University of Southampton

Faculty of Environmental and Life Sciences

School of Biological Sciences

Investigating Cellular Origins to Identify Peptide Vaccine Targets in two Independent Transmissible Tumours Circulating in the Tasmanian Devil (*Sarcophilus harrisii*)

DOI:

by

Rachel Siân Owen

ORCID ID: 0000-0001-6441-2213

Thesis for the degree of Doctor of Philosophy

September 2019

University of Southampton

Abstract

Faculty of Environmental and Life Sciences
School of Biological Sciences
Thesis for the degree of Doctor of Philosophy

Investigating Cellular Origins to Identify Peptide Vaccine Targets in two Independent Transmissible Tumours Circulating in the Tasmanian Devil (*Sarcophilus harrisii*)

by
Rachel Siân Owen

The Tasmanian devil (*Sarcophilus harrisii*) is under threat from two independent lineages of contagious cancer; Devil Facial Tumour 1 (DFT1) and Devil Facial Tumour 2 (DFT2), which are spread as allografts by biting and cause large tumours around the oral cavity of the animals. DFT1 is a Schwann cell derived tumour first identified in 1996 which persists in most wild devil populations, carrying a near 100% mortality and causing extreme population declines. DFT2 was identified in 2014 and is confined to a peninsula in the South-east of Tasmania, where its impact on devil populations has not yet been established. Despite indistinguishable gross phenotypes, evidence indicates that the tumours emerged independently in different animals. While DFT1 affected population declines have stabilised, DFT2 may be evolving to become a more widespread tumour, and further population declines threaten the future of the species.

This project is part of a larger study to identify candidate vaccine targets in DFT1 and DFT2. DFT1 evades immune detection by downregulating Major Histocompatibility Complex (MHC) class I, but expression can be restored with interferon gamma (IFNγ) and the presence of MHC class I on DFT1 cells increases their immunogenicity. In contrast, DFT2 cells maintain expression of MHC class I. We postulate that a vaccine against MHC class I restricted peptides could induce specific immune responses against both tumours. To identify potential tumour specific vaccine targets it is necessary to characterise and compare the peptides bound by MHC class I on tumour cells to those presented on healthy progenitor cells and other devil tissues. Currently, nothing is known about the binding properties of Tasmanian devil MHC class I molecules and the progenitor cells of DFT1 and DFT2 lack specific definition. In this thesis I aim to identify the cellular origins of DFT2 and define the peptide binding motif of a biologically significant MHC class I allele in order to identify specific vaccine targets.

Using a proteomics approach, I have demonstrated that DFT2 expresses proteins and genes specific to myelinating glial cells, indicating a similar origin to DFT1. However, DFT2 expresses lower levels of myelin associated proteins than DFT1 and shows an enrichment for developmental glial markers. These data indicate that DFT2 has emerged from an immature myelinating glial cell, whilst DFT1 emerged from a myelinating or pro-myelinating Schwann cell. Additionally, using immunohistochemical techniques I have demonstrated that the MHC class I expression profile of Schwann cells is potentially immune evasive, making these cells prone to transmissible malignant transformation. I also show that like DFT1, DFT2 upregulates MHC class I in response to IFNγ, but demonstrates a transcriptional response that is indicative of an oligodendroglial origin, and a distinct immunosuppressive gene expression profile which is largely unaffected by IFNγ.

I have generated a DFT2 cell line overexpressing a recombinant MHC class I allele which is present in the genomes of DFT1 and DFT2 and common in the wider devil population. This is, to our knowledge, the first stable transfection of DFT2 cells and these cell lines and experimental pipelines can be used for further study of the antigen presentation pathway in the Tasmanian devil and other non-model species. These data have indicated unusual properties of MHC class I peptide binding in the Tasmanian devil which were previously unknown.

Perhaps most importantly, the data generated in this thesis indicates that despite their independent origins, DFT1 and DFT2 may share neoantigens, thus these tumours may respond to a single vaccination strategy, a finding with significant scientific and economic interest for the DFT vaccination programme.

Table of Contents

Table of Tables

Table of Figures

List of Accompanying Materials

Owen, R. S., Siddle, H. V. S. (2019). Devil Facial Tumours: Towards a vaccine. *Immunological Investigations.* In press. DOI: 10.1080/08820139.2019.1624770

Research Thesis: Declaration of Authorship

Print name:	Rachel Siân Owen

Title of thesis:	Investigating cellular origins to identify peptide vaccine targets in two independent transmissible tumours circulating in the Tasmanian devil (*Sarcophilus harrisii*)

I 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.

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:

Owen, R. S., Siddle H.V.S. (2019). Devil Facial Tumours: Towards a vaccine. *Immunological investigations.* In press. DOI: 10.1080/08820139.2019.1624770.

Signature:		Date:	

Acknowledgements

It goes without saying that I could not have gotten to this point without the help and support of countless people over the past few years.

Firstly, I'd like to thank my primary supervisor Hannah Siddle, for your invaluable support at all stages of my PhD. Thank you for taking a chance on me and for pushing me to be the best scientist I can be. When I started this journey, I was terrified of public speaking and would not have believed that in little over 3 years I would be a 3MT UK National finalist, and that has only been possible thanks to your endless encouragement and belief in me. I feel like I'm finally starting to believe in myself as well! Thank you for your unwavering commitment to building my skills and my confidence, even when I make it difficult for you!

I would like to thank the rest of the Siddle lab, and members of the Skipp and Elliott labs both past and present, for providing help and support throughout my PhD. In particular I'd like to thank Annalisa Gastaldello for being a constant source of support and advice, as well as an outlet to grumble at when our experiments aren't working! I'd also like to thank Alistair Bailey and Ben Nicholas for their help with my immunopeptidome chapter, which thanks to the two of you has remarkably produced almost as much data as it did tears on my part! Additionally, I'd like to thank students Callum Daniel, Chloe Hand and Shannon Chilvers for their hard work in the lab which has helped generate some of the data presented in this thesis. I'd also like to thank my co-supervisor Paul Skipp for all your help and advice with my proteomics work, and Rachel Darley from the Elliott lab for all your help with flow cytometry and optimising my notoriously problematic transfections! I'd like to acknowledge my funders, The Leverhulme Trust, for supporting the work undertaken in this thesis. Finally, I'd like to thank our collaborators at the University of Tasmania, in particular Rodrigo Hamede and Manuel Ruiz-Aravena for giving me the opportunity to go out into the field trapping wild devils during my PhD.

I'd like to thank the handful of teachers throughout my life who saw something in me that I have only recently seen in myself, who encouraged me to work hard and pursue my interest in the biological sciences without whom I wouldn't have made it to this point. I want to particularly thank Mrs Bowdery for your unwavering support, encouragement and "tough love", even in the face of my refusal to do homework!

I can't write my acknowledgements without mentioning the people who formed the majority of my support network over the last three and a half year: the other PhD students! Friday nights at the Stags complaining about failed experiments and spending money we don't have are the highlight of my weeks and have kept me going even when it felt like nothing was going right! The Golden Summer 2018 will go down in history. Special mentions go out to my original PhD gang: Emelia, George, Katie, Charlie, Chris and Casey. It's been a pleasure to go through this crazy journey with you, thank you for your friendship and support. The table incident and Rogue December 2016 will never be

forgotten! Another special mention to Charlotte, Aleks, Ane, Monika and Evelyn. The midweek gals meals complete with too much wine and gossiping have been a highlight!

I also want to thank all of my friends from outside of the PhD bubble, without whom I probably would never have made it into the bubble in the first place. Alice, Steph, Kat and Grace: You are my sisters. Thank you for putting up with me for so many years, for always having my back and for always being my four woman hype squad whenever I've struggle to believe in myself. Molly, Kat and Shannon: Thank for the years of spontaneous adventures! Nonsense trips abroad, following tours on a whim and mass meet ups to cry at Marvel films have been tentpoles of my life over the last four years and I wouldn't have it any other way! Thank you for keeping me sane throughout my PhD! Finally, a shout out to Rowan and Kelly for your support, friendship and love over the last 7 years!

Aled, Iain, Phil, Shay and Joel: Every day your music and the people I met through it continue to shape my life. I know I've not always been easy to deal with but thank you for persevering and dragging me through my teens to become a semi-functional adult. I'm not sure I could've made it this far without you. When I'm old and grey I'll still be telling stories of weeklong trips to Sydney, rogue nights in Norwich and Frankie Sandford. Thank you for changing my life.

Thank you to my immediate and extended family for their endless, unconditional love and support in the face of all my questionable decisions over the last few years. To my dad Eric, I've been a nightmare and I know you'd rather I didn't shave my head and get tattoos, but I like to think this thesis goes some way towards an apology for that! Thank you to you and Lesley for supporting me at home as I've written up my thesis and for always encouraging me to be the best version of myself. Glyn, thanks for always keeping me grounded and for your love and encouragement. Special mention to the Rogan family for all your love over the last four years. I also want to thank Yvette for supporting me, in particular through the PhD rejections I fielded, and always encouraging me to keep going.

Finally, I want to thank my mum, Patricia Owen. Every single decision and action I've taken in my life has been for you, and this is no different. Thank you for raising me to be a strong, intelligent and confident woman. You gave me the tools to forge the life I wanted for myself, and I hope you're proud of what I've done with them. I think you would be.

This thesis marks the end of an almost four year journey that has seen me change significantly for the better, and that's thanks to all of you.

As they say: Part of the journey is the end. And everything is going to work out exactly the way it's supposed to.

Definitions and Abbreviations

Single letter amino acid codons and colour codes for physical properties:

Amino acid	Single letter codon	Category
Alanine	A	Hydrophobic
Cysteine	C	Polar neutral
Aspartate	D	Charged - Acidic
Glutamate	E	Charged - Acidic
Phenylalanine	F	Hydrophobic
Glycine	G	Hydrophobic
Histidine	H	Charged - Basic
Isoleucine	I	Hydrophobic
Lysine	K	Charged - Basic
Leucine	L	Hydrophobic
Methionine	M	Amphipathic
Asparagine	N	Polar neutral
Proline	P	Hydrophobic
Glutamine	Q	Polar neutral
Arginine	R	Charged – Basic
Serine	S	Polar neutral
Threonine	T	Polar neutral
Valine	V	Hydrophobic
Tryptophan	W	Amphipathic
Tyrosine	Y	Amphipathic

1ry/2ry – Primary/Secondary

8/9/10mer – Peptide of 8/9/10 amino acids in length

Ab – Antibody

ACTA2 – Actin Alpha 2, Smooth Muscle

AEBSF – 4-benzenesulfonyl fluoride hydrochloride

APC – Antigen presenting cell

APS – Ammonium persulfate

ATP – Adenosine triphosphate

ATP5F1A - ATP synthase F1 subunit alpha

ATP5F1B - ATP synthase F1 subunit beta

β_2-m – Beta-2-microglobulin

BCR-ABL – Breakpoint cluster region protein - Abelson murine leukaemia viral oncogene homolog 1 fusion (Philadelphia chromosome)

bp – Base pairs

BSA – Bovine serum albumin

C – Control

CD3 – Cluster of differentiation 3

CD8 – Cluster of differentiation 8

cDNA – Complementary DNA

CDR – Complementarity determining region

CHO – Chinese hamster ovary

c-Jun – Jun proto-oncogene, AP-1 transcription factor subunit

CKB – Brain-type creatine kinase

Class I/II – MHC class I/MHC class II

Class Ia/Ib = Classical MHC class I/Non-classical MHC class I

CMV - Cytomegalovirus

CNP – 2',3'-Cyclic Nucleotide 3' Phosphodiesterase

CNS – Central nervous system

COL1A1 – Collagen, type I, alpha 1

cont. – Continued

CSPG4 – Chondroitin sulfate proteoglycan 4 (Neural/glial antigen 2)

CTVT – Canine transmissible venereal tumour

DAPI – 4’,6-diamidino-2-phenylindole

dbCAMP – Dibutyryl-Cyclic adenosine monophosphate

ddH_2O – Double distilled water

DFT1 – Devil facial tumour 1

DFT2 – Devil facial tumour 2

DMEM – Dulbecco's Modified Eagle Medium

DMEM/F-12 – Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12

DMP - Dimethyl pimelimidate

DNA – Deoxyribonucleic acid

DNase – Deoxyribonuclease

dNTPs – Deoxyribonucleotide triphosphates

dPBS – Dulbecco's phosphate buffered saline

DPYSL2 – Dihydropyrimidinase Like 2

DPYSL3 – Dihydropyrimidinase Like 3

E. coli – *Escherichia coli*

EDTA – Ethylenediaminetetraacetic acid

EGR2 – Early growth response protein 2

ELISA – Enzyme-linked immunosorbent assay

ER – Endoplasmic reticulum

ERAP – Endoplasmic Reticulum Aminopeptidase

ERp57 – Protein disulfide-isomerase A3

FABP7 – Fatty acid binding protein 7

FACS – Fluorescence activated cell sorting

FAP – Fibroblast Activation Protein Alpha

FFPE – Formalin fixed, paraffin embedded

Fibs – Fibroblasts

Fig. – Figure

fmol – Femtomole

Funct. – Functional

G418 – Geneticin

GFAP – Glial fibrillary acidic protein

GFP – Green fluorescent protein

GO – Gene Ontology

GO:BP – Gene Ontology Consortium Biological Process

GO:CC – Gene Ontology Consortium Biological Process

H&E – Haematoxylin and Eosin histological stain

H_2O – Water

HCl – Hydrogen chloride

HF – High fidelity

HIV – Human immunodeficiency virus

HLA – Human leukocyte antigen

HPA – Human Protein Atlas

HPLC – High performance liquid chromatography

HPV – Human papilloma virus

ICC – Immunocytochemistry

IFNγ – Interferon gamma

IGEPAL-630 – Octylphenoxy poly(ethyleneoxy)ethanol

IHC – Immunohistochemistry

IL-1 – Interleukin 1

IL-2 – Interleukin 2

IL-10 – Interleukin 10

IL-12 – Interleukin 12

IL1R1 – Interleukin 1 receptor type 1

i-LSC – Immature-like Schwann cell

IP – Immunoprecipitation

iSC – Immature Schwann cell

IUCN – International Union for Conservation of Nature

kb – Kilobase pairs

KCl – Potassium chloride

kDa – Kilodaltons

kg – Kilograms

KLH – Keyhole limpet haemocyanin

LC-MS/MS – Liquid Chromatography-Tandem mass spectrometry

L1CAM – L1 cell adhesion molecule

LB – Lysogeny broth

M – Molar

MAG – Myelin associated glycoprotein

mAU – Milli Absorbance Units

MBP – Myelin basic protein

MC-38 – Murine colon adenocarcinoma derived cell line

MFI – Mean fluorescence intensity

Mg – Milligrams

MHC – Major histocompatibility complex

MHC-I – MHC class I

MHC-I+ – MHC class I positive

MHC-I- – MHC class I negative

mL – Millilitres

mM – Millimolar

MOG – Myelin oligodendrocyte glycoprotein

MPNST – Malignant peripheral nerve sheath tumour

MPZ – Myelin protein zero

Min – Minutes

miRNA – Micro RNA

mRNA – Messenger RNA

MSC – Myelinating Schwann cell

Na_2HPO_4 – Sodium phosphate dibasic

NaCl – Sodium chloride

NAOH – Sodium hydroxide

NCAM – Neural cell adhesion molecule

NDC – No cDNA control

NES – Nestin

NF1 – Neurofibromin 1

Ng – Nanograms

NG2 – Neural/glial antigen 2 proteoglycan

NGFR – Low affinity nerve growth factor receptor

NK – Natural killer

nM – Nanomolar

NMSC – Non-myelinating Schwann cell

NNP – Narawntapu National Park

NOTCH1 – Notch homolog 1, translocation-associated

NOTCH3 – Neurogenic locus notch homolog protein 3

Oct-6 – POU domain, class 3, transcription factor

OLIG2 – Oligodendrocyte transcription factor 2

OPC – Oligodendrocyte precursor cell

PAX3 – Paired box 3

PBS – Phosphate buffer saline

PCR – Polymerase chain reaction

PD-1 – Programmed death protein 1

PDGFRA – Platelet derived growth factor receptor alpha

PDI – Protein disulfide isomerase

PD-L1 – Programmed death-ligand 1

Pen/strep – Penicillin/Streptomycin solution

PFA – Paraformaldehyde

PI – Propidium iodide

PLC – Peptide loading complex

PLP – Proteolipid protein

Pmol – Picomoles

PMP22 – Peripheral myelin protein 22

PN – Peripheral nerve

PNS – Peripheral nervous system

PRX – Periaxin

QKI – Quaking I

RCF – Relative centrifugal force

REAC – Reactome

RNA – Ribonucleic acid

RNase – Ribonuclease

RNA-Seq – RNA Sequencing

ROBO – Roundabout receptor

RPL13A – Ribosomal protein L13a

RPM – Revolutions per minute

RPMI-1640 – Roswell Park Memorial Institute 1640 Medium

RT – Reverse transcriptase

RT-PCR – Reverse transcription polymerase chain reaction

S100 – S100 protein

Saha-UA/B/C – Tasmanian devil classical MHC class I

SCP – Schwann cell precursor

SDS – Sodium dodecyl sulfate

SDS-PAGE – Sodium dodecyl sulfate, polyacrylamide gel electrophoresis

SEM – Standard error of mean

SH – Stony Head

SN – Supernatant

SOX2 – Sex determining region-Y -box2

SOX10 – Sex determining region-Y -box10

STAT3 – Signal transducer and activator of transcription 3

T1/2/3 – Tumour 1/2/3

TAA – Tumour associated antigen

TAE – Tris, acetic acid and EDTA buffer

TAP – Transporter associated with antigen processing

TAPBPR – Tapasin-related protein TAPBRP

TBS – Tris buffered saline

TBST – Tris buffered saline with Tween20

TCR – T-cell receptor

TD – Tasmanian devil

TEAB – Tetraethylammonium bromide

TEMED – Tetramethylethylenediamine

TeO – Triethanolamine

TFA – Trifluoroacetic acid

TGFB1- Transforming growth factor beta 1

TGFB2 – Transforming growth factor beta 2

THY1 – Thy-1 cell surface antigen

TLR – Toll-like receptor

TRAMP-C1 – Murine prostate adenocarcinoma derived cell line

Tris – Trisaminomethane

TrisCl – Trisaminomethane chloride

TSA – Tumour specific antigen

TUBB3 – Tubulin Beta 3 Class III

U – Enzyme units

µg – Micrograms

µL – Microlitres

µM – Micromolar

UNT - Untransfected

UV – Ultraviolet

VDJ recombination – Variable diversity joining recombination

VEGF – Vascular endothelial growth factor

Chapter 1 Review of the Literature

1.1 Cancer and the immune system

1.1.1 Contagious cancers

With the exclusion of cancers with a known infectious causative agent, such as HPV-associated cervical cancer (Schiffman *et al*. 2016) and Epstein-Barr virus associated Burkitt's Lymphoma (Messick *et al*. 2019), cancer in humans is not a contagious disease. Seminal studies on tumour transplantation between hosts over a century ago first hypothesised that host mice possess genetic traits which would prevent tumour growth following transplantation unless matched with the traits of the tumour (Little and Tyzzer 1916). Just as with organ and graft rejection, cancer cells transplanted between individuals are not immunologically compatible, causing the recipient to mount an immune response against the cancer (Focosi *et al*. 2011). Transmission of cancer cells between individuals requires a high degree of immunological and genetic similarity between patients, a weakened immune system in the receiving patient, or mutations which result in immune escape.

There is no evidence of direct continual transmission of any human cancer, however there are extremely rare cases where a single transmission of cancer cells can occur between hosts, such as transplacental transmission of some cancers from mother to foetus during pregnancy (Yagasaki *et al*. 2011; Potter and Schoeneman 1970), vice-versa in the case of the rare but aggressive tumour choriocarcinoma (Kawamura *et al*. 2013), and following organ and blood transplantation from an affected donor into a cancer-free patient (Xiong *et al*. 2017). In one notable case, a man infected with HIV died after a cancer from a parasitic tapeworm spread to a number of his organs (Muehlenbachs *et al*. 2015). These rare cases involve either host tissue immune privilege or immunosuppression in the recipient (Fassati and Mitchison 2010), indicating a requirement for exceptional or unusual circumstances not normally seen in nature. Despite the immune boundaries blocking direct allograft transmission of cancer cells between individuals, there exists in nature eight cancer lineages which can evade the host immune system and are able to propagate throughout their respective species. These cancers are Canine Transmissible Venereal Tumour (CTVT) in dogs (Murgia *et al*. 2006), five lineages of disseminated neoplasia in four species of marine bivalve (Metzger *et al*. 2016; Metzger *et al*. 2015) and two distinct facial tumours (Devil Facial Tumour 1 (DFT1) and Devil Facial Tumour 2 (DFT2)) in Tasmanian devils (Pye *et al*. 2016b; Pearse and Swift, 2006).

CTVT is a transmissible cancer that was first described in 1810 by a vet in London and subsequently identified in dogs worldwide (Strakova and Murchison 2015; Blaine 1810). CTVT causes tumour growth predominantly around the external genitalia of domestic dogs, which spreads by allograft transmission of cancer cells during coitus and social sniffing and licking behaviour (Murgia *et al.*

2006). CTVT has been estimated as having first arisen in an early, genetically isolated dog population between 10000 and 13000 years ago, making it the world's oldest known cancer (Murchison *et al*. 2014; Murgia *et al*. 2006). CTVT is rarely fatal to host dogs, and often demonstrates spontaneous tumour regression (Frampton *et al.* 2018; Mukaratirwa *et al.* 2004; Pérez *et al.* 1998). Full genome sequencing of CTVT from two different dogs has indicated that CTVT has gained almost two million somatic mutations since it first appeared, compared with one to five thousand somatic mutations in most human cancers, reflecting the advanced age of the CTVT tumour (Murchison *et al*. 2014). This strikingly high number of somatic mutations makes determining the origin of CTVT difficult as little remains of the founder genome, however there is evidence to suggest that CTVT is derived from a myeloid cell, and shows evidence of a macrophage progenitor cell (Albanese *et al*. 2002; Sandusky *et al*. 1987).

Disseminated neoplasia in marine bivalves was first documented in the 1970s (Brown *et al*. 1978), but was only recently discovered to be transmissible and several distinct lineages have been identified (Metzger *et al.* 2016; Metzger *et al.* 2015). Interestingly, at least one of these cancers shows historical evidence of cross species transmission, while another shows the ability to propagate in another species following experimental inoculation (Metzger *et al*. 2016; Metzger *et al*. 2015). Though bivalves, like all invertebrates, lack an adaptive immune system, they possess a haemolymph containing immune factors and haemocytes, the invertebrate immune cells (Gestal *et al*. 2008). There is also evidence that invertebrates have systems of allorecognition (Bouallegui 2019; Furuta and Yamaguchi 2011; Nicotra *et al*. 2009) and how these neoplasia spread between hosts is still not fully understood. These neoplasia originate in the haemolymph, and are likely to be of haemocyte origin (Metzger *et al*. 2016).

The contagious cancer DFT1 (originally termed DFTD) was initially documented in a wild Tasmanian devil in 1996 (Pearse and Swift 2006). DFT1 is transmitted as an allograft between devils during normal social biting behaviour, and causes large, ulcerated and necrotic tumours predominantly around the face and mouth of the animals (Loh *et al.* 2006; Pearse and Swift 2006). DFT1 has a mortality rate of nearly 100%, causing death by organ failure or starvation within six to twelve months of disease development (Hamede *et al*. 2011), though some rare cases of remission have been observed (Pye *et al*. 2016a). Transcriptomic analysis and the presence of the Schwann cell specific marker periaxin on tumour cells shows that DFT1 evolved from a Schwann cell, although the differentiation state of the original Schwann cell remains unclear (Tovar *et al*. 2011; Murchison *et al*. 2010).

In 2014, a second transmissible tumour, designated DFT2, was identified in the Tasmanian devil (Pye *et al*. 2016b). DFT2 is grossly indistinguishable from DFT1, causing the same necrotic tumours around the face. DFT2 tumours do not stain for the DFT1 specific marker periaxin, and the cells are morphologically and genetically distinct to DFT1 (Stammnitz *et al*. 2018). Thus, the two cancers emerged and have evolved separately. Whilst the mortality rate is not known, all devils captured with the disease thus far have been euthanised due to the severity of the tumour. DFT2 is currently

geographically restricted to the Channel Peninsula region of South-East Tasmania (Pye *et al*. 2016b) but its prevalence and disease front are increasing (James *et al.* 2019). There is some evidence that DFT2 has emerged from the neuroectoderm (Stammnitz *et al*. 2018) though its progenitor cell type is unknown.

The eight cancers described here all share the remarkable ability to evade the anti-tumour and anti-graft immune responses of their host species and propagate throughout genetically diverse populations, something which has not been seen in human cancers. The existence of contagious cancers challenge our understanding of tumour and transplant immunology, thus presenting an interesting model for tumour immune evasion and metastasis in a naturally occurring system. Understanding the mechanisms behind the immune escape and consequent spread of these tumour lineages may provide insights into the fundamental immunology of human cancer. The recent emergence of DFT2 in particular provides a unique opportunity to examine the early stages of a transmissible tumour and track the evolution of immune evasion strategies in a cancer under extreme selective pressure. Additionally, the emergence of two transmissible tumours in the Tasmanian devil in a short time frame raises concerns for the future survival of an already endangered species, and presents a significant conservation problem. Understanding the immunological processes that have lead to the emergence of DFT1 and DFT2 is crucial to developing effective disease management strategies to support the remaining wild Tasmanian devil population.

1.1.2 CD8+ T-cell mediated tumour immunosurveillance and graft rejection

Under normal circumstances, the vertebrate immune system can identify and remove some tumour and precancerous cells in the body in a process called immunosurveillance (Baxevanis 2019). The theory of tumour immunosurveillance was first informed by the observation that individuals with suppressed or otherwise compromised immune systems were at significantly increased risk of spontaneous tumour development (Swann and Smyth 2007; Birkeland *et al*. 1995). The predominant effector cell involved in immunosurveillance and clearance of developing tumours is the CD8+ T-cell, although other immune cells including CD4+ T-cells and dendritic cells play a role (Sedlacek *et al*. 2019; Ostroumov *et al*. 2018; Gao *et al*. 2005). One mechanism of CD8+ T-cell mediated immunosurveillance involves interactions with Major Histocompatibility Complex (MHC) class I molecules on the surface of tumour cells. The T-cell receptor (TCR) binds to MHC class I-peptide complexes on the cell surface and can identify foreign MHC class I molecules and foreign or mutated bound peptides. Cells presenting mutated tumour-specific antigens are recognised as abnormal, and can be destroyed by activated CD8+ cytotoxic T-cells (Vitale *et al*. 2019; Rogel *et al*. 2017). The immune system can also detect cells lacking MHC class I molecules, and these cells are eliminated by natural killer cells (NKs) (Malmberg *et al*. 2017).

Despite these internal mechanisms to protect against neoplastic transformation, tumours can still form in the presence of a functional immune system. Indeed, immune evasion is considered a hallmark of cancer (Hanahan and Weinberg 2011). The ability of a tumour to evade the immune system is essential to its success, and is often due to specific mutations acquired by the cancer to evade detection and destruction by T-cells (Garrido *et al*. 2017). One way cancer cells can do this is by altering the MHC class I antigen presentation pathway, or the MHC class I molecule itself, preventing the presentation or detection of abnormal peptides on the surface of tumour cell. There is substantial evidence that reduction in MHC class I functionality in cancer cells is not only common, but results in aggressive tumour growth, reduced host immune response to the tumour and a poor prognosis (Imai *et al*. 2017; Simpson *et al*. 2010; Watson *et al*. 2006; Korkolopoulou *et al*. 1996). Tumour cells may also modulate their microenvironment with immunosuppressive factors (many such factors exist, with VEGF and IL-10 being just two examples) to avoid detection and destruction by the immune system (Joyce and Fearon 2015). Another major mechanism by which tumour cells can evade immune destruction is through the exploitation of regulatory T-cell (T-reg) function. T-regs are a subset of CD4+ T-cells with a predominantly immunosuppressive function which have a key role in maintaining peripheral immune tolerance and preventing autoimmunity (Sakaguchi *et al.* 2008). High levels of T-regs are often identified in the tumour microenvironment in human cancers, correlating with a poor prognosis, and recruitment of T-regs has been demonstrated to be an early event in the tumourigenesis of some tumour types (Fu *et al.* 2018; Shabaneh *et al.* 2018; Mougiakakos *et al.* 2010). T-regs recruited to the tumour microenvironment can effectively suppress the CD8+ T-cell mediated anti-tumour response and induce immune tolerance of tumours (Najafi *et al.* 2019). Additionally, there is evidence that inhibition of T-reg function or depletion of T-regs from

the tumour microenvironment is an effective emerging cancer treatment, confirming the role of these cells in reducing tumour immunosurveillance (Ou *et al.* 2018; Wang *et al.* 2018). Cancer cells are constantly under immune surveillance and must overcome significant pressure from several arms of the immune system (Figure 1.1), resulting in the evolution of neoplastic cells which can escape immune system mediated destruction and form a tumour.

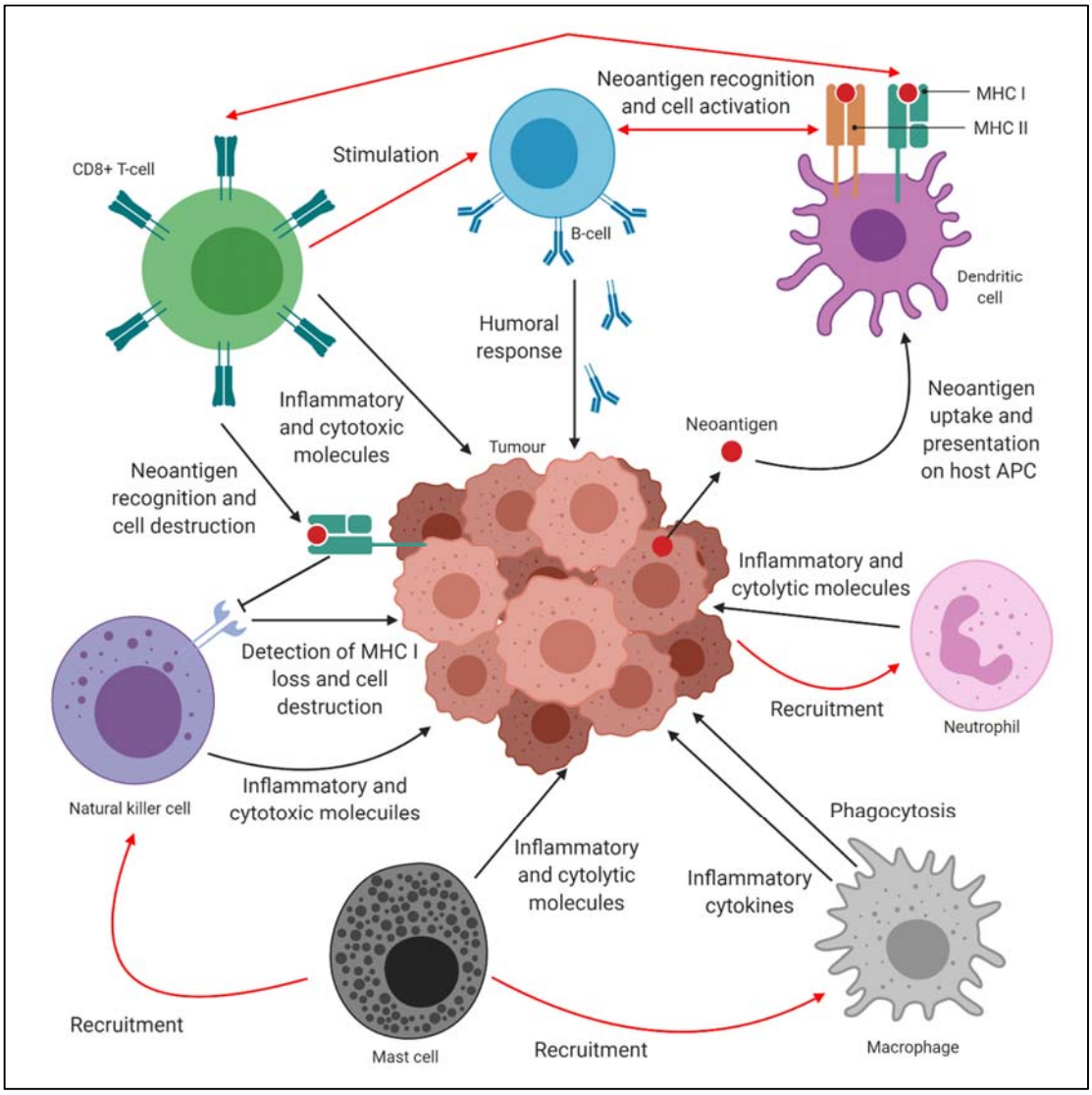


Figure 1.1: Tumour immunosurveillance is a complex process involving several immune system cell types.
Mutated tumour neoantigens can be detected by CD8+ T-cells when presented by MHC class I (MHC I) on tumour cells, or following uptake and presentation of neoantigens by dendritic cells on MHC class I and MHC class II (MHC II). Following tumour emergence and detection, several cell types are recruited and release cytotoxic, cytolytic and inflammatory molecules into the tumour microenvironment to destroy neoplastic cells. In order to survive and grow, a tumour must evolve mechanisms of immune escape. Adapted from (Chimal-Ramírez *et al.* 2013). Created with BioRender.com.

Graft rejection is a process by which transplanted organs, or in the case of contagious cancers, transplanted cells, are identified as foreign by the host immune system and subsequently destroyed. Graft rejection can be hyperacute, acute or chronic. Hyperacute rejection occurs immediately following transplantation and is induced by the presence of anti-donor alloantibodies in the host prior to the transplant being received, primarily due to mismatched MHC or ABO blood groups (Moreau *et al*. 2013). Acute rejection occurs shortly after grafting over a period of weeks to months, and is induced in a similar manner to hyperacute rejection, with alloreactive T-cells and B-cells recognising foreign antigens on the graft and inducing an immune response through cell-mediated or humoral mechanisms, which can be largely prevented through the use of immunosuppressive drugs post-transplant (Mengel *et al.* 2012; Fox and Chowdhury 2004). Chronic graft rejection is the most common type of graft rejection in human medicine (Moreau *et al.* 2013) and is characterised by the formation of tertiary lymphoid like structures within the transplanted organs (Cheng *et al.* 2011) and the presence of a primarily Th17 autoimmune response (Deteix *et al.* 2010). Chronic rejection appears to be induced by a predominantly humoral immune response rather than cell mediated (Sarma *et al.* 2012; Fukami *et al.* 2009).

The main cause of graft rejection is the recognition of allogeneic MHC class I and class II molecules and their associated antigens on the cell surface by reactive host T-cells or alloantibodies, resulting in a cytotoxic response from the host immune system against the graft. Foreign MHC molecules from donor cells can initiate a rapid immune response from recipient CD4+ and CD8+ T-cells through several mechanisms. First, direct recognition of foreign MHC molecules and bound antigens present on the grafted cells can be directly recognised by reactive host CD8+ T-cells and alloantibodies which recognise foreign MHC molecules and directly initiate a cytotoxic response against the transplanted cells (Lin and Gill 2016). Additionally, antigen presenting cells (APCs) present in the donor tissue are able to migrate to the host lymph node, where they present donor MHC and peptide to mature and developing host T-cells, activating reactive T-cells to target the graft tissue (Lakkis *et al*. 2000). Host APCs may also take up graft derived peptides through endocytosis (Harper *et al*. 2015) and present these on their cell surface either on MHC class II or cross-presented on their MHC class I (Figure 9.2) (Cruz *et al.* 2017; Ochando and Braza 2017; Steinman and Swanson 1995). Foreign peptides presented on host APC MHC class II will activate naïve CD4+ T-cells resulting in activation of B-cells and antibody production, or activation of naïve CD8+ T-cells against the graft tissue depending on the cytokine environment (Sokolovska *et al.* 2007; Mailliard *et al.* 2002). Foreign antigens which are cross-presented onto host APC MHC class I will directly activate host CD8+ T-cells against the grafted tissue (Sánchez-Paulete *et al.* 2017). Additionally, host B-cells can initiate anti-graft immune responses, both directly within the grafted tissue by cytokine production which activates surrounding immune cells to induce an anti-graft immune response, or by the production of alloantibodies following activation by APCs which can recognise antigens on donor cells (Haririan *et al*. 2009; Thaunat *et al*. 2005). Combined, these mechanisms of foreign cell rejection form a robust immune response which should prevent the continuous transmission of

tumours between genetically distinct hosts. This issue can be at least partially addressed by "matching" MHC alleles between donor and host, preventing the non-self MHC allele response and ensuring a similar antigen repertoire on the donor cell surface (Lakkis and Lechler 2013). However, even in MHC matched transplants, T-cells are also able to recognise peptide antigens derived from polymorphic proteins, known as minor histocompatibility antigens, and can still cause graft rejection between MHC genotype identical hosts in the absence of immunosuppressive drugs (Spierings 2014; Sebille *et al*. 2001).

Despite the mechanisms designed to prevent both tumour formation and allograft transmission, contagious cancers have evolved the remarkable ability to evade both. These cancers avoid the immune barriers to tumourigenesis and overcome both major and minor histocompatibility recognition to grow in MHC discordant hosts. One possible mechanism behind this immune evasion may involve modulating the expression of MHC molecules. Whilst both MHC class I and II molecules have a role in immunosurveillance and graft rejection, MHC class II are normally only expressed on professional antigen presenting cells such as dendritic cells (Zhao *et al*. 2017), whilst MHC class I molecules are expressed on all nucleated cells. Evidence indicates that MHC class II is more important in graft rejection within 6 months of transplantation, whilst MHC class I is more important in long term graft survival (Ayala Garcia *et al*. 2012), suggesting that MHC class I may be more relevant in terms of immune evasion in transmissible tumours that have persisted for a number of years in the population. Indeed, MHC class I has been shown to be downregulated in both CTVT and DFT1 (Siddle *et al*. 2013; Murgia *et al*. 2006), indicating that modulation of MHC class I expression plays a role in the development and spread of these remarkable tumours.

1.2 Major Histocompatibility Complex Class I

MHC molecules present fragments of proteins on the cell surface which can be used by the immune system to determine whether a cell is foreign, mutated or infected. There are two classes involved in antigen presentation, MHC class I and MHC class II. MHC class II molecules are primarily involved in the presentation of peptides derived from exogenous proteins following lysosomal degradation, such as bacterial proteins, whilst MHC class I molecules present peptides from endogenous proteins degraded by the ubiquitin-proteasome pathway (Leone *et al*. 2013). MHC class I molecules on the cell surface are interrogated by the CD8+ T-cell TCR to determine whether they are "non-self" or mutated. Cells presenting tumour-associated antigens, peptides of viral and "non-self" origin, and "non-self" MHC class I molecules are flagged for destruction and killed (Haworth *et al*. 2015). MHC class I are subcategorised as "classical" or "non-classical", which are structurally related to each other but differ in polymorphism, expression and function.

1.2.1 Classical MHC class I

Classical MHC class I molecules are ubiquitously expressed by all nucleated cells and are involved in classical antigen presentation, immune responses to infections and grafts and tumour immunosurveillance. Classical MHC class I are both polygenic and polymorphic, and humans possess three classical class I genes (HLA-A, HLA-B and HLA-C) each with hundreds to thousands of alleles within the population (Albrecht *et al*. 2017), though numbers of genes and alleles vary between species (Kelley *et al*. 2005). Human MHC expression is co-dominant, and human cells can express up to six different classical class I alleles (HLA-A, B and C from each parent). This high level of polymorphism allows MHC complexes to bind and present a wide range of peptides, and also gives sensitive and specific recognition of self vs foreign antigens (Janeway *et al*. 2001). Polymorphism of MHC class I has been associated with reduced viral immune escape at a population level (Schmid *et al.* 2008) and reduced MHC class I polymorphism has been linked with reduced ability of a species to respond to infectious disease (Maibach and Vigilant 2019; Ujvari and Belov 2011).

1.2.2 Non-classical MHC class I

Non-classical class I (Human HLA-E, HLA-G and HLA-F) are structurally related to classical class I molecules, but with several key genetic and functional differences. Non-classical class I molecules are less polymorphic than their classical counterparts (Hannoun *et al*. 2018; Sullivan *et al*. 2006) and show more variable and often restricted tissue expression than the classical molecules (Howcroft and Singer 2003; Braud *et al*. 1999). Some non-classical molecules are subject to recognition by the TCR, including HLA-E and HLA-G, but this interaction is significantly more restricted than seen between classical MHC class I and the TCR, and may result in unusual responses such as HLA-E induced T-reg expansion (Anderson *et al.* 2019; da Silva *et al.* 2018; Kochan *et al.* 2013; Jiang *et al.* 2010; Allen and Hogan 2001). One notable function of a non-classical MHC class I (HLA-G) is modulating the immune response in the placenta and invading trophoblasts of the early embryo during pregnancy to prevent immune rejection of the foetus by the mother (Hunt *et al*. 2000; Slukvin *et al*. 2000). HLA-E has an important role in regulating the function of NK cells, and there is evidence that a lack of these non-classical molecules results in a reduced NK cell response to cells with downregulated classical molecules, suggesting a wider role for non-classical molecules in innate immunity (Smyth *et al*. 2013). However, there is also evidence that human cancers can overexpress non-classical molecules, which confer the tumour cells with an immunosuppressive phenotype, as the non-classical molecules appear to reduce the cytolytic activity of a number of immune effector cells (Kochan *et al*. 2013).

1.2.3 The structure of MHC class I

The MHC class I complex consists of three subunits, the heavy chain, Beta-2-microglobulin (β_2-m) and a short (8-12 amino acids) peptide (Wieczorek *et al*. 2017). MHC class I molecules can only form a stable complex in the presence of all three components. In the absence of either the bound peptide or the β_2-m subunit, the heavy chain is unable to fold effectively and is instead exported empty to the cytoplasm and degraded by the proteasome (Albring *et al*. 2004). β_2-m is a single immunoglobulin domain which is ubiquitously expressed in all tissues (Le *et al*. 2017). The heavy chain is comprised of 3 extracellular domains, α1, α2 and α3 and the entire structure is anchored to the membrane by a transmembrane region with a cytoplasmic tail (Figure 1.1). Variation between class I alleles occurs primarily in the α1 and α2 domains whilst the α3 domain is relatively conserved across genes and alleles. It is the variation within the α1 and α2 domains of the MHC class I molecule, particularly within the peptide binding groove, which infers the specificity of peptide binding for each class I allele. As shown in Figure 1.2, there are several important residues within the binding groove whose side chains interact with the side chains of peptide residues, forming 'pockets' within the binding groove. The shape and chemical properties of these pockets determine the amino acid sequences that fit and stably bind the class I molecule. The corresponding residues on the peptide which are most essential for binding are known as "anchor residues", and the majority of class I molecules have "preferred" anchor residues that determine binding, referred to as the peptide binding motif (Meydan *et al*. 2013).

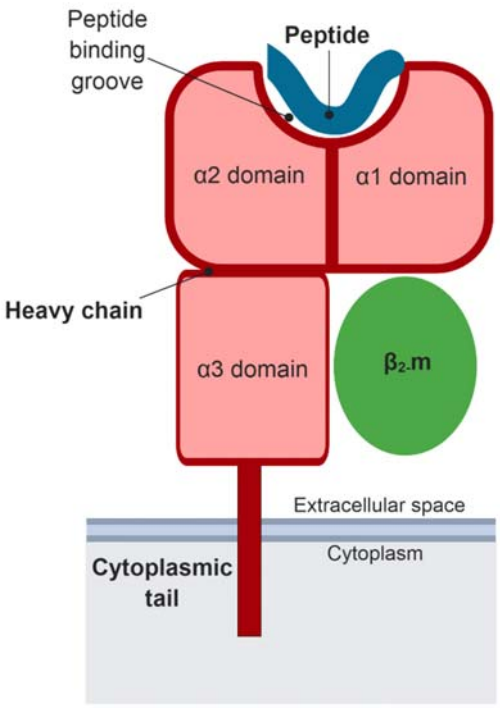


Figure 1.2: The basic structure of MHC class I.

MHC class I consists of a large heavy chain stabilised by β_2-m and a peptide. Allelic diversity is predominantly found in the α1 and α2 regions which contact the peptide, designated the peptide binding groove. Adapted from Parham (2009). Created with BioRender.com.

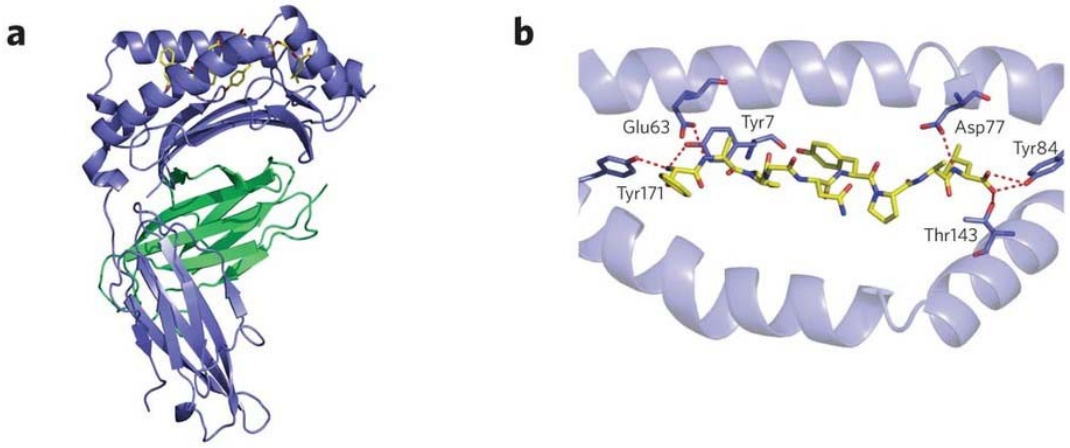


Figure 1.3: Ribbon structures of an MHC class I molecule interacting with a virally encoded peptide.

MHC class I – peptide (FAPGNYPAL) interaction shown from side view (a) and top view of the peptide binding groove (b). The heavy chain is shown in purple, β_2-m in green and the peptide in yellow. Interactions between peptide and MHC residues are indicated by dotted lines in (b). (Neefjes and Ovaa 2013).

