, constitutive TK activation could be due to the loss of the inhibitory juxtamembrane domain of FLT3, as reported for FIP1L1-PDGFRα gene rearrangement.[9] GOLGB1 has been recently reported as a fusion partner with PDGFRB in a t(3;5)(q13;q33) translocation in a male patient with MLN-Eos.[10] PDGFRB has also been reported to be fused with another golgin subfamily member, GOLGA4.[11] The other published FLT3 partners,

Accepted article preview online 31 October 2016; advance online publication, 2 December 2016