1.2.4 The antigen presentation pathway

The MHC class I complex presents peptides derived from intracellular proteins following proteolysis through the ubiquitin-proteasome pathway (Antoniou *et al*. 2003). These peptides are often derived from normal cellular proteins following the end of their life cycle in the cell, but can also present peptides from proteins which have misfolded or been damaged during production, mutated proteins (often in cancer cells) and virus-derived proteins following infection (Leone *et al*. 2013). Proteins are ubiquitinated, unfolded and translocated to the cytoplasm for proteolysis in the proteasome (Finley 2009). The resulting peptides are translocated into the endoplasmic reticulum (ER) for assembly with the class I heavy chain and β_2-m. This translocation is performed by the transporter associated with antigen processing (TAP) an ATP-binding-cassette transporter consisting of two subunits, TAP-1 and TAP-2 (van Hateren *et al*. 2017). Following TAP-mediated peptide translocation into the ER, the TAP-peptide complex associates with a number of proteins within the ER, including tapasin, ERp57, PDI and the MHC class I and β_2-m molecules, forming a "Peptide loading complex" (PLC) (Howe *et al*. 2009). At this point, several quality control measures exist to ensure that only stable peptide-MHC complexes are translocated to the cell surface. Tapasin not only stabilises the PLC, but is also involved in increasing the rate of peptide loading and dissociation, and for improving the discrimination between optimal and suboptimal loaded peptides (van Hateren *et al*. 2013). Peptides are also edited by chaperone proteins ERAP1 and ERAP2, which trim peptides to the optimal length for class I binding (James *et al*. 2013), and other chaperone proteins such as TAPBPR and calreticulin also play a role in optimising the final peptide binding to ensure a stable complex reaches the cell surface (Neerincx *et al*. 2017; Howe *et al*. 2009). Following successful and stable binding of a peptide into a binding groove, the chaperone proteins are released and the class I complex is released from the ER, travelling through the Golgi before migrating to the cell membrane, where the binding groove is exposed on the outside of the membrane for T-cell interrogation (Figure 1.3). Antigen presentation is a complex, multi-step pathway, and the dysregulation of any step can result in a reduced number or complete loss of class I molecules on the cell surface. Many cancers show evidence of dysregulated antigen processing, allowing for immune evasion (Leone *et al*. 2013).

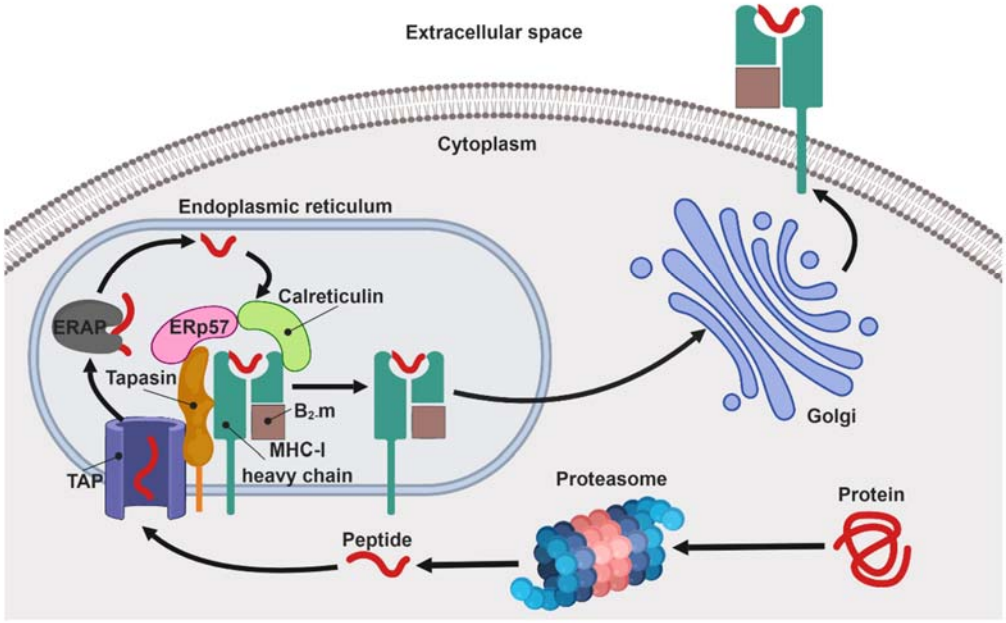


Figure 1.4: The MHC class I antigen processing pathway.

Proteins are degraded in the cytoplasm through the proteasome and transported into the endoplasmic reticulum by TAP for MHC class I peptide loading. Peptides are trimmed by ERAP and loaded onto the MHC class I heavy chain as part of the peptide loading complex. Accessory proteins ERp57, Tapasin and Calreticulin dissociate upon stabilisation of the heavy chain, peptide and β_2-m complex. The folded MHC class I complex is transported through the Golgi to the cell surface where it presents peptides for T-cell interrogation. Adapted from Murat and Tellam (2015). Created with BioRender.com.

1.2.5 MHC class I peptide binding

In humans, individuals have up to six unique MHC molecules presented on their cell surface, each with its own broad peptide binding repertoire, and may be presenting thousands of peptides at any one time on the cell surface (Melief and Kessler 2017). This means that the full cohort of presented peptides in an individual, or the immunopeptidome, is unlikely to be the same as that of any other individual. This variation is generated by the peptide specificity of each class I molecule, which is ultimately determined by the sequence of the binding groove. There are several common elements associated with peptide binding. The size of the binding pocket of class I molecules limits the size of potential peptides that can bind to it (Figure 1.2) usually between 8-12 amino acids, with a preference for 9mers in humans and mice (Leone *et al*. 2013), although this may vary between alleles and species. The peptide repertoire of a class I molecule is also dependent on the "binding motif" of the molecule. The binding motif is often more restricted towards the N and C termini of the peptide. The structural and electrochemical requirements in these areas (most frequently at positions two and nine on the peptide for human class I molecules) are determined by binding pockets that result in the peptide being tightly anchored within the binding groove. Each class I molecule will have a different binding motif and different electrochemical requirements within the binding groove, resulting in molecular specificity (Sidney *et al*. 2008). The limited length of MHC class I peptides, along with the presence of measurable binding motifs allows the use of computational methods to predict whether any given peptide will bind to a particular class I molecule (Figure 1.4). This is particularly important for production of peptide vaccines against emerging pathogens and potentially against a variety of cancers, and may also help with the treatment of autoimmune disorders (Meydan *et al*. 2013).

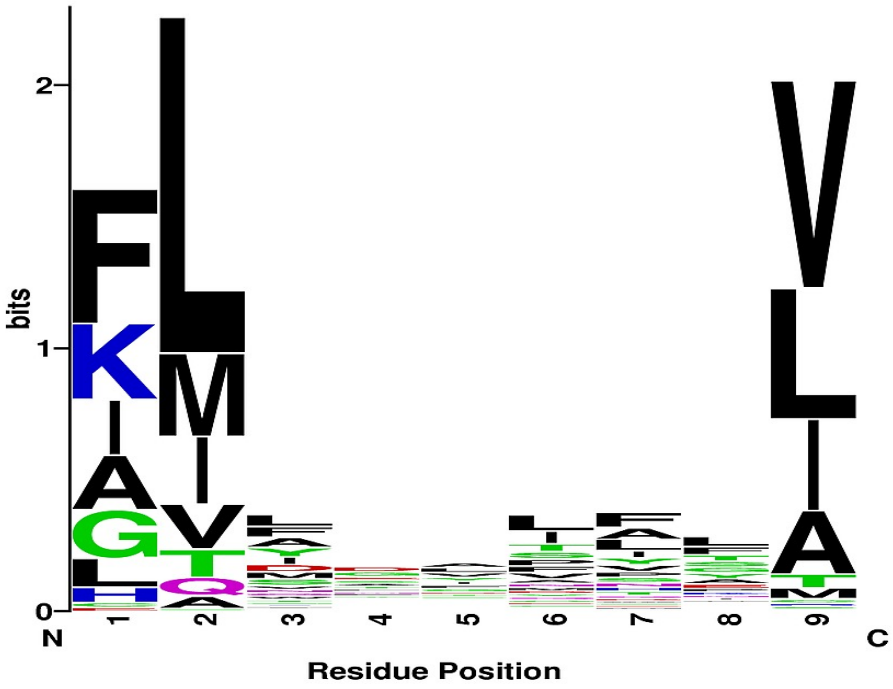


Figure 1.5: A visualisation of an example MHC class I binding motif.

X axis represents position of the amino acid along the peptide. Y axis represents amino acid frequency. Colours denote classes of amino acids. Letters represent the single letter codons for amino acids. The height of each letter indicates the conservation of that amino acid between peptides within the binding repertoire of the class I molecule (Binkowski *et al*. 2012).

1.2.6 MHC class I and T-cell receptor interaction

The immune response to peptide/MHC class I is instigated through the TCR. The TCR is formed in the thymus during thymic development through the process of Variable, Diversity, Joining (VDJ) recombination, which involves random genomic rearrangement of the germline TCR (Dash *et al*. 2017). From a limited number of TCR genes, around twenty million different functional TCRs can be produced. Each receptor contains six hypervariable regions known as complementarity-determining regions (CDRs) which are responsible for the range of peptides a TCR can recognise and interact with (Rossjohn *et al*. 2015). All six CDRs exist in the variable domains of the TCR, which are extracellular and able to interact directly with the MHC-peptide complex on the surface of other cells (Figure 1.5). TCR binding sites may have shallow binding surfaces which directly contact projecting peptide side chains, or deeper peptide binding pockets which can completely engulf large peptide side chains. The number of contacts between the TCR and peptide is limited, as TCRs will only contact the 5-7th amino acid in an 8-9 peptide chain on an MHC class I complex. This results in a high level of cross reactivity between any given MHC-peptide complex and multiple TCRs. Different peptides may generate the same activation signal in a TCR, although there is evidence that as little as one residue change in a peptide can prevent T-cell activation (Hennecke and Wiley 2001). Successful T-cell activation through the TCR is dependent on the presence and functionality of the T-cell co-receptor CD8, which augments a stimulatory signalling cascade through the CD3 complex on the T-cell surface following TCR activation, eliciting a cytotoxic response and resulting in the death of the infected cell (Wang *et al*. 2009). CD8 interacts directly with the MHC class I heavy chain, contacting the α2 and α3 domains of the heavy chain and stabilising the TCR/MHC interaction during antigen recognition (Devine *et al*. 1999; Salter *et al*. 1990). Certain peptide-MHC complexes are recognised by more TCRs than others, resulting in varying immunogenicity across peptide-MHC complexes. Large and aromatic amino acids in the presented peptide increased the immunogenicity of the complex, and virally encoded "non-self" peptides illicit a stronger T-cell response than "self" coded peptides (Calis *et al*. 2013). The difference in immunogenicity and the ability for T-cells to recognise healthy self-antigens allows for a targeted immune response against unhealthy cells to remove them from the system.

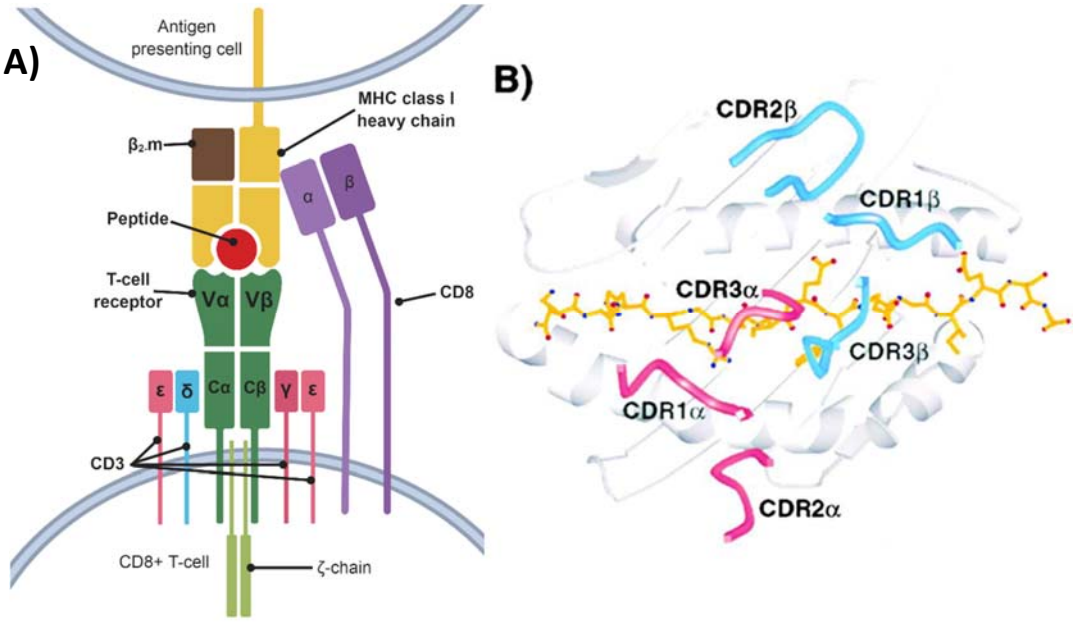


Figure 1.6: Interaction between the TCR and the peptide-MHC complex at the cell surface.

A) The TCR (green) and MHC-peptide (yellow/brown/red) complex. The variable regions of the TCR interact directly with the peptide and variable regions of the MHC class I. CD8 (purple) interacts with the α3 region of the MHC class I heavy chain, whilst the CD3 (pink/blue) complex interacts with the constant regions of the TCR. Adapted from Hennecke and Wiley (2001). Created with BioRender.com.

B) Interaction between the TCR CDRs and the MHC peptide binding groove. Note the TCR (Blue and pink) interacts with both the peptide (yellow) and the peptide binding groove (white). (Hennecke and Wiley 2001).

1.2.7 Tumour associated and tumour specific antigens

The peptidome of a given cell is the full repertoire of peptides produced from the genome of that particular cell. The immunopeptidome is the peptides within the peptidome which are presented on the cell surface by MHC molecules. The peptides expressed by MHC class I on the surface of a tumour cell may be tumour-specific or tumour associated. Tumour specific antigens (TSAs) are peptides exclusively expressed on tumour cells, whereas tumour associated antigens (TAAs) are present on other body cells but overexpressed in tumour cells (Offringa 2009). As a cancer evolves proteins within the cell will accumulate mutations which result in the phenotype and clinical outcome of the cancer. A cancer may contain thousands of mutated proteins within its proteome, and even within its peptidome, but only a very small number of these proteins will translate to the immunopeptidome (Figure 1.6). Yadav *et al*. (2014) illustrate this principle in their identification of tumour-specific antigens from two mouse tumour cell lines, MC-38 and TRAMP-C1. Both exomes were sequenced to identify tumour-specific mutations, and this data was used to predict the presence of unique mutated peptides expressed by MHC class I on the surface of each cell line. Unique tumour-specific antigens on the tumour cell surface were then identified experimentally using mass-spectrometry. Whilst 28,439 and 10,118 exome mutations were detected in MC-38 and TRAMP-C1 respectively, experimental detection identified only 3 immunogenic neoantigens on the cell surface of MC-38, and none in TRAMP-C1. This study neatly highlights the major challenge for the immune system when fighting tumour growth, the lack of tumour-specific antigens expressed on the surface of cancer cells, and also illustrates a potential work-flow for identifying these antigens.

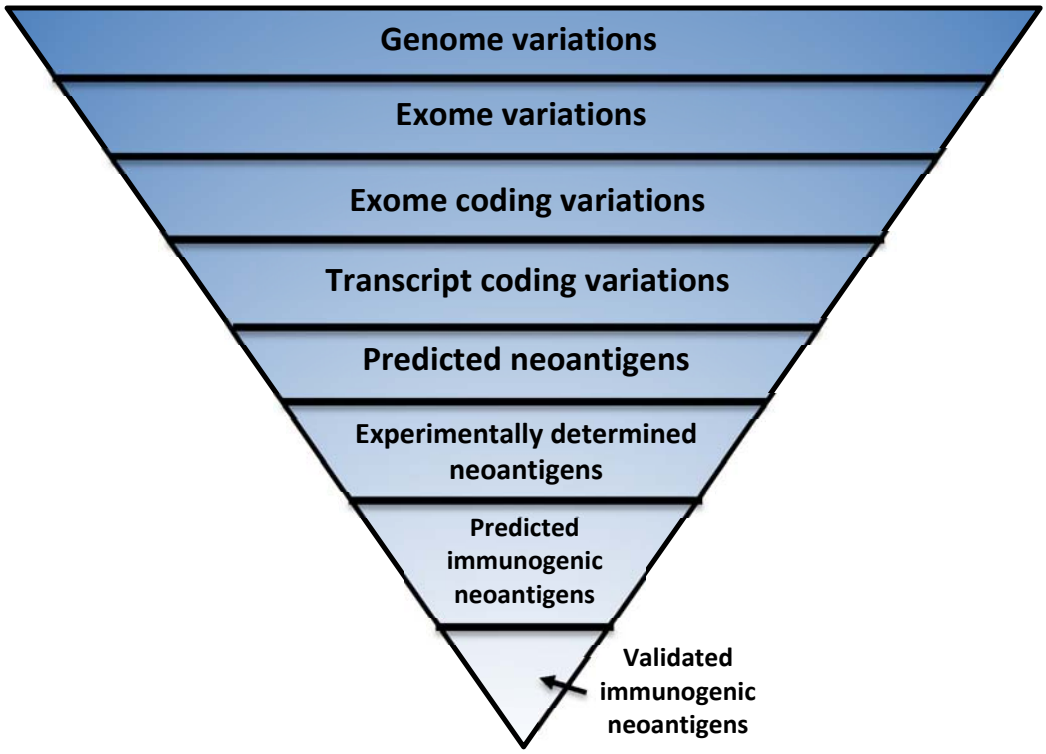


Figure 1.7: Very few mutations in a tumour are available on the cell surface for T-cell interrogation.
Schematic adapted from Yadav *et al*. (2014).

1.2.8 Cancer vaccines

A wealth of research has been undertaken to develop methods for designing vaccines to exploit MHC class I restricted tumour antigens and induce an immune response against tumours. These have included vaccinations using whole tumour cell preparations (Small *et al*. 2007; Berd *et al*. 2004) which doesn't require knowledge of specific tumour antigens but is more vulnerable to the general effects of self-antigen immune tolerance, as well as more targeted approaches using single TSAs as the vaccine target. The lack of highly specific antigens on the tumour surface, coupled with the degenerate nature of T-cell receptor (TCR) - MHC class I interactions (Wilson *et al*. 2004; Mason 1998) has proven challenging for researchers in the area of cancer immunotherapy and in some cases has initiated toxicity against healthy tissue (for example the MAGE-A3 antigens) (Linette *et al*. 2013; Morgan *et al*. 2013). Vaccines may also be targeted towards aberrantly expressed TAAs, which while not as specific as TSAs there is evidence that patients can harbour antibodies against these antigens that form an effective immune response without inducing autoimmunity (Marijt *et al*. 2018; Vella *et al*. 2009). When targeting a vaccine against MHC class I restricted peptides, it is also important to evaluate MHC class I expression and peptide binding properties to assess the potential for a given antigen to raise an effective and specific immune response against tumour cells (Zhao *et al.* 2013).

Two cancer vaccines are currently approved for use in patients; Oncophage, a patient specific heat shock protein-peptide complex vaccine approved for use in Russia to treat renal cell carcinoma, and in clinical trials for a number of other tumour types (Reviewed by Wood and Mulders (2009)) and Sipuleucel-T, a cell based immunostimulant therapy that involves stimulation of the patient dendritic cells to respond to an antigen from prostatic acid phosphatase, approved by the United States Food and Drug Administration for treatment of hormone refractory prostate cancer (Reviewed by Wei *et al*. (2015)). These vaccines are therapeutic rather than prophylactic, however prophylactic vaccines against human tumours are under development (Bautz *et al*. 2017; Keenan *et al*. 2014).

1.3 Cancer progenitors

Mutations which result in malignant cellular transformation are often in genes involved in the normal regulation of cellular processes such as cell survival, motility and proliferation. Dysregulation of these vital cellular functions can result in the development of a cancer (Hanahan and Weinberg 2011; Hanahan and Weinberg 2000). Cancers are a set of markedly different diseases which vary in genotype and phenotype and are initiated by malignant transformation of a single progenitor cell. Cancers from similar tissues demonstrate high variability, and indeed a single tumour will often show remarkable heterogeneity varying in cellular morphology, marker expression and even response to treatment. There is considerable debate about where this variation arises; whether it is the specific carcinogenic mutations which give rise to new sub-populations, or whether these sub-populations arise from progenitor cells at different stages in the lineage hierarchy of the cell type (Visvader 2011).

The progenitor cell of a cancer influences the phenotype of the tumours it produces. Cells in a dedifferentiated state, such as stem cells, often exhibit a higher proliferative capacity than their more highly differentiated counterparts (Sia *et al*. 2017), and a number of studies have concluded that the progenitor cell of a given cancer, and the differentiation state of this cell, impacts tumour growth, metastatic potential and ultimately its clinical outcome (Blaj *et al*. 2017; Goldstein and Witte 2013; Ince *et al*. 2007; Sainsbury *et al*. 2000). Understanding the origins of a cancer can provide clues as to the specific mechanisms of carcinogenesis in different cancers, and the best course of prevention and treatment for different cancer subtypes. Understanding the cellular origins of a cancer is also crucial for identifying peptide vaccine targets, as an immunopeptidome of the healthy cell type is needed to draw direct comparisons to the diseased state to identify mutated or over-expressed peptides that are specific to the cancer cells.

Two independent transmissible tumours have emerged in the Tasmanian devil in the last 25 years, DFT1 and DFT2. Understanding the origins of these tumours could provide a mechanism for why the species seems vulnerable to this sort of tumour and is also crucial for developing disease management strategies to supporting the remaining wild population. Work by Murchison *et al*. (2010) has identified a Schwann cell progenitor for DFT1, although the differentiation state of the progenitor is unknown. Little is known about the DFT2 although early evidence may indicate a neuroectodermal origin (Stammnitz *et al*. 2018).

1.3.1 Schwann cells

Schwann cells are the principle glial cells of the peripheral nervous system (PNS) and have important functions in the development of the PNS in neonates, as well as axonal growth and myelination, trophic support of the PNS and axonal regeneration following nerve injury in adults. They are also responsible for removing cellular debris within the PNS. Unlike neuronal cells, Schwann cells will retain the ability to divide and proliferate throughout their lifespan (Bhatheja and Field 2006). Schwann cells are also known to be capable of inducing remyelination in axons which have demyelinated due to illness or injury (Vallières *et al*. 2017) and can exist in a number of differentiation states, with two major terminal states known as myelinating (MSCs) and non-myelinating (NMSCs) Schwann cells. MSCs myelinate, support and repair large axons, whilst NMSCs perform equivalent functions for smaller axons, though do not produce myelin (Castelnovo *et al*. 2017). Schwann cells maintain the ability to dedifferentiate to an immature-like state in response to cellular injury and stress (Kim *et al*. 2014).

1.3.2 Schwann cell development and differentiation

The development of Schwann cells begins in the neural crest of the embryo, of which Schwann cells are a terminal cell type (Liu and Cheung 2016). Differentiation to the terminal Schwann cell phenotypes occurs in a stepwise manner, with a number of distinct intermediate stages which are detailed in Figure 1.7 and Table 1.1 (Saitoh *et al*. 2016).

The first distinct cell along the Schwann cell lineage is the Schwann cell precursor (SCP), a cell which differs extensively from both neural crest cells and mature Schwann cells (Kim *et al*. 2017). SCPs express glial fibrillary acidic protein (GFAP), a nervous system cell marker which is not expressed in neural crest cells and show distinctly increased motility compared to mature Schwann cells. SCPs do not express some markers normally associated with Schwann cells, such as S100. SCPs die rapidly by controlled apoptosis in culture lacking axonal contact, though these cells can be saved by culture in neuron-conditioned media, suggesting a role of axonally produced factors in survival and development (Jessen *et al*. 1994).

SCPs can differentiate into immature Schwann cells (iSCs) which express the low affinity nerve growth factor receptor (NGFR) and maintain a high proliferative capacity (Kobayashi *et al*. 2012). iSCs also express high levels of the transcription factor sex determining region Y (SRY)-box 2 (SOX2), a factor expressed in a number of multipotent cells in different tissue types. This reflects the ability of iSCs to differentiate further into MSCs and NMSCs depending on external signals (Liu *et al*. 2015). The levels of SOX2 expression drops as the cells undergo further differentiation to their terminal states, reflecting the reduction in multipotency (Finzsch *et al*. 2010).

The progression of iSCs to MSCs is reliant on the expression of transcription factor Oct-6, which is initially expressed in pro-myelinating Schwann cells and induces expression of the transcription factor early growth response protein 2 (EGR2) prior to terminal differentiation (Figure 1.7) (Liu *et al*. 2015). This results in the induction of a number of genes associated with myelin production, including myelin basic protein (MBP) and Periaxin (PRX) (Bacallao and Monje 2015; Gillespie *et al*. 1994). MSCs, once differentiated, become associated with a single large axon, wrapping repeatedly around the axon and laying down myelin associated proteins to form the myelin sheath (Torii *et al*. 2015).

iSCs can also differentiate into NMSCs. NMSCs associate with the smaller diameter axons of the PNS, and one NMSC will wrap around several of these smaller axons to form a bundle, known as Remak's fibres, separating individual axons with small cytoplasmic processes from the body of the cell (Corfas *et al*. 2004). There are many more NMSCs in the PNS than MSCs, reflecting the numbers of unmyelinated axons in the PNS (Griffin and Thompson 2008). NMSCs also exist in the neuromuscular junction, where they are known as perisynaptic Schwann cells, and have a role in upkeep of the neuromuscular junction and modulation of synaptic transmission (Sugiura and Lin 2011). NMSCs do not express the proteins involved in myelin production such as MBP and PRX, and share many markers with their less differentiated counterparts (Liu *et al*. 2015). Evidence indicates that the direct interaction of iSCs with large axons results in the repression of early differentiation markers such as neural cell adhesion molecule (NCAM), resulting in the induction of myelin-associated proteins and differentiation into the MSC phenotype (Clark *et al*. 2017; Jessen *et al*. 1990).

MSCs and NMSCs retain the ability to de-differentiate back to immature-like Schwann cells (i-LSCs) in response to cellular stress, particularly following axonal damage and Wallerian degeneration of the nerve fibre (Monje *et al*. 2010). Whilst this process is simple for NMSCs, which retain a number of iSC markers, MSCs require the cessation of myelin production and all proteins associated with it. The inhibition of myelin production and initiation of myelin loss is achieved through the activation of a number of transcription factors, including paired box-3 (PAX3) (Doddrell *et al*. 2012), notch homolog 1, translocation associated (NOTCH1) and neurogenic locus notch homolog protein 3 (NOTCH3) (Woodhoo *et al*. 2009) and jun proto-oncogene, AP-1 transcription factor subunit (c-Jun) (Parkinson *et al*. 2008). These are all transcription factors involved in mitogenic signalling and cell proliferation, and dedifferentiation is often accompanied by an increase in cell growth. However, it has been shown that whilst the two events often occur together, dedifferentiation and re-entry into the cell cycle are uncoupled cellular processes (Monje *et al*. 2010). Once dedifferentiated, i-LSCs are then able to demyelinate the damaged nerve by increasing lysosomal and proteasomal activity, working alongside inflammatory cells to break down the myelin sheath and remove the debris from the damaged nerve (S. Y. Jang *et al.* 2017). Dedifferentiated Schwann cells also produce growth factors and extracellular matrix proteins which both support survival of neurons and actively encourage regeneration of the axon (Hyung *et al.* 2019). Interestingly, this ability is not demonstrated by iSCs, suggesting that the dedifferentiated state, whilst very similar to an iSC, is a distinct intermediate cell type (Jessen and Mirsky 2016).

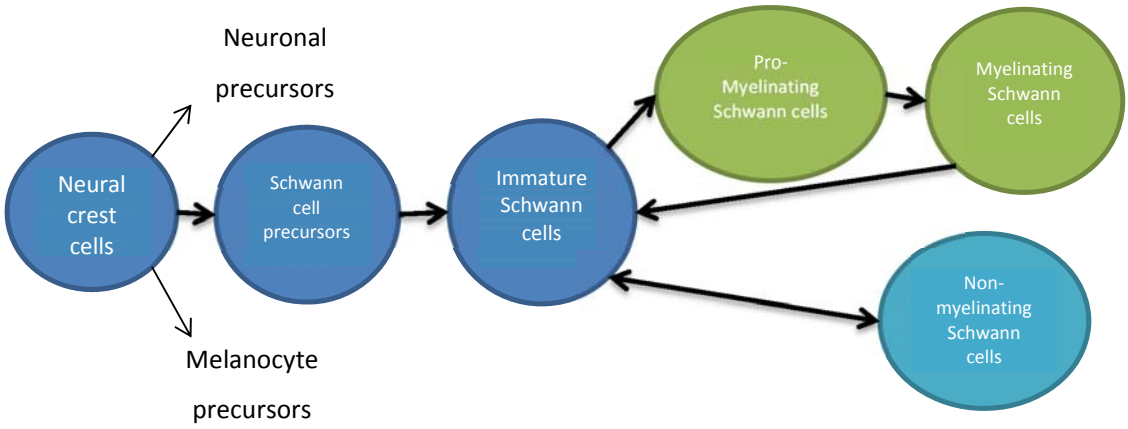


Figure 1.8: Schwann cell differentiation.

Mature Schwann cells retain the ability to cycle between distinct differentiation states in response to nerve injury throughout adult life. Adapted from Jessen and Mirsky (2005).

Table 1.1: The major identifying markers for Schwann cells at different stages of differentiation.

An X denotes that a marker is present in a specific cell type. Adapted from Liu *et al*. (2015).

	Schwann cell precursors	Immature Schwann cells	Pro-myelin Schwann cells	Myelinating Schwann cells	Non-myelinating Schwann cells
SOX10	X	X	X	X	X
S100		X	X	X	X
MPZ	X	X		X	
MBP				X	
GAP43	X	X			X
Oct6	X	X	X	X	X (Low levels)
EGR2			X	X	
NGFR		X			X
SOX2	X	X			
PRX				X	

1.3.3 Schwann cells in culture

As is the case with many cell types, Schwann cells may not behave the same in culture as they do *in vivo*. The importance of axonally derived signals and contact has already been discussed, and these contacts are often missing in culture and must be simulated with supplements associated with neuronal growth (Corfas *et al*. 2004). Given the complex differentiation process of Schwann cells, it is important to understand how the expression profile of the differentiation state markers changes in continuous culture. Liu *et al*. (2015) performed a comprehensive study of the changes in marker expression in Schwann cells *in vitro*. They found that S100, whilst widely considered to be the most accurate marker for Schwann cells, is not reliable enough *in vitro* to be used as a marker alone, as levels of the marker appear to fluctuate unexpectedly within cell populations. Interestingly, this study also noted that *in vitro*, Schwann cells appear to express a marker profile that correlates very closely with that of *in vivo* iSCs, with some additional myelin associated protein expression. This is perhaps to be expected, as nerve damage sustained during isolation is likely to result in dedifferentiation of mature cells.

1.3.4 Oligodendrocytes

Oligodendrocytes are the myelinating glial cells of the central nervous system (CNS), which produce a myelin sheath to speed up signal transmission throughout the brain and spinal cord (Nave 2010), as well as providing trophic and metabolic support to neurons (Lee *et al*. 2012). Although oligodendrocytes perform equivalent functions to Schwann cells, they are a distinct cell type with several key differences to Schwann cells. Firstly, whilst Schwann cells myelinate axons in a 1:1 ratio, oligodendrocytes extend processes outwards to wrap around and myelinate multiple axons at once (Matthews and Duncan 1971)(Figure 9.1). Oligodendrocyte myelin differs from Schwann cell myelin in its component proteins and lipids, notably lacking in MPZ, the main protein component of Schwann cell myelin. Instead, CNS myelin is highly enriched for PLP, its major myelin component protein (Morell and Quarles 1999). Oligodendrocytes also do not dedifferentiate following nerve injury. Instead, a population of oligodendrocyte precursor cells (OPCs) exists in the adult brain, which can differentiate into oligodendrocytes (or astrocytes) to remyelinate and repair axons following nerve injury (Dent *et al*. 2015). These cells are distinguished by their expression of the markers PDGFRA, NG2 and OLIG2, which drastically reduces as the cells differentiate and start producing myelin (Moyon *et al*. 2015; Horiuchi *et al*. 2010; Rivers *et al*. 2008; Liu *et al*. 2007). Remarkably, evidence indicates that these CNS resident OPCs are able to differentiate into Schwann cells, implying that despite arising from distinct lineages these two cell types are able to converge (Zawadzka *et al*. 2010).

1.3.5 MHC expression by Schwann cells and oligodendrocytes

There is evidence that inflammatory demyelinating neuropathies such as Chronic inflammatory demyelinating polyneuropathy and Guillain–Barré syndrome are caused by an autoimmune attack of the host Schwann cells, indicating that Schwann cells present reactive antigens on the cell surface (Kwa *et al*. 2003). It has been shown that human and rodent Schwann cells express low levels of MHC class I on their cell surface, but generally do not express MHC class II. Interestingly, following stimulation with the inflammatory cytokine interferon gamma (IFNγ), human and rat Schwann cells show a marked increase in MHC class I expression, as well as the induction of MHC class II expression (Lilje and Armati 1997; Samuel *et al*. 1987). Early evidence suggested that murine Schwann cells did not upregulate MHC molecules in response to inflammation (Hirsch *et al*. 1983), but more recent work has shown that murine Schwann cells can respond to IFNγ treatment in a similar manner to their rat and human counterparts, suggesting a conserved role of the Schwann cell in antigen presentation across different species (Meyer Zu Hörste *et al*. 2008). This role of the Schwann cell as a conditional antigen presenting cell, with the ability to regulate MHC class I and II molecules in response to external signals, leaves Schwann cells with a mechanism for both immune avoidance and autoimmunity in different disease states. Similarly, oligodendrocytes express low levels of MHC class I *in vitro* and *in vivo*, and normally do not express MHC class II molecules (Redwine *et al.* 2001; Jurewicz *et al.* 1998). There is a significant upregulation of MHC class I and MHC class II in response to infection and inflammation in both mature oligodendrocytes and oligodendrocyte precursor cells, indicating that oligodendroglial lineage cells are potential targets for CD8+ T-cell destruction regardless of their differentiation state (Falcão *et al.* 2018). Interestingly, MHC class II, but not MHC class I, upregulation in oligodendrocytes is crucial to their ability to both demyelinate and remyelinate nerve axons, indicating an important role in responses to infections in the CNS and the role of inflammation in the generation of mature, myelin-producing oligodendrocytes (Arnett *et al.* 2003).

1.4 Devil facial tumour disease

1.4.1 The Tasmanian devil

The Tasmanian devil (*Sarcophilus harrisii*) is the world's largest extant carnivorous marsupial, a distinction held since the extinction of the Tasmanian tiger (*Thylacinus cynocephalus*) in 1982 (McKnight 2008). The Tasmanian devil is a dark coated, stocky animal which can reach up to 30cm in length, weigh up to 12 kg and has one of the most powerful bites relative to size in the animal kingdom. It is well known for its distinctive disturbing growl and aggressive behaviour during mating and feeding. Tasmanian devils are primarily scavengers choosing to feed on carrion rather than hunt, though they are capable of hunting small prey including birds and fish. Devils are nocturnal and primarily solitary animals, interacting only during mating or whilst feeding on a communal carcass (Owen 2005). The species was once prevalent across mainland Australia but now exists in the wild only on the Australian island state of Tasmania, though the species is widespread across the island. Despite historic persecution of the animals by farmers, the population across the island in the early 1990s was estimated to be up to 150,000 individuals (Hawkins *et al*. 2008).

1.4.2 The emergence of DFT1

For the last twenty years, the wild Tasmanian devil population has been in steep decline, which was estimated at over a 60 % reduction in 2006, and has continued to fall, with most recent estimates indicating an overall average decline of 77% (Lazenby *et al*. 2018; McCallum *et al*. 2009; McCallum *et al*. 2007; Hawkins *et al*. 2006). This resulted in the reclassification of the Tasmanian devil from "Least concern" to "Endangered" by the IUCN red list in 2008 (Hawkins *et al*. 2008). This dramatic fall in devil numbers is primarily due to the emergence of Devil Facial Tumour 1 (DFT1) (McCallum *et al*. 2007). In 1996 a photographer captured the first photographs of wild Tasmanian devils with large, disfiguring tumours present around the face and mouth, which has since been retroactively assigned as likely cases of DFT1 (Hawkins *et al*. 2006). Since this first emergence of the disease in North-eastern Tasmania, it has rapidly spread throughout the rest of the island, and now covers the majority of the geographical range of the devil (Figure 1.8) (Lazenby *et al*. 2018; McCallum *et al*. 2009; McCallum *et al*. 2007).

DFT1 is characterised by aggressive and poorly differentiated neoplasms which most commonly form on mucosal surfaces around the face and mouth and frequently metastasise to distant organs. These tumours can grow very large, preventing the devils from eating and swallowing, and are estimated to cause almost 100 % mortality among devils (Loh *et al*. 2006). Chromosome analysis of the tumour cells show that the cells comprising these tumours have undergone major chromosomal rearrangements, including chromosome loss and the presence of four marker chromosomes not seen in healthy devil tissue (Pearse and Swift 2006). These grossly abnormal chromosomal patterns were highly conserved across tumour cells from different devils, all of which show little to no similarity to their host devil. This evidence led to the discovery that DFT1 is a transmissible cancer, initially appearing as a tumour in a single host devil, before being transmitted as an allograft during devil biting behaviour throughout the species (Pearse and Swift 2006).

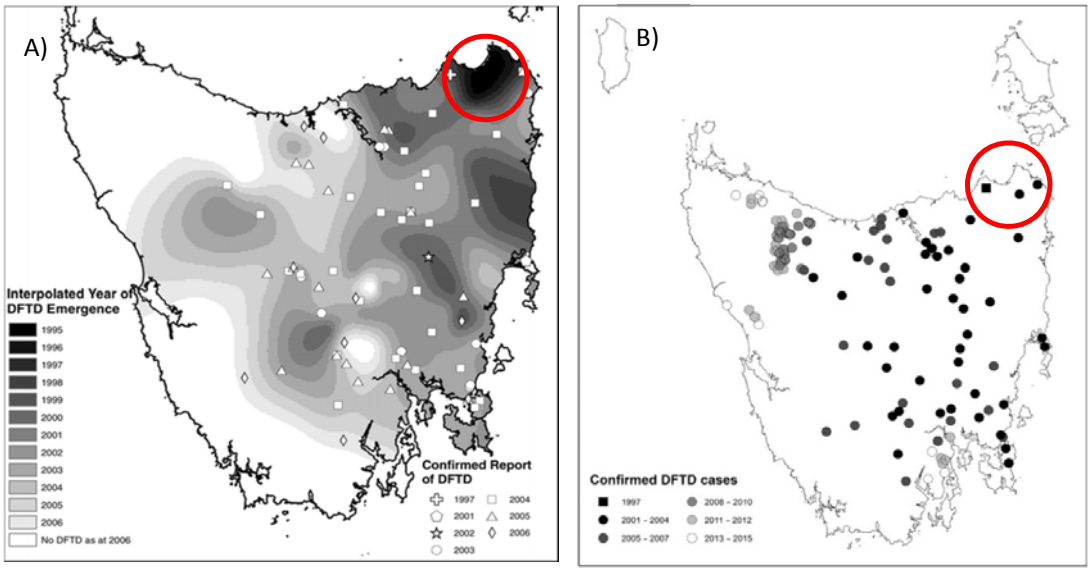


Figure 1.9: DFT1 has spread across Tasmania and now exists across most of the island.

A) The spread of DFT1 across Tasmania between 1995 and 2006. Contour colouring indicates year of DFT1 emergence. White shapes indicate confirmed reports of DFT1, and the year is denoted in the key (McCallum *et al*. 2007).

B): Confirmed DFT1 cases from 1997-2015. Greyscale circles indicate confirmed DFT1 cases, with a graded colour scale where lighter colours indicate more recently diagnosed cases (detailed in key on map) (Lazenby *et al*. 2018).

The red circle on both maps denotes the location DFT1 was first identified.

1.4.3 The origins of DFT1

A transcriptomic analysis of DFT1 by Murchison *et al*. (2010) revealed that DFT1 is likely of Schwann cell origin. Initial miRNA sequencing showed that DFT1 miRNAs have a unique profile which is conserved across DFT1 samples, and distinct from that of ten different tissue types. The tissue with the closest miRNA expression profile to DFT1 was that of the brain. In addition, transcriptome analysis of DFT1 and host tissue revealed the expression of twenty transcripts expressed at greater than twice the level in DFT1 as they were in healthy testis tissue. Nine of these were proteins involved in the myelination pathway, including structural myelin proteins such as MBP and PRX, as well as transcription factors associated with Schwann cell differentiation such as SOX10 and SOX2. Clustering of gene expression data for thirty-one tumour-enriched genes grouped DFT1 cells with peripheral nerves (Figure 1.9), where Schwann cells are primarily found. In addition, DFT1 cells stain with antibodies for several Schwann cell markers, including S100. Interestingly, whilst DFT1 is a cytologically undifferentiated tumour, many of the Schwann cell markers it expresses are associated with the highly differentiated MSC. It has been widely documented that PRX is a highly specific marker for DFT1 cells, and both tumour samples and cell lines stain strongly for the protein (Tovar *et al*. 2011). Although DFT1 expresses markers associated with MSCs, it is currently unclear exactly what differentiation state the progenitor Schwann cell was in when it underwent malignant transformation. Additionally, whilst DFT1 has no cytogenetically detectable sex chromosomes (Pearse and Swift 2006) genomic studies have demonstrated two alleles for X-linked genes and no evidence of any Y chromosome genetic material (Deakin *et al*. 2012; Murchison *et al*. 2012), indicating the tumour emerged in a female host. Thus far there has been no evidence to suggest that DFT1 was the result of a viral transformation (Stammnitz *et al*. 2018; Murchison *et al*. 2012).

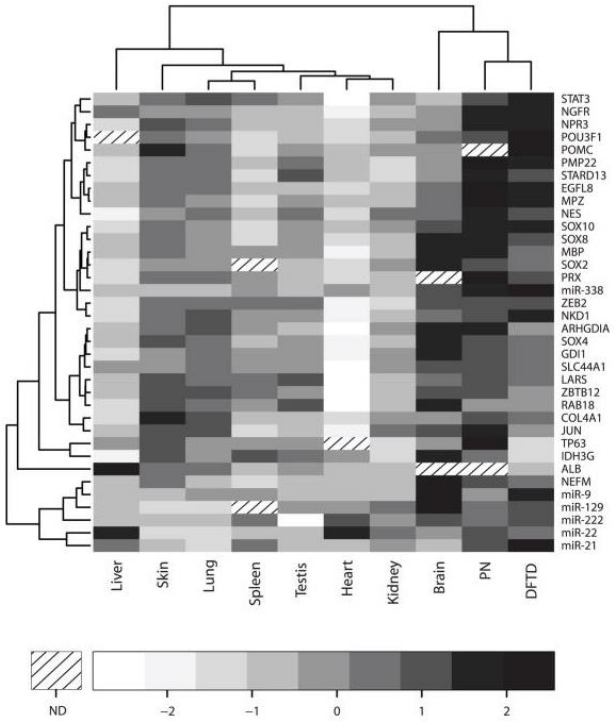


Figure 1.10: DFT1 gene expression clusters most closely with peripheral nerve.

Heat map of semiquantitative RT-PCR gene profiling of 31 genes across a panel of devil tissue samples. PN indicates a Peripheral nerve sample. Colour scale indicates mean gene expression level standardised across tissues with black indicating the highest expression and white indicating the lowest, whilst scored boxes indicate no detection. The top dendrogram indicates hierarchical clustering of gene expression using Pearson's correlation statistic. ND = Not determined. (Murchison *et al*. 2010).

1.4.4 Immune responses to DFT1

As outlined in section 1.1.2, the immune system should detect foreign cells transplanted as an allograft and mount an immune response against them. Despite the presence of a functional immune system in the devil (Brown *et al*. 2011; Kreiss *et al*. 2008; Woods *et al*. 2007), protective immune responses against DFT1 are rare (Pye *et al*. 2016a; Kreiss *et al*. 2015). Due to the restricted and isolated geographical range of the devils, they have undergone a number of genetic bottlenecks, resulting in significantly reduced genetic diversity between animals (Jones *et al*. 2004). While this low genetic diversity, particularly the reduced polymorphism of MHC class I, may contribute to the ability of a contagious cancer to emerge and spread (Siddle *et al.* 2007a), skin allograft experiments have shown that devils are still capable of mounting a full immune response against foreign allografts, with all successful allografts being rejected by the host devils within 3 weeks (Kreiss *et al*. 2011). Understanding how DFTs avoid detection by the immune system is essential to understanding the mechanisms behind its evolution.

Given the importance of MHC class I in both tumour immunosurveillance and transplant rejection, changes in MHC class I both in terms of structural mutations and expression levels may play a role in the ability of contagious cancers to avoid immune destruction. It has previously been shown that another contagious cancer, CTVT, downregulates surface expression of MHC class I, allowing it to spread between and grow in MHC class I discordant hosts (Yang *et al*. 1987). Low surface MHC class I expression in CTVT cells is associated with progressive growth of the tumours. Interestingly, CTVT tumours often stop growing or regress. It has been demonstrated that the "stationary" and "regression" phases of the tumour are associated with an increase in the levels of surface MHC class I expression on the tumour cells and infiltration of lymphocytes into the tumour mass (Hsiao *et al*. 2002; Yang *et al*. 1987).

It has been shown that DFT1 cells lack functional MHC class I molecules on the cell surface, allowing them to avoid detection by the T-cells of the host immune system (Figure 1.10, Figure 1.11)(Siddle *et al*. 2013). Furthermore, it has been demonstrated that this lack of functional MHC class I protein is due to downregulation of genes involved in the antigen processing pathway, such as β_2-m, TAP1 and TAP2, all of which are required for successful peptide loading and presentation on the cell surface. Importantly, this downregulation is due to epigenetic factors, rather than structural mutations in the affected genes. This means that DFT1 cells can still express full transcripts of these genes, and retain the ability to produce functionally sound proteins which can be induced with treatment of the cells with interferon-gamma (IFNγ) (Figure 1.11). Despite this upregulation of MHC class I in response to IFNγ, inflammation does not necessarily result in an immune response to DFT1 tumours, as these have been demonstrated to upregulate PD-L1 in response to IFNγ (Flies *et al*. 2016). This is a well-defined immune evasion tactic in human tumours which prevents IFNγ mediated tumour clearance (Gato-Cañas *et al*. 2017). Interestingly, in rare cases DFT1 cells towards the outside of the tumour mass in stained biopsies are β_2-m positive, and this is associated with an

infiltration of CD3+ lymphocytes, indicating that the devil immune system can sometimes overcome these barriers and raise an immune response to DFT1 tumours *in vivo* (Figure 1.12)(Siddle *et al*. 2013). Under normal circumstances, it would be expected that NK cells would detect this loss of MHC class I and initiate cell death (Waldhauer and Steinle 2008). Devils do have a detectable cytotoxic NK cell response, but it is unknown why these NK cells are unable to recognise the lack of MHC class I on the DFT1 cell surface (Brown *et al*. 2011). It has been demonstrated that devil NK cells are able to kill DFT1 cells *in vitro*, indicating functionality and an ability for the innate immune system to target these tumours, though it is unclear why this doesn't occur in wild animals (Brown *et al*. 2016).

There is some evidence that in rare cases devils are able to mount an effective immune response against DFT1. Pye *et al.* (2016a) have identified antibodies against DFT1 cells in the sera of 6 wild devils captured between 2011 and 2015 in West Pencil Pine, central North Tasmania, which correlated with tumour regression in at least four animals, one of which remained disease free for at least 3 years. In one animal without observed tumour regression, the production of serum antibodies corresponded with the infiltration of T-lymphocytes into the tumour mass. Interestingly, the antibodies present are specific for MHC class I positive DFT1 cells, indicating that this may be an MHC restricted immune response against the tumours.

There is also evidence that DFT1 has exerted selective pressure on the wild devil population, resulting in the evolution of immune or cancer related genes in the devil. Five of these genes are associated with host immunity to CTVT in dogs, indicating that host devils may be evolving to better identify and respond to DFT1 tumour cells. This is an incredibly rapid evolutionary response to an emerging pathogen that has rarely been documented in a wild species (Epstein *et al*. 2016). Additionally, there is evidence that wild devils are becoming more capable of living with the tumour over time, with a striking sex bias towards females (Ruiz-Aravena *et al*. 2018).

These studies paint a picture of a species rapidly evolving to combat a novel pathogen, which is now capable of raising a previously undetected immune response against the tumour. These responses combined with the limited success of early vaccine trials indicates that it may be possible to support the wild population through a prophylactic vaccine programme.

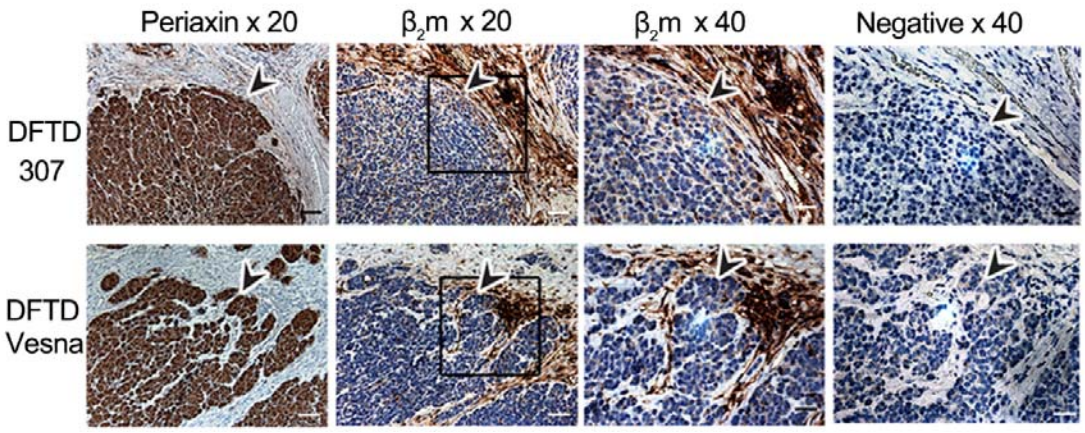


Figure 1.11: DFT1 tumours does not express functional MHC class I.

Two DFT1 tumour biopsies stained by immunohistochemistry (IHC) for PRX to confirm tumour cell identity, and for β_2-m. Brown staining indicates the protein of interest. Blue staining indicates nuclei. Arrowheads indicate DFT1 tumour cells. Boxes indicate 40x zoom area. (Scale bars: 20x magnification, 50 μm; 40x magnification, 20 μm.) (Siddle *et al*. 2013).

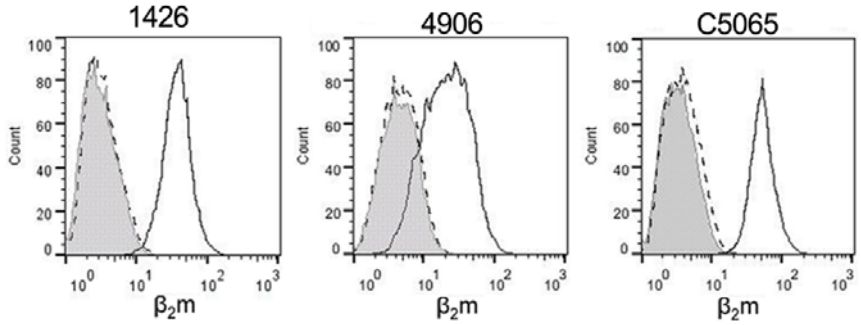


Figure 1.12: DFT1 cells upregulate surface β_2-m expression in response to IFNγ treatment *in vitro*.

Three DFT1 cell lines treated with IFNγ and interrogated by flow cytometry for cell surface β_2-m expression. X axis represents fluorescence intensity. Y axis represents cell count. Shaded peak represents IFNγ treated cells stained with pre-immune serum as a negative control. Black dashed line represents untreated cells stained with β_2-m antibody. Black solid line represents IFNγ treated cells stained with β_2-m antibody (Siddle *et al*. 2013).

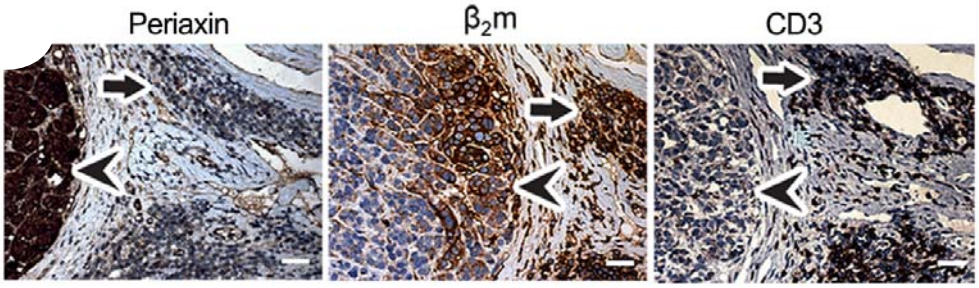


Figure 1.13: Devils can raise an immune response to DFT1 *in vivo*.

Serial sections of a DFT1 tumour stained by IHC for PRX, β_2-m and CD3. Brown staining indicates the protein of interest. Blue staining indicates nuclei. Arrowheads represent tumour cells, arrows represent CD3 positive infiltrating lymphocytes. (Scale bars: 20x magnification, 50 μm; 40x magnification, 20 μm.) (Siddle *et al*. 2013).

1.4.5 The discovery and origins of DFT2

In 2014, two wild devils were recorded in South-east Tasmania with facial tumours grossly indistinguishable from DFT1 tumours. Histological assessment of the tumours from these devils by Pye *et al*. (2016b) showed atypical features. Whilst DFT1 cells are normally round cells which form bundles, these tumours were arranged in sheets and showed cells which were stellate, fusiform and amorphic. IHC staining showed that these tumours were negative for PRX, a specific DFT1 marker (Figure 1.13) (Pye *et al*. 2016b; Tovar *et al*. 2011). Further, these tumour cells have a distinct karyotype compared to DFT1 that is conserved between tumour cells from different animals, suggesting another clonally transmissible cancer. Notably, these tumours were missing the 4 marker chromosomes typical of DFT1 and showed evidence of a Y chromosome. This new tumour has been termed DFT2. Microsatellite analysis of DFT2 has shown that it is as different from DFT1 as it is from healthy devil cells (Pye *et al.* 2016b)(Figure 1.14), and the two tumours have similar but distinct mutational signatures (Stammnitz *et al.* 2018). Early evidence has demonstrated that DFT2 likely emerged very recently from the neuroectoderm although the cell type of origin is unknown (Stammnitz *et al*. 2018).

At the time of writing, DFT2 has been formally identified in at least 23 devils in the Channel region of south east Tasmania and remains geographically restricted to this region, although the prevalence and disease front are increasing (James *et al.* 2019). The emergence of a second contagious cancer in the species is troubling given the current rate of population decline.

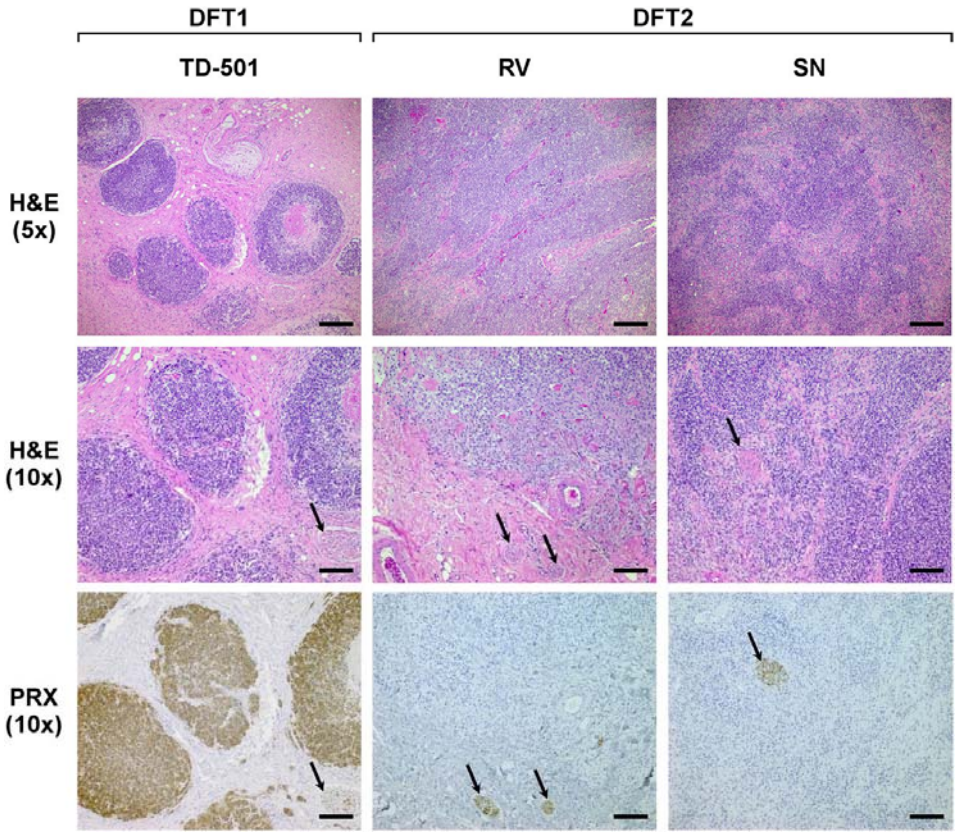


Figure 1.14: Histological analysis of DFT1 and DFT2 tumours.

H&E and PRX staining by IHC of one DFT1 and two DFT2 tumour sections. Brown staining indicates the protein of interest. Blue staining indicates nuclei. Arrowheads denote peripheral nerve bundles, which are positive for periaxin due to their Schwann cell component (Pye *et al*. 2016b).

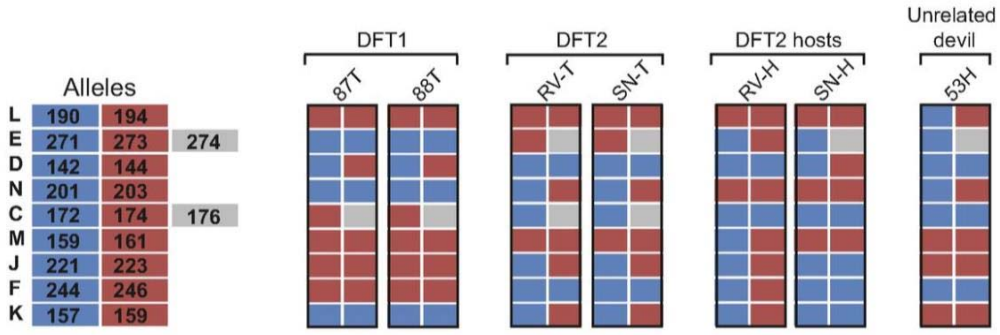


Figure 1.15: Microsatellite marker analysis of DFT2 tumours.

Polymorphic microsatellite marker sequencing of two DFT1 and two DFT2 tumours, two DFT2 host devils and one unrelated devil. Markers are grouped by rows and colours indicate the alleles present in each sample (Pye *et al*. 2016b).

1.4.6 Immune responses to DFT2

Surprisingly, and in striking contrast to DFT1, work from our lab published in Caldwell *et al*. (2018) has demonstrated that DFT2 cells express MHC class I. β_2-m, a component of the MHC class I complex, can be detected on the surface of all available DFT2 cell lines, in direct contrast to DFT1 cell lines which are negative, though the levels of β_2-m on the surface of DFT2 cells are lower than on healthy fibroblast control cells. Classical and non-classical MHC class I heavy chain and β_2m transcripts are also detectable in DFT2 cell lines and biopsies, and qPCR indicates that DFT2 cell lines express significantly more of these transcripts than DFT1. Interestingly, DFT2 expresses the non-classical MHC class I gene Saha-UK. Both classical and non-classical MHC class I protein can be detected in DFT2 tumour biopsies with IHC staining (Figure 1.15), however the expression of these proteins varies between tumours, with one tumour (812T1) appearing completely negative for classical class I whilst maintaining variable expression of Saha-UK. Expression of MHC class I is also heterogenous within tumours, with variable expression by tumour cells within a single biopsy. Classical MHC class I expression appears largely localised to the cell membranes, indicating cell surface expression, whilst non-classical expression shows more cytoplasmic localisation in some tumours. Interestingly, whilst CD3 lymphocytes can be detected infiltrating the tumour mass in some biopsies, this is not restricted by MHC class I expression, as the largely negative 812T1 shows infiltrating immune cells.

How DFT2 can spread between hosts whilst expressing MHC class I is not fully understood. One possible mechanism may be through its expression of non-classical MHC class I. Some non-classical MHC class I molecules are known to have immunosuppressive functions in humans (Persson *et al*. 2017) and overexpression of these molecules is correlated with poor prognosis in some cancers (Li *et al*. 2012). The monomorphic nature of Saha-UK may also reduce the likelihood of an immune response between devils. Another possibility that must be revisited is that DFT2 is able to pass between individuals due to depleted genetic diversity in the Tasmanian devil population and the prevalence of particular MHC class I alleles. Caldwell *et al*. (2018) sequenced MHC class I alleles in DFT1, DFT2 and host devil tissue samples, and has demonstrated that DFT2 appears to have a MHC genotype very similar to that of the devils it has infected (Table 1.2). Importantly, DFT2 appears to express some of the most common MHC class I alleles in the devil population (Lane *et al*. 2012). Both mechanisms likely contribute to the continued transmission of DFT2. Disturbingly, this work indicates that DFT2 tumours may be losing MHC class I expression as they contact MHC class I disparate hosts, suggesting that DFT2 may eventually lose MHC class I expression entirely and follow a similar trajectory to DFT1.

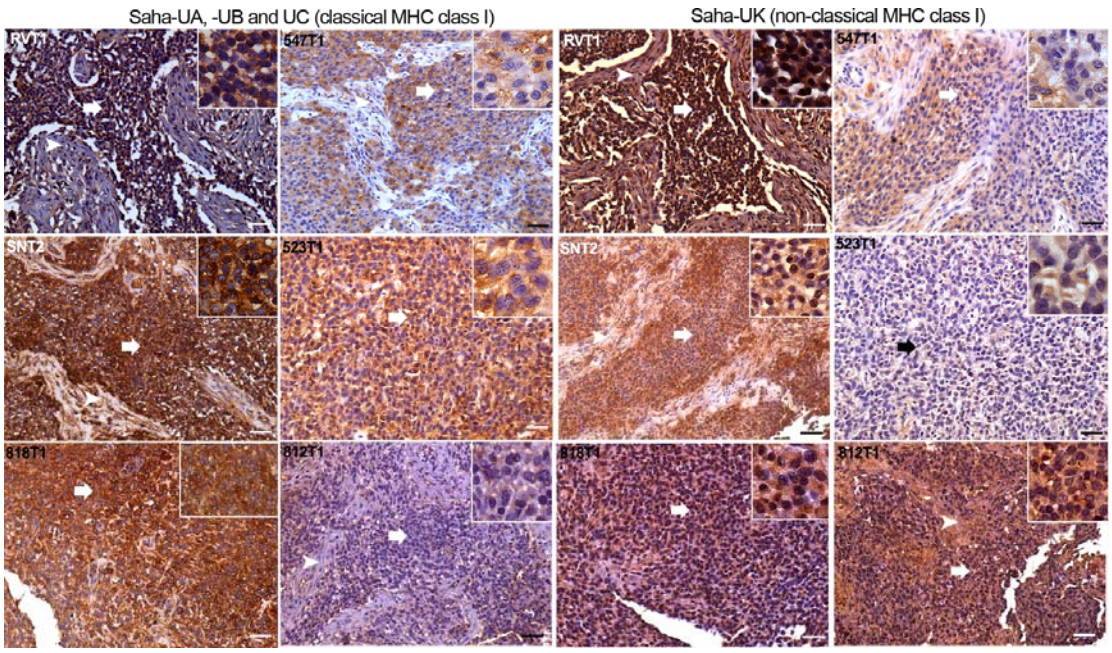


Figure 1.16: DFT2 tumours express variable levels of MHC class I *in vivo*.

Six primary DFT2 tumours stained by IHC for Classical MHC class I and Non-classical MHC class I (Saha-UK) expression imaged at 200x magnification. Brown staining indicates the protein of interest. Blue staining indicates nuclei. Arrows indicate tumour cells, arrowheads indicate stroma. Boxed images of indicated tumour cells are at 600x magnification. Scale bars represent 50 µm.

Table 1.2: DFT2 tumour and hosts express similar MHC class I alleles.

MHC class I sequencing of one DFT1 and one DFT2 cell line, and three DFT2 host spleens. Grey boxes indicate an allele was identified in a sample. Numbers indicate the number of clones of each allele identified following transformation (Caldwell *et al.* 2018).

NCBI allele name	DFT1	DFT2	TD_RV	TD_818	TD_SN
Sahal*46					
Sahal*27		22	22	14	2
Sahal*27–1		13		25	13
Sahal*74/88		9			
Sahal*35		13		10	
Sahal*90		4			1
Sahal*49/82				7	
Sahal*97			6		2
Sahal*33			6		2
Sahal*36			2		
Sahal*37			5		1
Sahal*29				2	2
Sahal*32(UD)					
Saha-UK					

1.4.7 Vaccination strategies for DFTs

Modelling of the long term effects of DFT1 indicates that only 21% of predicted scenarios result in the extinction of the Tasmanian devil, and the tumour itself is over twice as likely to disappear entirely from the wild population (Wells *et al*. 2019). However, these models do not factor in the potential effects of DFT2 if the new tumour spreads beyond its current range and it is not clear whether the emerging adaptations against DFT1 in the devil population will also be beneficial against DFT2. Were DFT2 to follow a similar trajectory to DFT1, it has the potential to cause irreparable damage to the remaining wild devil populations, reigniting concerns that this vulnerable species may struggle to survive in the wild. A protective vaccine against DFTs is an attractive solution to preserve devils in the wild, enabling the release of vaccinated devils from captive insurance populations into areas affected by DFTs, and vaccination of wild devils in these populations to help establish herd immunity.

The ability of wild host devils to raise an immune response to DFT1, albeit infrequently, and the seemingly MHC class I restricted nature of the DFT1 immune response has informed several vaccination programmes. These studies have primarily used a single DFT1 cell line that was established in culture and is defined by Pearse *et al*. (2012) as strain 3 (DFTD1-C5065, RRID:CVCL_LB79), which was derived from the tumour of a wild Tasmanian devil trapped in 2005. In the studies discussed below this is the cell line used for all experiments unless otherwise stated. The vaccine preparations and associated immune responses used in devils are summarised in Table 1.3.

Preliminary vaccination studies in Tasmanian devils using irradiated DFT1 cells failed to induce cytotoxic or humoral immune responses (Brown *et al*. 2011). However, it has been demonstrated that DFT1 cells can be used to induce an immune response in mice when DFT1 cells are inoculated as a xenograft (Pinfold *et al*. 2014). Subsequently, Kreiss *et al*. (2015) trialled combinations of inactive DFT1 cell preparations and adjuvants to assess devil immune responses to vaccination (summarised in Table 1.3). Three devils (TD2, TD4 and TD5) demonstrated significant anti-DFT1 antibody responses in response to immunisation. TD2 and TD4 produced antibodies in response to two different vaccine preparations and demonstrated significant cytotoxicity against DFT1 cells following multiple immunisations. Two devils (TD1 and TD3) produced low titres of antibody which were shorted lived in the serum and resulted in low levels of cytotoxicity. TD2 was challenged with the same DFT1 cells used in the vaccine preparation and remained disease free for 189 days, though developed a tumour 154 post challenge with a different DFT1 strain. This study demonstrated that whilst variable, devils can produce specific antibodies against DFT1 following immunisation with inactive cell preparations in the presence of an adjuvant. Crucially, this study also demonstrated that antibody production following vaccination can be protective and can increase disease free survival following DFT1 challenge.

Further vaccination and immunotherapy studies by Tovar *et al*. (2017) have shown that immunisation with inactive DFT1 cell preparations can result in antibody production and associated tumour regression or lack of tumour engraftment. This study has also demonstrated that vaccination using MHC class I positive DFT1 cells (MHC-I+) is more effective than vaccination with MHC class I negative DFT1 cells (MHC-I-), and that sonicated or irradiated whole cell lysates are more effective at initiating an immune response than protein extract alone (Table 1.3). Devils immunised with inactivated MHC-I+ DFT1 cells induced higher levels of serum antibodies against both MHC-I+ and MHC-I- DFT1 cells than devils immunised with MHC-I- cell preparations or adjuvant only controls. However, this alone was not enough to initiate effective anti-tumour responses after DFT1 challenge. Remarkably, one devil (TD4-Mm) immunised with sonicated and irradiated MHC-I+ DFT1 cells produced a strong antibody response against MHC-I+ and MHC-I- DFT1 cells which persisted for at least 11 months post immunisation. This devil remained tumour free after DFT1 challenge for 189 days before being euthanised for age related reasons. Serum levels of MHC-I+ and MHC-I- DFT1 antibodies were still elevated at the point of euthanasia, indicating that the immunisation protocol may have delayed or even prevented tumour engraftment. This study also demonstrated that MHC-I+ DFT1 cells can be used as an effective immunotherapy in devils. Three devils (TD1-My, TD6-Tp and TD7-Sy) were injected with live MHC-I+ DFT1 cells following the engraftment of DFT1 tumours after challenge. This immunotherapy resulted in elevated serum levels of antibodies against MHC-I+ and MHC-I- DFT1 cells which persisted for at least one week post treatment, increased immune cell infiltrate into the tumour mass and regression of the tumours. This therapy was not effective at treating the non-immunised control devil, and neither was immunotherapy using inactivated MHC-I+ DFT1 cells. Crucially, these vaccination trials have also demonstrated that MHC class I increases the immunogenicity of DFT1 cells, indicating the immune response is MHC class I restricted. However, immunotherapy using live MHC class I positive DFT1 cells is a high risk treatment, and in two cases resulted in the formation of small tumours at the immunotherapy site.

The strongest induction of antibody responses seen in the study by Tovar *et al*. (2017) included the adjuvants PolyI:C and CpG oligonucleotides (Table 1.3), two ligands of Toll-like receptors (TLR), a target which has been heavily utilised by adjuvants in human vaccination studies (Galluzzi *et al*. 2012). ISCOMATRIX™ with Poly I:C and CpG oligonucleotides has proved to be a highly effective adjuvant combination in human studies (Silva *et al*. 2015), however the efficacy of this combination in devils has not been fully elucidated. A study by Patchett *et al*. (2017) assessed the effect of stimulating the TLR on immune activation in the Tasmanian devil, to inform further improvement of adjuvant combinations in the vaccination programme. *In vitro* stimulation of Tasmanian devil peripheral blood mononuclear cells (PBMCs) with the TLR ligands PolyICLC and imiquimod and IL-2 induced IFNγ production from predominantly CD3+CD4- cells, indicating a cell mediated immune response. The effect of these adjuvants on devil immune responses *in vivo* was then assessed by vaccination with PolyICLC and imiquimod as adjuvants using a model antigen (Keyhole limpet haemocyanin (KLH)) (Table 1.3). Devils immunised with PolyICLC as an adjuvant exhibited a

strong KLH-specific IgG antibody response which was comparable across both preparations, indicating that PolyICLC alone is sufficient to induce an antigen-specific antibody response in Tasmanian devils. Serum anti-KLH antibody levels decreased in the four devils immunised with PolyICLC adjuvant and were at primary response levels by 6 months post immunisation. Exposure of these devils to KLH without adjuvants 10 months post immunisation resulted in a rapid antibody response in all four devils, indicating that PolyICLC induces long term immunological memory in Tasmanian devils. This study has demonstrated that TLR ligands, particularly TLR3 ligands like PolyI:C and PolyICLC, are able to induce significant long term immune responses when used as vaccination adjuvants in the Tasmanian devil. This is particularly striking given the devils used in this study are considered old (3-6 years), and evidence has demonstrated that Tasmanian devils suffer a significant decline in T-cell receptor diversity after the age of 1, which is expected to reduce their ability to respond to pathogens including DFTs (Cheng *et al.* 2019). The ability of older devils to raise specific immune responses in this way is promising for the future of the vaccination programme. It is however worth nothing that the age and sex balance of the devils immunised in this work is not always equal across preparations, with the imiquimod only preparation being given to two male devils (3 years) and the female devils used in the remaining preparations being 2 years old their male counterparts. Given the described disparity in how long female and male devils survive with DFT1 in the wild (Ruiz-Aravena *et al.* 2018) and the apparent bias towards female devils in the documented cases of wild tumour regression (Pye *et al.* 2016a), it is likely that the sex of the devils plays a role in their ability to raise an immune response to DFTs. Thus, it is important to consider age and sex imbalances in the procedure when drawing conclusions from all vaccination studies. Additionally, whilst devils were still able to raise anti-KLH antibodies upon KLH challenge 10 months post-immunisation, any wild vaccination programme would need a significantly longer period of protection, and the long term ability of devils to raise antibodies following immunisation must be fully assessed.

Table 1.3: Summary of the vaccination protocols and immune responses demonstrated in four major studies of immunisation against DFT1.

NA indicates that this parameter was not assessed. +, ++ and +++ is a qualitative scale indicating increasing levels of serum antibody and/or cytotoxic responses, with + indicating lowest levels and +++ indicating the highest levels.

Study	Vaccine/adjuvant	Vaccine programme	Animals	Antibodies/ Cytotoxicity	Prophylactic
Brown *et al.* 2011	10^8 irradiated DFT1 cells (Strain 2, Pearse and Swift 2006) with Montanide adjuvant	Four doses at monthly intervals	TD1 TD2	None/None None/none	NA
Kreiss *et al.* 2015	10^8 freeze/thawed DFT1 cells (Strain 2: ½ Pea and 2112, Pearse and Swift, 2006) with Montanide adjuvant	Three doses at weekly intervals with final dose 8 weeks later. TD2 received pre-challenge booster of the same preparation.	TD1 TD2	+/NA +++/NA	No Only against strain 2
	10^8 irradiated DFT1 cells with Montanide adjuvant and CpG 1668 oligonucleotides	Three doses at four weekly intervals	TD3 TD4	+/++ +++/+++	NA
	10^8 irradiated DFT1 cells with Montanide adjuvant and CpG 1585 oligonucleotides	Three doses at four weekly intervals	TD5 TD6	++/None None/None	NA
	10^8 sonicated DFT1 cells with Montanide adjuvant and CpG 1668 oligonucleotides	Three doses at four weekly intervals	TD2 TD4	++/+++ ++/+++	NA
Tovar *et al.* 2017	400 µg protein extract from heat-treated DFT1 cells with ISCOMATRIX™ adjuvant	Three doses at monthly intervals. 1000µg protein booster 6 months post immunisation	TD1-My	None/None	No
	$2x10^6$ – $1.5x10^7$ freeze/thawed DFT1 cells pre-treated with Trichostatin A (TD2-GA) or cytokine enriched medium (TD3-Ty) with ISCOMATRIX™ adjuvant	Two doses at monthly intervals	TD2-GA TD3-Ty	++/NA + (MHC-I$^+$ DFT1 only)/ NA	No No
	A) $3x10^7$ sonicated DFT1 cells pre-treated with IFNγ B) 10^6 irradiated DFT1 cells pre-treated with IFNγ Both preparations included ISCOMATRIX™, Poly I:C and CpG (1585, 2395) oligonucleotide adjuvants	Two doses of A at monthly intervals Two doses of B at monthly intervals Booster: 2 doses of B 4 months post immunisation. TD4-Mm received additional double dose of B 7 months post booster 1.	TD4-Mm TD5-Br	++/NA ++/NA	Yes NA
	A) $2x10^6$ irradiated DFT1 cells pre-treated with IFNγ B) $2x10^7$ sonicated DFT1 cells pre-treated with IFNγ Both preparations included ISCOMATRIX™, Poly I:C and CpG (1585, 2395) oligonucleotide adjuvants	Three doses of A at fortnightly intervals One dose of B one month after final dose of A Booster with A 6 months post final immunisation	TD6-Tp TD7-Sy	++/NA +++/NA	No No
Patchett *et al.* 2017	100 µg model antigen KLH	One dose followed by a double dose 42 days later	TD8-Pe TD9-Au	+/NA (Anti-KLH) +/NA	NA NA
	100 µg model antigen KLH with 100 µg imiquimod	One dose followed by a double dose 42 days later	TD10-Ka TD11-Jo	+/NA (Anti-KLH) +/NA	NA NA
	100 µg model antigen KLH with 100 µg PolyICLC	One dose followed by a double dose 42 days later TD13-Ad received a 100 µg KLH booster immunisation 262 days post second dose	TD12-Gw TD13-Ad	+++/NA +++/NA	NA NA
	100 µg model antigen KLH with 100 µg imiquimod and 100 µg PolyICLC	One dose followed by a double dose 42 days later TD15-Sp received a 100 µg KLH booster immunisation 262 days post second dose	TD14-No TD15-Sp	+++/NA +++/NA	NA NA

Vaccination studies in the Tasmanian devil have been limited by small sample sizes due to the restricted use of devils in research, and often use older animals which may need to be euthanised due to age related health problems before the trials are complete. The implementation of the Wild Devil Recovery Project by the Tasmanian Government in 2015 presented an opportunity to trial vaccination strategies on a younger, larger cohort of devils. Pye *et al*. (2018) immunised a total of 52 devils from the disease free insurance populations prior to their wild release in two separate trials. Both vaccination trials used a combination of sonicated and irradiated MHC-I+ DFT1 cells (DFT1-C5065, RRID:CVCL_LB79) and TLR agonists as adjuvants, and booster immunisations used only irradiated MHC-I+ DFT1 cells with TLR agonist adjuvants. The details of both immunisation and monitoring programmes are summarised in Table 1.4, although it is worth noting that due to the nature of trapping wild devils, not all animals received all immunisations, and this has been accounted for during data analysis. 19 devils released in the first trial in Narawntapu National Park (NNP) received two to four monthly immunisations consisting of sonicated and irradiated MHC-I+ DFT1 cells, followed by a final booster immunisation four months later prior to wild release. 33 devils released in the second trial at Stony Head (SH) a year later received only two immunisations followed by a 5 month booster immunisation post wild release. Sera was collected from immunised devils at regular intervals throughout the primary vaccination course (See Table 1.4 for detail) and assessed for antibodies against MHC-I+ and MHC-I- DFT1. There was no significant difference in the levels of antibodies detected between the two, and vaccination increased average serum levels of anti-DFT1 antibodies. The primary immunisation protocol had no significant impact on levels of antibody in the sera, but both age and sex significantly affected the outcome in the NNP cohort, with juvenile and female devils producing higher levels of antibody. Interestingly, sex did not have a significant impact on the antibody response of the SH devil cohort. Of the total 52 devils vaccinated, 50 were producing antibodies by the end of the primary immunisation course. Post release trends in the NNP population indicated antibody levels reverted to baseline levels 12 weeks post booster, whereas in the SH cohort many devils retained the antibody levels seen after their second primary immunisation 5 months later, prior to their booster vaccine.

This study also performed MHC microsatellite analysis to determine if MHC diversity is linked to vaccination outcomes. This analysis was performed by screening 12 polymorphic MHC linked loci (previously characterised in Cheng and Belov (2012)) in vaccinated devils using Tasmanian devil specific primers to amplify and sequence the microsatellite region of interest. The association between MHC microsatellite genotype and antibody response was analysed by multiple linear regression factoring in age, sex and population to establish a model to determine the effect of specific MHC class I and MHC class II markers on the antibody response to vaccination. This analysis identified three MHC class I linked loci with highly significant (MHC-I_10, p-value < 0.005) or significant (MHC-I_02 and MHC-I_08, p-value < 0.05) associations with strong antibody responses to vaccination. The three microsatellite markers with significant effects on antibody responses were not present in the DFT1 cell line used for immunisations, indicating that MHC diversity results in a

stronger immune response against DFT1, although this has never been demonstrated in unvaccinated devils. The antibody production against MHC-I- cells indicates that whilst the immune responses arising from vaccination are MHC linked, the immune system can also respond to non-MHC linked markers on the DFT1 cells, a promising finding for generating a vaccine against a tumour that normally does not express MHC.

Table 1.4: Summary of the vaccination protocols and serum collection time points from large scale vaccination studies by Pye *et al.* 2018.

All vaccine preparations use the DFT1_C5065 cell line (see Table 2.1) which was established from a wild devil with DFT1 which was captured in 2005. Not all devils received the full immunisation course and not all devils were recaptured for serum collection at all timepoints.

Release site	Number of devils	Primary immunisation course	Booster immunisations	Adjuvants	Serum collection timepoints
NNP	19	A) $2x10^7$ sonicated MHC-I^+ DFT1 cells B) $2x10^6$ irradiated MHC-I^+ DFT1 cells Two doses of A followed by two doses of B at monthly intervals	$2x10^6$ irradiated MHC-I^+ DFT1 cells Four months post immunisation, immediately pre-release	ISCOMATRIX™ with PolyI:C and CpG oligonucleotides (1585 and 2395)	Fortnightly throughout primary course 2 weeks post primary course On the day of booster 2 weeks post booster 2, 6 and 12 weeks post release
SH	33	A) $2x10^7$ sonicated MHC-I^+ DFT1 cells B) 2x106 irradiated MHC-I^+ DFT1 cells One dose of A followed by one dose of B at monthly intervals	$2x10^6$ irradiated MHC-I^+ DFT1 cells Five months post immunisation, four months post-release	ISCOMATRIX™ with PolyICLC and imiquimod	4 weeks post first immunisation 6 weeks post second immunisation 6 weeks post primary course Regularly during 4 months post release 5 months post release

These studies have provided a wealth of evidence that it is possible to stimulate an effective immune response *in vivo* against DFT1, However, responses in devils are variable and the antigens on DFT1 cells that are targets for the immune system are not understood. While there is some evidence that these are MHC restricted, other antigens may also be playing a role. In the same way that humans have variable responses to vaccination, some of this variability derives from normal differences in the host devil immune systems (Kimman *et al*. 2007). Current vaccination strategies in humans are moving towards specific epitope targets rather than whole organism preparations in an attempt to vaccinate against pathogens which have resisted traditional vaccination methods and reduce antigenic load and adverse effects such as vaccine induced autoimmunity, changes to the host microbiome resulting in increased infection and pathogen load of non-target organisms and off-target effects resulting in toxicity to other organ systems (Chauhan *et al.* 2019; Jain and Baranwal 2019; Mina 2018; Ahmad *et al.* 2019; Linette *et al.* 2013; Shimizu *et al.* 2012). Similarly, there is a drive to develop vaccines against cancer based on single peptide targets to generate a strong and specific immune response against malignant cells (Zilberberg *et al*. 2015). At present, DFT1 vaccination trials have demonstrated inconsistent immune responses which vary significantly between animals, a requirement for multiple booster immunisation which is not feasible for treatment of wild populations, and as yet there is no concrete evidence that any DFT1 vaccination trial has been prophylactic. A prophylactic vaccine which can induce consistently strong immune responses against both DFT1 and DFT2 is needed to prevent their continued spread, and a more targeted peptide vaccine approach may help to improve current outcomes.

1.4.8 MHC class I in Tasmanian devils

The MHC class I genes are a rapidly evolving gene family which has undergone species specific gene duplications and expansions throughout mammalian evolution, making it difficult to identify and define orthologous genes between species if they are not closely related (Adams and Parham 2001; Flajnik and Kasahara 2001; Rogers 1985). Characterisation of the MHC class I gene family of the devil is essential to understanding the immune response, or lack thereof, of the devil to the tumours. A combination of studies utilising cDNA, BAC libraries and genome sequence have defined a number of classical and non-classical MHC class I genes (Cheng and Belov 2014; Deakin *et al*. 2012; Siddle *et al*. 2010; Siddle *et al*. 2007b). These genes consist of three classical genes, Saha-UA, Saha-UB and Saha-UC, and five non-classical genes, Saha-UD, Saha-UK, Saha-UM, Saha-MR1 and Saha-CD1 (Table 1.5). Based on a study of 29 healthy and 22 diseased devils in the West Pencil Pine population, Saha-UB is the most polymorphic class I gene, with ten distinct alleles, followed by Saha-UC with seven, Saha-UA with six and Saha-UD with four (Lane *et al*. 2012). The alleles for the three classical loci show very little sequence divergence, particularly between Saha-UA and Saha-UB, with a nucleotide sequence similarity of over 97%. A deletion of 1.646kbp at that start of Saha-UA has also been identified in the population, rendering Saha-UA an inactive pseudogene (Cheng *et al*. 2012). Interestingly, the non-classical gene Saha-UD is closely related to the three classical genes, and phylogenetic analysis clusters Saha-UD with the classical genes, unlike the remaining non-classical genes, which cluster with similar genes from other marsupial species (Cheng and Belov 2014). As with all classical class I genes, Saha-UA, Saha-UB and Saha-UC are ubiquitous, whilst almost all the non-classical genes show restricted tissue expression within the body. The exception to this is Saha-MR1, though an equivalent allele is also ubiquitously expressed in humans (Cheng and Belov 2014).

MHC class I genotyping in wild devils across Tasmania has been undertaken by Siddle *et al*. (2010). 25 distinct MHC class I genotypes were identified from 387 devils across 15 sampling sites. There is evidence that DFT1 is able to spread more easily in areas of Tasmania where the devil populations have a less diverse MHC class I genotype. The east of the island, where DFT1 now covers the entire geographical area, shows markedly reduced MHC class I diversity than the west, where many populations are still not affected by DFT1. In the east of the island, 24% of devils share an MHC class I genotype (Type A), and 30% of devils share an MHC class I genotype with DFT1 (Type 1) (Figure 1.16). This correlation between low MHC class I diversity and DFT1 spread demonstrates the importance of understanding MHC class I in the species and in the context of the disease.

Our lab has previously identified expression of the non-classical class I molecule Saha-UK in both DFT1 and DFT2 (Caldwell *et al*. 2018). This work has also identified a Saha-UC allele, SahaI*27, which is expressed at high levels in DFT2, and is present in both DFT1 and all sequenced host devils for DFT2, as well as a number of DFT1 hosts (Lane *et al.* 2012). The differential expression of these molecules in the context of DFTs suggests that they may play an important role in their immune

evasion, and understanding the peptides bound by these molecules may be key to identifying mechanisms for immune evasion in DFT2, as well as potentially identifying tumour specific antigens for vaccine development.

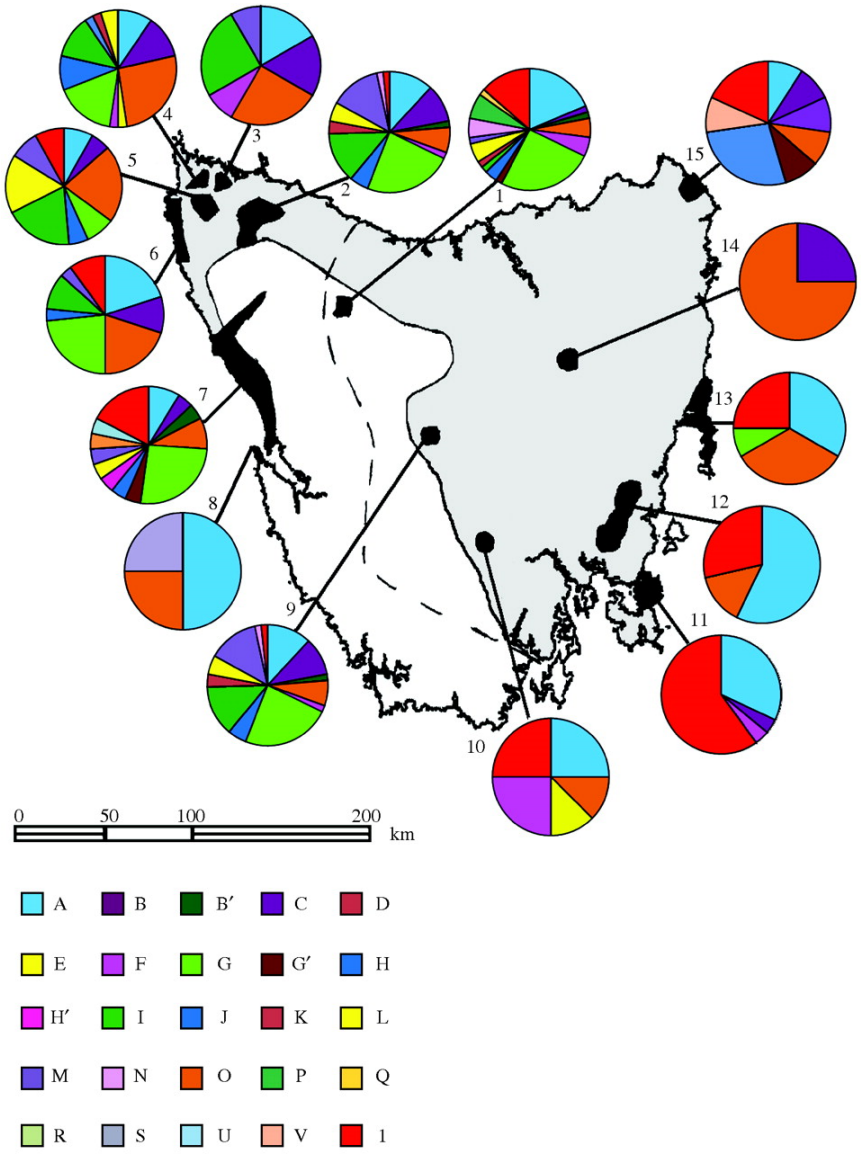


Figure 1.17: MHC class I genotype diversity across Tasmania.

Each genotype is represented by a colour in the key. The dashed line represents the DFT1 disease front (as of 2010), with DFT1 existing in populations to the right of the line. The block grey area of the map represents the geographical range hosting the main populations of devils on the island. (Siddle *et al*. 2010).

Table 1.5: Summary table of MHC class I in the Tasmanian devil.

Data compiled from (Cheng *et al*. 2012). Polymorphic refers to polymorphism of the MHC class I heavy chain gene in the wider Tasmanian devil population. Low refers to <4 alleles in the general population.

Classical/Non-classical	Gene name	Polymorphic	Tissue expression
Classical	Saha-UA	Yes	Ubiquitious
	Saha-UB	Yes	Ubiquitious
	Saha-UC	Yes	Ubiquitious
Non-classical	Saha-UD	Low	Restricted
	Saha-UK	No	Restricted
	Saha-UM	No	Restricted
	Saha-MR1	No	Ubiquitious
	Saha-CD1	Low	Restricted

1.5 Aims and Objectives

This PhD project is part of a larger study to identify peptides that would be suitable candidates for a prophylactic peptide vaccine against DFT1 and DFT2. In particular, we are interested identifying tumour specific and tumour associated antigens which are bound to MHC class I in the context of DFT1 and DFT2, which could be used alongside effective adjuvants to initiate a strong and specific anti-tumour immune response in healthy Tasmanian devils, preventing the continued spread of the DFT1 and DFT2 throughout the island and supporting the remaining wild population.

In order to successfully produce a peptide vaccine the peptide binding repertoire of both healthy and diseased cells must be well understood in order to ensure an effective and specific response and to reduce off target effects. For the devil and DFTs, this requires definition of the cellular origins of both tumours, characterisation of the peptide binding repertoire of devil MHC class I molecules and an understanding of MHC class I expression across healthy and diseased tissues. Conversely, this characterisation may help in the identification of MHC class I/peptide complexes which induce immune tolerance in the devil, resulting in DFT immune evasion and spread. This information may prove particularly useful in the context of DFT2, which continues to evade immune destruction despite surface expression of class I.

This project aims to fill in some of these gaps in information by elucidating the likely progenitor cell type of the newly emerged DFT2 and characterising MHC class I expression in the DFT1 and DFT2 progenitor cells. This project also aims to generate an allele specific immunopeptidome and the peptide binding motif of Saha-UK and SahaI*27-1 (a *Saha-UC* allele) as these proteins are expressed on both DFT1 (after treatment with IFNγ) and DFT2 cells as well as being broadly present in the devil population.

This project has the following objectives:

1. To determine the cellular origins of the recently emerged DFT2, using a proteomic and transcriptomic approach to identify enriched biological pathways of interest which may point to the cell type of origin.

2. To further define the differentiation state of DFT1's Schwann cell progenitor.

3. To determine how MHC class I expression has changed in both DFT1 and DFT2 from their respective progenitor cells.

4. To characterise the response of DFT2 to IFNγ treatment to identify similarities and differences with DFT1, further defining both the cellular origins of DFT2 and the MHC class I system in these tumours.

5. To characterise the peptide binding properties of biologically significant MHC class I alleles in the context of DFT2 for generation of binding motifs and identification of presented peptides which may be suitable as vaccine targets.

Chapter 2 Methods and Materials

2.1 Cell culture

2.1.1 Cell lines and culture conditions

The DFT cell lines used in this project were all derived from primary tumours. The three DFT1 cell lines were derived from tumours in three different animals and have been described in Deakin *et al.* (2012). The three DFT2 cell lines were derived from tumours in three different animals and have been described in Pye *et al.* (2016b) and Caldwell *et al.* (2018). The fibroblast cell line was derived from a skin sample of captive bred female animal and is in Murchison *et al.* (2012). All devil derived cells (summarised in Table 2.1) were grown at 35 ^{o}C and 5 % CO_2 and were split 1:3 using TrypLE Express (Gibco, Cat no. 12605010) at 80 % confluency. A Chinese Hamster Ovary (CHO) cell line stably transfected with recombinant devil IFNγ (Siddle *et al.* 2013) was grown at 37 ^{o}C and 5 % CO_2 and were split 1:10 using TrypLE Express at 80 % confluency. All media components are given in Table 2.1. IFNγ treatment of cell lines was carried out using the supernatant from CHO_IFNγ diluted 1:2 with DFT media for 16 hours according to Siddle *et al.* (2013).

Cells were harvested by washing with Dulbecco's phosphate buffered saline (dPBS) (Gibco, Cat no. 15326239), followed by addition of TrypLE Express. After 5-10 min incubation at 35 ^{o}C (DFT) or 37 ^{o}C (CHO) media was added to inactive the trypsin and cells were transferred to falcon tubes. Cells were spun in a centrifuge at 350 RCF for 5 min at room temperature. Cells were resuspended in 1 mL media, diluted 1:2 with trypan blue (Gibco, Cat no. 15250061) and viable cells were counted using a haemocytometer.

When expanding cells for immunoaffinity purification cells were grown in HYPERFlask® cell culture vessels (Corning, Cat no. CLS10030). ~10^7 cells were seeded into a hyperflask and rotated gently to ensure even distribution. One layer of cells was visualised to check confluency. At ~90% confluency, media was discarded and cells were detached using dPBS and tapping the flask sharply followed by centrifugation at 800 RCF for 5 min at 4 ^{o}C, resuspension in dPBS and counting. ~10^7 cells were reseeded into each flask. Remaining cells were washed three times with PBS, with centrifugation at 800 RCF for 10 min at 4 ^{o}C, and either used directly in experiments or frozen as pellets at -80 ^{o}C for later use.

Table 2.1: Cell lines and culture conditions used in this PhD thesis.

Cell line	Growth media	Sex	Culture established
DFT1 – 4906 (Deakin *et al.* 2012) (RRID:CVCL_LB78)	RMPI-1640 + glutamine (500 mL, Gibco, Cat no. 61870044), 10 % heat inactivated FBS (500 mL, Gibco, Cat no. 10500-064), 50 µg/mL penicillin, 50 µg/mL streptomycin (Pen/strep, 5000 U/mL, Gibco, Cat no. 15070063)	Female	2005
DFT1 – 1426 (Deakin *et al.* 2012) (RRID:CVCL_LB76)	RMPI-1640 + glutamine, 10 % heat inactivated FBS, 50 µg/mL penicillin, 50 µg/mL streptomycin	Female	2005
DFT1 – C5065 (Deakin *et al.* 2012) (RRID:CVCL_LB79)	RMPI-1640 + glutamine, 10 % heat inactivated FBS, 50 µg/mL penicillin, 50 µg/mL streptomycin	Female	2005
DFT2 - RV (Pye *et al.* 2016b)	RMPI-1640 + glutamine, 10 % heat inactivated FBS, 50 µg/mL penicillin, 50 µg/mL streptomycin	Male	2014
RV_UC	RMPI-1640 + glutamine, 10 % heat inactivated FBS, 50 µg/mL penicillin, 50 µg/mL streptomycin, 0.8 mg/mL G418 (Gibco, Cat no. 11811023)	Male	2014 (Transfected 2017)
RV_UK	RMPI-1640 + glutamine, 10 % heat inactivated FBS, 50 µg/mL penicillin, 50 µg/mL streptomycin, 0.8 mg/mL G418	Male	2014 (Transfected 2017)
DFT2 – SN (Pye *et al.* 2016b)	RMPI-1640 + glutamine, 10 % heat inactivated FBS, 50 µg/mL penicillin, 50 µg/mL streptomycin	Male	2014
DFT2 – TD549 (Caldwell *et al.* 2018)	RMPI-1640 + glutamine, 10 % heat inactivated FBS, 50 µg/mL penicillin, 50 µg/mL streptomycin	Male	2015
Fibroblasts – Salem (Murchison *et al.* 2012)	DMEM + glutamine (500 mL, Gibco, Cat no. 10566-016), 10 % FBS (500 mL, Gibco, Cat no. 10099141), 50 µg/mL penicillin, 50 µg/mL streptomycin	Female	2005
CHO (Provided by J. Kaufman)	DMEM/F-12 + glutamine (500 mL, Gibco, Cat no. 11320033), 10 % FBS, 50 µg/mL penicillin, 50 µg/mL streptomycin	N/A	N/A
CHO_IFNγ (Siddle *et al.* 2013)	DMEM/F-12 + glutamine, 10 % FBS, 50 µg/mL penicillin, 50 µg/mL streptomycin	N/A	N/A

2.2 Tumour and tissue samples

2.2.1 Sample collection

Samples in this project are derived from captive bred Tasmanian devils euthanised for health or age related causes, and from wild Tasmanian devils either trapped during monitoring trips or found dead due to road strike or other causes. Briefly, devils are trapped in custom made pipe traps (constructed from 30 cm diameter PVC pipe) and the captured animal is transferred by gently sliding it from the culvert pipe trap into a hessian sack and handled by the investigator. The animal's eyes are covered during the handling process to reduce stress. Tumour biopsies were taken post-mortem or from live animals which were released immediately after processing. Tissue samples were collected post-mortem. Tissue samples were either fixed in 10 % neutral buffered formalin for paraffin embedding or preserved in RNAlater for gene expression analysis. Samples were collected by collaborators at the University of Tasmania. All animal procedures were performed under a Standard Operating Procedure approved by the General Manager, Natural and Cultural Heritage Division, Tasmanian Government Department of Primary Industries, Parks, Water and the Environment or under University of Tasmania Animal Ethics Committee Permit A0014976.

2.3 Proteomics

2.3.1 Sample preparation

DFT1_4906 cells treated with IFNγ (described above), DFT2_RV cells and Salem cells were washed 3x with 1X PBS and stored at -80 ºC until lysis. Pellets were thawed and resuspended in lysis buffer (0.1 % SDS (Sigma-Aldrich, Cat no. 71725) + 0.1 M TEAB (Sigma-Aldrich, Cat no. 86600)). Cells were lysed on ice using a fine probe sonicator. Protein concentration was determined using a Direct Detect® Infrared Spectrometer (MerckMillipore) and 25 µg of protein was made up to 20 µL in water. Protein was reduced for 1 hr at 60 ºC with 1 mM Di-thiothreitol (Melford, Cat no. MB1015) and alkylated for 45 min at room temperature in the dark with 5.5 mM Iodoacetamide (Sigma-Aldrich, Cat no. I1149). Samples were digested overnight at 37 ºC with 2 µg Trypsin gold (Promega, Cat no. V5280). Digested samples were dehydrated and resuspended in 0.1 % trifluoroacetic acid (TFA) (Sigma-Aldrich, Cat no. T62200). A C18 clean-up was performed using an Empore High Performance Extraction disk plate. Methanol was added to plate to wet the silica. Plate was washed with elution solvent (80 % acetonitrile (Fisher scientific, Cat no. A998-1) and 0.5 % acetic acid (Fisher scientific, Cat no. 10171460)) followed by wash solvent (0.5 % acetic acid), with centrifugation at 100 RCF for 2 min. Samples were added and centrifuged at 250 RCF for 1 min. The plate was washed 2x with wash solvent before elution with elution solvent. Samples were dehydrated and stored at -80 ºC.

2.3.2 Mass spectrometry

Cell pellets (10^9 cells per replicate, grown by Annalisa Gastaldello) were shipped on dry ice to Sri Ramarathinam at Monash University in Melbourne, Australia. Samples were fractionated into seven fractions to increase proteome coverage. All proteomic analysis was performed in triplicate.

2.3.3 Protein identification and quantification

Proteins were identified from the mass-spectrometry output using Peaks software. Peaks combines classical database searching with de-novo mass spectra peptide identification to map tryptic peptides from the spectra back to their predicted source protein, as described in Zhang *et al.* (2012). Peptide identification was performed using a custom-made database from Alistair Bailey which combines the Tasmanian devil genome assembly available on Ensembl (Hunt *et al.* 2018) with transcriptome data for DFT1_4906 (+IFNγ) and DFT2_RV to identify mutant proteins as well as normal. An R script was kindly written by Alistair Bailey to clean up the data and remove duplicates and contaminants. Proteins were kindly quantified using label-free methods by Alistair Bailey in Peaks. Briefly, peptides identified in the raw mass spectrometry data for each individual replicate were filtered to remove low quality and confidence peptides from analysis and corrected for inconsistencies in HPLC separation following fractionation. Peptide sequences are then assigned across samples and signal intensity ratio was calculated from the raw MS/MS data to give a peptide ratio for each identified peptide. The top few peptide ratios from each identified protein were then averaged to give a protein ratio. Expression values were then quantified based on the calculated ratio of an internal standard protein of known concentration. The significance of identification was evaluated based on the quality of the peptide identification process and on identification in multiple replicates. Proteins were then annotated against the custom identification database to identify Ensembl transcript, protein and gene IDs for analysis. Only proteins identified and quantified in all three replicates of each cell line are included in final analysis.

2.3.4 Data analysis

For unquantified data, Ensembl gene IDs were filtered to remove duplicates and directly compared. For quantified datasets, proteins were manually annotated to fill in missing Official Gene Symbols, using homology searching where the protein is not fully annotated in Ensembl. A significance cut threshold of $p < 0.01$ was used to ensure only high confidence proteins were included. Direct comparison of protein expression values between cell lines was performed in R (R Core Team 2013). Briefly, replicate raw expression values were transformed to a log2 scale and normalised based on median expression across cell lines. Missing values were imputed based on a downshifted mean and standard deviation. Hierarchical clustering was performed on imputed expression datasets to identify similarities between cell lines using the hclust function within the heatmap visualisation function in R. R script for normalisation, imputation and hierarchical clustering of protein expression is presented in Appendix B.1 (Data used presented in Appendix C.2).

Proteins were ranked in each cell line based on average peptide intensity across all three replicates, from highest to lowest. The ranked protein list for each cell line was then uploaded individually to the g:Profiler g:GOSt web application (Reimand *et al.* 2016; Reimand *et al.* 2007), which carries out functional enrichment analysis of gene lists against multiple functional databases. The databases used

in this analysis are; Gene Ontology (GO) Consortium Biological Process (GO:BP) database, Gene Ontology Consortium Cellular Compartment (GO:CC) database (The Gene Ontology Consortium 2018; Ashburner *et al.* 2000), C) Reactome (REAC) database (Fabregat *et al.* 2018) and Human Protein Atlas (HPA) database (Thul *et al.* 2017) (Human Protein Atlas available from www.proteinatlas.org). Analysis was run using a human background to avoid loss of information due to poor functional annotation of the Tasmanian devil genome. Functional analysis used a significance cut off of p-value < 0.05, where the p-value is a significance value corrected for multiple testing using the g:SCS algorithm, a custom made algorithm which considers hierarchical nature of functional analysis terms which has been validated in Reimand *et al.* (2016) and Reimand *et al.* (2007). Functional enrichment terms were also normalised, imputed and clustered using the same R code as protein expression to compare broad functional enrichment (Appendix B.1, Appendix C.3). Functional analysis was also performed in the exact same manner on A) Proteins uniquely quantified in a single cell line (for DFT1, DFT2 and fibroblasts), B) Proteins quantified in DFT2 which were not quantified in fibroblasts, C) Proteins quantified in DFT1 which were not quantified in fibroblasts and D) Proteins quantified in all three cell lines plus proteins unique to fibroblasts.

2.4 Antibodies

Table 2.2: Antibodies used in the generation of this PhD thesis.

U/A/B/C_15-25-18 probes for all devil classical MHC class I proteins. UK_15-29-1 probes for the Saha-UK non-classical MHC class I protein.

Antibody (Cat. no)	1ry or 2ry	Obtained from	Working dilution
S100 (Z0311)	Primary	Dako	1:400
Myelin basic protein - MBP (A0623)	Primary	Dako	1:200
Periaxin (HPA001868)	Primary	Sigma-Aldrich	1:300
Myc (9E10) (13-2500)	Primary	Thermo scientific	1:100 1:1 ratio with Protein A beads for IPs and Immunoaffinity
β_2-m_13-34-48	Primary	Dr Hannah Siddle and Prof. Karsten Skjødt	1:10 of purified 0.5 mg/mL 1:2 of hybridoma supernatant for flow cytometry
β_2-m_13-34-38	Primary	Dr Hannah Siddle and Prof. Karsten Skjødt	1:1 ratio with Protein A beads for IPs and Immunoaffinity
UA/B/C_15-25-18 (Tasmanian devil classical MHC class I heavy chain)	Primary	Dr Hannah Siddle and Prof. Karsten Skjødt	Neat hybridoma supernatant
UK_15-29-1 (Saha-UK heavy chain)	Primary	Dr Hannah Siddle and Prof. Karsten Skjødt	0.5 mg/mL purified neat
FABP7 (51010-1-AP)	Primary	ProteinTech	1:100
PDGFRA (ab124392)	Primary	Abcam	1:100
Goat anti-rabbit Alexa Fluor 488 (R37116)	Secondary	Life technologies	1:1000
Goat anti-mouse Alexa Fluor 488 (A11001)	Secondary	Life technologies	1:1000
Goat anti-mouse Alexa Fluor 647 (A-21235)	Secondary	Invitrogen	1:1000
Anti-mouse 680 (925-68070)	Secondary	LiCor	1:15000
Anti-rabbit 680 (925-68071)	Secondary	LiCor	1:15000

2.5 Immunocytochemistry (ICC)

2.5.1 ICC

Cell lines were seeded into 8 well chamber slides at $1x10^5$ cells/mL and incubated at 35 ºC overnight in standard media as described in Table 2.1. Cells were washed in 1X PBS for 5 min, immersed in ice cold 1:1 methanol-acetone for 10 min, and washed in 1X PBS for 5 min. Cells were covered with blocking buffer (PBS + 10 % goat serum (Biowest, Cat no. S2000-500) + 0.1 mM EDTA (Fisher scientific, Cat no. AM9260G)) and incubated at 37 ºC for 30 min to prevent non-specific antibody interactions. Primary antibodies were diluted to the required concentration (Table 2.2) in blocking buffer and cells were incubated for 30 min at room temperature. Cells were washed in 1X PBS for 3x 5 min. Secondary antibodies were diluted to the appropriate concentration in blocking buffer and cells were incubated in the dark at room temperature for 30 min before washing in 1X PBS for 3x 5 min. Coverslips were mounted onto the slides using Fluoroshield with DAPI (Sigma-Aldrich, Cat no. F6057-20ML) and left to dry in the dark. Slides were imaged on the Zeiss Axioplan microscope using appropriate filters. Images were analysed in ImageJ.

2.6 Immunohistochemistry (IHC)

2.6.1 Sectioning of tissue samples

Formalin-fixed, paraffin embedded (FFPE) samples were sectioned into 10 µm thick sections (unless otherwise specified) using a microtome. Sections were mounted onto microscope slides and dried at 37 ^{o}C for 48 hours.

2.6.2 Deparaffinisation and antigen retrieval

Sections were deparaffinised in xylene (Fisher scientific, Cat no. X3P-1GAL) for 2x 10 min, then rehydrated through successive 5 min immersions in graded ethanol (Fisher scientific, Cat no. 64-17-5) (2x 100 %, 95 %, 80 %, 70 %). Sections were washed 2x 3 min in ddH_2O before immersion in citrate buffer (10 mM citric acid (VWR, Cat no. 84841.290) + 25 mM NaOH (Fisher scientific, Cat no. 1310-73-2) in ddH_2O) and incubated at 95 ^{o}C for 45 min. Sections were left in citrate buffer to cool for 15 min then washed under running water (1x 3 min) and 2x 3 min ddH_2O.

2.6.3 IHC

Sections were covered in Dual endogenous enzyme block (Dako – EnVision + Dual link system kit, Cat no. K4063) for 10 min then washed in running water for 5 min, dipped in ddH_2O, then washed in 1X PBS 3x 5 min. Sections were incubated in 10 % goat serum in 1X PBS for 30 min followed by immediate incubation with diluted primary antibody (Table 2.2) at 4 ^{o}C overnight, then washed 4x 3 min with 1X PBS. Sections were incubated in labelled polymer (Dako kit) in the dark at room temperature for 30 min, then washed 3x 5 min with 1X PBS. Sections were incubated with substrate-chromagen (Dako kit) for 5 min at room temperature and immediately washed in ddH_2O for 5 min then under running water for 5 min. Slides were immersed in Gill's haematoxylin (Sigma-aldrich, Cat no. GHS1128) for 4 min and immediately rinsed under running water for 5 min. Slides were placed in 0.2 % ammonia water for 2 min and rinsed in running water for 5 min.

2.6.4 Dehydration and coverslip mounting

Sections were dehydrated through successive 5 min immersion in graded ethanol (70%, 80%, 95%, 2x 100%) followed by 2x 10 min immersion in xylene and air dried before a coverslip was mounted with Vectamount mounting agent (Vector, Cat no. H-5000). Imaging was carried out using the QImaging Retiga 2000R colour imaging system.

2.7 cDNA generation

2.7.1 Nucleospin RNA extraction

RNA was extracted from cell lines from pellets of up to $5x10^6$ cells and from up to 30 mg of devil tissue samples frozen in RNA-later using a Nucleospin RNA isolation kit (Macherey-nagel, Cat no. 740955.50) and following the manufacturer's instructions. Briefly, the cell pellet was lysed with Buffer RA1 + β-mercaptoethanol (National diagnostics, Cat no. EC-603) and filtered through a NucleoSpin filter, then precipitated with ethanol and run through an RNA column. The column membrane was de-salted, contaminating DNA removed with DNase and the column was washed and dried to remove ethanol from membrane. RNA was eluted from the column in RNase free water and the concentration and purity (260/280) of the RNA was measured on a Nanodrop 2000 before storage at -80 ^oC.

2.7.1 Trizol RNA extraction

RNA was extracted from up to 100 mg of devil sciatic nerve tissue using TRI Reagent (Sigma-aldrich, Cat no. T9424) as per manufacturer's instructions. Briefly, tissue was homogenised in TRI reagent in a glass homogeniser and centrifuged to remove insoluble material. Tissue lysate was then phase separated with chloroform and centrifuged, and the clear RNA phase was precipitated using isopropanol and washed in 75 % ethanol before air drying to remove ethanol residue and resuspension in ddH_2O. The concentration and purity (260/280) of the RNA was measured on a Nanodrop 2000 before storing at -80 ^oC.

2.7.2 Reverse transcription

1000 ng of RNA was reverse transcribed to cDNA using the following reaction;
1X Revertaid premium RT buffer (Thermoscientific, 5X stock, Cat. no. EP0441), 200 U Revertaid premium Reverse Transcriptase (Thermoscientific, 200 U/µL stock, Cat. no. EP0441), 1000 ng total RNA, 20 µM of Oligo $(dT)_{15}$ primer (Promega, Cat no. C1101), 0.5 mM dNTPs (VWR) in a total volume of 15 µL with RNase free H_2O.
The RNA, oligo $(dT)_{15}$, dNTPs and H_2O were mixed and heated to 65 ^oC for 5 min and immediately put on ice. The Reverse Transcriptase and 5X buffer were added and the reaction was incubated at 60 ^oC for 30 min, followed by deactivation of the enzyme at 80 ^oC for 15 min.

2.8 PCR

2.8.1 Primer design

Primers were designed for several genes of interest, including MHC class I genes, Schwann cell associated genes and immune associated genes selected based on their relevance to Schwann cells, cancer or immune function in other species, and the presence of the gene in the devil genome and/or proteome. Primers were designed to overlap exon boundaries to allow use in qPCR. Primers were designed from the DEVIL7.0 assembly of the devil genome (GCA_000189315.1) accessed via Ensembl. Primer chemical properties were checked using OligoCalc (Kibbe 2007) and primers were ordered from and made by Eurofins Genomics. A full list of primers used in this project is included in Appendix B.2.

2.8.2 RT-PCR

A master mix containing the following reagents was made up on ice for each primer set;
1X HF buffer (Thermoscientific, 5X stock, Cat no. F518L), 200 µM dNTPs, 0.4 µM forward primer, 0.4 µM reverse primer, 0.016 U/µL Phusion *Taq* polymerase (Thermoscientific, 2 U/µL stock, Cat no. F530S) in a total volume of 25 µL per reaction ddH_2O.
Ribosomal protein L13a (RPL13A) was used as a housekeeping control. 250 ng cDNA was added separately to each individual reaction.
Tubes were sealed and put on a thermal cycler for the following programme:

98 ^oC for 30 sec, 98 ^oC for 10 sec, Optimised primer annealing temperature (55-65 ^oC) or optimisation gradient for 30 sec, 72 ^oC for 30 sec, Return to step 2 for 25-35 cycles, 72 ^oC for 5 min.

2.8.3 Gel electrophoresis

PCR products were analysed on a 1.2 % agarose gel prepared with 1X TAE and 1X Gel red (Biotium, 10000X stock, Cat no. #41003)). 1X loading dye (6X stock) was added to each PCR reaction and 10-15 µL of each reaction was loaded into each well on the gel alongside a 100 kb DNA ladder (Invitrogen, Cat. no. 15628-050) to estimate amplicon size. The gel was run at 90-100 V for 25-50 min in 1X TAE running buffer depending on gel size and required resolution and imaged on a Syngene G:BOX.

2.8.4 Extraction of DNA from agarose gels

The required DNA band on an agarose gel was identified based on molecular weight and the DNA was excised from the gel under a UV light box. The DNA was extracted using a PCR clean-up and gel extraction kit (Macherey-nagel, Cat. no. 740609.50) according to the manufacturer's instructions.

Briefly, agarose gel was dissolved in Buffer NTI and run through a clean-up column, washed twice and dried to remove ethanol from the membrane before elution in 15 µL of Buffer NE and purified. DNA was quantified on a NanoDrop 2000c (Thermoscientific).

2.9 MHC class I gene cloning

2.9.1 Amplification of MHC class I genes

Two MHC class I alleles (*Saha-UK* and *SahaI*27-1*) were selected for further investigation based on their high expression levels in DFT1 and DFT2. cDNA was generated from DFT1 cells treated with IFNγ as described in section 2.1.1. Two initial PCR reactions were performed using primers that amplify *Saha-UA*, *Saha-UB* and *Saha-UC* and *Saha-UK*. To amplify the full coding sequence of the class I heavy chain, the primers were designed in regions where there is little variation between the different MHC class I genes, meaning the PCR reaction amplifies all classical MHC class I alleles (*Saha-UA/B/C*) in the cDNA sample due to high sequence similarity. The amplicon from the first round of PCR was extracted from the agarose gel (Section 2.8.4) and a second round of PCR was performed, using a forward primer specific for *Saha-UA*, *Saha-UB* and *Saha-UC* or specific for *Saha-UK* and a reverse primer universal for all genes with a myc tag at its 5' terminus. The forward primers include an *NheI* restriction site and the reverse includes an *EcoRI* restriction site. Primer sequences are shown in Table 2.3. The product of the second round of PCR was extracted from the agarose gel and cloned into the pJET cloning vector (Section 2.9.2). Inserts were screened for the alleles of interest by sequencing as described in Section 2.10.1.

Table 2.3: Devil MHC class I primer sequences used during cloning.

Primers were designed using the available devil genome from Ensembl. Primers are designed to amplify the full length of the gene, including all exons. Red text indicates *NheI* restriction site, blue text indicates *EcoRI* restriction site. Myc tag is indicated in bold.

Primer	Annealing temp	Sequence
UA/B/C forward	60°C	AGTTA**GCTAGC**GCCACCATGGGCTCTCCGGCGCG CGC
UA/B/C reverse		GTCTCTCTGACAGCCAAAGAACAAAAACTTATTTC
UK forward	59°C	AGTTA**GCTAGC**GCCACCATGGAACCTTTGAAGCTC
UK reverse		GAAATAAGTTTTTGTTCAGCTTTGGCTGTCAGAGA G
Common class I reverse	As for forward primer	ATAGC**GAATTC**TCTAGAT**CAGATCTTCTTCAGAA ATAAGTTTTTG**

2.9.2 Blunt end cloning into pJET

Genes were cloned into pJET using the CloneJET PCR cloning kit (Thermoscientific, Cat no. K1231). The amount of PCR product required for the ligation reaction was calculated using the following equation:

$$\frac{ng\ of\ vector\ x\ length\ of\ insert\ (kB)}{length\ of\ vector\ (kB)}\ x\ \frac{3}{1} = ng\ of\ insert$$

On ice, 1X reaction buffer (2X stock), calculated ng of PCR product, 50 ng pJET1.2/blunt Cloning Vector (50 ng/µL stock) and 5 U T4 DNA Ligase (5 U/µL stock) were combined, made up to 20 µL with autoclaved H_2O and incubated at room temperature for 20 min before use in the transformation reaction.

2.9.3 Sticky end cloning into pIRES and pcDNA3.0 vectors

The amount of PCR product required for the ligation was calculated using the equation described in Section 2.9.2. Ligation was performed using the following reaction:

1X Rapid ligation buffer (Promega, 2X stock, Cat. no. C671A), 50 ng of digested vector, the calculated amount of insert and 1-3 U of T4 DNA ligase (Promega, 1-3 U/µL stock, Cat. no. M180A) and made up to 12 µL with autoclaved H_2O.

The reaction was mixed and incubated at room temperature for 20 min then used directly in a transformation into *E. coli*, with resulting colonies screened for recombinant construct and the pIRES or pcDNA3.0 construct isolated.

2.9.4 Restriction enzyme digest

For cloning into pIRES, plasmid and vector were digested with *EcoRI* and *NheI*. For cloning into pcDNA3.0, plasmid and vector were digested with *XhoI* and *XbaI* (See Figure 2.1).

1 µg of DNA or vector was added to 1X Multi-core buffer (Promega, 10X stock, Cat. no. R999A), 5 µg BSA (Promega, 10 mg/mL stock, Cat no. R396E), at least 1 U enzyme per 10ul reaction (*EcoRI* (Promega, 12 U/µL stock, Cat no. R601A), *NheI* (Promega, 10 U/µL stock, Cat. no. R650A), *XhoI* (Promega, 10 U/µL stock, Cat no. R616A), *XbaI* GQ (Promega, 10 U/µL stock, Cat no. R618A)), made up to 50 µL with autoclaved H_2O in a PCR tube and incubated at 37 ^oC for 90 min, then at 65 ^oC for 15 min to heat inactivate the enzymes. Restriction digests were then run on an agarose gel and the required band (either the digested vector or the digested insert) was cut out and purified.

2.9.5 Sequencing

15 µL of plasmid between 50-100 ng/mL is required for each sequencing reaction. To sequence the entire gene, two separate reactions with primer T7 in the plasmid and primer 85 (Appendix B.2) from exon 2 were run. 30 µL of each isolated plasmid was sent along with primer 85 to Eurofins Genomics for sequencing. Sequence data was then analysed using CLC Viewer and the identity of the gene was confirmed.

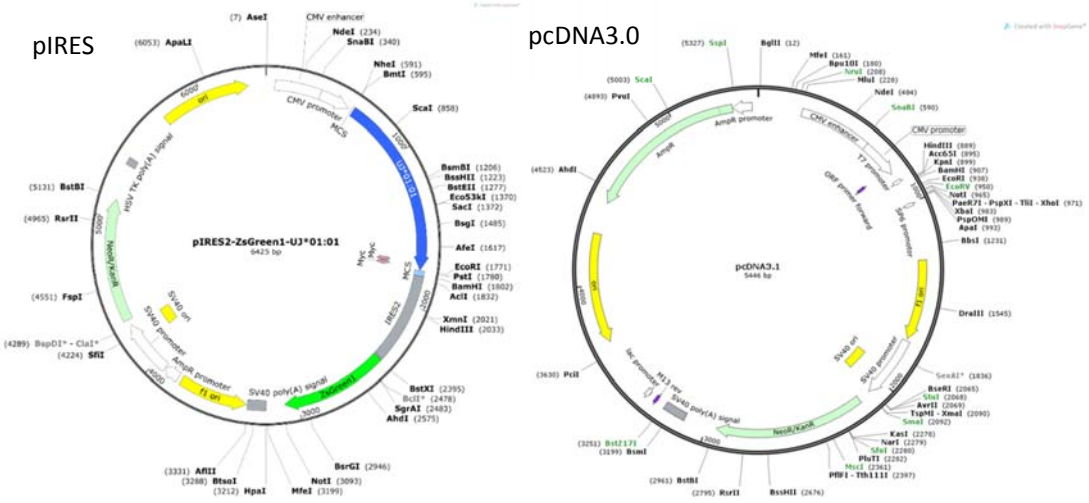


Figure 2.1: Schematics of the two mammalian expression vectors used in this project, pIRES and pcDNA3.0

Plasmids contain a cytomegalovirus (CMV) promoter prior to the insertion site to induce gene expression and an ampicillin resistance gene for selection.

2.10 Site-directed mutagenesis

2.10.1 Site-directed mutagenesis

Site-directed mutagenesis was used to insert an additional myc tag into the pIRES constructs. Mutagenesis primers containing an extra myc tag were designed using NEBaseChanger (Table 2.4). Site-directed mutagenesis was performed using the Q5 Site-directed mutagenesis kit (New England Biolabs, Cat. no. E0554S). In a PCR tube, 1X Q5 hot start high-fidelity master mix (2X stock) was combined with 5 µM of each primer, 1-25 ng of purified plasmid and made up to 25 µL with autoclaved H_2O. The reaction underwent the following PCR cycles: 98 ^oC for 30 sec, 98 ^oC for 10 sec, 62 ^oC for 30 sec, 72 ^oC for 30 sec, Return to step 2 for total 25 cycles, 72 ^oC for 2 min.

1 µL of this reaction was combined with 1X KLD Reaction Buffer (2X stock), 1X KLD Enzyme Mix (10X stock), made up to 10 µL with autoclaved H_2O and incubated at room temperature for 5 min before transformation into E. coli (Section 2.11). Colonies were screened by PCR using primers that flanked the myc insertion site to identify larger inserts (Section 2.8.2) and then sequenced to identify the second myc tag (Section 2.9.5).

Table 2.4: Primers used to perform site-directed mutagenesis in this PhD thesis.

Primer	Annealing temp	Sequence
347 F	62^oC	AGCGAAGAAGATCTGATCTAGAGAATTCGCTATATCTTTC
348 R		AATCAGTTTCTGTTCCAGATCTTCTTCAGAAATAAGTTTTTG

2.11 Transformation of *E. coli*

2.11.1 Transformation and growth of *E. coli*

2 µL of product from Section 2.9.2 or 2.9.3 was added to 50 µL of recently thawed, competent DH5α *E. coli* cells, gently mixed and incubated on ice for 20 min, then heat shocked at 42 ºC for 1 min and put immediately back on ice for 2 min. 450 µL of room temperature LB Broth (Media kitchen) was added and incubated with shaking at 37 ºC for an hr. Ampicillin (AppliChem) was added to melted LB Agar to a final concentration of 100 µg/mL and poured into sterile culture plates. Cells were removed from the incubator, 120 µL spread on each LB Agar plate before incubation at 37 ºC overnight. Single colonies were picked using a pipette tip and incubated in 50 µL LB Broth with 100 µg/mL ampicillin and incubated for two hours at 37 ºC.

2.11.2 Screening of transformed colonies by PCR

Colonies were screened for the insert using a colony PCR with the following reaction;
1X Red *Taq* DNA polymerase master mix (VWR, 2X stock, Cat no. 733-2130), 0.5 µM of each primer (Primer 85 and 86, Appendix B.2), 1 µL of cultured colony cells, to 10 µL reaction volume with autoclaved H_2O.
The PCR cycling conditions were as follows: 95 ºC for 2 min, 95 ºC for 30 sec, Optimised primer temperature (55-65 ºC) for 30 sec, 72 ºC for 30 sec, return to step 2 25-30x, 72 ºC for 5 min.

An agarose gel was run to identify colonies with the transformed gene.

2.11.3 Purification of plasmid DNA

A glycerol stock of each colony was made by adding 850 µL of culture to 150 µL of glycerol (Fisher scientific, Cat no. BP229-1) and stored at -80 ºC. Remaining cells were centrifuged at 4600 rpm for 10 min and all supernatant removed. The plasmid was isolated from the cells using the Plasmid DNA purification kit (Macherey-nagel, Cat. no. 740499.50) according to the manufacturer's instructions. Briefly, pellets were resuspended and lysed, and the lysate clarified using a clarification column and run through a plasmid column. Plasmid column was washed twice and dried to remove ethanol from the membrane before elution of the plasmid in 50 µL of Buffer AE and measurement of plasmid concentration using a Nanodrop 2000.

2.12 Transfection of mammalian cells

2.12.1 Transfection

Transfection into RV cells was optimised using green fluorescent protein (GFP), and then performed on RV, 4906 and CHO cells using the completed pIRES and pcDNA3.0 constructs. Cells were plated into 6 well cell culture plates at $1x10^6$ cells/mL (RV and 4906) or $3x10^5$ cells/mL (CHO) and grown to ~80 % confluency. The DNA was mixed with Fugene HD transfection reagent (Promega) at a range of ratios during optimisation (4:1, 3:1, 2:1) in RV, and eventually used at a 2:1 ratio in RV and 4906, and a 3:1 ratio in CHO cells for the final transfections, then incubated at room temperature for 10 min. This reaction was added to each well of cells to give a final DNA concentration of 2 μg per well and incubated at 35 ºC for 24 hours before the media was discarded and replaced with fresh media containing 0.8 mg/mL G418 (Gibco, Cat no. 11811023) to select for transfected cells. During optimisation using GFP, transfection success was determined by identification of fluorescent cells under a microscope. Transfection success of experimentally transfected cells was determined by the presence of live cells in culture following 1 week of G418 treatment and confirmed by Western blot and flow cytometry (Sections 2.13 and 2.14).

2.13 Flow cytometry

2.13.1 Cell preparation

Cells were harvested, counted and washed 3x in 1X PBS before incubation with primary antibody (Table 2.2) for 30 min on ice. Cells were washed 3x in FACS buffer (1 % BSA powder (Sigma-aldrich, Cat no. 05482) in PBS) and resuspended in secondary antibody (Table 2.2) before incubation in the dark on ice for 30 min. Cells were washed 3x in FACS buffer and resuspended in FACS buffer for analysis. For cell death analysis, propidium iodide (PI) solution was added to samples at a final concentration of 0.1 mg/mL, incubated on ice in the dark for 15 min and immediately run on the flow cytometer. PI fluoresces upon DNA binding but is unable to pass through an intact cell membrane, so only fluoresces in dead cells with ruptured membranes. Cells being stored for more than two hours prior to flow analysis were resuspended in FACS buffer + 2 % PFA to fix the cells.

2.13.2 Flow cytometry analysis

Cells were analysed on the FACSCalibur flow cytometry system (BD Biosciences) or the Merck Guava® easyCyte™ benchtop system. Briefly, control cells were used to correctly set up the detection and gating system for forward scatter, side scatter and an appropriate laser for the secondary antibody to ensure all cell populations were captured during the analysis. 10000 cells were analysed per sample, and a representative cell gating strategy is presented in Figure 2.2. Data was analysed using FlowJo (FACSCalibur) or Guava® InCyte software.

2.13.3 Fluorescence activated cell sorting

Following staining with β_2-m and appropriate gating systems set up using control cells, transfected cells were sorted with the assistance of Richard Jewell on the FACS AriaII (BD Biosciences) based on mean fluorescence intensity. Briefly, cells with higher fluorescence intensity than an untransfected control population were sorted into sterile PBS (Figure 7.7). Fresh, sterile media was added to sorted cells before cells were plated into 24 well cell culture plates (Corning, Cat no. CLS3527). Cells were maintained and expanded in G418 selection media as described in Section 2.1.1.

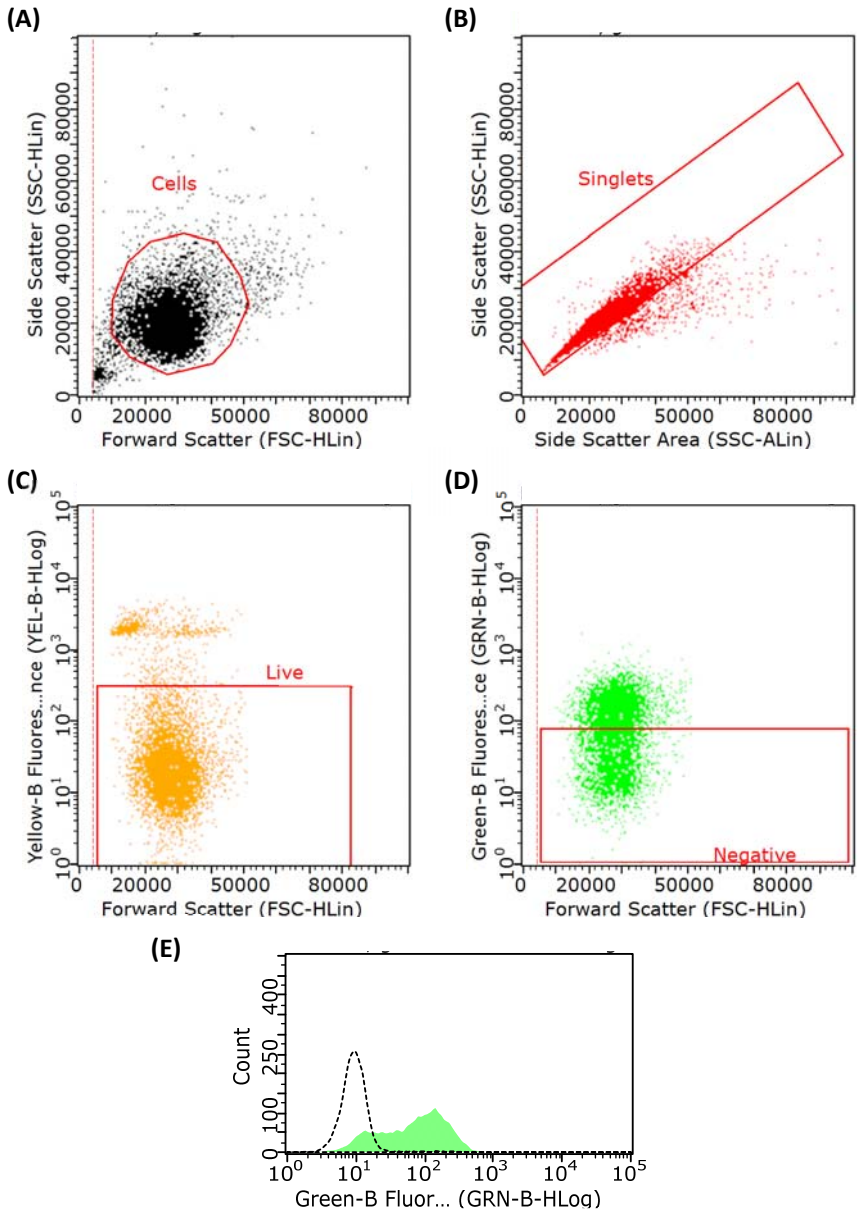


Figure 2.2: Representative gating strategy for flow cytometry analysis of DFT cell lines.

(A) Debris and dead or abnormal cells are removed from the analysis through gating based on forward and side scatter properties. Y axis represents laser side scatter, X axis represents forward scatter. **(B)** Plot gated on A. Droplets containing more than one cell (doublets) removed from analysis using side scatter properties. Y axis represents the same side scatter parameter as the Y axis in A), X axis represents the area under the side scatter curve which is proportional to the time taken for a droplet to pass through the detection laser. **(C)** Plot gated on B. Dead cells are removed from analysis based on PI fluorescence. Live population determined from unstained control cells. Y axis represents PI fluorescence on a log scale, X axis represents forward scatter properties. **(D)** Plot gated on C. β_2-m expression is measured using fluorophore fluorescence. Negative population identified from unstained control cells. Y axis represents fluorophore fluorescence, hence β_2-m expression on a log scale, X axis represents forward scatter properties **(E)** Expression in different samples is compared by histogram. Black dashed line is an unstained negative control sample. Green peak is a stained sample. Y axis represents cell count, X axis represents β_2-m expression (Green-B fluorescence) on a log scale.

2.14 Western blot

2.14.1 Buffers

Table 2.5: Buffers used for Western blot analysis in the preparation of this thesis.

2X Lysis buffer	**0.3 M NaCl, 0.1 M TrisCl, 2 mM MgCl2 in ddH_2O**
Lysis buffer with 1 % digitonin	2X Lysis buffer 1:2 with 2 % digitonin (Calbiochem, Cat no. 300410), 0.1 mM AEBSF (Sigma-aldrich, Cat no. A8456)
2X Laemmli sample buffer	120 mM Tris pH 6.8, 20 % glycerol, 4 % SDS, 0.02 % bromophenol blue (Sigma-aldrich, Cat no. 115-39-9) in ddH_2O
Loading buffer	2X Laemmli sample buffer 1:2 with 5 % β-mercaptoethanol in ddH_20
10X Running buffer	25 mM Tris, 192 mM glycine, 1 % SDS in ddH_2O
10X Transfer buffer	25 mM Tris, 192 mM glycine, 3.75 % SDS in ddH_2O
1X Transfer buffer	10X transfer buffer 1:10 with 20 % methanol and ddH_2O
10X TBS	150 mM Trizma hydrochloride (Sigma-aldrich, Cat no. 1185-53-1), 1.37 M NaCl in ddH_2O buffered to pH 7.6 with pure HCl.
TBST	10X TBS 1:10 with ddH_2O + 0.1 % Tween20 (Sigma-aldrich, Cat no. 9005-64-5)

2.14.2 Lysate preparation

Cells were harvested and the pellets washed 3x in 1X PBS, then freeze-thawed on dry ice and resuspended in Lysis buffer with 1 % digitonin, vortexed and incubated at 4 ºC with gentle mixing for 30 min. Lysates were centrifuged for 10 min at 11000 RCF and 4 ºC and the supernatant collected. Protein concentration in the lysates was determined by Bradfords assay using Pierce™ Coomassie (Bradford) Protein Assay Kit (Thermofisher Scientific, Cat no. 23200) as per manufacturer's instructions.

2.14.3 Western blot

An SDS resolving gel and stacking gels were made according to Table 2.6 and Table 2.7 with a higher percentage gel used for resolving smaller proteins. 30 µg of sample was made up to 25 µL with loading buffer and put on a heat block at 95 ºC for 10 min, then loaded into the centre wells of the gel along with a protein ladder (Geneflow, Cat no. S6-0024). 5 µl loading buffer was added to any empty wells, and the gel was then run in 1X running buffer in ddH_2O at 200 V until the ladder was well separated. 4x blotting paper, 2x sponges and 1x nitrocellulose membrane per gel were soaked in 1X transfer buffer for 30 min then stacked on the black side of the stacking case in the following order: 1x sponge, 2x blotting paper, gel, nitrocellulose membrane, 2x blotting paper, 1x sponge. The proteins were transferred to membrane in 1X transfer buffer at 100 V for 70 min at 4 ºC. Nitrocellulose membrane was incubated in 5 % milk powder (Marvel) in TBST for 1 hr at room temperature and washed 3x 10 min in TBST on a roller at room temperature. Primary Ab (Table 2.2) was diluted in TBST and incubated with the membrane overnight at 4 ºC. The membrane was washed 3x 10 min in TBST before incubation in secondary antibody (Goat α mouse 680 (Table 2.2)) at 1:1000 in 5 % milk at room temperature for 1 hr, then washed 3x 10 min with TBST and imaged on the Li-cor scanner.

Table 2.6: Reagents required to make 10mL of SDS-PAGE resolving gel.

Reagents	8% gel	12 % gel	15 % gel
Protogel 30 % (National diagnostics, Cat no. EC-890)	2.7 mL	4.0 mL	5.0 mL
Protogel resolving buffer (4X) (National diagnostics, Cat no. EC-892)	2.5 mL	2.5 mL	2.5 mL
ddH2O	4.8 mL	3.4 mL	2.4 mL
APS (Fisher chemical, Cat no. A/P470/46)	50 µL	50 µL	50 µL
TEMED (National diagnostics, Cat no. EC-503)	10 µL	10 µL	10 µL

Table 2.7: Reagents required to make 5mL of 4% SDS-PAGE stacking gel.

Reagents	Volume
Protogel 30 % (National diagnostics, Cat no. EC-890)	0.65 mL
Protogel stacking buffer (National diagnostics, Cat no. EC-893)	1.25 mL
ddH2O	3.05 mL
APS (Fisher chemical, Cat no. A/P470/46)	50 µL
TEMED (National diagnostics, Cat no. EC-503)	10 µL

2.15 Immunoprecipitation

2.15.1 Buffers

Table 2.8: Buffers used to perform immunoprecipitation.

1X Lysis buffer	**0.5 % IGEPAL 630 (Sigma-aldrich, Cat no. 9002-93-1), 50 mM Tris (pH 8), 150 mM NaCl, 10 µL/mL Halt Protease Inhibitor cocktail (100 X, Thermo scientific, Cat no. #1862209) in ddH_2O.**
Wash buffer	0.005 % IGEPAL 630, 50 mM Tris (pH 8), 150 mM NaCl, 5 mM EDTA, 10 µL/mL Protease inhibitor in ddH_2O.
Solution A	0.1 M Boric acid (Sigma-aldrich, Cat no. 10043-35-3), 0.1 M KCl in ddH_2O.
Solution B	0.1 M NaOH in ddH_2O.
Borate buffer	50 mL Solution A, 3.97 mL Solution B, 46.03 ml ddH_2O (pH 8).

2.15.1 Lysate preparation

Cells were harvested as described in section 2.1, resuspended in PBS and counted before being centrifuged at 400 RCF for 5 min and resuspended in 1 mL IP lysis buffer per 10^7 cells. Protein concentration was determined by Bradford assay as in section 2.14.2. Lysates were stored at -20 ^oC until required.

2.15.2 Immunoprecipitation

50 µL of rProtein A Sepharose Fast Flow beads (GE healthcare, Cat no. 17-1279-01) were washed with PBS 4x at 4 ^oC then incubated with 10 µg of antibody for 1.5 hours on a rotor at 4 ^oC. Beads were spun down at 11000 rpm for 2 min and the supernatant removed. A small amount of the supernatant and the beads were saved to check efficiency of antibody binding by Western blot. Beads were washed with Borate buffer (1X), followed by 50 mM Tris (pH 8) (1x) followed by Wash buffer (3x), then incubated overnight on rotor at 4 ^oC with 100 µg protein lysate, topped up to 300 µL with lysis buffer. Beads were spun down at 9000 rpm and the supernatant removed. A small amount of supernatant was saved to check protein binding efficiency by Western blot. Beads were washed with Wash buffer 3x, with supernatant from the first two washes saved to check binding efficiency. Beads were resuspended in 30 µL of Western blot loading buffer (Table 2.5) and stored at -20 ^oC until Western blot analysis.

2.16 Immunoaffinity purification of MHC class I complex

Protocol based on W6/32 immunoaffinity purification written by Ben Nicholas at University of Southampton.

2.16.1 Buffers

Table 2.9: Buffers used to perform immunoaffinity purification in this thesis.

Borate buffer	**0.1 M Boric acid, 0.1 M KCl, 4 mM NaOH in ddH_2O**
TeO buffer	0.2 M Triethanolamine (TeO) (Sigma-aldrich, Cat no. 102-71-6) in ddH_2O, adjusted to pH 8.2 with HCl
DMP crosslinker	40 mM Dimethyl pimelimidate (DMP) (Sigma-aldrich, Cat no. 58537-94-3) in TeO buffer, adjusted to pH 8.3 with NaOH
Citrate buffer	0.1 M Citric acid monohydrate (Fisher scientific, Cat no. 5949-29-1) in ddH_2O (pH 3.0)
0.2M TrisCl	2.42 g Tris in ddH_2O adjusted to pH 8 with HCl
Lysis buffer/ Wash buffer 1	20 mM TrisCl, 150 mM NaCl, 0.5 % (v/v) IGEPAL-630, 0.25 % (w/v) Sodium deoxycholate (Sigma-aldrich, Cat no. 302-95-4), 0.2 mM iodoacetamide (Sigma-aldrich, Cat no. 144-48-9), 1 mM EDTA, protease inhibitors (Sigma-aldrich, cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail, Cat no. 4693159001) in ddH_2O
Wash buffer 2	20 mM TrisCl, 150 mM NaCl in ddH_2O
Wash buffer 3	20 mM TrisCl, 400 mM NaCl in ddH_2O
Wash buffer 4	20 mM TrisCl in ddH_2O

2.16.2 Immunoaffinity column preparation

2 mL of rProtein A Sepharose Fast Flow beads were equilibrated to room temperature and washed with PBS, then centrifuged at 3000 RCF and resuspended in 10 mL PBS, then incubated with 2 mg antibody (β_2m (13-34-38) or myc – Table 2.2) for 60 min at 4 ºC with constant agitation. A 1.5 x 10 cm Econo-Column® Chromatography Column (Bio-rad, Cat no. #7371512) was washed with 10 % acetic acid followed by PBS and the Ab-bead mixture was transferred to the column and the flow through discarded. The beads were washed with borate buffer, followed by TeO buffer, then cross linked to the antibody using DMP crosslinker incubated at room temperature for 60 min. The reaction was terminated with ice cold 0.2M TrisCl which was allowed to run through the column. The beads were then washed with citrate buffer, followed by 50 mM TrisCl, and the flow through discarded. The beads were either used immediately for immunoaffinity or stored in a small amount of 50 mM TrisCl at 4 ºC for up to a week.

2.16.3 Lysate preparation

$4x10^8$ RV_UC cells were grown in hyperflasks as described in section 2.1.1, and stored as pellets at -80 ºC before use. 10^8 cells per replicate were thawed on ice and resuspended in 5 mL lysis buffer with repeated pipetting to break up the pellets, then incubated with agitation for 60 min at 4 ºC to fully lyse the cells. Lysate was transferred to 1.5 mL Eppendorf tubes and centrifuged at 2000 RCF for 10 min at 4 ºC to pellet aggregates. The lysate was transferred to fresh Eppendorf tubes and centrifuged at 15000 RCF for one hour at 4 ºC. Clarified supernatant was transferred to a fresh 15 mL falcon tube and kept on ice for immediate use in immunoaffinity purification.

2.16.4 Immunoaffinity purification

Myc antibody beads were resuspended in 2 mL of wash buffer 4 and added directly to the lysate. Lysate and beads were incubated at 4 ºC with rotation for 16 hours to capture the recombinant myc-tagged SahaI*27-1 construct. An Econo-column was cleaned with 10 % acetic acid followed by PBS, and β_2-m antibody beads were resuspended in 2 mL of wash buffer 4 in a fresh 15 mL falcon tube. The mycAb/lysate mix was added to the column, and the flow through collected directly into the falcon tube containing the β_2-m antibody, which was incubated for 16 hours with rotation at 4 ºC to capture endogenous MHC class I. Immediately following incubation, both myc and β_2-m beads were washed sequentially with wash buffers 1, 2, 3 and 4. Captured proteins were eluted from the columns with 5 mL 10 % acetic acid and collected into 3 x 2 mL Protein LoBind Eppendorf tubes (Eppendorf, Cat no. 0030108116. Samples were dried overnight at room temperature in centrifugal evaporator and stored at 4 ºC until separation by HPLC.

2.17 Immunopeptidomics

Protocols written by and carried out with the help of Ben Nicholas and Alistair Bailey at University of Southampton

2.17.1 High performance liquid chromatography (HPLC)

Dried samples from immunoaffinity were resuspended in 100 µL of 1 % TFA/1 % acetonitrile and collated into a single Eppendorf per sample. Used tubes were washed with 1 % TFA/1 % acetonitrile and added to the combined tube to a final total of 440 µL per sample. Samples were clarified by centrifugation at 15000 RCF for 10 min at 4 ºC and transferred to glass HPLC vials. Samples were run on a Dionex UltiMate 3000 standard HPLC system (Thermofisher scientific, Cat no. IQLAAAGABHFAPBMBEX) with a Chromolith® RP-18 endcapped column (MerckMillipore, Cat. no. 1021290001) of 100 mm length and 4.6 mm internal diameter with a biomodal pore structure (13 nm – 2 µm pore size) for sensitive size discrimination, using a pre-defined 30 min gradient run of 1 mL/min. 0.5 mL fractions were collected throughout the run. An example HPLC trace of the MHC class I complex is shown in Figure 2.3. The MHC class I peptides are not always visible on a trace, but elute just before β_2-m. The 16 fractions preceding the β_2-m peak were collected and odd and even fractions were combined into LowBind Eppendorf tubes (2x odd fractions, 2x even fractions) and dried overnight at room temperature in a centrifugal evaporator. Samples were stored at 4 ºC until mass spectrometry analysis.

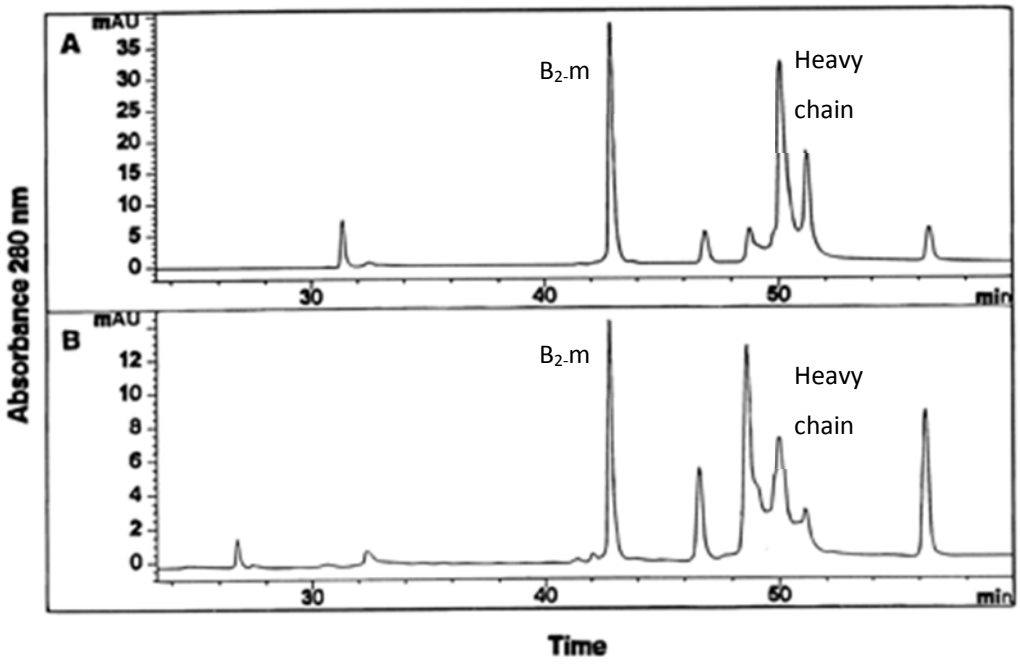


Figure 2.3: Example HPLC trace of an MHC class I fractionation.
X axis represents retention time, Y axis represents absorbance at 280nm. The β_2-m and heavy chain peaks are indicated. Peptides are found in the fractions preceding the β_2-m peak (Hörig *et al.* 1999).

2.17.2 Mass spectrometry

Samples were resuspended in 1 % formic acid and clarified by centrifugation. Mass spectrometry was performed by Dr Paul Skipp using a Waters G2-S Synapt HDMS mass spectrometer operating in MSe mode. Data was acquired from 50 to 2000 m/z with ion mobility enabled using alternate low and high collision energy (CE) scans. Low CE was 5 V and elevated, ramped from 20-40 V. The lock mass Glu-fibrinopeptide, (M+2H)+2, m/z = 785.8426) was infused at a concentration of 200 fmol/µL was infused at 300 nL/min and acquired every 13 seconds. The raw mass spectra were processed using ProteinLynx Global Server Ver 3.0 enabled through Symphony (Waters, Manchester, UK) and the data processed to generate reduced charge state and de-isotoped precursor and associated product ion mass lists. These mass lists were searched against a custom made Tasmanian devil database to determine peptide identity.

2.17.3 Peptide identification

Peptides were identified from the mass-spectrometry output using Peaks software as described in Section 2.3.3 against Alistair Bailey's custom-made Tasmanian devil database to identify wild type and mutant peptides and their source proteins. An R script was kindly written by Alistair Bailey to remove duplicates and contaminants. Peptides were quantified using label-free methods by Alistair Bailey in Peaks as described in Section 2.3.3.

2.17.4 Data analysis

Reproducibility of peptide sequences between replicates was calculated in Excel. Peptide length distributions and standard error of means were calculated in Excel for each immunoaffinity experiment across three replicates. Sequence alignment and simultaneous clustering of all peptides detected across all three replicates was initially performed using the GibbsCluster-2.0 server using the automatic standard input parameters (Andreatta *et al.* 2017; Andreatta *et al.* 2012). An MHC class I binding motif was produced manually by calculating the frequency of every amino acid at every position along all peptides detected in all three replicates using an Excel script written by Nathan Croft at Monash University in Melbourne, Australia. Analysis focusses on peptides of 8-10 amino acids in length. Amino acids with a frequency at a given position along the peptide of over 10 % are considered moderately enriched, over 20 % is considered strongly enriched and over 30 % is considered dominant. Peptide binding motifs were graphed manually based on this categorisation. Positive selection motifs were generated using the genomic frequencies of each amino acid in the Tasmanian devil followed by a Chi-squared test with Yates' and Bonferroni correction using an Excel script written by Ralf Schittenhelm at Monash University in Melbourne, Australia. Amino acids with calculated positive selection of >100 % are considered moderately enriched, over 200 % is considered strongly enriched and over 300 % is considered dominant. Only positively selected

amino acids which were significantly enriched following both Yates and Bonferroni correction were plotted.

Chapter 3 A molecular comparison of two clonally distinct transmissible tumours, DFT1 and DFT2

3.1 Introduction

Given the rarity of naturally occurring contagious cancers, the emergence of a second tumour with the ability to spread as an allograft in the Tasmanian devil is remarkable and raises questions about the origins and similarities between the two tumours.

DFT2 tumours have highly similar gross phenotypic features to DFT1, causing lesions predominantly around the face and oral cavity which grow quickly and become ulcerated and necrotic (Figure 3.1). However, early analysis of DFT2 biopsies found that these tumour cells do not stain for the Schwann cell marker PRX, a widely used specific diagnostic marker for DFT1 (Pye *et al.* 2016b). Genomic and karyotypic studies from Pye *et al.* (2016b) show that DFT2 is a genetically distinct transmissible cancer, likely to have formed separately to DFT1 (Figure 1.13, Figure 1.14).

Despite their differences, the phenotypic similarity of these tumours and the similar transmission mechanics suggest that they may share similar cellular origins. Indeed, a comparison of broad marker expression by DFT1 and DFT2 has indicated that DFT2 has arisen from the neuroectoderm (Stammnitz *et al.* 2018). A detailed understanding of these cellular origins of DFT2 is crucial for understanding the evolutionary mechanisms behind the emergence and immune evasion capabilities of this new tumour (Blanpain 2013). Importantly, without an understanding of the cellular origins of these tumours, identifying mutated or aberrantly expressed proteins which may be useful for developing treatments and disease management strategies is difficult.

We hypothesise that DFT2 has emerged from a similar cell type to DFT1. To test this hypothesis, I have generated proteomes for DFT1, DFT2 and devil fibroblast cell lines to compare protein expression between the tumours. There is well documented disparity between continuous cell lines and primary samples (Kaur and Dufour 2012), but cell lines reduce the possibility of contamination from host (self) cells and increases consistency across samples. A proteomics approach was undertaken due to the described poor correlation between mRNA and protein expression, with high levels of mRNA expression not always resulting in high levels of translated protein (Liu *et al.* 2016). A proteomic analysis of these tumours is therefore likely to be more functionally informative about their cellular origins.

Figure 3.1: DFT1 and DFT2 tumours are phenotypically indistinguishable.
Photographs of wild Tasmanian devils exhibiting characteristic facial tumours caused by A) DFT1 (Provided by Rodrigo Hamede) and B) DFT2 (Stammnitz *et al.* 2018)**.**

3.2 Aims and objectives

In this chapter I will generate proteomes for DFT1, DFT2 and fibroblasts cells, and analyse the similarities and differences in protein expression between the three cell types. I will make a direct comparison of DFT2 to DFT1 to identify similar proteins and biological processes which are enriched in both tumours and may indicate DFT2s likely cell type of origin. I will also identify major differences between the two tumours, which may represent differences in cellular origin, or different mutational events as the tumours have formed. The data generated in this chapter characterising the cellular origins of DFT2 will feed directly into further work on identifying DFT2 diagnostic biomarkers (Chapter 6) and tumour specific vaccine targets (Chapter 8).

This chapter has the following objectives:

1. To generate whole cell proteomes in triplicate from representative DFT1 and DFT2 cell lines and a representative healthy fibroblast cell line.

2. To identify similarities and differences in the protein expression profiles of DFT1 and DFT2.

3. To make direct comparisons between the functionally enriched biological pathways in DFT1 and DFT2, using fibroblasts as a healthy control cell for comparison.

4. To identify enriched biological pathways in DFT2 which may be indicative of the tumour's cell type of origin.

3.3 Results

3.3.1 The proteome of DFT2 is more similar to DFT1 than to fibroblasts

Three non-quantitative whole cell proteomes were generated for DFT2_RV, DFT1_4906 and fibroblasts. Proteomes were experimentally derived from mass spectrometry analysis on the whole cell lysates of each cell line in triplicate (Full triplicate proteome data presented in Appendix C.1). Duplicate Ensembl protein IDs within a single replicate were removed from analysis. The unique Ensembl protein IDs detected in each replicate for each cell line were compared to assess the reproducibility of the experiments. In all three cell lines, over two thirds of proteins were detected in all three replicates (Figure 3.2). The remaining third of detected proteins in all three samples is split almost equally between proteins detected in one or two replicates. This was considered to be reasonably good reproducibility based on published proteomics data and taking into account the use of a non-model species (Collins *et al.* 2017; Tabb *et al.* 2010). To ensure any conclusions generated from this data were robust, all further analysis presented here includes only proteins which were identified in all three replicates for each cell line, and remaining data was discarded from analysis.

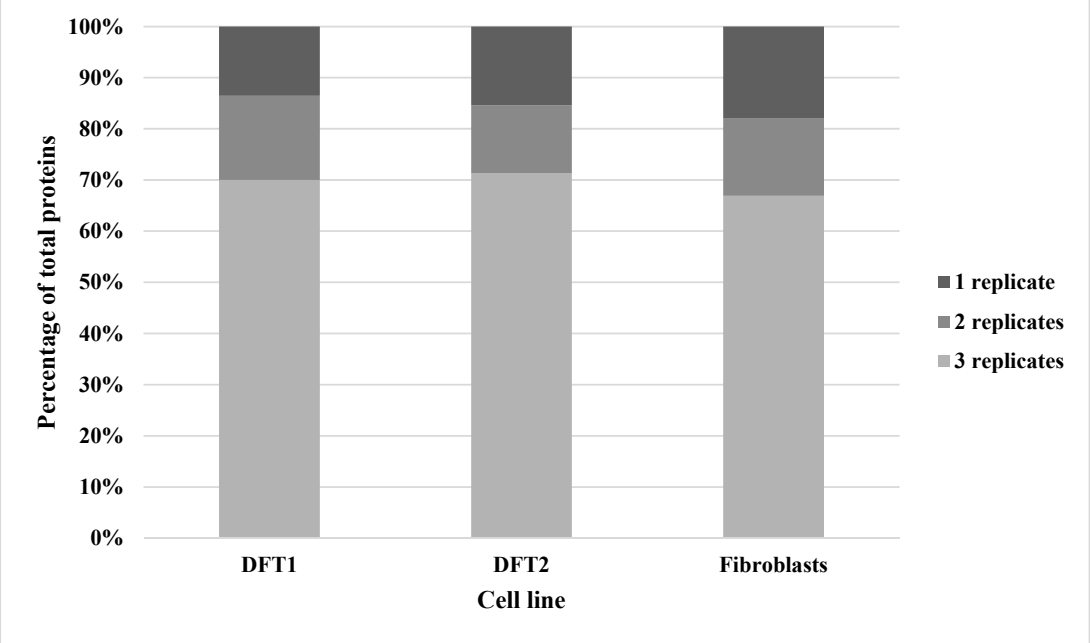


Figure 3.2: Over two thirds of proteins detected in DFT1, DFT2 and fibroblasts were detected in all three experimental replicates.

Stacked column charts indicating percentage of total unique proteins identified in each cell line that were identified in multiple experimental replicates. Y axis represents the percentage of the total number of unique Ensembl protein IDs identified in all replicates of a given cell line.

The resulting proteome sizes containing proteins identified in all three replicates of a given cell line were 4776 proteins identified in DFT2, 5169 in DFT1 and 4287 in fibroblasts. The majority of the experimentally determined proteins are shared between all three cell lines (Figure 3.3). There are 696 proteins found in all replicates of both DFT1 and DFT2 which are not identified in fibroblasts, as well as 481 proteins unique to DFT2 which may be of use for identifying DFT2s origins. DFT2 shares more experimentally detected proteins with DFT1 than with fibroblasts.

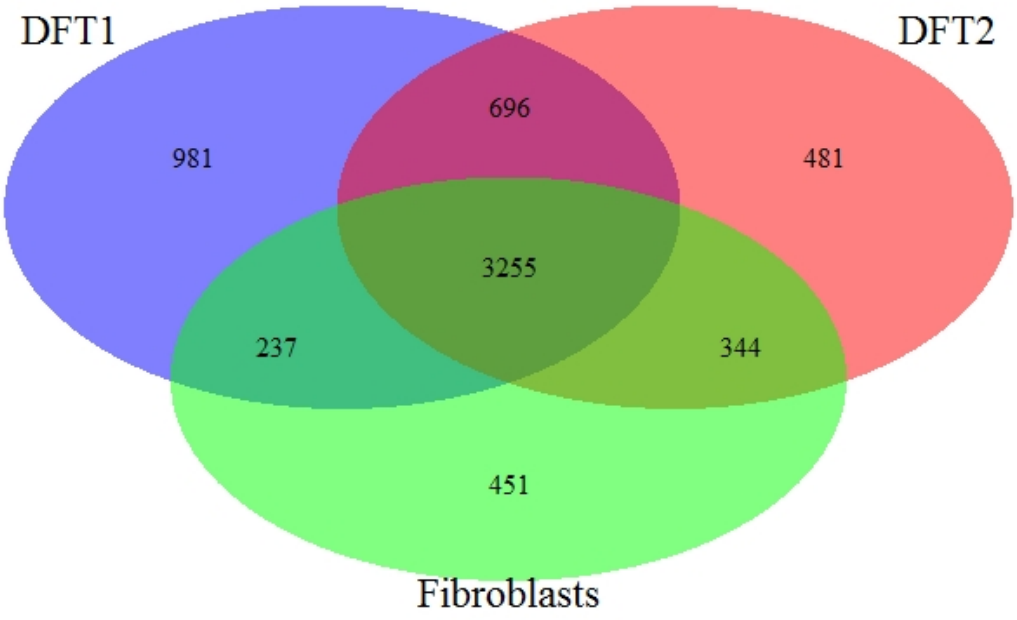


Figure 3.3: DFT1, DFT2 and fibroblasts share most experimentally detected proteins.
Venn diagram indicating the overlap of all experimentally detected proteins in DFT1 (blue), DFT2 (red) and fibroblasts (green). These proteins were not quantified. All proteins were detected in all replicates of each cell type.

A subset of proteins from each cell line was quantified using label-free methods in PEAKS mass spectrometry software (Full quantified proteome data presented in Appendix C.2) (Table 3.1). All quantified proteins were detected and quantified in all three replicates of a given cell line. Proteins detected in all three replicates of a given cell line were quantified using a significance cut off of 0.01. For DFT2, 1550 proteins were quantified of which 1453 were unique Ensembl peptide IDs. Ensembl peptide IDs refer to a specific isoform of a protein, and one protein ID may have multiple peptide IDs associated with it. Thus, 1453 unique peptide IDs corresponded to 1333 unique proteins. Due to a lack of functional annotation for multiple isoforms in the Tasmanian devil, all analysis was performed using the unique protein IDs. For DFT1, 1033 proteins were quantified, of which 970 were unique Ensembl peptide IDs corresponding to 904 unique proteins. For fibroblasts, 730 proteins were quantified, of which 701 were unique Ensembl peptide IDs corresponding to 657 unique proteins. Of these, 633, 283 and 169 proteins were uniquely quantified in DFT2, DFT1 and fibroblasts respectively, whilst 182 proteins were quantified in all 3 cell lines (Figure 3.4). DFT2 shares more experimentally quantified proteins with DFT1 than fibroblasts.

Table 3.1: Summary table of proteins detected and quantified in all three replicates of each cell line.
Total proteins indicates the number of unique Ensembl protein IDs identified in three replicates of the non-quantified proteome. Total quantified indicates the total number of proteins quantified using label free methods in PEAKS. Unique protein IDs indicates the number of unique Ensembl protein IDs identified in the total quantified proteome. All cell lines indicates the number of unique protein IDs quantified in all three cell lines. Two cell lines indicates the number of unique protein IDs quantified in the indicated cell line and one additional cell line. One cell line indicates the number of unique protein IDs quantified in a single cell line only.

Cell line	**Total proteins**	**Total quantified**	**Quantified: Unique protein IDs**	**Quantified: all cell lines**	**Quantified: two cell lines**	**Quantified: one cell line**
DFT2	4776	1550	1333	182	517	633
DFT1	5169	1033	904	182	439	283
Fibroblasts	4287	730	657	182	306	169

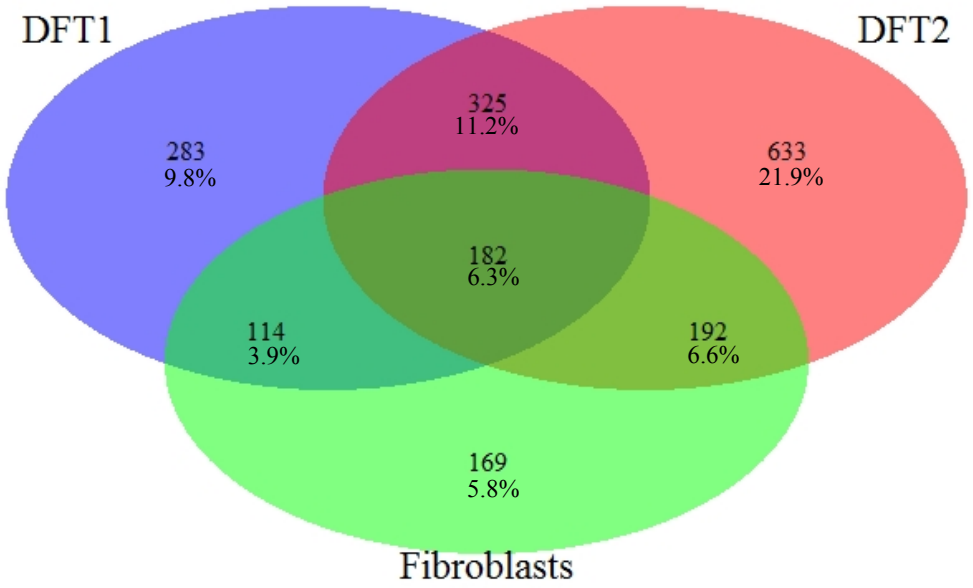


Figure 3.4: The majority of proteins quantified in DFT2 are unique to DFT2.
Venn diagram of all unique proteins quantified in PEAKS analysis and the overlap of these proteins between cell lines. All unique proteins quantified were detected in all replicates of each cell lines.

Expression levels of quantified proteins across replicates and across cell lines were compared using a heatmap. Expression values for proteins not detected in a cell line were imputed using R (code presented in full in Appendix B.1) to allow for hierarchical clustering of all quantified proteins across all cell lines. Imputation was performed based on the normalised distribution of protein expression in each proteome, and imputed values assume a narrowed and downshifted distribution. Additionally, the average expression across replicates of the 182 proteins quantified in all three cell lines were compared without additional imputation of data. All hierarchical clustering was performed in R using the hclust function (Chapter 2.3.4, Appendix B.1). Hierarchical clustering is indicated on the heat maps by dendrograms.

Protein expression across replicates for each cell line cluster closely together indicating good experimental reproducibility (Figure 3.5a). Between cell lines, DFT1 clusters more closely with fibroblasts than with DFT2, although the distance between clusters is large and indicates weak clustering. Fibroblasts and DFT1 have significantly more imputed values than DFT2 which may be skewing the data, so the expression of proteins detected in all three cell lines was compared to avoid the need for data imputation. For this analysis, the represented expression value is an average across replicates for each cell line. Hierarchical clustering of the average expression of proteins quantified in all three cell lines indicates that DFT2 is more closely clustered to DFT1 cells than fibroblasts cells, indicating that the expression of common proteins in these datasets is more similar between DFT1 and DFT2, although this clustering is weak (Figure 3.5b). The relatively small degree of overlap between cell lines makes direct comparison of protein expression at this level difficult, and additionally this analysis provides no biological information on the proteins being expressed.

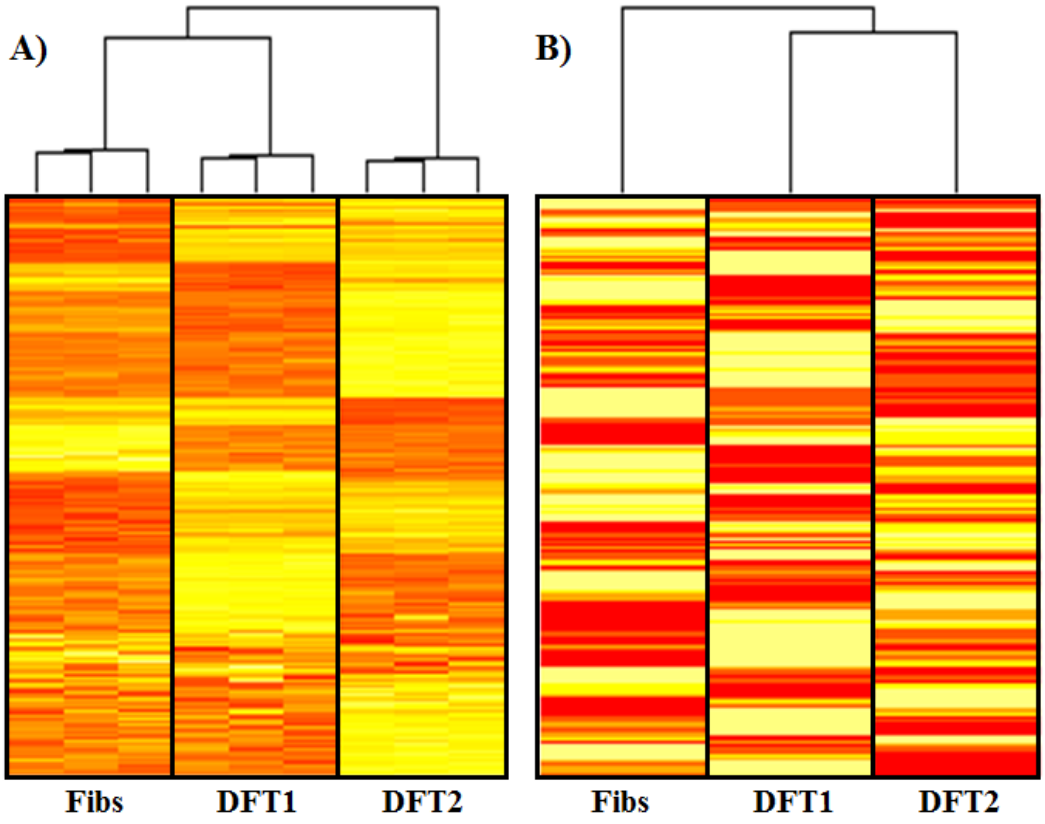


Figure 3.5: Hierarchical clustering of protein expression across cell lines is inconclusive.

Heat maps comparing quantified protein expression between DFT2, DFT1 and fibroblasts (Fibs).

A) Total quantified protein expression in each of three replicates per cell line. Each row represents one protein. Each column represents an experimental replicate, and triplicate results for each cell line are grouped into columns by black lines. Missing values for proteins not detected in all cell lines for were imputed using R for cluster analysis (Appendix B.1)

B) Proteins identified and quantified in all three cell lines. Each row represents one protein. Each column represents the average expression across triplicate experiments for each cell line. Cell lines are separated by black lines. No imputation was required for this dataset, as all proteins were quantified in all cell lines.

Incremental colour scale represents expression values on a log2 scale. Pale yellow indicates highest protein expression, yellow to orange indicates reducing protein expression and red represents lowest expression. Hierarchical clustering of expression between cell types was performed in R using the hclust function and is indicated by the dendrogram at the top of the heatmaps. Dendrograms which are closer to the heatmap indicate more closely related clusters than dendrograms which are far from the heatmap.

Gene set enrichment and functional analysis was performed on the quantified proteins for each cell line ranked from highest expression to lowest using multiple functional databases through the g:Profiler g:GOSt application (Reimand *et al.* 2016; Reimand *et al.* 2007). The databases used in this analysis are; A) Gene Ontology (GO) Consortium Biological Process (GO:BP) database, which informs of the broad biological functions enriched in the proteome, B) Gene Ontology Consortium Cellular Compartment (GO:CC) database, which informs which part of the cell the proteins in the proteome function (The Gene Ontology Consortium 2018; Ashburner *et al.* 2000) C) Reactome (REAC) database (Fabregat *et al.* 2018), which informs of the enrichment of signalling and metabolic pathways within the proteome and D) Human Protein Atlas (HPA) database (Thul *et al.* 2017; Human Protein Atlas available from www.proteinatlas.org), which informs of cell and tissue type specific signatures in the proteome. All functional analysis was run against a human background dataset to include as many proteins as possible. Functional analysis on full quantified proteomes is presented in Appendix C.3.

P-values for functional terms not enriched in all three cell lines were imputed in R to allow for direct comparison and clustering of functional analysis between cell lines. Imputation was performed assuming a narrowed and downshifted distribution (Appendix B.1). Hierarchical clustering was performed in R on each enrichment set to identify similarities between the three cell lines. Note that this clustering does not consider biological function, only similarities in enrichment of specific functional terms between cell lines. Functional enrichment is visualised using heatmaps with dendrograms to represent clustering.

Clustering of gene set enrichment analysis against the GO:BP database indicates that DFT1 and fibroblasts cluster together more closely than with DFT2 (Figure 3.6a), whilst clustering of analysis against the GO:CC database indicates that DFT2 and fibroblasts cluster more closely than with DFT1 (Figure 3.6b), although clustering is weak in both datasets. Clustering of gene set enrichment analysis against the HPA and Reactome databases demonstrates DFT2 clusters more closely with DFT1 than with fibroblasts, and this clustering is stronger than seen in the GO:BP and GO:CC datasets. This data broadly indicates that the DFT2 proteome is enriched for more unique biological pathways than DFT1 and fibroblasts, although this clustering is weak and difficult to interpret due to imputed values. Similarly, this data indicates some similarity between the enriched cellular compartments in the DFT2 and fibroblasts proteomes, but this clustering is weak.

Enrichment against both the Reactome and the HPA databases demonstrate moderate clustering between DFT1 and DFT2 (Figure 3.6c, Figure 3.6d), indicating that the signalling and metabolic pathways enriched in DFT1 and DFT2 are more similar to each other than to those enriched in fibroblasts, and that the DFT1 and DFT2 proteomes are more similar to the same human tissue samples than fibroblasts. The lack of biological information in this analysis combined with the lack of strong similarities between datasets due to missing data makes the results difficult to interpret. A deeper analysis of the biology of the enriched pathways within each cell line, and analysis of cell line specific pathways is needed to identify the cell type of origin of DFT2.

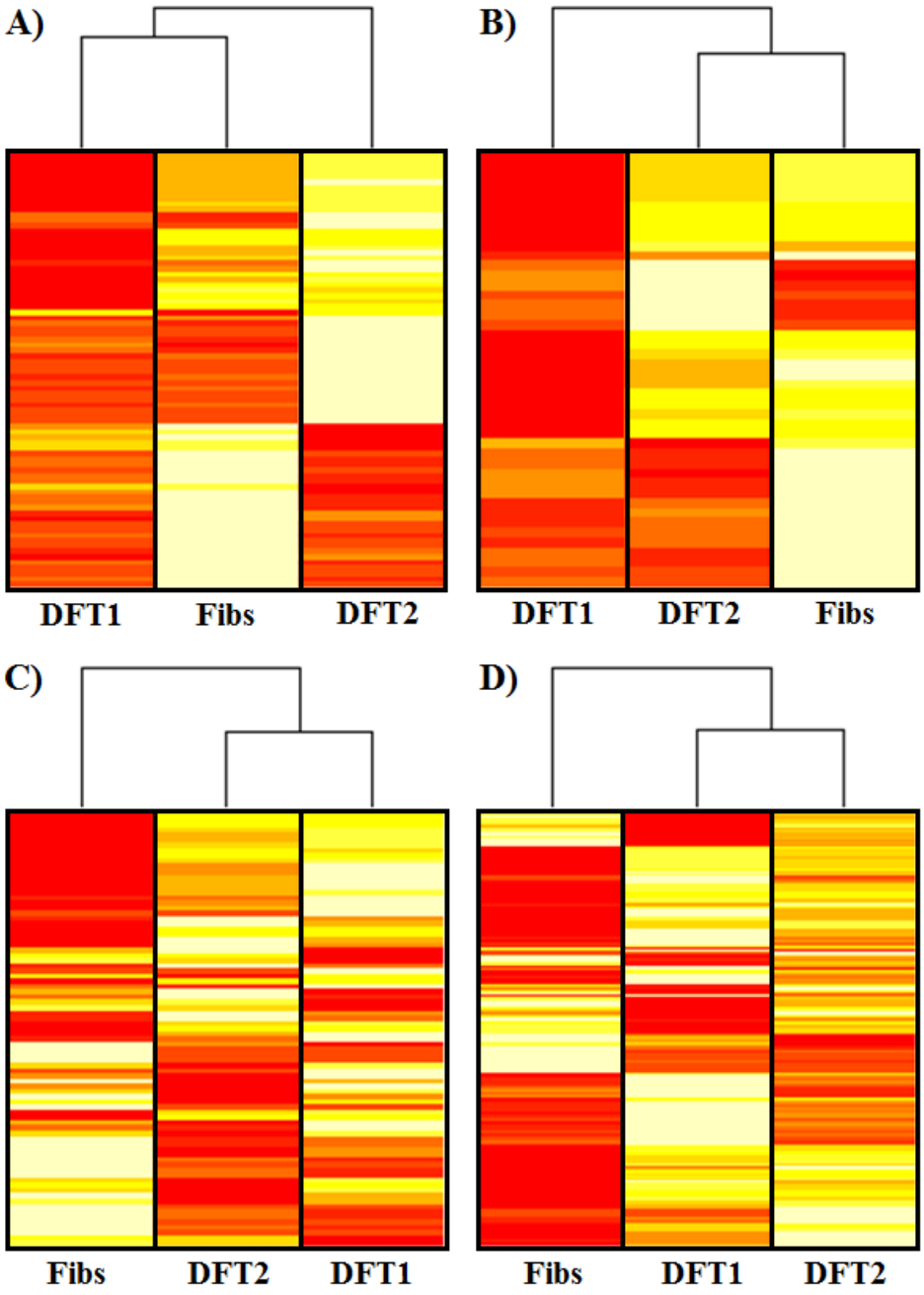


Figure 3.6: The DFT1 and DFT2 proteomes are enriched for similar metabolic and signalling pathways and show similar tissue specific protein signatures compared to fibroblasts.

Heatmaps indicating functional enrichment within the proteomes of DFT1, DFT2 and fibroblasts using the A) GO: Biological process, B) GO: Cellular compartment, C) Human protein atlas and D) Reactome databases. Each row represents a specific functional term. Incremental colour scale represents -log10(enrichment p-value). P-values have been adjusted for multiple testing. Light yellow indicates highest enrichment in the proteome, yellow to orange indicates reducing enrichment and red is the lowest enrichment. Hierarchical clustering of expression between cell types was performed in R using the hclust function and is indicated by the dendrogram at the top of the heatmap. Dendrograms which are closer to the heatmap indicate more closely related clusters than dendrograms which are far from the heatmap. Full details of functional analysis are presented in Appendix C.3.

3.3.2 DFT1 and DFT2 are enriched for proteins and biological processes which occur within the nervous system

Broad analysis and comparison of functional enrichment in the DFT2 proteome to DFT1 and fibroblasts does not indicate any strong similarities to either DFT1 or fibroblasts. The specific functional terms enriched in each quantified proteome were analysed on a term by term basis to identify cell-type specific functions which could indicate DFT2s cell type of origin.

The large input lists and the degeneracy of functional terms and protein function makes identifying functional enrichment related to specific cell types from the full quantified proteome datasets difficult. Analysis of GO:BP indicates enrichment for pathways which give no cell type specific information, and whilst these may be biologically interesting, in the context of identifying the cell type of origin of DFT2 these terms do not give any relevant information. Functional terms that did not indicate any cell-type specific function or specialisation, such as "translation" and "RNA processing", were therefore discarded from deeper analysis of the enriched terms but are presented in full in Appendix C.3 and comparisons between cell lines are presented in Appendix C.4.

Only functional terms which indicate a cellular specialisation or are specific to a cell or tissue type, and therefore can provide insight into the cellular origins of DFT2 were analysed and full comparisons between these processes in DFT1, DFT2 and fibroblasts were made. Additionally, a Sheep fibroblast proteome dataset previously published by Karbiener *et al.* (2017) has been analysed in the same way as the devil proteomes as a negative control. The percentage of functional terms indicating a cellular specialisation was calculated on a cell line basis; the total number of specialised terms for a given cell line was counted and divided by the total number of enriched terms in the same cell line. 10.7%, 7.5% and 13.4% of functional terms were deemed to indicate a cellular specialisation in DFT2, DFT1 and fibroblasts respectively. These cell-type specific terms have been grouped into broad functional categories which are summarised in Table 3.2.

21 cell-type specific GO:BP terms in DFT2 are biological processes linked to immune functions, 3.8% of the total enriched functional terms. This is comparable with 4.1% in DFT1 and 4% in fibroblasts, indicating that all three cell lines are enriched for pathways involved in immune responses and activation. A further 20 enriched, cell-type specific biological processes in DFT2 are associated with host response to viruses and other pathogens, 3.6% of the total enriched processes. This is over twice as enriched as viral and pathogen response terms are in both DFT1 and fibroblasts (1.5% and 1.8% respectively).

DFT2 is additionally enriched for 6 processes associated with cell motility (1.1% of total processes) which is comparable with Fibroblasts (1% of total processes), but no enrichment for cellular motility processes is identified in the DFT1 proteome, indicating DFT2 cells may be more motile. DFT2 is further enriched for one process each associated with wound healing, nervous system development, muscle development and gland development, two processes associated with liver development and function and three processes associated with drug metabolism. Of these cell-type specific terms, four

are uniquely enriched in DFT2; Gland development, glial cell development, muscle cell cellular homeostasis and positive regulation of endothelial cell migration. These four terms all originate from different cell types and it is unclear which, if any, relate to the origins of DFT2.

DFT1 is enriched for three processes each associated with wound healing and drug metabolism and two processes associated with liver development. The Schwann cell tumour DFT1 is not enriched for any nervous system associated processes, whilst the non-nervous fibroblast cell line is enriched for five, demonstrating the difficulty of drawing solid conclusions on cellular origins from such a large dataset. A more focussed approach is needed to identify cell type specific functions which are currently not defined. By discarding proteins which are present in all three cell lines from the functional analysis, and instead focussing on proteins only identified in a single cell type, functional analysis should highlight fewer housekeeping processes and more processes which are specific to the cell type being analysed.

Table 3.2: Summary table of the enriched, cell-type specific Biological Processes in the full quantified DFT2, DFT1 and fibroblast proteomes.

Numbers indicate the number of GO:BP terms significantly enriched in each cell line which are associated with each broad functional category. Full data and comparison analysis is presented in Appendix C.3 and C.4.

	Number of enriched terms		
Broad functional category	**DFT1**	**DFT2**	**Fibroblasts**
Immune process	17	21	20
Response to pathogens	6	20	9
Cell movement	0	6	5
Wound healing	3	1	8
Muscle development and function	0	1	9
Nervous system development and function	0	1	5
Liver development and function	2	2	0
Drug metabolism	3	3	3
Fertilisation	0	0	2
Bone development and function	0	0	1
Gland development and function	0	1	0

GO:CC analysis indicates that most compartments enriched in each cell type are ubiquitous and give no information on cell type of origin. Cellular compartments giving no cell-type specific information were discarded from further analysis (as during analysis of GO:BP enrichment) but are presented in full in Appendices C.3 and C.4. A summary table of all the enriched cell-type specific cellular compartments sorted into broad functional groups is presented in Table 3.3. DFT2 is enriched for 10 cellular compartments associated with the nervous system, 3.9% of all enriched cellular compartments in DFT2. This is comparable with fibroblasts (4.1%) and higher than seen in DFT1 (2.1%). DFT2 is additionally enriched for five compartments associated with cell motility, higher than DFT1 (one compartment) but fewer than seen in fibroblasts (seven compartments), again indicating that DFT2 may be a more motile tumour than DFT1. DFT2 is additionally enriched for three compartments associated with muscle cells, two associated with immune cells and two associated with viral responses.

The most highly enriched cell-type specific GO:CC category in the DFT2 proteome is the nervous system. Fibroblasts are also enriched for nervous system compartments despite being a non-nervous cell type, a somewhat unexpected result. All the nervous system GO:CC compartments and associated p-values are compared in Table 3.4. Five of the nervous system associated cellular compartments enriched in Fibroblasts are associated with the synapse, including postsynaptic regions, compartments which exist in non-neuronal cell types. Additionally, the sheep fibroblast proteome analysed in the same way is also enriched for nervous system specific GO:CC compartments. This demonstrates that the enrichment of these compartments in fibroblasts is not due to contamination, and further illustrates the difficulty deciphering cellular origins from larger datasets, confirming the need for a cell specific approach.

Table 3.3: Summary table of the enriched, cell-type specific Cellular compartments in the full quantified DFT2, DFT1 and fibroblast proteomes.

Numbers indicate the number of GO:CC terms significantly enriched in each cell line which are associated with each broad functional category. Full data is presented in Appendix C.4.

	Number of enriched terms		
Broad functional categories	**DFT1**	**DFT2**	**Fibroblasts**
Immune cells	1	2	1
Viral, bacterial or parasite response	0	2	0
Cell movement	1	5	7
Muscle cells	1	3	11
Nervous system cells	5	10	9
Epithelium	0	0	2
Embryo	0	0	1

Table 3.4: Analysis of the full quantified proteomes demonstrates enrichment for nervous system specific cellular compartments in fibroblasts as well as tumour cells.

Condensed summary table indicating all nervous system GO:CC terms identified as enriched by functional analysis of the full quantified proteomes of DFT1, DFT2 and devil fibroblasts, and additionally a previously published sheep fibroblast proteome (Karbiener *et al.* 2017). Raw data can be found in Appendix C.3 and Appendix C.4. P-values are corrected for multiple testing. NA indicates the functional term was not significantly upregulated in that cell type.

Cellular component	**DFT1 Adjusted p-value**	**DFT2 Adjusted p-value**	**Fibroblasts Adjusted p-value**	**Sheep fibroblasts Adjusted p-value**
myelin sheath	1.20E-32	1.47E-26	5.96E-36	4.25E-30
neuron part	0.016727	9.23E-07	1.33E-05	1.25E-08
neuron projection	NA	0.00031	0.01743	3.95E-05
axon part	NA	0.000553	NA	0.014836
axon	NA	0.001154	NA	0.000486
distal axon	NA	0.005358	NA	NA
postsynapse	NA	0.028618	1.27E-05	6.20E-07
dendrite	NA	0.031228	NA	0.004175
dendritic tree	NA	0.033128	NA	0.004492
somatodendritic compartment	NA	0.038583	NA	0.001902
postsynaptic density	NA	NA	0.001339	0.02649
asymmetric synapse	NA	NA	0.001603	0.032055
neuronal cell body	NA	NA	0.023106	0.044604
npBAF complex	2.21E-05	NA	NA	NA
nBAF complex	0.000382	NA	NA	NA
glutamatergic synapse	0.000567	NA	0.015243	0.005476
neuron to neuron synapse	NA	NA	0.000543	0.016326

To reduce the amount of information which is not specific to the cell-type thus is not helpful for identifying cellular origins, and to analyse cell type specific functions, gene set enrichment and functional analysis was performed on proteins which were uniquely quantified in a single cell line (633, 283 and 169 proteins in DFT2, DFT1 and Fibroblasts respectively). Unique proteins were ranked based on expression levels from highest to lowest for analysis. The results of this analysis are presented in full in Appendix C.5. As before, results for GO:BP, GO:CC and Reactome analysis which were not cell-type specific and did not provide information on cellular origins were not considered in further analysis, but a comparative analysis of all terms is presented in Appendix C.6.

GO:BP analysis of proteins uniquely expressed in each cell line was performed. After removal of non-cell-type specific terms that give no information on cellular origins, DFT2 is exclusively enriched for immune system and pathogen associated cellular processes (Table 3.5). A similar immune process enrichment is seen in DFT1 but is almost entirely absent in fibroblasts. These processes are likely derived from the fact that DFT1 and DFT2 and transmissible tumours which are modulating immune response pathways to evade immune detection rather than being indicative of a cellular origin. DFT1 is additionally enriched for one nervous system process, two drug metabolism processes and two liver associated processes.

Analysis of the GO:CC terms giving cell-type specific information from unique proteins in each cell line demonstrated that DFT1 and DFT2, but not fibroblasts, are enriched for cellular compartments specific to the nervous system (three and six respectively) (Table 3.6). Additionally, DFT2 is enriched for four compartments associated with cell motility, compared with one in DFT1 and five in fibroblasts. DFT2 is further enriched for two compartments associated with epithelial cells, both of which are enriched in DFT1 and fibroblasts. Interestingly. DFT2 no longer demonstrates enrichment for immune cell compartments, in contrast to both DFT1 and fibroblasts.

Table 3.5: Summary table of the enriched, cell-type specific Biological Processes following functional analysis of unique proteins in the DFT2, DFT1 and fibroblast proteomes.

Numbers indicate the number of GO:BP terms significantly enriched in each cell line which are associated with each broad functional category. Full data is presented in Appendix C.5 and C.6.

	Number of terms		
Broad functional category	**DFT1**	**DFT2**	**Fibroblasts**
Immune process	14	9	1
Viral, bacterial or parasite response	5	5	2
Cell movement	0	0	0
Wound healing	0	0	5
Muscle development and function	0	0	2
Nervous system development and function	1	0	0
Liver development and function	2	0	0
Drug metabolism	2	0	0

Table 3.6: Summary table of the enriched, cell-type specific Cellular Compartments following functional analysis of unique proteins in the DFT2, DFT1 and fibroblast proteomes.

Numbers indicate the number of GO:CC terms significantly enriched in each cell line which are associated with each broad functional category. Full data is presented in Appendix C.5 and C.6.

	Number of enriched terms		
Broad functional categories	**DFT1**	**DFT2**	**Fibroblasts**
Immune cells	4	0	1
Cell movement	1	4	5
Muscle cells	1	0	9
Nervous system cells	3	6	0
Epithelium	2	2	3
Embryo	0	0	1

The broad functional category with the most enriched GO:CC terms following functional analysis of proteins uniquely quantified in DFT2 is the nervous system. Specific analysis of the nervous system cellular compartments enriched in both DFT1 and DFT2 demonstrates a significant enrichment for the myelin sheath, a highly specialised cell compartment expressed only in Schwann cells and oligodendrocytes. Previous strong enrichment for the myelin sheath seen in fibroblasts is no longer present. DFT2 is enriched for neuron and axon parts whilst DFT1 is enriched for the synapse and Schwann cell microvillus (Table 3.7).

The strong and unique enrichment for nervous system associated cellular compartments in DFT2 indicates the tumour is enriched for proteins associated with the nervous system. Analysis of the enriched Reactome pathways in the unique proteins from each cell line indicates a stronger enrichment for nervous system associated pathways in both DFT1 and DFT2 than fibroblasts (Table 3.8). DFT2 is only enriched for Reactome pathways involved in the nervous and immune systems. DFT1 is also enriched for pathways from the nervous and immune systems, and additionally weakly enriched for smooth muscle contraction. Fibroblasts are enriched for three nervous system pathways at significantly lower levels than DFT1 and DFT2, and additionally for wound healing, muscle development and immune processes.

It is worth noting that the DFT1 and fibroblast cell lines used in this project have been in cell culture for nine years longer than the DFT2 cell line (see Table 2.1), which is likely to alter the cellular proteome substantially. Additionally, factors such as depth of the original tumour biopsy the cell lines were established from, infection status of the original tumour, location of the tumour on the body and the sex of the host devil may significantly impact the proteome of the derived cell lines particularly with regards to the activation of immune mediated processes in the tumour and host. Thus, it is difficult to attribute the cause of the differential enrichment of immune and infection associated biological processes between the three cell lines, which may fluctuate during long or short term cell culture, or reflect the specific host immune environment rather than the tumour cell of origin. In contrast, myelination and mature nervous system pathways would not be expected to be upregulated in cell culture or in response to differential infection in host devils, thus the consistent enrichment for the myelin sheath in both DFT2 and the known Schwann cell tumour DFT1, combined with the consistent and high levels of enrichment for additional nervous system associated pathways in both DFT1 and DFT2 may be indicative of a shared origin for these two tumours.

Table 3.7: Unique proteins expressed by DFT1 and DFT2 demonstrate functional enrichment for cellular compartments of the nervous system.

Condensed table indicating all nervous system GO:CC terms identified following analysis of proteins identified and quantified in a single cell type. Raw data can be found in Appendices C.5 and C.6. P-values are corrected for multiple testing. NA indicates the functional term was not significantly upregulated in that cell type.

Cellular compartment	DFT1 adjusted p-value	DFT2 adjusted p-value	Fibroblasts adjusted p-value
myelin sheath	5.11E-26	0.00020413	NA
neuron part	NA	0.000863251	NA
neuron projection	NA	0.001346872	NA
axon	NA	0.018452846	NA
distal axon	NA	0.018573388	NA
axon part	NA	0.019703751	NA
synapse	0.024163741	NA	NA
Schwann cell microvillus	0.04614155	NA	NA

Table 3.8: **DFT1 and DFT2 are more strongly enriched for Reactome pathways associated with nervous tissue than fibroblasts.**

Condensed table indicating all cell-type specific Reactome pathways identified following analysis of proteins identified and quantified in a single cell type. Raw data can be found in Appendices C.5 and C.6. P-values are corrected for multiple testing. NA indicates the functional term was not significantly upregulated in that cell type. Nervous system processes are indicated with yellow highlight.

Reactome pathway	DFT1 adjusted p-value	DFT2 adjusted p-value	Fibroblasts adjusted p-value
Viral mRNA Translation	2.39E-11	3.17E-10	1.42E-05
Influenza Life Cycle	1.29E-12	4.45E-10	4.37E-05
Influenza Infection	5.70E-12	9.51E-10	8.32E-05
Signalling by ROBO receptors	5.80E-10	8.07E-09	0.000455
Influenza Viral RNA Transcription and Replication	9.06E-10	1.07E-08	0.000199
Regulation of expression of SLITs and ROBOs	8.06E-12	5.78E-08	0.000234
Axon guidance	1.61E-06	5.11E-05	0.000812
Axonal growth stimulation	NA	0.000745	NA
Disease	0.010455	0.001315	NA
Axonal growth inhibition (RHOA activation)	NA	0.004464	NA
Cross-presentation of soluble exogenous antigens (endosomes)	0.000389	0.004626	NA
p75NTR regulates axonogenesis	NA	0.005578	NA
Host Interactions of HIV factors	3.37E-05	0.006523	NA
HIV Infection	0.016855	0.012464	NA
Antigen processing-Cross presentation	0.006939	0.021646	NA
Platelet degranulation	0.00386	NA	NA
Downstream signalling events of B Cell Receptor (BCR)	0.000916	NA	NA
Activation of NF-kappaB in B cells	0.00087	NA	NA
Virus Assembly and Release	0.018609	NA	NA
Assembly of Viral Components at the Budding Site	0.018609	NA	NA
TCR signalling	0.002165	NA	NA
Downstream TCR signalling	0.001232	NA	NA
Smooth Muscle Contraction	0.022146	NA	6.74E-07
Interleukin-12 family signalling	0.000593	NA	NA
Neutrophil degranulation	4.41E-06	NA	NA
Dissolution of Fibrin Clot	NA	NA	0.007403
Platelet Adhesion to exposed collagen	NA	NA	0.043193
Gene and protein expression by JAK-STAT signalling after Interleukin-12 stimulation	0.003761	NA	NA
Interleukin-12 signalling	0.000227	NA	NA
Interleukin-1 signalling	0.009462	NA	NA

3.3.3 Protein expression in DFT1 and DFT2 indicates close similarity to nervous system tissues and cell types

DFT2 has demonstrated a significant enrichment in its proteome for cellular compartments and signalling pathways associated with the nervous system comparable with the enrichment seen in the known Schwann cell tumour DFT1, most of which are not enriched in the fibroblast cell line, indicating that DFT2 may have emerged from the nervous system, similar to DFT1.

The HPA database compares the protein lists from each cell type to gene expression in a variety of human tissues. As demonstrated in Figure 3.6d, the HPA tissue expression profiles of the full quantified DFT1 and DFT2 proteomes cluster together, indicating they reflect similar human tissues. Three of the top 5 tissue types enriched in the full quantified DFT2 proteome are nervous, and both DFT2 and the Schwann cell derived DFT1 are significantly more enriched for these tissues than fibroblasts (Table 3.9a, Table 3.9b). None of the top 5 terms enriched in the full quantified fibroblast proteome are from nervous tissue (Table 3.9c). A comparison of all nervous tissue terms in all three cell lines demonstrates that DFT1 and DFT2 are more enriched for all identified nervous cells and tissues than fibroblasts (Figure 3.7). Three nervous tissues were not identified to have any enrichment within the fibroblast proteome (Appendix C.3, Appendix C.4).

The unique quantified proteins in DFT1, DFT2 and fibroblasts were also analysed against the HPA database. Fibroblasts showed no significant enrichment for any tissues, though this may be due to the lower numbers of proteins (Appendix C.5). DFT1 and DFT2 both remain enriched for neuronal tissues. DFT2 is significantly enriched for only six tissues, two of which are from the nervous system (Table 3.10a). DFT1 is significantly enriched for 167 terms and the top term is a neuronal tissue (Table 3.10b).

Table 3.9: Protein expression in DFT2 and DFT1 is most similar to that of nervous system tissues.

Top 5 tissue types enriched in A) DFT2, B) DFT1 and C) Fibroblasts following analysis of all quantified proteins against the Human Protein Atlas database. Raw data is presented in Appendix C.5. P-values are corrected for multiple testing. Nervous system tissues are indicated with yellow highlight.

A) DFT2 top 5 Human protein atlas tissue type	**DFT1 adjusted p-value**	**DFT2 adjusted p-value**	**Fibroblasts adjusted p-value**
cerebral cortex; neuronal cells	9.14E-22	8.21E-20	0.000675
cerebellum; Purkinje cells	9.22E-20	4.98E-19	0.011261
pancreas; exocrine glandular cells	3.81E-17	1.29E-17	0.003168
urinary bladder; urothelial cells	1.18E-16	6.19E-17	0.01334
caudate; neuronal cells	5.51E-13	8.43E-17	0.027983

B) DFT1 top 5 Human protein atlas tissue type	**DFT1 adjusted p-value**	**DFT2 adjusted p-value**	**Fibroblasts adjusted p-value**
cerebral cortex; neuronal cells	9.14E-22	8.21E-20	0.000675087
stomach 2; glandular cells	7.23E-21	1.02E-14	2.88E-05
cerebellum; Purkinje cells	9.22E-20	4.98E-19	0.011260851
stomach 1; glandular cells	1.91E-19	1.26E-16	1.41E-06
rectum; glandular cells	1.43E-18	3.83E-16	0.002324742

C) Fibroblasts top 5 Human protein atlas tissue type	**DFT1 adjusted p-value**	**DFT2 adjusted p-value**	**Fibroblasts adjusted p-value**
stomach 1; glandular cells	1.91E-19	1.26E-16	1.41E-06
endometrium 2; glandular cells	5.25E-11	2.37E-13	6.32E-06
salivary gland; glandular cells	1.19E-15	1.97E-14	6.39E-06
stomach 2; glandular cells	7.23E-21	1.02E-14	2.88E-05
appendix; lymphoid tissue	1.12E-11	1.65E-09	3.54E-05

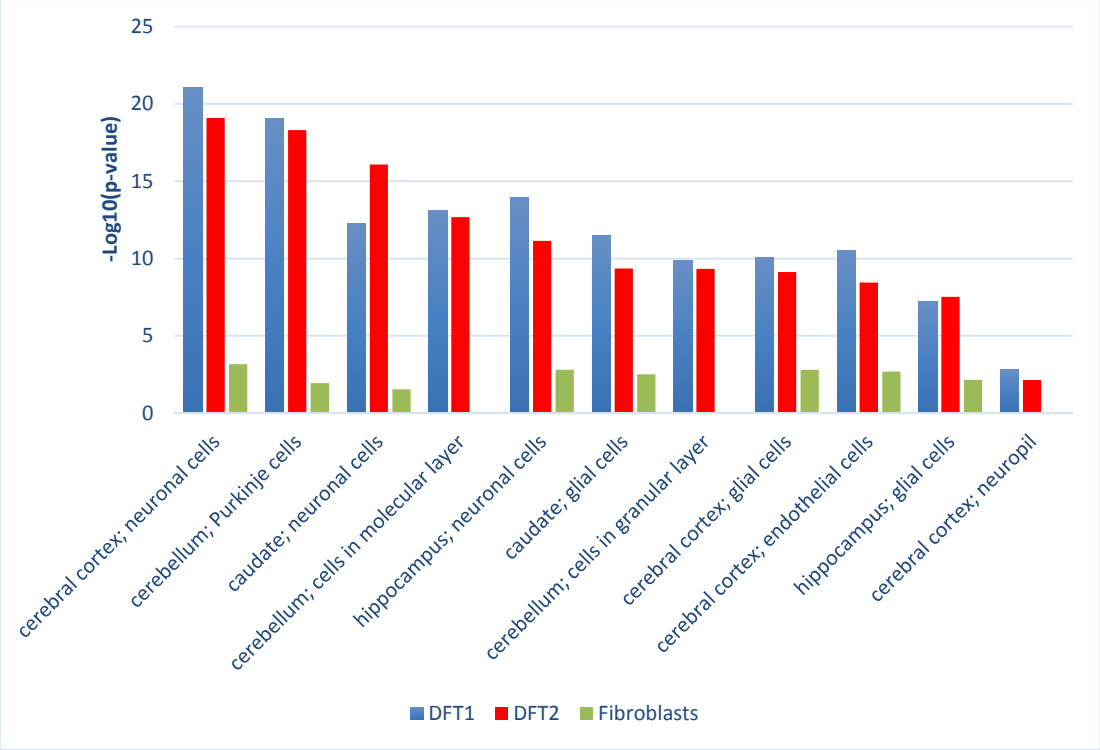


Figure 3.7: DFT1 and DFT2 are more similar to nervous system tissues than fibroblasts.
Comparison of enrichment values for all nervous system associated tissues across DFT1, DFT2 and fibroblasts following functional analysis using all quantified proteins in each sample against the Human Protein Atlas background. P-values were corrected for multiple testing prior to log transformation. Comparisons presented in Appendix C.6.

Table 3.10: Unique proteins in DFT1 and DFT2 are similar to nervous system tissue types.
Tables indicating the top 6 significantly enriched tissues and adjusted p-values for the unique proteins in A) DFT2 and B) DFT1 following analysis against the Human Protein Atlas. Nervous system tissues are indicated with yellow highlight. Comparisons presented in Appendix C.6.

A)	***Human tissue type***	**Adjusted p-value**
	soft tissue 2; fibroblasts	0.002343
	placenta; decidual cells	0.00543
	cerebellum; cells in granular layer	0.006513
	testis; cells in seminiferous ducts	0.014277
	caudate; neuronal cells	0.022492
	bone marrow; hematopoietic cells	0.041168

B)	***Human tissue type***	**Adjusted p-value**
	cerebral cortex; neuronal cells	2.65E-22
	stomach 1; glandular cells	4.76E-18
	nasopharynx; respiratory epithelial cells	5.00E-18
	seminal vesicle; glandular cells	6.83E-18
	stomach 2; glandular cells	9.44E-18
	pancreas; exocrine glandular cells	1.53E-17

It is likely that both tumours share functionally relevant proteins which are not present in fibroblasts. Analysis of the proteins uniquely quantified in DFT2 (633 proteins) and proteins quantified in DFT2 and DFT1 but not in fibroblasts (325 proteins) (958 proteins total, hereon referred to as DFT2 + tumour specific proteins) demonstrated significant enrichment for 145 tissues, the most significant of which is the cerebral cortex (Table 3.11a). Analysis of the proteins uniquely quantified in DFT1 (283 proteins) and proteins quantified in DFT1 and DFT2 but not fibroblasts (325 proteins) (608 proteins total, hereon referred to as DFT1 + tumour specific proteins) demonstrated significant enrichment for 172 tissues, and the most significant is also cerebral cortex (Table 3.11b), indicating both tumours are expressing proteins, shared and unique, which infer a nervous tissue phenotype. Analysis of the proteins unique to fibroblasts (169 proteins) combined with the proteins present in all 3 cell lines (182 proteins) (351 proteins total, hereon referred to as Fibroblasts + ubiquitous proteins) demonstrates significant enrichment for 103 tissues, with none of the top 6 being nervous tissues (Table 3.11c). Full analysis is presented in Appendix C.7.

Table 3.11: Protein expression by both tumours is most similar to nervous system tissue and cell types.
Top 6 significantly enriched tissue types following analysis against the Human Protein Atlas.
A) DFT2 + tumour specific proteins, B) DFT1 + tumour specific proteins and C) Fibroblasts + ubiquitous proteins. Nervous system tissues are indicated with yellow highlight. Full data is presented in Appendix C.7

A)	***Human tissue type***	**Adjusted p-value**
	cerebral cortex; neuronal cells	1.52E-13
	pancreas; exocrine glandular cells	5.26E-13
	urinary bladder; urothelial cells	7.23E-13
	cerebellum; Purkinje cells	6.23E-12
	placenta; trophoblastic cells	2.94E-11
	caudate; neuronal cells	3.85E-11

B)	***Human tissue type***	**Adjusted p-value**
	cerebral cortex; neuronal cells	2.83E-17
	adrenal gland; glandular cells	2.81E-16
	placenta; trophoblastic cells	3.22E-16
	cerebellum; Purkinje cells	8.91E-16
	stomach 2; glandular cells	9.28E-16
	rectum; glandular cells	2.55E-15

C)	***Human tissue type***	**Adjusted p-value**
	spleen; cells in white pulp	1.70E-07
	liver; bile duct cells	3.06E-06
	breast; myoepithelial cells	3.25E-06
	endometrium 2; glandular cells	9.59E-06
	tonsil; germinal centre cells	1.51E-05
	salivary gland; glandular cells	1.86E-05

A comparison of all nervous system associated tissues significantly enriched in DFT2 + tumour specific proteins, DFT1 + tumour specific proteins and fibroblasts + ubiquitous proteins demonstrated that tumour specific protein expression is more similar to nervous system tissues than fibroblast specific and ubiquitous proteins (Figure 3.8), indicating that there is a tumour specific enrichment for nervous system proteins that is not present in the proteome of fibroblasts, or in the proteins common to all three cell types.

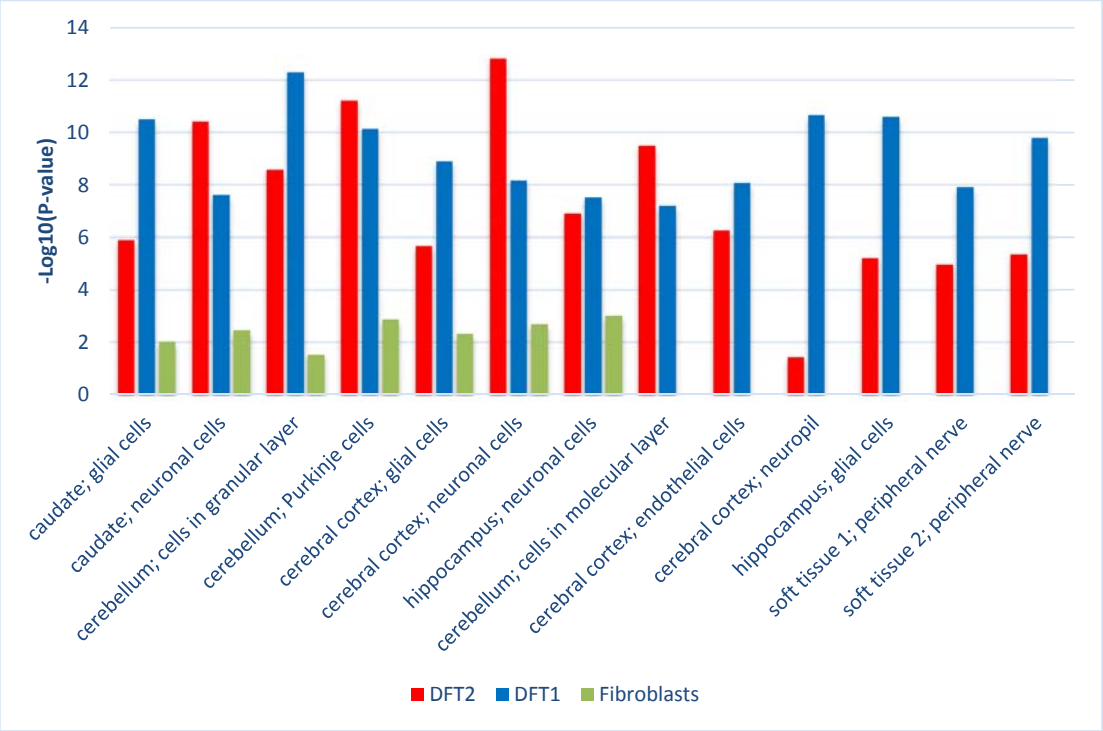


Figure 3.8: Tumour proteomes are consistently more strongly enriched for nervous system proteins than fibroblasts.

Comparison of all nervous system associated tissues found to be enriched in the proteomes of: DFT2 + tumour specific proteins (Labelled DFT2, red), DFT1 + tumour specific proteins (Labelled DFT1, blue) and Fibroblasts + ubiquitous proteins (Labelled Fibroblasts, green). X axis indicates human tissue type from the HPA database. Y axis indicates -log10 enrichment adjusted p-value.

3.3.4 DFT1 and DFT2 express protein markers of neuronal and glial cells

Whilst broad analysis of protein expression and functional enrichment is inconclusive, deeper specific analysis of the functional pathways and cellular compartments enriched in the DFT2 proteome indicates a functional enrichment for proteins from the nervous system. DFT2 is consistently enriched for cell compartments specific to the nervous system to a higher degree than the known nervous system tumour DFT1, and significantly more enriched for nervous system processes and cellular compartments than the non-neuronal fibroblast proteome. Additionally, DFT1 and DFT2 have similar protein expression profiles when analysed against the Human protein atlas database, demonstrating close similarity to several nervous system tissue types. Based on this similarity and the strong enrichment within DFT2 for nervous system associated signalling pathways and cellular compartments, in particular for the Schwann cell and oligodendrocyte restricted myelin sheath, I hypothesise that DFT2 has also emerged from a myelinating glial lineage.

The top 50 most highly expressed proteins in DFT2 were evaluated to identify any proteins with cell-type specific expression profiles which may support or reject this hypothesis (Table 3.12). All of the top 50 highest expressed proteins in DFT2 are widely expressed in different tissues in the body. 45 proteins had no significant cell-type specific expression pattern or restrictions and did not provide any information on the cellular origins of DFT2. 5 of these proteins (nestin (NES), dihydropyrimidinase like 2 (DPYSL2), creatine kinase b (CKB), dihydropyrimidinase like 3 (DPYSL3) and low affinity nerve growth factor receptor (NGFR)) are normally overexpressed in nervous system tissue. One of these, NGFR, is unique to DFT2 and not detected in DFT1 or fibroblasts. NES and DPYSL3 are detected in both DFT1 and fibroblasts but at lower levels than in DFT2. DPYSL2 is detected only in fibroblasts but at lower levels than in DFT2. CKB is detected in DFT1 at lower levels than in DFT2. This analysis indicates that DFT2 is overexpressing proteins which are normally overexpressed in the nervous system.

Not all proteins associated with a seemingly cell type specific functional term demonstrate a cell type specific expression pattern, and many are ubiquitous and widely expressed across different tissues. The specific proteins identified within each gene functional term were therefore evaluated individually to identify whether these datasets are truly enriched for nervous system restricted proteins and pathways.

Table 3.12: The top 50 most highly expressed proteins in DFT2 includes proteins which are overexpressed in the nervous system.

Values indicate the Log2 average raw expression value across replicates for each cell line. Values are colour coded based on expression values where light yellow is the highest expression, yellow – orange indicates decreasing expression values and red is the lowest. White and NA indicates a protein was not quantified in that cell line. Tissue expression of proteins was evaluated using the GeneCards database. Full data presented in Appendix C.2.

Protein	Expression	DFT2	DFT1	Fibroblasts
VIM	Ubiquitous	36.75533	NA	37.35678
HIST1H4A	Ubiquitous	35.77755	36.16101	37.06728
ANXA2	Ubiquitous	35.61617	NA	NA
PPIA	Ubiquitous	35.54697	34.6909	NA
NES	Primarily neurons, overexpressed in foetal brain	35.18522	29.52589	29.89302
TUBB	Ubiquitous	35.06995	34.15788	34.66112
ANXA1	Ubiquitous	34.60517	32.99945	NA
TUBB6	Ubiquitous	34.52379	NA	31.84346
HIST1H2AA	Widely expressed	34.34581	34.69782	NA
HIST1H2AB	Widely expressed	34.34581	NA	NA
HIST1H2AD	Widely expressed	34.34581	NA	NA
HIST1H2AG	Widely expressed	34.34581	34.69782	NA
PKM	Ubiquitous	34.32362	29.6764	29.29205
HIST1H1E	Widely expressed	34.23125	35.35678	34.77018
HMGB1	Ubiquitous	34.19992	NA	34.11458
S100A11	Ubiquitous	34.16289	31.99609	33.93591
SERPINH1	Widely expressed	34.13512	33.08849	NA
PGK1	Ubiquitous	34.07527	33.17335	NA
CALR	Ubiquitous	33.99889	33.59038	35.32138
GSN	Ubiquitous	33.87005	NA	33.75865
TPI1	Ubiquitous	33.86698	NA	35.17781
GPI	Ubiquitous	33.79136	NA	34.15537
DPYSL2	Ubiquitous, overexpressed in nervous system	33.72178	NA	32.92887
ARHGDIA	Ubiquitous	33.69021	NA	NA
CKB	Widely expressed, highest in brain and retina	33.68395	31.07793	NA
MDH1	Ubiquitous	33.68395	NA	32.39049
ALDOA	Ubiquitous	33.66642	32.9899	34.65402
ATP5F1B	Ubiquitous	33.60554	NA	34.16788
MAP4	Widely expressed	33.60112	NA	NA
CWC15	Widely expressed	33.57204	32.74535	NA
TAGLN2	Ubiquitous	33.54582	NA	34.74201
HNRNPA1	Ubiquitous	33.52652	33.58628	32.72788
CFL1	Ubiquitous	33.48352	33.3733	34.37417
MSN	Ubiquitous	33.46903	30.70196	33.68882
SOD1	Ubiquitous	33.44989	NA	33.80424
DPYSL3	Widely expressed, high in foetal brain	33.42384	31.90335	31.67275
PPIB	Widely expressed	33.40282	NA	33.96774
RPL12	Ubiquitous	33.39262	NA	33.58702
LMNA	Widely expressed	33.35635	34.71668	34.20963
RPS5	Ubiquitous	33.28416	NA	33.62709
UBA1	Ubiquitous	33.27262	NA	33.6724
FLNA	Widely expressed	33.25772	31.74134	35.63522
ANXA6	Ubiquitous	33.24926	33.95625	32.96143
ANXA5	Ubiquitous	33.2106	31.3845	34.07261
PHGDH	Widely expressed	33.17385	32.52807	28.94652
PRDX2	Ubiquitous	33.14579	NA	31.38792
EIF4A1	Widely expressed	33.10733	NA	33.15637
NGFR	Predominantly nervous system	33.10368	NA	NA
MDH2	Ubiquitous	33.09742	33.58292	33.69471
VAT1	Ubiquitous	33.07421	NA	31.78358

Nervous system associated functional terms identified in the unique quantified proteomes of each cell line (Section 3.3.2) were interrogated and the proteins associated with each functional term were extracted (Appendices C.5 and C.6). Each protein was searched against the GeneCards database to identify which proteins were largely ubiquitous or widely expressed and which would normally be specific to or overexpressed in nervous tissue in order to validate the functional enrichment of nervous system processes within DFT1 and DFT2.

Proteins were identified in the enriched nervous system functional terms in both DFT1 and DFT2 which are specific to or overexpressed in nervous system tissue, whilst none were identified in fibroblasts (Table 3.13). 22% of the proteins associated with neuronal ontologies (listed in Table 3.7 and Table 3.8) in the unique DFT2 proteome are either overexpressed or restricted to central and/or peripheral nervous tissue. Five of these proteins are specific markers for immature glial cells (FABP7, L1CAM, NGFR), pan-maturity glial cells (S100B) or neurons (TUBB3). Additionally, the quantified DFT2 proteome uniquely contains PDGFRA and CSPG4, marker proteins for oligodendrocyte progenitor cells and QKI, a marker of mature oligodendrocytes (Appendix C.2). Similarly, 25% of the proteins associated with neuronal functions in DFT1 are overexpressed in nervous system tissue, for example, CNP is a protein that is a specific marker of differentiating myelin producing glial cells. The quantified DFT1 proteome also uniquely expresses myelin component proteins MAG and MOG (Appendix C.2). Additionally, the non-quantified proteomic data for DFT1 (Appendix C.1) confirms the specific expression of PRX, which is detected in all DFT1 replicates but is not detected in any DFT2 or fibroblast replicates. Fibroblasts express no neuronal specific markers in this analysis, but the fibroblast specific markers FAP, COL1A1 and ACTA2 are uniquely quantified in the fibroblast proteome (Appendix C.2).

Table 3.13: DFT1 and DFT2 express proteins which are specific to or overexpressed in nervous system tissues.

Tables containing all proteins associated with nervous system functional enrichment categories (described in Section 3.3.2) for A) DFT1, B) DFT2 and C) Fibroblasts. Yellow highlight indicates proteins which are overexpressed in or specific to nervous system tissues and cell types. * indicates proteins which are a specific marker for neuronal or glial cells (Proteins extracted from analysis in Appendix C.5, functional terms of interest detailed in Appendix C.6)

A) DFT1							
ACO2	CCT3	EZR	MYH10	PSMA3	RBM8A	RPL8	SPTBN1
ACTB	CFL1	GNB1	MYH14	PSMB7	RDX	RPS11	TCP1
AKR1B1	CKB	GNB2	MYL6	PSMC1	RPL10A	RPS13	TLN1
AP2M1	CLTC	HSP90AA1	NDUFV2	PSMC3	RPL14	RPS20	TUFM
ATP1B1	CNP*	HSPA5	NME2	PSMC4	RPL17	RPS27L	UBA52
ATP5F1A	CSNK2A1	HSPA9	PAK2	PSMC5	RPL18A	RPS29	UCHL1
ATP5PB	DLAT	IDH3A	PDHA1	PSMC6	RPL23	RPSA	UQCRC1
ATP5PD	DLD	ITGA5	PHGDH	PSMD12	RPL26	SDHA	UQCRC2
ATP6V1B2	DPYSL3	MDH2	PKM	PSMD2	RPL5	SLC25A3	VCP
CANX	ELOB	MSN	PRDX3	PSMD7	RPL7A	SPTAN1	VDAC2

B) DFT2							
AAK1	ARL3	CSNK2B	FKBP15	KIF13B	PRKCA	RPL10	S100B*
ABI1	ARPC3	CUL2	FKBP4	L1CAM*	PSMA4	RPL13A	SDCBP
ABLIM1	ARPC5	CYFIP1	FLRT3	MAP4	PSMB2	RPL18	SLC25A4
ACTN4	ATP5F1C	DBNL	GAB1	MME	PSMB3	RPL22	SMN2
ACTR3	BAG3	DCLK1	GIGYF2	MPP1	PSMB6	RPL23A	SNAP23
AHCY	BASP1	DST	GNA11	MTPN	PSMD3	RPL29	STAU1
AHCYL2	BLVRB	DYNLL2	GNAI2	NCSTN	PSMD6	RPL34	STMN1
ANXA2	CAD	DYNLT1	GNAS	NDRG1	PSMD9	RPL37A	STXBP1
AP3D1	CKB	EEA1	GSK3B	NDRG2	PTBP2	RPL9	TKT
APP	COPS4	EIF4A3	HDAC6	NGFR*	PTK2	RPLP2	TMOD2
ARF4	COPS5	EIF5A	HNRNPH2	NQO1	PURA	RPS12	TRIO
ARHGDIA	CORO1A	FABP7*	ITGA1	PAFAH1B1	RAB13	RPS15	TUBB3*
ARHGEF2	CRYAB	FBXO7	ITPR3	PFN2	RAB5B	RPS21	TWF2
				PRKAR2A	RANBP1	RPS25	WDR1

C) Fibroblasts							
RPS23	RPL6	RPL31	RPS6	PSMD1	PSMA7	PFN1	ITGB1
RPS3	RPL3	RPS8	RPL27	PSMD11	RPL21	MYL9	PDLIM7

3.4 Discussion

Transmissible tumours are rare in nature, and the emergence of a second transmissible tumour in the Tasmanian devil within twenty years of the first was unprecedented. The gross phenotype of DFT2 tumours is indistinguishable from DFT1, and the cells transmit in a similar manner. Thus, it has been postulated that DFT2 may have emerged from DFT1, either as a rare subclone or due to a fusion event. Cytogenic and genetic analysis has since demonstrated that these two tumours are as distinct from each other as they are from their host devils, and there is no evidence that DFT2 has emerged as a DFT1 clone or from a fusion with DFT1 (Stammnitz *et al.* 2018; Pye *et al.* 2016b). Despite this, it has been shown that DFT1 and DFT2 have similar mutational signatures and broadly similar expression of neuroectodermal markers (Stammnitz *et al.* 2018) which indicates that despite emerging independently, these two tumours may share a common cellular origin.

In this project I have performed a whole cell proteome analysis on a DFT1, DFT2 and a fibroblast cell line in order to identify proteins and biological functions which further define the cellular origins of DFT2. The known cellular origins of both DFT1 and fibroblasts were used to control for the degenerate and non-specific nature of gene set enrichment analysis to prevent bias of the DFT2 dataset. Broad analysis of protein expression in the three cell lines indicates that most experimentally detected proteins are present in all three cell lines (Figure 3.3). In order to analyse protein expression levels, a subset of the whole cell proteomes for each cell line was quantified. Only a small number of proteins were quantified in all three cell lines (Figure 3.4), and there are significantly more proteins quantified in the DFT2 cell line than in DFT1 or fibroblasts, and additionally the DFT2 proteome has over twice as many unique quantified proteins as DFT1, and four times as many unique quantified proteins as fibroblasts. This may be due to differences in how long each cell line has been in culture. DFT2_RV was established in 2014 (Pye *et al.* 2016b) and has been in culture for 5 years, whilst both DFT1_4906 and Salem_Fibroblasts have been in culture since 2005 (Pearse *et al.* 2012). Continuous cell culture has been widely documented to have significant effects on gene and protein expression (Mouriaux *et al.* 2016; Wenger *et al.* 2004; Briske-Anderson *et al.* 1997), and indeed the fibroblast cell line used here has been documented as developing trisomy 6 during continuous culture (Murchison *et al.* 2012). Thus it can be inferred that the increased time in culture of DFT1_4096 and Salem_Fibroblasts has contributed to their significantly lower levels of unique and quantifiable proteins, which may be due to lower numbers of unique proteins in the cell lines, or more likely lower levels of these proteins which may fall below the limits of detection or quantification by mass spectrometry (Armbruster and Pry 2008). Lower protein expression seen in DFT1 compared to DFT2 may also relate to tumour evolution, and DFT1 may have downregulated the expression of many unique proteins during intense selective pressure as it spread through the environment, whilst DFT2 has had less time in the population to lose this protein expression. Other factors which may influence the protein expression discrepancies seen here include how the cells were isolated from the original tumour or tissue samples, the depth of tumour biopsy the tumour cells were isolated from, and how

long the tumour has been growing in the individual at the point of biopsy, although this information is not available.

Both the DFT1 and fibroblast proteomes required far more data imputation prior to hierarchical clustering due to the large number of unique proteins quantified in DFT2. This has resulted in weak clustering which is difficult to properly interpret (Figure 3.5). Clustering of expression of the 182 proteins quantified in all three cell lines does indicate closer similarity between DFT2 and DFT1 than fibroblasts, but this clustering is still weak and provides no clear conclusions. Additionally, this analysis provides no information on biological function, thus a more detailed analysis of the specific proteins and pathways enriched in these proteomes was required to define the origins of DFT2.

Gene set enrichment analysis was performed on the quantified proteomes of all three cell lines using a human background from four functional databases: GO: Biological process to identify functional biological pathways which are enriched in the dataset, GO: Cellular compartment to identify which part of the cell proteins in the dataset have derived from, Reactome to identify specific signalling and metabolic pathways enriched in the dataset and finally Human Protein Atlas to identify similarities between the dataset and protein expression in defined human tissues and cell types (Figure 3.6). However, clustering of these analyses was weak and difficult to interpret, in part due to data imputation which may skew the results of these non-normal datasets despite normalisation. This cluster analysis also gives no information on the biology of the functional enrichment in any of the datasets, just the similarity of enrichment between the three cell lines. Therefore, it is difficult to interpret the broad comparison of the three proteomes at this level and impossible to draw any conclusions regarding biological origins of DFT2.

A difficulty with interpreting gene set enrichment results is the degeneracy and the loose hierarchy of functional terms. One functional term has multiple ‘parent’ terms, which are usually nonspecific and hide cell type specific enrichment. Additionally, proteins have numerous functions within different cell types. Any given protein may function within a highly cell-type specific pathway, but also functions outside of that pathway in other cell types. This is highlighted by the strong enrichment for the highly neural specific “myelin sheath” cell compartment in the fibroblast proteome. One of the proteins within this cellular compartment in the fibroblast proteome is ATP5F1B, a subunit of mitochondrial ATP synthase, a ubiquitous protein complex which is expressed by almost all cell and tissue types (Collinson *et al.* 1996). As it has been demonstrated that the myelin sheath has its own ATP synthase (Ravera *et al.* 2011), an ubiquitous protein like ATP5F1B is additionally assigned a cell type specific functional term. Cases like this result in false positives during enrichment analysis can make enrichment data difficult to interpret.

Functional analysis was performed on the proteins uniquely quantified in a single cell type to reduce false positives and define DFT2 specific biological functions. Enrichment of nervous system specific GO:CC terms previously present in the fibroblasts analysis are not identified as enriched when analysing unique fibroblast proteins (Table 3.6), indicating that degenerate functional information

has been removed. Additionally, proteins unique to the Schwann cell derived DFT1 demonstrate significant enrichment for “myelin sheath”, “synapse” and “Schwann cell microvillus”, indicating that the analysis is defining cell-type specific functionality. The GO:CC functional category with the highest number of enriched terms following analysis of unique proteins in DFT2 is the nervous system. In particular, DFT2 is significantly enriched for “myelin sheath”, a cell compartment only present in Schwann cells or oligodendrocytes (Nave and Werner 2014). As myelin production is usually associated with cessation of cell growth (Monje *et al.* 2010), myelin sheath associated proteins are unlikely to be activated during carcinogenesis, and this enrichment may be a good indicator of the DFT2 progenitor cell.

DFT2 is highly enriched for signalling and metabolic pathways associated with the nervous system when analysed against the Reactome database (Table 3.8). The proteins unique to DFT2 show significant enrichment for 6 signalling pathways directly involved in the development and function of the nervous system, including “SLIT and ROBO signalling” and “regulation of axonal growth”. Slit-ROBO signalling is a crucial component of nervous system development through its control of axon guidance (Brose *et al.* 1999). However, axon guidance is a complex process mediated by extracellular signals from multiple origins (Myers *et al.* 2011), and is an early developmental function associated with organ development and cell proliferation (Blockus and Chédotal 2016) indicating these signalling pathways may have been activated during carcinogenesis or conversion to a cell line, as has happened with fibroblasts.

When analysed against the HPA database, DFT1 and DFT2 are consistently more similar to nervous system tissue and cell types than fibroblasts. This is true when looking at the full proteomes (Table 3.9, Figure 3.7), unique proteins (Table 3.10) or tumour-specific proteins (Table 3.11, Figure 3.8), indicating that there is a consistent pattern of similarity within both tumours to human nervous tissue..

As previously explained, the degeneracy of functional analysis means it is difficult to interpret these results in a biologically meaningful way without analysis of specific proteins within each functional term. To validate the functional specificity of the pathways enriched in this analysis, I manually analysed the expression profile of each protein that was associated with nervous system restricted functional terms following analysis of the proteins unique to a single cell line. Within these functional terms, DFT1 and DFT2 expressed 20 and 23 proteins respectively that are either restricted to or overexpressed in nervous system tissues such as the brain or peripheral nerve, whilst fibroblasts express none (Table 3.13), indicating the functional enrichment of the nervous system in DFT2 is likely valid.

Four proteins unique to DFT2 (FABP7, NGFR, L1CAM and S100B) (Table 3.13) are specific cell markers for cells of a myelinating glial lineage. FABP7 is a fatty acid binding protein associated with development of the radial glial fibre in the brain (Furuhashi and Hotamisligil 2008) which is specifically expressed in neural stem cells, oligodendrocyte precursor cells and immature Schwann cells, and is associated with increased cell growth which can lead to tumorigenesis in nervous tissue

(Sharifi *et al.* 2013; De Rosa *et al.* 2012; Sharifi *et al.* 2011; Miller *et al.* 2003). L1CAM is a neural adhesion molecule expressed by neurons, Schwann cells and oligodendrocytes (Mikulak *et al.* 2012) which is associated with cell development and growth. It is expressed at high levels in immature and precursor cell types, not in terminally differentiated myelinating cells, and is upregulated to induce dedifferentiation and promote cell growth and nerve regeneration following axon injury (Mirsky *et al.* 2008). NGFR, or p75NTR, is a neurotrophin receptor which regulates neuronal cell processes and is a marker for immature and non-myelinating Schwann cells (Saadipour *et al.* 2017; Liu *et al.* 2015). NGFR expression is associated with cell growth and it is upregulated in response to nerve injury or disease in both the central and peripheral nervous systems (Tomita *et al.* 2007; Zuliani *et al.* 2002). S100B is a transcription factor involved in differentiation and myelination which is a widely used pan-maturity marker for Schwann cells, oligodendrocytes and astrocytes, though evidence suggests it is more highly expressed in myelinating cells (Liu *et al.* 2015; Fujiwara *et al.* 2014; Hachem *et al.* 2005; Mata *et al.* 1990). The unique, high and consistent expression of these markers in the DFT2 proteome across three independent biological replicates indicates they are constitutively expressed by the tumour.

These markers are all associated with tumours of the central and peripheral nervous system in humans, but may also be upregulated in other tumour types (Morihiro *et al.* 2013; Tagliavacca *et al.* 2013; Rodriguez *et al.* 2012; Spyra *et al.* 2011), indicating that these markers may represent tumour evolution rather than the cellular origins. Indeed, all perceived similarities between DFT1 and DFT2 may be due to being derived from cancers whilst the fibroblast cell line is not. I would postulate that the expression of multiple glial markers by DFT2 and DFT1, combined with functional enrichment for the nervous system and the expression of proteins associated with increased expression in the nervous system makes this unlikely. Whilst not identified during functional analysis, the DFT2 proteome additionally uniquely expresses CSPG4 and PDGFRA. CSPG4, also known as NG2, combined with PDGFRA are widely used markers for oligodendrocyte precursor cells *in vivo* (Tripathi *et al.* 2010; Rivers *et al.* 2008). Interestingly, it has been demonstrated that DFT2 has a copy number variant at its PDGFRA locus (Stammnitz *et al.* 2018), and PDGFRA mutations are extremely common in glioma and glioblastomas (Koschmann *et al.* 2016; Ozawa *et al.* 2010). DFT2 also uniquely expresses the RNA binding protein QKI, which is an essential component of central nervous system myelination and is an oligodendrocyte marker (Chen *et al.* 2007). This detailed proteome analysis combined with the functional data indicate that DFT2 has most likely emerged from a myelinating glial cell type.

3.5 Conclusions and future direction

The data presented in this chapter indicate that DFT2 expresses similar proteins and is functionally similar to DFT1. Additionally, the expression of multiple specific markers and nervous system associated proteins strongly indicates that DFT2 has emerged from a similar cell type to DFT1, a myelinating glial cell. This analysis indicates that DFT2 uniquely expresses common oligodendrocyte markers and this new tumour may be oligodendroglial rather than Schwann cell, although further studies are needed to confirm this. This is a remarkable finding which has implications for the future of DFT research. Two transmissible tumours in just 20 years from similar cell types may indicate a propensity for myelinating glia in the Tasmanian devil to form tumours which can cross histocompatibility barriers, raising concerns that more of these tumours could emerge. However, it also implies that these two tumours may be expressing similar neoantigens, and may respond to a single vaccination strategy, something which would be more financially viable than two independent vaccines and would provide a higher level of protection for the wild species.

The next steps are to fully validate this cell type of origin through confirmation of further markers at an RNA and protein level in more DFT2 cell lines. I will then validate the expression of these markers in tumour samples to confirm these results are still applicable to the tumour *in vivo*. Both tumours can then be compared directly to their cell type of origin to identify neoantigens which may be useful in vaccine production.

Chapter 4 A comparison of two independent transmissible tumours to their progenitor cell type

4.1 Introduction

Comparison of the DFT1 and DFT2 proteomes has indicated that DFT2 is enriched for biological processes associated with the nervous system and has likely emerged from a similar cell type to its Schwann cell derived counterpart DFT1 (Murchison *et al.* 2010).

Myelinating glial cells exist in the adult body in multiple differentiation states, and in the case of Schwann cells, retain the ability to de-differentiate throughout an adult's life (Kim *et al.* 2014; Zhang *et al.* 1999). At present, little information is available on Tasmanian devil Schwann cells or oligodendrocytes, although the discrete stages of myelinating glial differentiation are well studied in other species. It is currently unknown the specific differentiation state of the DFT1 and DFT2 progenitor cells, although the data generated in Chapter 3 indicates they may have been different, with the DFT2 tumour likely deriving from a more immature myelinating cell. Understanding the differentiation state of the tumour progenitors is key to understanding the evolution of the tumours, and for identifying tumour specific targets which could be used to develop an anti-tumour vaccine.

The emergence of two tumours from such similar cell types indicates there may be an underlying propensity for myelinating glial cells in the Tasmanian devil to undergo malignant transformation and avoid immune detection. Indeed, there is evidence that DNA methylation in DFT1 is remarkably similar to that of devil peripheral nerve and is significantly higher in both than in other devil tissues (Ujvari *et al.* 2013), which may indicate an increased risk of cancer formation in devil Schwann cells than other tissue types (Asada *et al.* 2015; Kulis and Esteller 2010). Another potential risk factor in the evolution of immune escape is MHC class I expression, which is currently not defined in devil myelinating glial cells. DFT1 avoids the immune system by downregulating surface expression of MHC class I (Siddle *et al.* 2013), whilst DFT2 expresses both classical MHC class I and a non-classical MHC class I allele Saha-UK (Caldwell *et al.* 2018). Understanding MHC expression by the progenitor cells will determine whether the ability to cross histocompatibility barriers is an adaptation that emerged during tumour evolution or the successful exploitation of characteristics already present in the progenitor cell. Understanding MHC class I expression in the healthy progenitor cell is also important for designing an effective peptide vaccine strategy against tumour specific antigens presented by MHC class I molecules.

In this chapter, I will investigate further the hypothesis that DFT2 and DFT1 have emerged from the same progenitor cells in different differentiation states. I will also test my hypothesis that the emergence of two transmissible tumours from a single cell type is at least partly due to aberrant MHC class I expression in devil myelinating cells. Due to the challenges of sample collection (See Chapter

2.2.1) predominantly carried out in the field on wild animals, obtaining brain samples for analysis of oligodendrocytes is difficult. However, sciatic nerve samples are more easily obtained, and are highly enriched for Schwann cells, the known DFT1 progenitor. For this reason, the work in this project focuses solely on characterising devil Schwann cells, and comparing both DFT1 and DFT2 to this cell type.

4.2 Aims and objectives

In this chapter I will assess the expression of myelinating cell markers from different differentiation states, namely immature Schwann cells or oligodendrocyte precursor cell markers and myelinating Schwann cell and oligodendrocyte markers in DFT1 and DFT2 to ascertain the differentiation state of their original progenitor cell, using devil sciatic nerve samples as a Schwann cell enriched control. I will also assess the expression of MHC class I in devil sciatic nerve, and by proxy devil Schwann cells, to ascertain whether MHC class I expression in DFTs is an adaptive evolutionary response or an artefact of the progenitor cell.

This chapter has the following objectives:

1. Assess by RT-PCR and IHC the expression of a variety of myelinating cell associated markers in Tasmanian devil sciatic nerve samples, and DFT1 and DFT2 cell lines to identify the differentiation state of the DFT1 and DFT2 progenitors.

2. Assess the expression of MHC class I components on a transcript and protein level in Tasmanian devil sciatic nerve samples.

4.3 Results

4.3.1 DFT1 is more enriched for markers of the glial restricted myelin sheath than DFT2

Proteomic analysis of DFT1 and DFT2 cell lines identified proteins present in both tumours that are associated with the nervous system, in particular myelinating glial cells (See Chapter 3.3.4). Gene expression of nervous system associated markers in 3 DFT2 and 3 DFT1 cell lines was assessed alongside devil sciatic nerve and a fibroblast cell line by RT-PCR. cDNA concentration across cell lines was kept constant and RPL13A was used as a control. Due to the variation in tissue quality and cell types within the sciatic nerve sample this PCR is not considered to be quantitative. As indicated in Figure 4.1, RT-PCR shows that sciatic nerve is positive for MPZ, MBP, PDGFRA and PMP22 transcripts and weakly positive for FABP7 transcripts. Sciatic nerve would normally be expected to be positive for NFGR, L1CAM and MAG, however no transcripts were detected. Fibroblasts are either negative or weakly positive for all transcripts compared to the DFT2 cell lines. All DFT2 cell lines express transcripts of all of these markers, whereas DFT1 is negative for PDGFRA, FABP7 and L1CAM, markers of immature Schwann cell differentiation states. DFT1 cell lines are consistently more enriched for myelin associated markers MAG, MPZ, MBP and PMP22 than DFT2 cell lines, indicating a myelinating Schwann cell phenotype in DFT1.

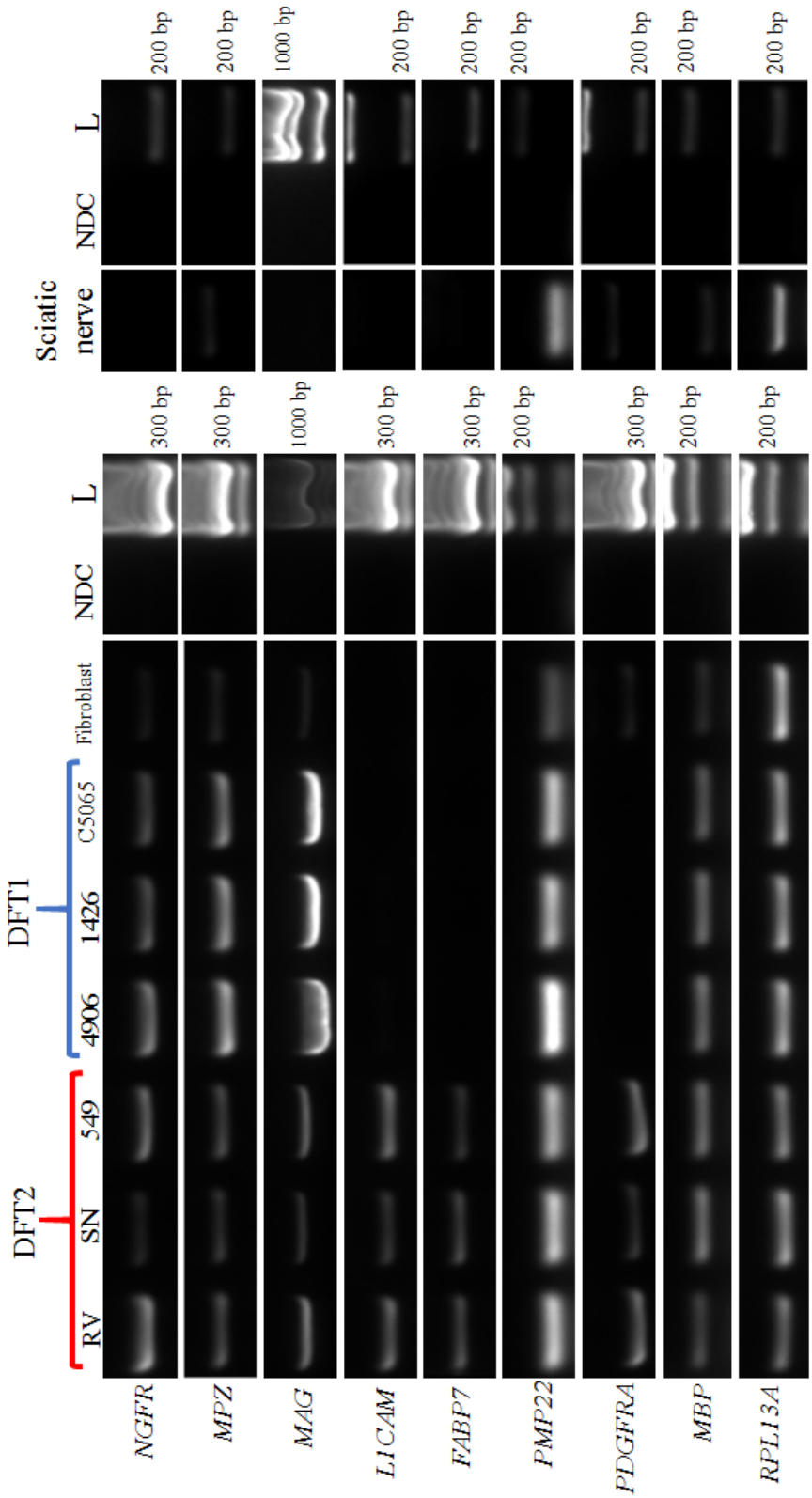


Figure 4.1: RT-PCR of Schwann cell markers in DFT1 and DFT2 cell lines and on a formalin fixed sciatic nerve sample.

Each PCR reaction contains 250 ng cDNA, and RPL13A is a housekeeping control gene. Sciatic nerve PCR reactions were run separately with an extended cycling programme and are not considered quantitative. NDC denotes a no cDNA negative control. L denotes a DNA ladder of known size. Ladder size is marked.

IHC was performed on DFT1 and DFT2 tissue sections to assess expression of three Schwann cell markers, PRX, S100 and MBP, alongside sciatic nerve sections to further assess their expression (Figure 4.2). IHC analysis is restricted to these three markers due to the difficulties in finding cross-reactive antibodies for the Tasmanian devil. Positive Schwann cells can be identified within the nerve sections, staining for PRX, S100 and MBP. PRX and MBP staining appears predominantly localised to the cell membrane with some cytoplasmic staining, whilst S100 staining is cytoplasmic, which fits with expected cellular localisation of these proteins. DFT1 tumour cells are positive for all markers, and DFT1 stroma is positive for MBP and S100 and negative for PRX, a widely used DFT1 diagnostic marker (Tovar *et al.* 2011). PRX and S100 staining of DFT1 tumour cells is predominantly cytoplasmic with some membrane localisation, expected localisations for these proteins. MBP staining in DFT1 tumour cells is cytoplasmic with regions of membrane localisation, particularly towards the outer edges of tumour bundles. As has been previously described, DFT2 is negative for PRX, and here we show it is weakly positive for S100 and positive for MBP. Both tumour cells and stroma are positive for MBP, whilst only tumour cells are positive for S100 but expression is minimal is and highly variable. Both MBP and S100 appear predominantly cytoplasmic with some regions of membrane localisation.

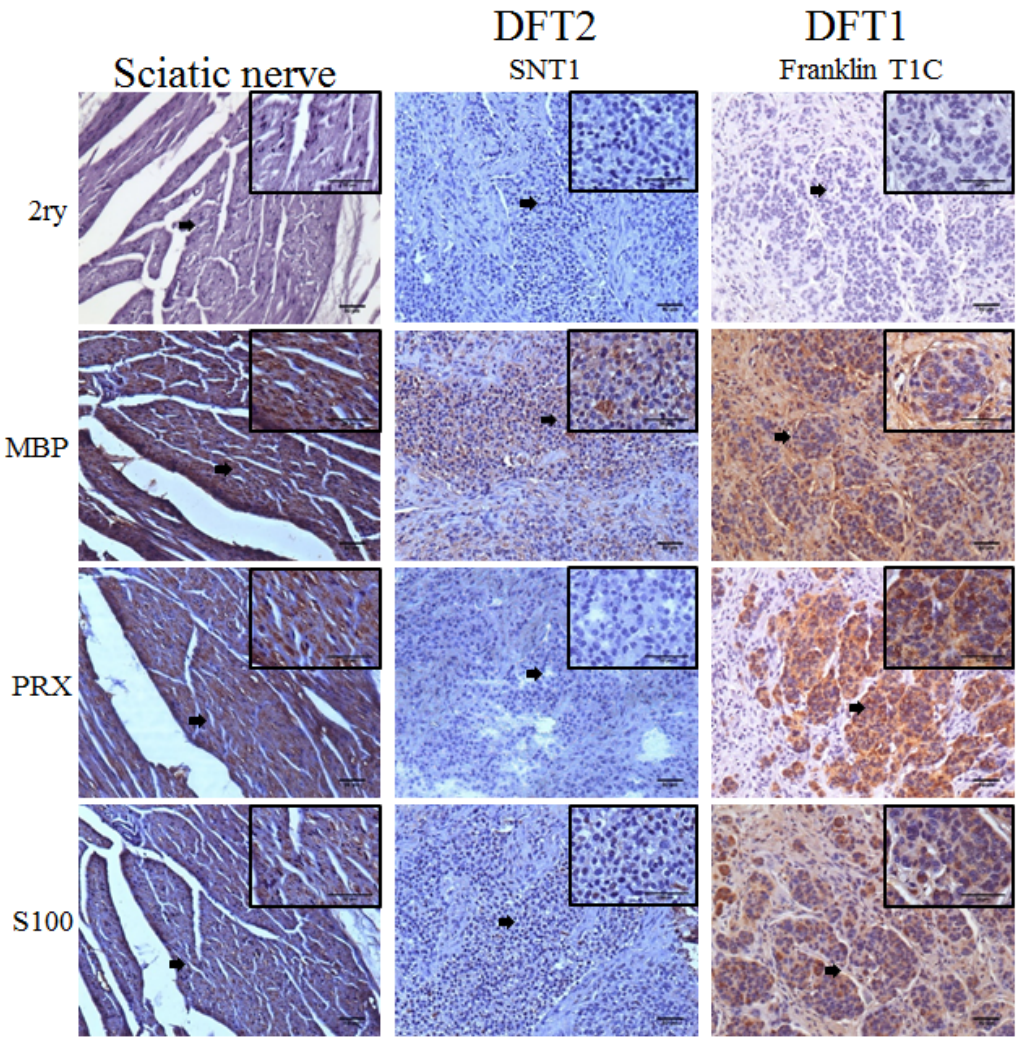


Figure 4.2: IHC assessing the expression of Schwann cell marker proteins in sciatic nerve, DFT1 and DFT2 tissue samples.

Brown staining indicates the protein of interest. Blue staining indicates nuclei. Main images taken at 200x magnification. Boxed images are taken at 400x magnification and represent cells of interest as indicated by black arrows. Sections are cut at 7 µm thickness. Scale bars represent 50 µm.

4.3.2 Tasmanian devil Schwann cells only express the non-classical MHC class I gene *Saha-UK*

The expression of MHC class I associated genes in devil sciatic nerve was assessed by RT-PCR and compared to DFT1 (DFT1_4906) and DFT2 (DFT2_RV). A single primer set was used to amplify the three classical MHC class I heavy chain genes, *Saha-UA/B/C,* which due to sequence similarity can only be amplified together. Gene specific primers were used to amplify two non-classical MHC class I heavy chain genes, *Saha-UK* and *Saha-UD*, and the protein complex component β_2-m (all primers are described in Caldwell *et al.* (2018) and Appendix B.2.

Sciatic nerve expresses high levels of a non-classical MHC class I heavy chain transcript, Saha-UK, and of the accessory protein β_2-m, but no transcripts of the classical MHC class I genes, *Saha-UA/B/C* or the non-classical *Saha-UD* were detected (Figure 4.3). In comparison, DFT2_RV expresses transcripts of classical MHC class I genes and of the non-classical *Saha-UK*, along with high levels of β_2-m and no detectable *Saha-UD*, whilst DFT1_4906 expresses much lower levels of β_2-m and *Saha-UK*, trace levels of *Saha-UD* and no detectable *Saha-UA/B/C*. Note that comparison of expression levels between sciatic nerve and the two cell lines is not quantitative due to variation in the tissue and RNA quality from formalin fixed sciatic nerve samples.

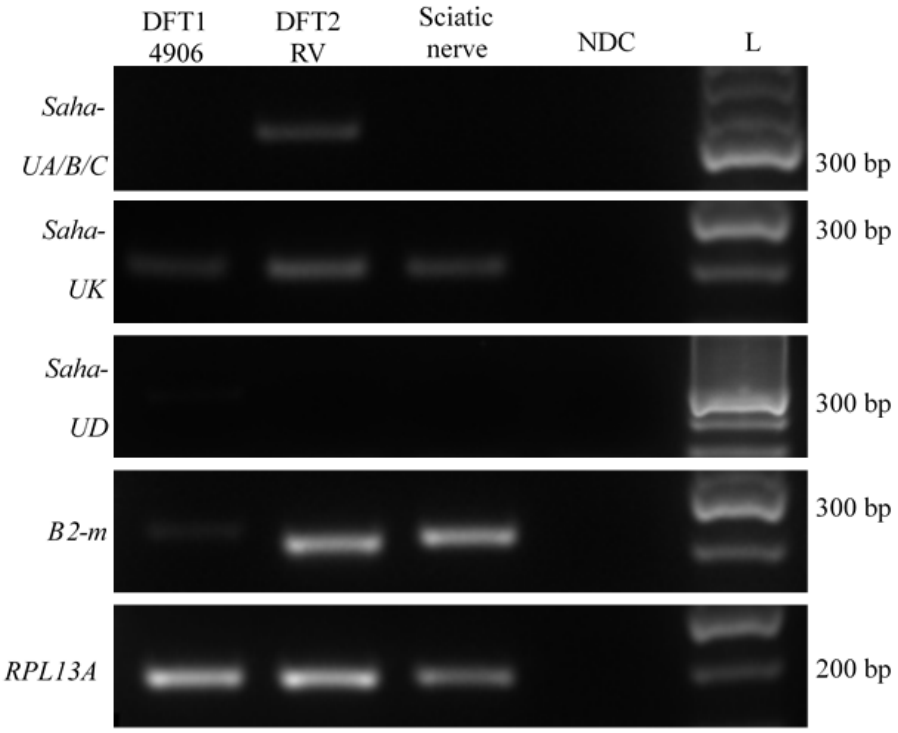


Figure 4.3: RT-PCR comparing the expression of MHC class I associated genes in devil sciatic nerve samples.

250 ng cDNA used per reaction. NDC denotes a no cDNA negative control. L donates a DNA ladder of known molecular size. Ladder DNA marker sizes are indicated

IHC was performed to assess the levels of MHC class I protein in formalin fixed paraffin embedded sciatic nerve samples. IHC was performed using two antibodies previous generated in our laboratory in collaboration with Professor Karsten Skjødt at the University of Southern Denmark, one of which detects Saha-UK (UK_15-29-1) and the other which detects all three classical MHC class I heavy chains (UA/B/C_15-25-18). Antibodies are described in Caldwell *et al.* (2018) and Table 2.2.

Sciatic nerve is positive for Saha-UK, with cytoplasmic and membrane localisation of the protein surrounding nerve bundles indicating cell surface expression of the protein on Schwann cells. Conversely, sciatic nerve is almost entirely negative for Saha-UA/B/C, supporting the RT-PCR data (Figure 4.4), and demonstrates predominantly cytoplasmic rather than membrane staining in the few positive cells. IHC staining using the same antibodies on DFT1 (Figure 1.10) and DFT2 (Figure 1.15) samples have previously demonstrated results that support the RT-PCR data shown in Figure 4.3 (Caldwell *et al.* 2018; Siddle *et al.* 2013)**,** however this is the first comparison to sciatic nerve.

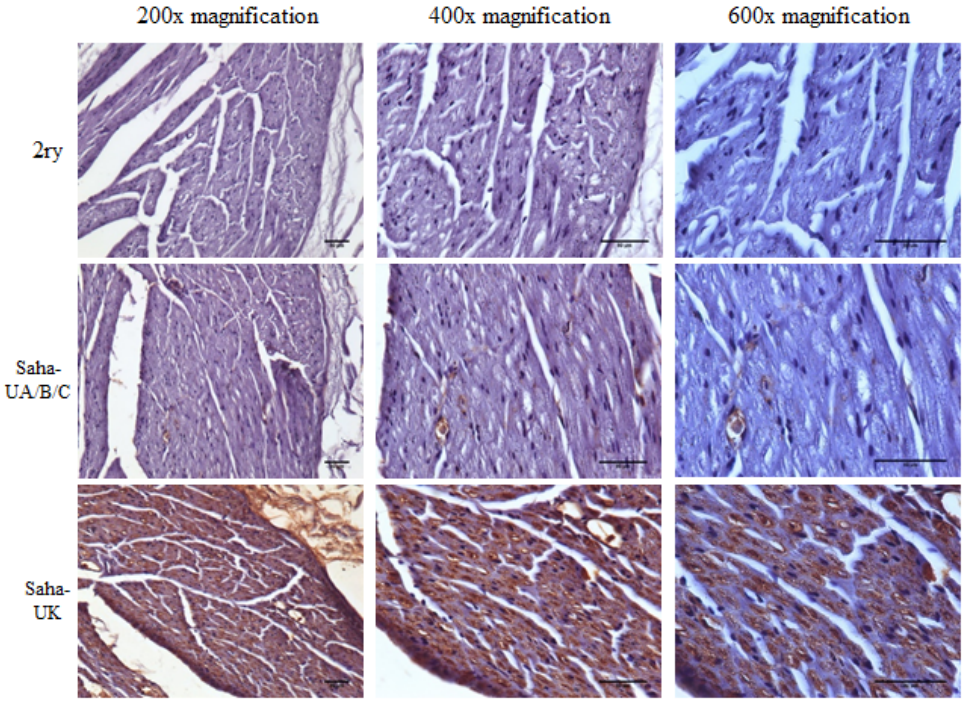


Figure 4.4: IHC assessing MHC class I protein expression in sciatic nerve samples.
Brown staining indicates the protein of interest. Blue staining indicates nuclei. Sections are cut at 7 µm thickness. Scale bar indicates 50 µm

4.4 Discussion

As discussed in Chapter 3, DFT2 has likely originated from a myelinating glial cell, though it is difficult to define whether DFT2 is of Schwann cell or oligodendrocyte origin, and the differentiation state is unclear.

Ideally, DFT2 and DFT1 could be compared to both oligodendrocyte and Schwann cell samples at a range of differentiation states, to define the specific origins of both tumours. Initial work was undertaken to establish a devil Schwann cell culture system for this analysis, which was optimised in adult mice (Appendix D.1). Extreme species specificity of Schwann cell culture (Wang *et al.* 2013; Kaewkhaw *et al.* 2012), the restriction of using senior animals due to the expected lack of neonatal samples for isolation (Andersen *et al.* 2016), and slow, passage restricted growth of the isolated cells indicated that the establishment of a continuous culture for this work was not viable to set up within the time and financial constraints of the project. As an alternative analysis was performed on tissue samples. Due to sampling occurring primarily in the field, brain and oligodendrocyte samples from devils are extremely difficult to obtain, so this work uses sciatic nerve samples preserved in RNA later and formalin as a progenitor cell comparison.

The sciatic nerve is highly enriched for Schwann cells, however there are other cell types present in the tissue sample, including neurons and fibroblasts, as well as an immature and non-myelinating Schwann cell population, which will reduce the relative expression of Schwann cell markers in the tissue sample (Stierli *et al.* 2018), and must be considered when interpreting protein and gene expression data. For example, Schwann cells associated with a large axon such as the sciatic nerve are expected to be myelinating Schwann cells (Spiegel *et al.* 2007) but the RNA isolated from devil sciatic nerve shows low expression of peripheral myelin associated genes MPZ, MBP and MAG (Figure 4.2). NGFR and L1CAM are markers of immature Schwann cells, and this cell type is likely not abundant enough in the tissue sample for their gene expression to be detected (Klein *et al.* 2014). Three proteins associated with the myelination pathway and known to be markers of myelinating Schwann cells, MBP, PRX and S100, (B. Jang *et al.* 2017; Liu *et al.* 2015; Gillespie *et al.* 1994) can be identified in tissue sections of the nerve localised around the outside of the nerve bundles, indicating their expression by Schwann cells within the sample. This data suggests the enrichment of devil sciatic nerve for Schwann cells and demonstrates that Schwann cell markers normally seen in other species are also present in the devil.

Schwann cells exist in a number of differentiation states throughout adult life ranging from precursor and immature Schwann cells, which express genes associated with growth and development, through to the terminal states of myelinating and non-myelinating Schwann cells (Section 1.3.2). Myelination pathway components are only expressed at high levels in myelinating and pro-myelinating Schwann cells, and increased production of myelin components is normally associated with growth cessation (Monje *et al.* 2010), making them a useful measure of differentiation state, as they are highly cell type specific and unlikely to be upregulated during tumourigenesis. As demonstrated in both Figure

4.1 and Figure 4.2, DFT1 cell lines and tumour samples are consistently more enriched for myelin associated genes and proteins than DFT2, whilst also lacking in glial developmental genes FABP7, PDGFRA and L1CAM (Schaefer and Frotscher 2012; Rivers *et al.* 2008; Miller *et al.* 2003), indicating that DFT1 likely emerged from a more differentiated progenitor, a myelinating or a pro-myelinating Schwann cell. This data is supported by the fact that DFT1 marker protein PRX is a highly specific marker of myelinating Schwann cells (Tovar *et al.* 2011).

Protein and gene expression analysis indicates that DFT2 cell lines and tumours express myelin associated markers (Figure 4.1, Figure 4.2, Table 3.12, Table 3.13), indicating it emerged from a myelin competent cell lineage, supporting analysis by Stammnitz *et al.* (2018) indicating that DFT2 emerged from the neuroectoderm. Expression levels of myelin associated markers in DFT2 tumour and cell lines are consistently lower than seen in DFT1, indicating that DFT2 emerged from a less differentiated progenitor cell, such as an immature Schwann cell. It is still unclear whether DFT2 has emerged from a Schwann cell or an oligodendrocyte lineage cell in the CNS. MPZ, a myelin glycoprotein specific to peripheral myelin (Kilfoyle *et al.* 2006) is expressed in DFT2 cell lines suggesting a Schwann cell lineage, but DFT2 cell lines additionally express high levels of PDGFRA, which is often used along NG2 as a marker of oligodendrocyte precursor cells (OPCs) (Tripathi *et al.* 2010). As demonstrated in Chapter 3.3.4 (Appendix C.2), NG2 is uniquely detected in the DFT2 proteome, and this combination of markers may instead indicate an OPC origin. Interestingly, there is evidence to suggest that PDGFRA+ NG2+ OPCs are able to differentiate into MPZ+ Schwann cells in response to CNS demyelination caused by injury or disease (Assinck *et al.* 2017; Zawadzka *et al.* 2010). The seemingly dual profile of DFT2 may be a reflection of this cross differentiation process, and DFT2 may have originated from an OPC immediately following injury, resulting in a Schwann cell profile as the tumour re-differentiates.

The emergence of two transmissible tumours in the Tasmanian devil in just 20 years from a similar cell type is particularly unexpected due to the fact that malignant and metastatic tumours from Schwann cells and oligodendrocytes are reasonably rare occurrences in humans. Two major tumour types emerge from Schwann cells in humans, Schwannomas and Neurofibromas. Schwannomas, which account for approximately 5% of soft tissue neoplasms in humans, are benign encapsulated tumours which consist entirely of Schwann cells, with less than 1% of tumours progressing to malignancy (Biswas *et al.* 2007). Neurofibromas are more complex tumours, incorporating several local cell types including fibroblasts, nerve cells and immune cells, though progression is driven by non-myelinating Schwann cells (Ortonne *et al.* 2018; Zheng *et al.* 2008). Two types of neurofibroma exist; dermal neurofibromas which form in the nerves of the skin and have shown no evidence of malignant progression in humans, and plexiform neurofibromas which form in the deeper internal nerves, around 5% of which progress to malignancy (Rodriguez *et al.* 2012; Perry *et al.* 2001). When these tumours become malignant they are known as Malignant Peripheral Nerve Sheath Tumours (MPNSTs), a rare but highly aggressive tumour type which accounts for 3-10% of soft tissue neoplasms diagnosed in humans each year (Goertz *et al.* 2014). Interestingly, evidence has

demonstrated that MPNSTs avoid the immune system by dysregulating the expression of MHC class I and class II and other antigen presentation components, a similar mechanism of immune evasion to DFT1 (Lee *et al.* 2006). Similarly, oligodendrocyte tumours are uncommon in humans, accounting for approximately 3% of primary brain tumours (Cancer Research UK 2019; McCarthy *et al.* 2011), and whilst these tumours are often malignant, metastatic disease is incredibly rare (Zhou and Reddy 2014). Thus, the emergence of two aggressive and metastatic tumours from these cell types in the devil contrasts with the literature surrounding human myelinating cell neoplasms, and may reflect an increased propensity of myelinating glia in the devil to form aggressive tumours compared to humans, or may indicate the existence of an exogenous factor which has induced multiple malignant transformations in Tasmanian devil myelinating glia.

Tasmanian devil sciatic nerve and Schwann cells do not appear to express classical MHC class I on a transcript or protein level. Schwann cells are known to express low but usually detectable levels of MHC class I (Meyer zu Horste *et al.* 2010; Meyer Zu Hörste *et al.* 2008). MHC class I expression is similar in oligodendrocytes, with normally low levels which increase in response to infection and inflammation (Redwine *et al.* 2001). Tasmanian devil sciatic nerve expresses no detectable classical MHC class I transcripts, and almost negligible classical MHC class I heavy chain protein, with few cells in the section appearing positive. In contrast, sciatic nerve expresses both heavy chain protein and high transcript levels for the non-classical MHC class I, Saha-UK. DFT1 avoids the anti-graft response between hosts at least in part by lack of MHC class I expression (Siddle *et al.* 2013) and the data presented here indicates that this mechanism of immune evasion may be a pre-existing characteristic of the DFT1 progenitor cell. Caldwell *et al.* (2018) has demonstrated that DFT2 expresses MHC class I (Figure 1.15) and the tumour is currently restricted to animals with shared MHC class I alleles (Table 1.2). DFT2 tumours express the non-classical MHC class I allele Saha-UK, and it has been speculated that this molecule may have an immunosuppressive function similar to those documented in human non-classical MHC class I molecules (Smyth *et al.* 2013). The overexpression of Saha-UK seen in DFT2 may be a remnant of its glial predecessor which has provided a selective advantage to the tumour as it has continued to spread. This MHC class I profile implies that devil Schwann cells have a naturally immune evasive phenotype, which may be contributing to the repeated formation of transmissible tumours in the species.

4.5 Conclusions and future directions

The data gathered here has confirmed that DFT2 has emerged from a myelin competent cell lineage in the devil, similar to DFT1, though it is likely that DFT2 emerged from a less differentiated cell than DFT1. It is possible that DFT2 may have originated from an oligodendrocyte precursor cell in the CNS, although further work is needed to determine the exact cell type of origin.

Interestingly, the MHC class I expression phenotype in devil sciatic nerve and Schwann cells appears to be primed for immune evasion. This tissue does not express classical MHC class I protein or transcript, something which has been carried over by DFT1 during its evolution. Similarly, sciatic nerve shows strong protein and transcript expression of a non-classical MHC class I molecule, Saha-UK. This molecule has been implicated in aiding the spread and immune evasion of DFT2, and its overexpression was likely carried over from the DFT2 progenitor.

The emergence of myelinating glial derived tumours in humans is a reasonably uncommon occurrence, and it is rare for these tumours to progress to metastatic disease. Thus, the emergence of two independent transmissible tumours in the Tasmanian devil from Schwann cells and/or oligodendrocytes is a surprising event, and may be due to inherent characteristics of these cell types in the devil, or may indicate an as yet undetected exogenous cause of malignant transformation. Further study of the mechanisms of immune evasion and evolution in these unique tumours may provide insight into how the rare cases of metastatic myelinating glial disease emerge in humans, informing the fundamental biology behind Schwann cell and oligodendrocyte interactions with the immune system during disease progression. Indeed, the identification in this chapter of an immune evasive MHC class I phenotype in devil sciatic nerve and Schwann cells may present a potential mechanism for malignant transformation and metastasis of human Schwann cell tumours.

The similar but independent evolutionary origin for DFT2 and DFT1, coupled with an immune evasive MHC class I phenotype in mature devil sciatic nerve raises some concerns that more of these tumours may form, putting the future of the species at risk. However, the shared origins of the tumours also present the possibility that these two independent tumours may respond to a single treatment or prevention plan, which would make preventing their spread in the wild population easier and more economically viable.

The next steps will be to analyse the cell type specific markers that are shared between DFT1 and DFT2, particularly to identify proteins which may be mutated or otherwise unique to the tumours, in order to identify viable vaccine targets which could be effective at preventing both. The beginnings of this work are detailed in Chapters 7 and 8.

Chapter 5 An investigation of the cellular response of DFT2 to IFNγ

5.1 Introduction

Interferon-gamma (IFNγ) is an inflammatory cytokine with a well-documented role in tumour immunosurveillance and clearance of neoplastic cells from the system (Wang *et al.* 2014; Shankaran *et al.* 2001). IFNγ is produced by a large array of immune cells, including CD4+ and CD8+ T-cells (Matsushita *et al.* 2015; Corthay *et al.* 2005) and NK cells (Keppel *et al.* 2015) in response to several mitogens and cytokines including IL-2 and IL-12 (Ohteki *et al.* 1999; Kasahara *et al.* 1983). IFNγ mediates the transcription and expression of many genes, including components of inflammatory and apoptotic signalling pathways and the cell cycle (de Veer *et al.* 2001). IFNγ has been demonstrated to have a role in initiating immune responses against tumour cells (Street *et al.* 2002) and guiding immunoreactive T-cells to the tumour site (Nakajima *et al.* 2001), but there is also evidence that it can have an immunosuppressive function by preventing clearance of tumour cells by antitumour immune responses (S.-J. Lee *et al.* 2006).

It has been previously demonstrated that treatment of DFT1 with IFNγ causes an upregulation of surface MHC class I (Figure 1.11) (Siddle *et al.* 2013). Schwann cells, which normally express low levels of MHC class I, are also known to upregulate expression of these molecules following stimulation with IFNγ (Meyer Zu Hörste *et al.* 2008), indicating that the response seen by DFT1 is consistent with that of its cellular progenitor. It has also been demonstrated that DFT1 cell lysates pre-treated with IFNγ are more effective at raising an antibody response in host devils than untreated DFT1 lysates, and IFNγ treated cells can be used as immunotherapy against MHC class I negative DFT1 (Tovar *et al.* 2017) indicating that treatment with IFNγ makes DFT1 cells more immunogenic. IFNγ induced MHC class I upregulation and increased anti-tumour immunity has been documented in other tumour models (Martini *et al.* 2010). Nothing is known about how DFT2 responds to IFNγ treatment, and whilst it currently expresses MHC class I, evidence indicates the tumour is losing this expression as it spreads (Caldwell *et al.* 2018), and it may eventually lose expression entirely. In this case, understanding how MHC class I expression can be modulated in these cells will be essential for developing an effective vaccine against the tumour.

Different cell types respond differently to IFNγ treatment (O'Neil *et al.* 1999; Yang *et al.* 1995). In this thesis I have presented data showing that DFT2 likely emerged from a similar cell type to DFT1, either a Schwann cell or an oligodendrocyte. As both Schwann cells and oligodendrocytes upregulate MHC class I in response to IFNγ (Höftberger *et al.* 2004), I hypothesise that DFT2 will have a similar IFNγ response to DFT1 and should upregulate MHC class I following treatment.

5.2 Aims and Objectives

In this chapter I will assess the cellular response of DFT2 cell lines to stimulation with the inflammatory cytokine IFNγ. I will focus on how the expression of MHC class I and associated antigen processing pathway components is affected by IFNγ treatment in DFT2, and how this compares to DFT1, to establish whether IFNγ treatment is likely to increase the immunogenicity of DFT2. I will also assess how the expression of some immune and tumour associated genes are affected by IFNγ treatment in DFT2 to identify any differences between the tumours which may be aiding DFT2s spread (Table 5.1). Finally, I will look at how the expression of some myelin associated genes is affected by IFNγ treatment in both DFT1 and DFT2, and whether this correlates with what would be expected from healthy Schwann cells or oligodendrocytes (Table 5.1).

This chapter has the following objectives:

1. Assess the cell surface expression of MHC class I component protein β_2-m before and after IFNγ treatment in DFT2 cell lines.

2. Analyse by RT-PCR any changes in expression of specific MHC class I genes in response to IFNγ by DFT2 cell lines.

3. Analyse by RT-PCR the expression of a panel of genes associated with immune signalling and tumour formation in response to IFNγ in both DFT1 and DFT2 cell lines.

4. Analyse by RT-PCR the expression of a panel of genes associated with myelin and the DFT1 and DFT2 cell type of origin in response to IFNγ.

Table 5.1: Non-MHC markers assessed in this chapter and their relevance to the study.

Marker	Normal function	Relevance to DFTs
TGFB1	Immunoregulatory cytokine	Immune evasion mechanisms
TGFB2	Immunoregulatory cytokine	Immune evasion mechanisms
PD-L1	Suppression of T-cell mediated cell death	Immune evasion mechanisms
STAT3	Transcriptional activator	Mechanisms of tumourigenesis
IL1R1	Inflammatory signalling receptor	Immune evasion mechanisms
MAG	Myelin structural protein	Tumour origins
MBP	Myelin structural protein	Tumour origins
MPZ	Myelin structural protein	Tumour origins
NGFR	Neuronal growth receptor	Tumour origins
L1CAM	Neural-specific cell adhesion molecule	Tumour origins
FABP7	Neural-specific fatty acid binding protein	Tumour origins
PDGFRA	Oligodendrocyte precursor marker	Tumour origins

5.3 Results

5.3.1 DFT2 upregulates MHC class I expression in response to IFNγ

Three DFT2 cell lines were treated in triplicate with devil IFNγ for 16 hours alongside one DFT1 cell line (protocol previously described in Siddle *et al.* (2013)). Cells were harvested and either live-stained with an antibody against β_2-m to assess surface expression of MHC class I by flow cytometry or RNA was extracted for gene expression analysis by RT-PCR. β_2-m expression is presented as the mean fluorescence intensity (MFI) of each sample.

Flow cytometry analysis demonstrates that all DFT2 cell lines upregulate cell surface β_2-m in response to IFNγ treatment (Figure 5.1). All three DFT2 cell lines express β_2-m to a higher level than DFT1 post-IFNγ, with DFT2_SN elevating β_2-m to the highest level. Quantification of the average β_2-m expression in DFT2 and DFT1 cells pre and post IFNγ treatment demonstrates that β_2-m is significantly upregulated in all three DFT2 cell lines post IFNγ treatment, with the highest levels demonstrated by DFT2_SN (MFI = 207.22), followed by DFT2_RV (MFI = 146.81) then DFT2_549 (MFI = 131.65) (Figure 5.2a). Comparison of β_2-m expression between cell lines both pre and post IFNγ treatment demonstrates that despite expressing higher levels of β_2-m than DFT1 pre-IFNγ, all DFT2 cell lines upregulate β_2-m significantly more than DFT1 post IFNγ treatment (Figure 5.2b).

RT-PCR analysis of treated and untreated DFT2 cells indicates that all DFT2 cell lines upregulate transcripts of all classical MHC class I, the non-classical MHC class I *Saha-UK* and β_2-m in response to IFNγ treatment in the same manner as DFT1 (Figure 5.3). DFT2_SN and DFT2_549 marginally upregulate the non-classical MHC class I *Saha-UD* in response to IFNγ. DFT2 cell lines upregulate classical MHC class I and β_2-m transcription in a comparable manner to DFT1 and appears to upregulate *Saha-UK* at a slightly lower level than DFT1. DFT2_549 and DFT2_SN appear to upregulate *Saha-UK* more than DFT2_RV. DFT2_SN upregulates transcription of β_2-m more than DFT2_RV and DFT2_549, whilst levels of β_2-m transcript post treatment in DFT1 are comparable with pre-treatment levels in DFT2 cell lines, supporting the data seen by flow cytometry.

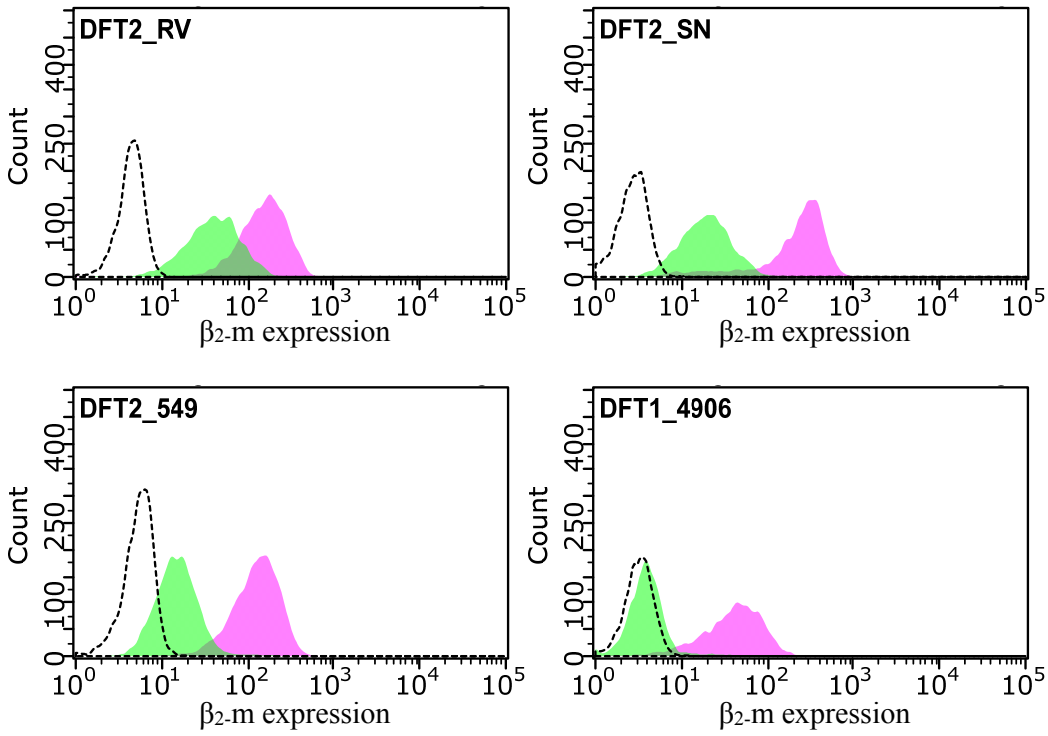


Figure 5.1: Increased cell surface β2-m expression following IFNγ treatment of DFT2 cell lines.
Representative histograms indicating the mean fluorescence intensity of AlexaFluor488 bound to cell surface β_2-m before and after treatment with IFNγ. Cell line is indicated in the top left hand corner of each graph. Y axis is cell count, X axis is fluorescence intensity. Black dashed line indicates secondary only negative control cells. Green peaks indicate untreated cells stained for cell surface β_2-m. Pink peaks indicate cell lines treated with IFNγ for 16 hours and stained for cell surface β_2-m. Gating strategy represented in Figure 2.2 and Appendix E.1. Data analysed using Guavasoft InCyte 3.3 software. Triplicate data presented in Figure 5.2 and Appendix E.1 (Figure S 3 – S 6).

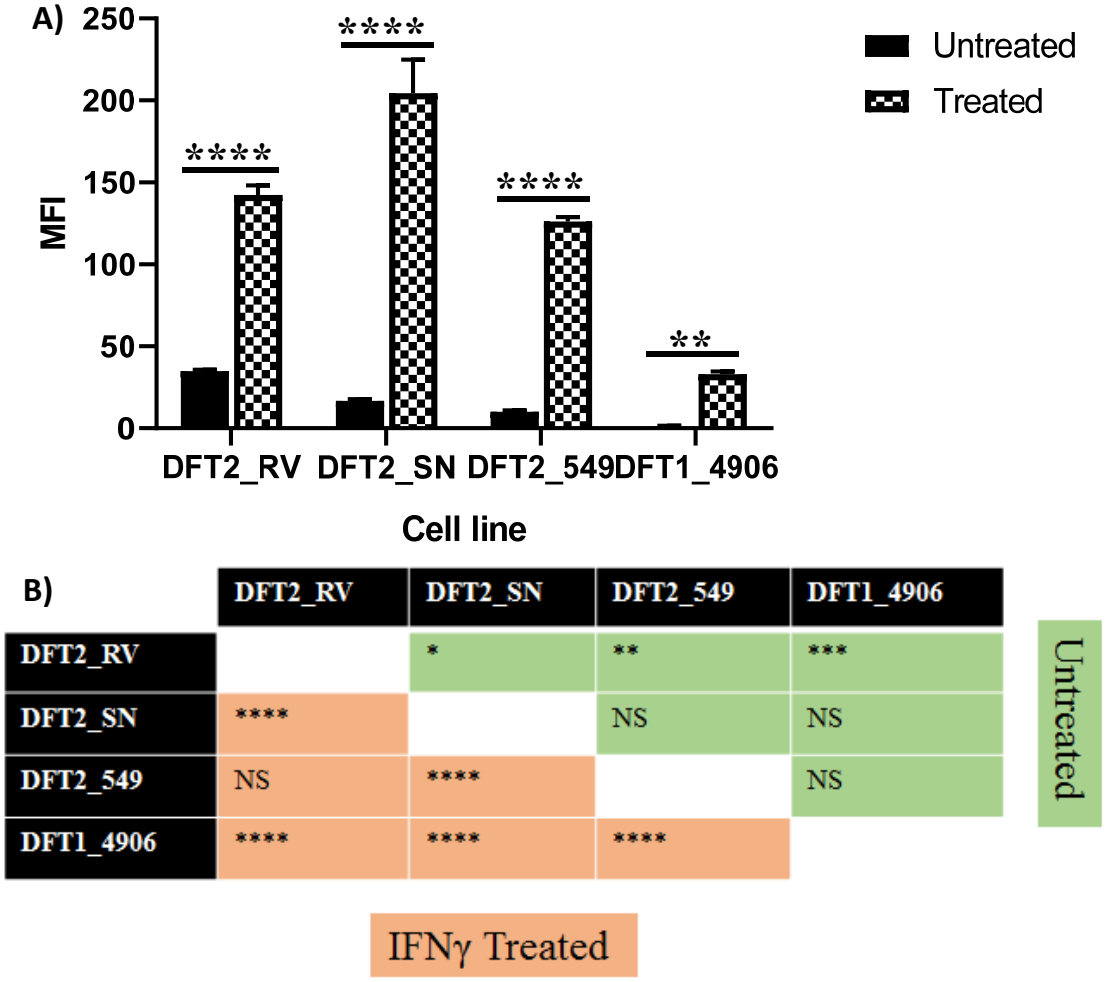


Figure 5.2: DFT2 cell lines upregulate β_2-m expression significantly more than DFT1.

A) Bar graph indicating mean upregulation of surface β_2-m following triplicate IFNγ treatment in three DFT2 cell lines and one DFT1 cell line. Error bars indicate standard deviation of mean. X axis indicates cell lines with and without IFNγ treatment, Y axis represents mean fluorescence intensity values for β_2-m expression. Stars indicate statistically significant β_2-m upregulation following treatment (**** = p-value < 0.0001, ** = 0.001 < p-value < 0.01). Data was analysed using two-way ANOVA and a post-hoc Sidak's multiple comparison test.

B) Table indicating difference in β_2-m expression between cell lines with and without IFNγ treatment. Green boxes indicate comparisons between untreated cell lines. Red boxes indicate comparisons between cell lines following IFNγ treatment. Stars indicate statistical significance of identified differences in β_2-m expression between cell lines both pre and post treatment (**** = p-value < 0.0001. *** = 0.0001 < p-value < 0.001. *** = 0.001 < p-value < 0.01. * = 0.01 < p-value < 0.05. NS = p value > 0.05, not significant). Data was analysed using two-way ANOVA, which indicated a significant (****) effect of cell line, IFNγ treatment and an interaction of both (Appendix E.1, Figure S 7) followed by a post-hoc Tukey's multiple comparison test.

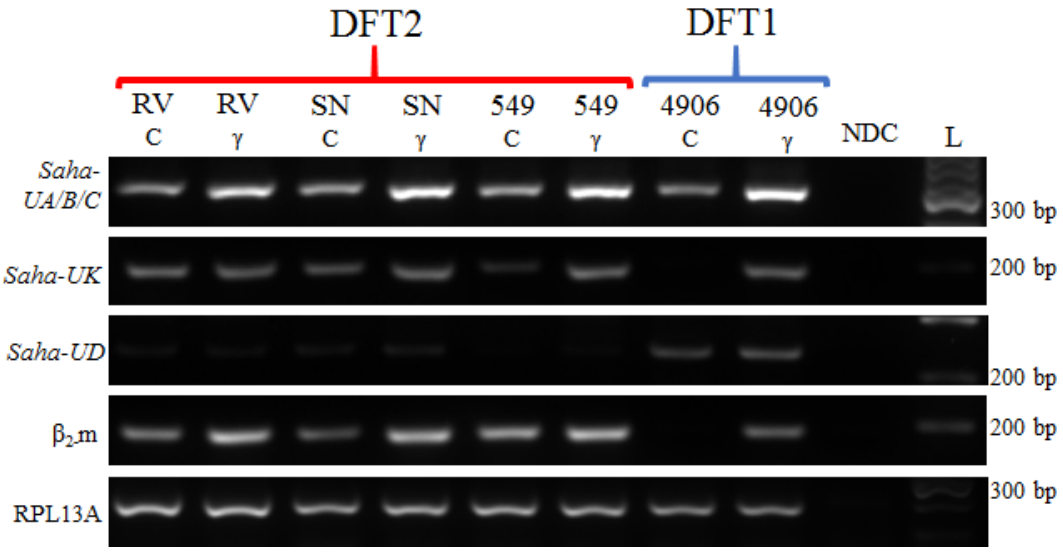


Figure 5.3: MHC class I associated genes are upregulated in DFT1 and DFT2 cell lines following treatment with IFNγ.

Treatments were performed in triplicate (Appendix E.2 Figure S 8a). Red bracket indicates DFT2 cell lines, blue bracket indicates DFT1 cell line. C denotes untreated control sample, γ denotes IFNγ treated sample. Each PCR reaction uses 250 ng cDNA which is controlled for using housekeeping gene RPL13A. NDC denotes no cDNA negative control. L denotes a DNA ladder of known sizes. Marker sizes are indicated. Saha-UA/B/C primers amplify all three classical MHC class I genes.

5.3.2 DFT2 modulates the immune response in a distinct manner to DFT1

A small panel of immune genes were screened by RT-PCR to identify any broad immune responses occurring in DFT2 before and after IFNγ treatment which may be aiding in the spread of DFT2 (Figure 5.4). All DFT2 cell lines upregulate transcripts of PD-L1 to a similar level to DFT1. Additionally, DFT2 cell lines upregulate STAT3 in response to IFNγ, with DFT2_SN showing the strongest upregulation, whilst DFT1 does not appear to upregulate STAT3 transcripts. All three DFT2 cell lines show constitutive transcription of immunosuppressive cytokines TGFB1 and TGFB2 which are unaffected by treatment with IFNγ. In contrast, DFT1 expresses negligible TGFB1 transcripts, and expresses low levels of TGFB2 which are upregulated following IFNγ treatment. Additionally, whilst both DFT2 and DFT1 cells upregulate transcription of the interleukin receptor IL1R1 in response to IFNγ, DFT2 expresses low to negligible levels of this receptor pre-treatment. DFT2_RV expresses the highest levels of IL1R1 post-treatment of the DFT2 cell lines and this is comparable with pre-treatment DFT1 levels.

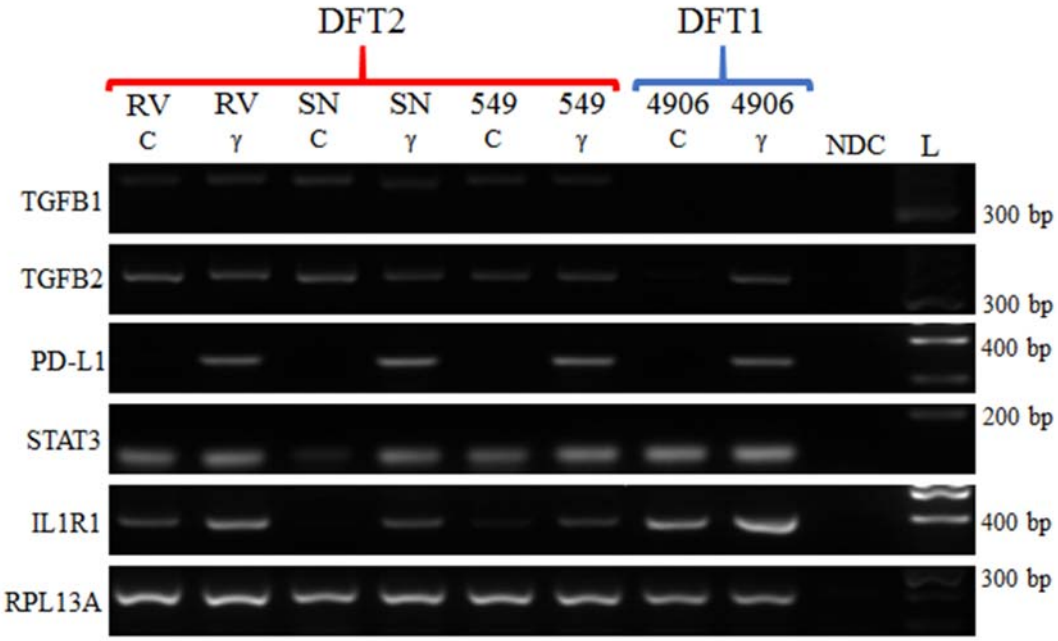


Figure 5.4: DFT2 cell lines have a distinct immune profile to DFT1 cells pre and post IFNγ treatment. Treatments were performed in triplicate (Appendix E.2, Figure S 8b). Red bracket indicates DFT2 cell lines, blue bracket indicates DFT1 cell line. C denotes untreated control sample, γ denotes IFNγ treated sample. Each PCR reaction uses 250 ng cDNA which is controlled for using housekeeping gene RPL13A. NDC denotes no cDNA negative control. L denotes a DNA ladder of known sizes. Marker sizes are indicated.

5.3.3 IFNγ treatment has variable effects on glial associated genes in DFT1 and DFT2

A selection of markers associated with the myelinating glial progenitor cells of DFT1 and DFT2 were assessed by RT-PCR pre and post IFNγ treatment (Figure 5.5). Immature glial markers previously only identified in DFT2 cell lines FABP7 and PDGFRA are unaffected by IFNγ treatment. L1CAM expression is marginally increased in all DFT2 and DFT1 cell lines following IFNγ treatment. Myelin component gene MBP expression is reducing following IFNγ treatment in both DFT1 and DFT2. Peripheral myelin protein MPZ is marginally decreased in DFT2 cell lines following IFNγ treatment but is unaffected in DFT1. MAG expression is increased in DFT2 cell lines following IFNγ treatment but decreases in DFT1 following treatment.

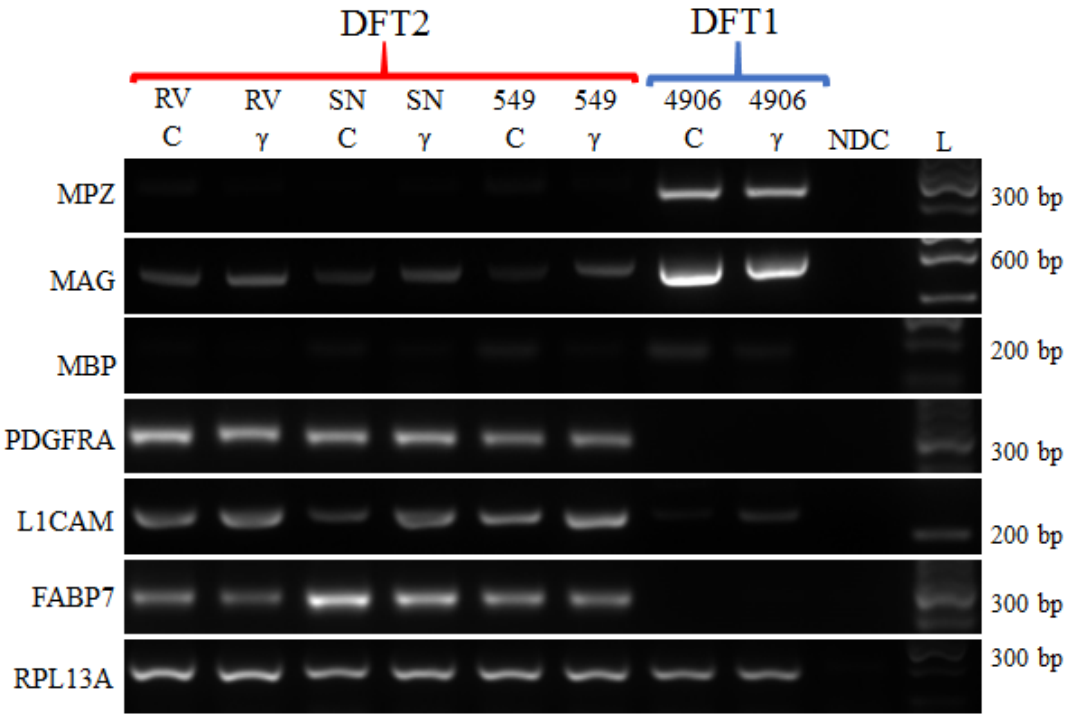


Figure 5.5: Glial associated marker genes are differentially affected by IFNγ in DFT1 and DFT2. Treatments were performed in triplicate (Appendix E.2, Figure S 8c). Red bracket indicates DFT2 cell lines, blue bracket indicates DFT1 cell line. C denotes untreated control sample, γ denotes IFNγ treated sample. Each PCR reaction uses 250 ng cDNA which is controlled for using housekeeping gene RPL13A. NDC denotes no cDNA negative control. L denotes a DNA ladder of known sizes. Marker sizes are indicated.

5.4 Discussion

Early vaccination studies have exploited the upregulation of MHC class I in response to IFNγ by DFT1 (Siddle *et al.* 2013) in order to initiate an immune response in host devils (Reviewed in Section 1.4.7, Summarised in Table 1.3 and Table 1.4). Tasmanian devils vaccinated with killed whole cell preparations of IFNγ treated, MHC class I positive (MHC-I+) DFT1 cells can raise antibodies against DFT1 which may result in prophylactic protection, although this needs further study (Pye *et al.* 2018; Tovar *et al.* 2017; Kreiss *et al.* 2015). Additionally, immunotherapy with MHC-I+ DFT1 cells can induce rejection of MHC class I negative (MHC-I-) DFT1 cells *in vivo* (Tovar *et al.* 2017), further demonstrating that inducing MHC class I expression on DFT1 cells with IFNγ treatment increases the ability of the devil immune system to detect and destroy DFT1 tumour cells. Remarkably, some wild devils have been found with detectable antibody responses against MHC-I+ DFT1 cells, which have correlated with tumour regression and disease free survival (Pye *et al.* 2016a). Given the association between surface MHC class I and improved immune responses to DFT1 cells, understanding how this system is modulated in DFT2 is crucial for developing vaccination strategies against this new tumour. Additionally, the mechanism of DFT2s response to IFNγ may help to further define its origins.

These results demonstrate that IFNγ treatment of DFT2 cell lines results in a significant and detectable upregulation of cell surface β_2-m expression (Figure 5.1, Figure 5.2). β_2-m is a component of the MHC class I complex (Grey *et al.* 1973) that is highly unlikely to be detected on the cell surface outside of a functional MHC class I molecule (Karlsson *et al.* 1980). Thus, an increase in surface β_2-m indicates an increase in functional MHC class I molecules (Garrido *et al.* 2012). β_2-m upregulation is not directly proportional to initial β_2-m expression, as demonstrated by DFT2_SN upregulating β_2-m to significantly higher levels than DFT2_RV despite expressing significantly lower initial levels. All three DFT2 cell lines upregulate β_2-m expression significantly more than DFT1 in response to IFNγ treatment, likely due to a recently identified hemizygous deletion of β_2-m in the genome of DFT1 (Stammnitz *et al.* 2018). DFT2 upregulates transcription of both classical and non-classical MHC class I heavy chain genes in response to IFNγ treatment, and appears to upregulate transcription of the non-classical MHC class I allele *Saha-UK* less strongly than DFT1. It has been previously demonstrated that *Saha-UK* is a dominant class I allele in DFT2 (Caldwell *et al.* 2018), and this may explain the reduced transcriptional upregulation following IFNγ treatment. Two DFT2 cell lines additionally show marginal upregulation of another non-classical MHC class I gene *Saha-UD*. Non-classical MHC class I genes have been widely implicated in tumour immune evasion and immune privilege (Haynes-Gilmore *et al.* 2014; Chiang and Stroynowski 2004; Ishitani *et al.* 2003; Slukvin *et al.* 2000) and the overexpression and upregulation of these non-classical alleles in response to inflammation may be indicative of such a role in the context of DFT2.

The ability of DFT2 to spread between MHC class I discordant hosts in the presence of MHC class I indicates that this tumour is using distinct methods of immune evasion to its older counterpart. This

project has assessed the transcriptional response of a small panel of immune associated genes to IFNγ treatment in DFT1 and DFT2 to identify alternate mechanisms of immune evasion. PD-L1 is a checkpoint molecule which is often strongly upregulated in response to IFNγ (Mimura *et al.* 2018) and acts as an inhibitory molecule, binding to the PD-1 receptor on activated T-cells and blocking their cytotoxic effects (Butte *et al.* 2007). PD-L1 has been implicated in the mechanisms behind several cancers (Gato-Cañas *et al.* 2017; Maekawa *et al.* 2014), and is an effective target of checkpoint blockade immunotherapy approaches (Shen and Zhao 2018). Flies *et al.* (2016) have previously demonstrated that PD-L1 is not constitutively expressed on DFT1 cell lines *in vitro* but is strongly upregulated in response to IFNγ. Here I show that all three DFT2 cell lines demonstrate a similar strong upregulation of PD-L1 in response to IFNγ, which is a mechanism which must be considered for vaccination strategies as PD-L1 effects may dampen the increased immunogenicity demonstrated post IFNγ treatment. STAT3 is a transcriptional activator implicated in tumorigenesis and immune signalling pathways (Nguyen *et al.* 2013; Takeda *et al.* 1998). It has been previously demonstrated that the ERBB/STAT3 axis is a major driver of DFT1, and STAT3 is constitutively activated in DFT1 (Kosack *et al.* 2019). The data presented here supports previous data from Kosack *et al.* (2019) showing transcription of STAT3 is not upregulated in DFT1 in response to IFNγ, however in contrast at least one DFT2 cell line, DFT2_SN, upregulates expression of STAT3 in response to IFNγ treatment. IL1R1 is a receptor for the cytokine IL-1 which is involved in the pro-inflammatory modulation of immune responses to tissue damage and autoimmunity (Lachmann *et al.* 2009; Reddy *et al.* 2009). All three DFT2 cell lines and DFT1 upregulate IL1R1 transcripts in response to IFNγ, but levels are consistently lower in all three DFT2 cell lines than DFT1. This indicates that DFT2 may have attenuated and dysregulated signalling through the IL1R1 receptor compared to DFT1, which may reduce the ability of host immune systems to raise proinflammatory responses to the tumour. Finally, the TGF-β signalling pathway is a complex signalling pathway associated with many cellular processes, although it has been strongly linked with immunosuppression (Oh and Li 2013). TGFB1 has been implicated in modulation of cytotoxic T-cell differentiation and growth, suppression of antigen specific T-cell responses, immune tolerance and tumour escape from immune detection (Donkor *et al.* 2012; Li *et al.* 2007; Gilbert *et al.* 1997), whilst TGFB2 has been associated with the alteration of antigen presentation by professional APCs and tumour invasion (Ouhtit *et al.* 2013; Takeuchi *et al.* 1998). Blockade of both cytokines has been implicated in improved immune clearance of tumour cells (Terabe *et al.* 2017) indicating that expression of these cytokines plays a role in tumour immune evasion. This work demonstrates that all three DFT2 cell lines constitutively express the cytokines TGFB1 and TGFB2 and this expression is unaffected by IFNγ treatment, in contrast to DFT1 which expresses negligible levels of TGFB1 pre and post treatment, and only expresses significant levels of TGFB2 post IFNγ treatment. The overexpression of these cytokines in DFT2 and their lack of responsiveness to IFNγ indicates that DFT2 may cultivate an immunosuppressive environment to prevent the detection of MHC class I antigens on the cell surface by host cytotoxic T-cells.

I have also assessed transcriptional regulation of some glial cell markers in response to IFNγ treatment. In Schwann cells, inflammatory signals should promote demyelination and dedifferentiation, resulting in the reduction of the expression of terminal myelin markers and an increase in the expression of more immature markers (Stone *et al.* 2018; Balabanov *et al.* 2007; Lisak *et al.* 2001; Lisak *et al.* 1998; Schneider-Schaulies *et al.* 1991). In contrast, inflammation induced demyelination causes activation and differentiation of oligodendrocyte precursor cells (OPCs) into mature, myelin producing oligodendrocytes (Glezer *et al.* 2006; Fancy *et al.* 2004). Thus, we were interested to see whether DFT1 and DFT2 have different responses to IFNγ with respect to myelin associated genes. There is a consistent reduction in myelin component gene MBP following IFNγ treatment in both DFT1 and DFT2 cell lines, indicating a Schwann cell dedifferentiation signature. MPZ is marginally reduced in the DFT2 cell lines post treatment and unaffected in DFT1, although MPZ is also a marker for immature Schwann cells (Liu *et al.* 2015). MAG is downregulated in DFT1 following treatment, indicative of a normal Schwann cell inflammatory response (Schneider-Schaulies *et al.* 1991), but is consistently upregulated in all three DFT2 cell lines following IFNγ treatment, a response more indicative of an oligodendroglial origin. FABP7 expression is not affected in DFT2, and not induced in DFT1 following IFNγ treatment, indicative of a remyelinating oligodendrocyte phenotype (Kipp *et al.* 2011). PDGFRA expression is also not affected by IFNγ treatment, which contradicts the literature that IFNγ reduces PDGFRA expression in OPCs (Tanner *et al.* 2011). However, mutations in PDGFRA have been demonstrated in human cancers (Velghe *et al.* 2013), and additionally analysis of the DFT2 genome has demonstrated that the tumour has a copy number variant of the PDGFRA gene resulting in two extra PDGFRA alleles in the tumour which may be affecting gene transcription (Stammnitz *et al.* 2018). L1CAM expression is increased following IFNγ treatment in both DFT1 and DFT2, an expected response as L1CAM is both an early glial cell marker (Schaefer and Frotscher 2012) and involved in the initiation of OPC remyelination after CNS injury (Barbin *et al.* 2004). Overall, the IFNγ response of glial associated genes in DFT2 may be more indicative of an oligodendroglial response than a Schwann cell response, although further work is needed to confirm this.

5.5 Conclusions and future directions

This work has demonstrated that it is possible to increase the expression of surface MHC class I on DFT2 cells in the same way as DFT1 through treatment with IFNγ, and the upregulation of MHC class I in DFT2 is significantly stronger than in DFT1. Both classical and non-classical MHC class I is upregulated following treatment, and this upregulation is detectable on the cell surface by β_2-m staining, indicating a functional overexpression.

Despite the similar MHC class I responses between the two tumours, there are some significant differences in how these tumours respond to IFNγ, demonstrating that these are two distinct tumours with distinct cellular and immune processes. In particular, DFT2 shows evidence of cultivating an immunosuppressive microenvironment and of attenuated inflammatory signalling, adaptations which may be aiding the tumour in its spread despite the presence of MHC class I on its cell surface.

The immune signalling pathways which are seemingly dysregulated in DFT2, such as TGFB and IL-1 signalling, are commonly altered in human cancers to avoid immune detection. The use of these immune evasion mechanisms by DFT2 indicates that despite their rarity, the fundamental biology and immunology underlying the emergence and spread of transmissible cancers is consistent with our current knowledge of cancers within a single organism, and transmissible tumours may act as an interesting disease model for studying tumour immune evasion mechanisms.

Additionally, this analysis has demonstrated that DFT2 may have a distinct cellular origin from DFT2, namely the oligodendrocyte rather than the Schwann cell. The modulation of glial associated markers following treatment of DFT2 with IFNγ appears more indicative of oligodendrocyte precursor cell differentiation than of Schwann cell dedifferentiation, whilst DFT1 shows a signature of an injured and dedifferentiating Schwann cell. Overall this work has demonstrated both similarities and significant differences in the immune responses underlying DFT1 and DFT2, knowledge which will be essential moving forward with vaccination programmes.

Chapter 6 Identification of a diagnostic biomarker for DFT2

6.1 Introduction

Whilst the myelinating Schwann cell specific protein PRX has been identified as highly specific marker for DFT1 tumour cells (Tovar *et al.* 2011) there is currently no effective marker for identifying DFT2 tumour cells. This makes identification of tumour cells in DFT2 sections difficult and is compounded by the variable tumour architecture demonstrated by DFT2 tumours (Caldwell *et al.* 2018; Pye *et al.* 2016b). A PCR based diagnostic test for DFT1 and DFT2 tumours has been developed by Kwon *et al.* (2018) (Figure 6.1) but the ability to confidently identify tumour cells in a tissue section following IHC is important for understanding protein expression in tumour cells compared to stromal cells.

As described in Chapters 3 and 4, DFT2 has emerged from a myelin competent cell type. Myelin and glial specific proteins are unlikely to be expressed in the tumour stroma, and one of these proteins may represent a useful biomarker for identifying DFT2 cells in tumour sections allowing for confident identification of tumour cells in tumour biopsies, making analysis of protein expression changes throughout the tumour easier.

In this chapter, I will assess the suitability of previously identified glial protein markers of interest as diagnostic biomarkers for DFT2 tumours *in vivo*.

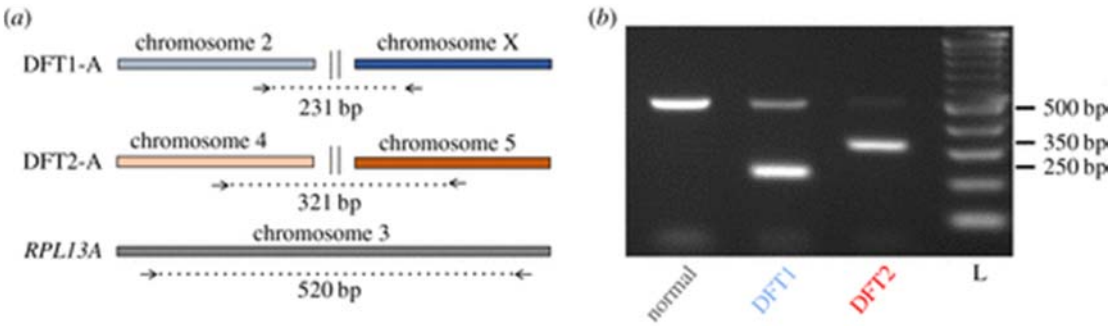


Figure 6.1: Tasman-PCR is a robust diagnostic PCR for DFT1 and DFT2.

A) Schematic indicating the chromosomal breakpoints amplified by Tasman-PCR. The triplex PCR assay has primers against DFT1-A, DFT2-A and housekeeping gene RPL13A. Presence of a band at 231 or 321 bp is considered diagnostic.

B) Gel electrophoresis of Tasman-PCR on DFT1, DFT2 and healthy tissue. The DFT1-A amplicon is only detected in the DFT1 sample, and the DFT2-A amplicon is only detected in the DFT2 sample (Kwon *et al.* 2018).

6.2 Aims and Objectives

Using proteomic and gene expression data generated and discussed in Chapters 3, 4 and 5, I aim to identify genes and proteins which are unique or overexpressed in DFT2 and may present useful biomarkers for the tumour. This work has a specific focus on markers which are likely derived from DFT2s progenitor cell, as these are more likely to be cell type specific and less likely to be found in the tumour stroma. I will then validate the specificity and consistency of these potential biomarkers across various cell line and tissue samples at a transcript and protein level.

This chapter has the following objectives:

1. Assess the specificity of the glial cell derived markers (Table 6.1) previously identified as unique to DFT2 by RT-PCR on a panel of devil tissues and cell lines.

2. Assess the specificity and consistency of validated markers in DFT2 tumour samples on a transcript and protein level.

Table 6.1: Genes assessed as potential DFT2 biomarkers in this chapter.

Proteome columns indicate whether a marker was identified in the quantified DFT1 and DFT2 proteomes.

Marker	Normal function	Relevance to DFTs	DFT2 proteome	DFT1 proteome
MAG	Myelin structural protein	Tumour origins	No	Yes
MBP	Myelin structural protein	Tumour origins	No	No
MPZ	Myelin structural protein	Tumour origins	No	No
PMP22	Myelin structural protein	Tumour origins	No	No
NGFR	Neuronal growth receptor	Tumour origins	Yes	No
L1CAM	Neural-specific cell adhesion molecule	Tumour origins	Yes	No
FABP7	Neural-specific fatty acid binding protein	Tumour origins	Yes	No
NF1	Schwann cell specific tumour suppressor	Tumour origins/mechanisms	No	No
PDGFRA	Oligodendrocyte precursor marker	Tumour origins/mechanisms	Yes	No
TGFB1	Immunosuppressive cytokine	Tumour immune evasion	No	No
TGFB2	Immunosuppressive cytokine	Tumour immune evasion	No	No
THY1	Neural enriched tumour suppressor	Tumour mechanisms	No	No

6.3 Results

6.3.1 Immature glial cell marker FABP7 is specific to DFT2 and PDGFRA is unique to tumour cells.

Potential markers for DFT2 identified as unique in the proteome (Chapter 3), unique at a transcript level (Chapters 4 and 5) or Schwann cell associated markers (Chapter 4 plus additional) were screened against DFT1 and DFT2 cell lines and tumour biopsies along with a panel of other devil tissue biopsies to identify markers which are specific to tumours and not widely expressed in devil tissues (Table 6.1). Early glial markers PDGFRA and FABP7, and the immunosuppressive cytokine TGFB2, were identified as being specific to DFT2 cell lines, though transcripts were not detected in either DFT2 tumour screened (Figure 6.2). DFT2 tumour samples screened in this experiment were negative for all markers and demonstrated poor expression of RPL13A relative to other samples indicating poor sample quality. The number of PCR cycles was increased from 28 to 30 for all further analysis of the DFT2 tumours by RT-PCR to improve detection sensitivity. At least one DFT2 cell line was positive for all markers assessed except MAG. Expression of MBP is largely universal but weak in non-glial samples, indicating it would not be an appropriate marker. NGFR is expressed by DFT2 cell lines and very strongly expressed by one DFT1 tumour, though is not expressed in the DFT1 cell lines.

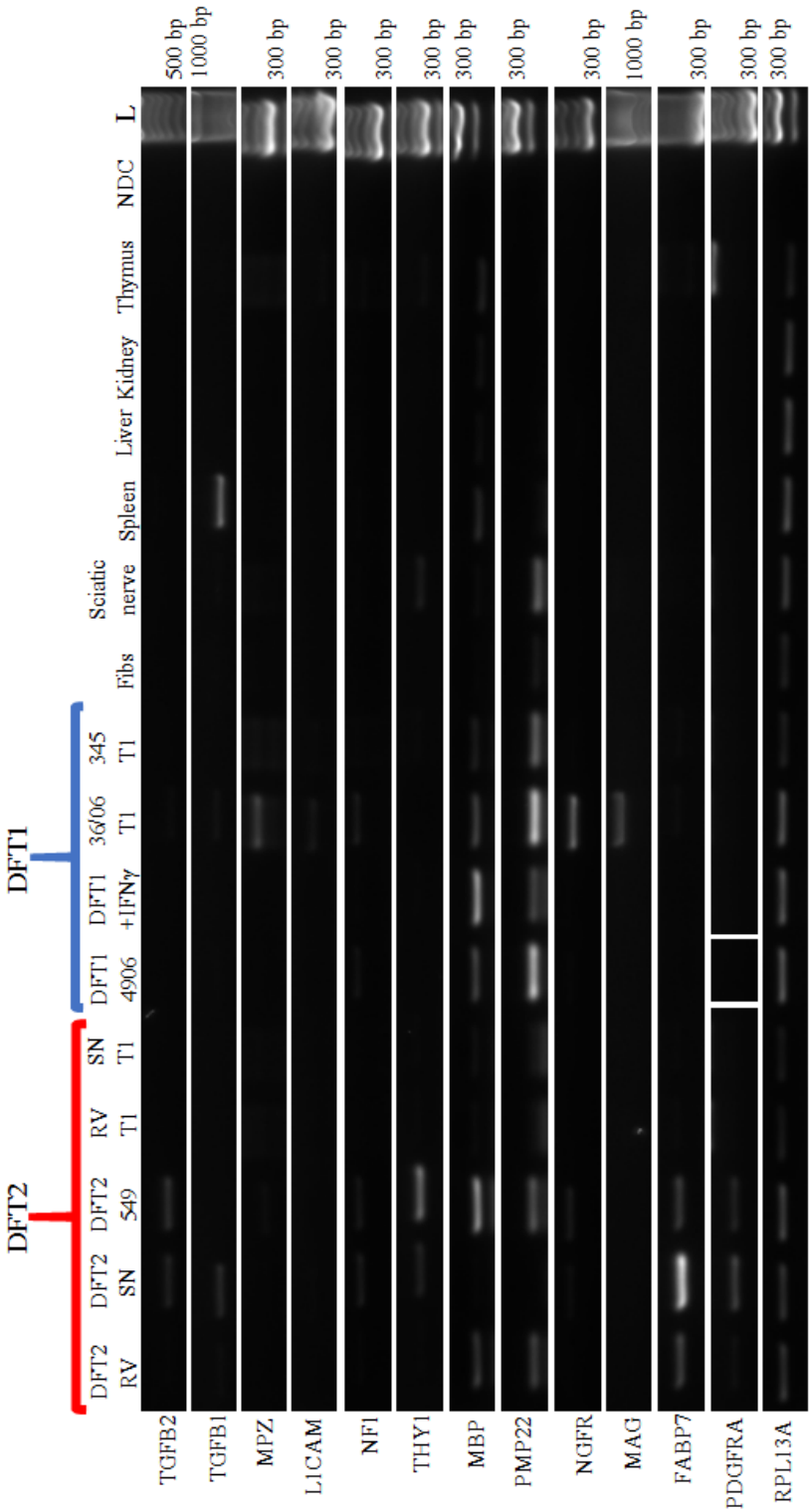


Figure 6.2: FABP7 and PDGFRA are specific to DFT2 cell lines but are not present in tumour samples.
Red bracket indicates DFT2 cell lines and tumour samples. Blue bracket indicates DFT1 cell lines and tumour samples. Fibs denotes the fibroblast cell line used in Chapter 3. 250 ng cDNA used per reaction. NDC denotes a no cDNA negative control. L donates a DNA ladder with known molecular sizes. Marker sizes are indicated. Note: Amplicon are shifted 100-200bp for all primer sets in this experiment due to documented migratory effects of GelRed DNA stain and disparity between loading dye used for ladder and samples. Amplicon identity was confirmed by sequencing (not shown) and identification of shifted genomic DNA. RPL13A is a housekeeping gene used to control for cDNA concentration and quality.

Western blot analysis of whole cell lysate from three DFT2 cell lines, DFT1_4906 and a fibroblast cell line was used to identify whether FABP7 and PDGFRA protein is specific and detectable in DFT2 cell lines (Figure 6.3). Blotting for PDGFRA detected no bands in fibroblast whole cell lysate but a ~140 kDa band is present in DFT2_SN and DFT1_4906, and a ~150 kDa band is present in DFT2_RV, DFT2_SN and DFT2_549. Devil PDGFRA is predicted to have a molecular weight of 125 kDa but the PDGFRA antibody used reliably detects human PDGFRA (122 kDa) at ~150 kDa (ab124392, see Table 2.2), and the band present at ~150 kDa in all three DFT2 cell lines is likely to represent PDGFRA protein. The slightly lower band at ~140 kDa present in DFT2_SN and DFT1_4906 may be either a non-specific reaction or may represent a distinct isoform of PDGFRA.

Blotting for FABP7 detects no bands in fibroblast or DFT1_4906 cell lysate and detects bands at ~15 kDa in DFT2_RV, DFT2_SN and DFT2_549 lysate. Devil FABP7 has a molecular weight of 15 kDa and these bands present in DFT2 cell lysate are likely to represent FABP7.

Western blot analysis of FABP7 and PDGFRA expression indicates that FABP7 is expressed in all three DFT2 cell lines and at least one isoform of PDGFRA is unique to DFT2 and expressed in all three cell lines. A possible second isoform of PDGFRA is detected in DFT1_4906 and DFT2_SN.

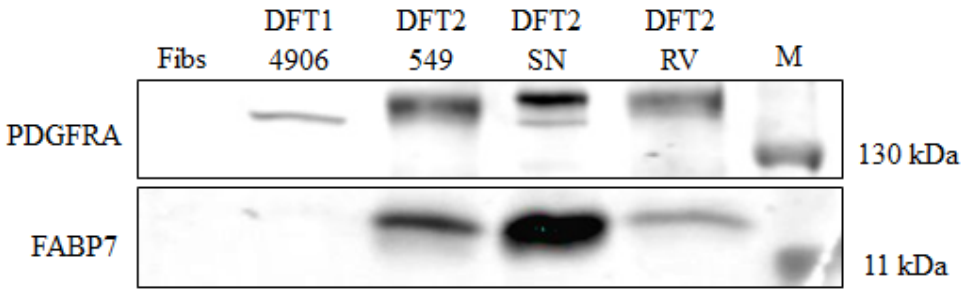


Figure 6.3: FABP7 protein is specific to DFT2 cell lines, whilst PDGFRA is restricted to both tumours. Western blot analysis on whole cell lysates of the fibroblast (Fibs), DFT1 (4906) and DFT2 (RV) cell lines used in the proteomic analysis detailed in Chapter 3, plus two further DFT2 cell lines (SN and 549), blotted for PDGFRA and FABP7. M denotes a protein marker of known size, with sizes indicated.

6.3.2 FABP7 and PDGFRA expression in primary DFT2 tumours is highly variable

Marker expression across tumour samples was assessed to ascertain the consistency of any potential markers, particularly PDGFRA and FABP7, at a higher number of PCR cycles to improve sensitivity of transcript detection in heterogenous tissue samples. RT-PCR was performed on a selection of tumour samples which have matched FFPE samples. All of the markers assessed showed high variability between tumour samples (Figure 6.4). The only marker present at high levels in all tumour samples was PMP22, but this is also expressed by most other cell lines and tissues previously assessed and is unlikely to be an appropriate marker. PDGFRA was also present in every tumour assessed, though levels are significantly lower in RV_T1 and SN_T1. FABP7 transcripts were not present in two tumours: SN_T1 and 818_T1. Tumour 812_T1 is positive for the most markers, whilst SN_T1 is positive for the fewest.

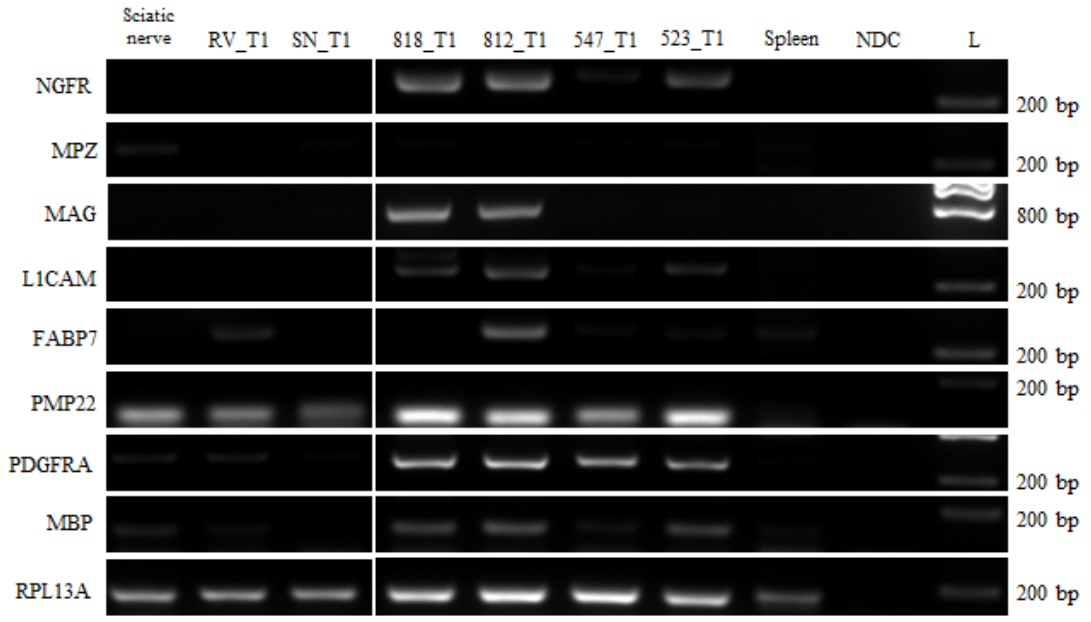


Figure 6.4: Marker expression is highly variable between DFT2 tumour samples.
250 ng cDNA used per reaction. NDC denotes a no cDNA negative control. L denotes a DNA ladder with known molecular sizes. Marker sizes are indicated. RPL13A is a housekeeping gene used to control for cDNA concentration and quality.

IHC staining on matched FFPE tumour samples for FABP7 (Figure 6.5) demonstrates extremely high variability between tumour samples, with some tumours (523_T2 and 812_T1) highly positive for FABP7, and others almost entirely negative (SN_T2, 523_T1). Transcript expression by RT-PCR on matched tumour samples was not correlated with protein expression, as demonstrated by SN_T1 which is negative for FABP7 transcript but stains strongly for FABP7 protein. Marker expression also varies dramatically within some tumours, such as RV_T1, where some strongly positive cells are surrounded by negative cells. Additionally, the stroma of some DFT2 tumours stains positively for FABP7 (SN_T1 and 523_T2) whilst in other tumours staining is specific to tumour cells (547_T1 and 812_T1). Staining is predominantly cytoplasmic, although RV_T1, SN_T1 and 818_T1 show strong nuclear staining, indicating normal expression of the protein. Two DFT1 tumours and a spleen sample are also positive for FABP7, and this staining appears cell specific rather than background and predominantly cytoplasmic.

Three DFT2 biopsies were stained by IHC for PDGFRA by an MSci student, Callum Daniel, under supervision (Figure 6.6). All three tumours were positive for PDGFRA, which is consistent with transcript data generated by RT-PCR for these tumours (RV_T1, 523_T1 and 547_T1). Staining for PDGFRA is more consistent across tumours than FABP7 and appears to specifically stain tumour cells rather than stroma. PDGFRA staining appears largely localised to the cell membrane, an expected localisation for a receptor protein. One DFT1 tumour and one spleen biopsy were stained for PDGFRA and both were positive. DFT1 staining was highly variable between tumour cells and large areas of tumour cells were negative. Spleen stroma was negative for PDGFRA but most cells in the sample were positive, although larger cells showed a more cytoplasmic localisation of staining than the smaller cells. PDGFRA positive DFT2 tumours were not necessarily FABP7 positive, with both RV_T1 and 523_T1 appearing largely negative or very weakly positive for FABP7 but positive for PDGFRA. 547_T1 is positive for both FABP7 and PDGFRA.

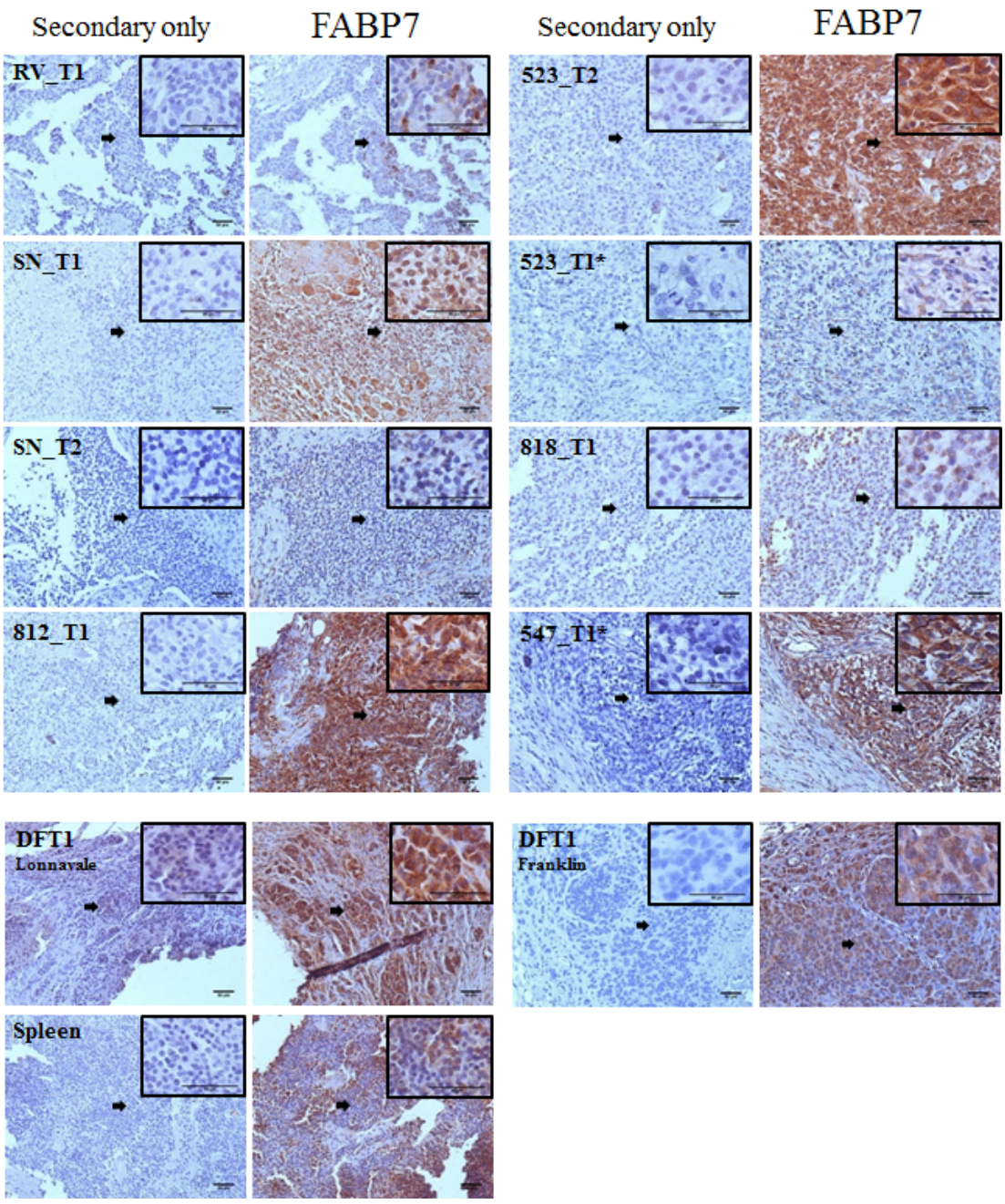


Figure 6.5: FABP7 expression in primary DFT2 biopsies is highly variable and not tumour specific.
IHC staining of 8x DFT2 tumours, 2x DFT1 tumours and 1x spleen biopsies for FABP7. DFT1 and spleen biopsies are labelled as such, all other biopsies are DFT2 tumours. Sections were all cut at 10 µm thickness. Main images taken at 200 x magnification. Boxed images taken at 600 x magnification. Arrows indicate the tumour cells imaged in boxed region. Positive cells are stained brown, nuclei are stained blue. * next to tumour name indicates staining was performed by an MSci student, Callum Daniel, under supervision. All staining imaged by Rachel Owen. Scale bar indicates 50 µm.

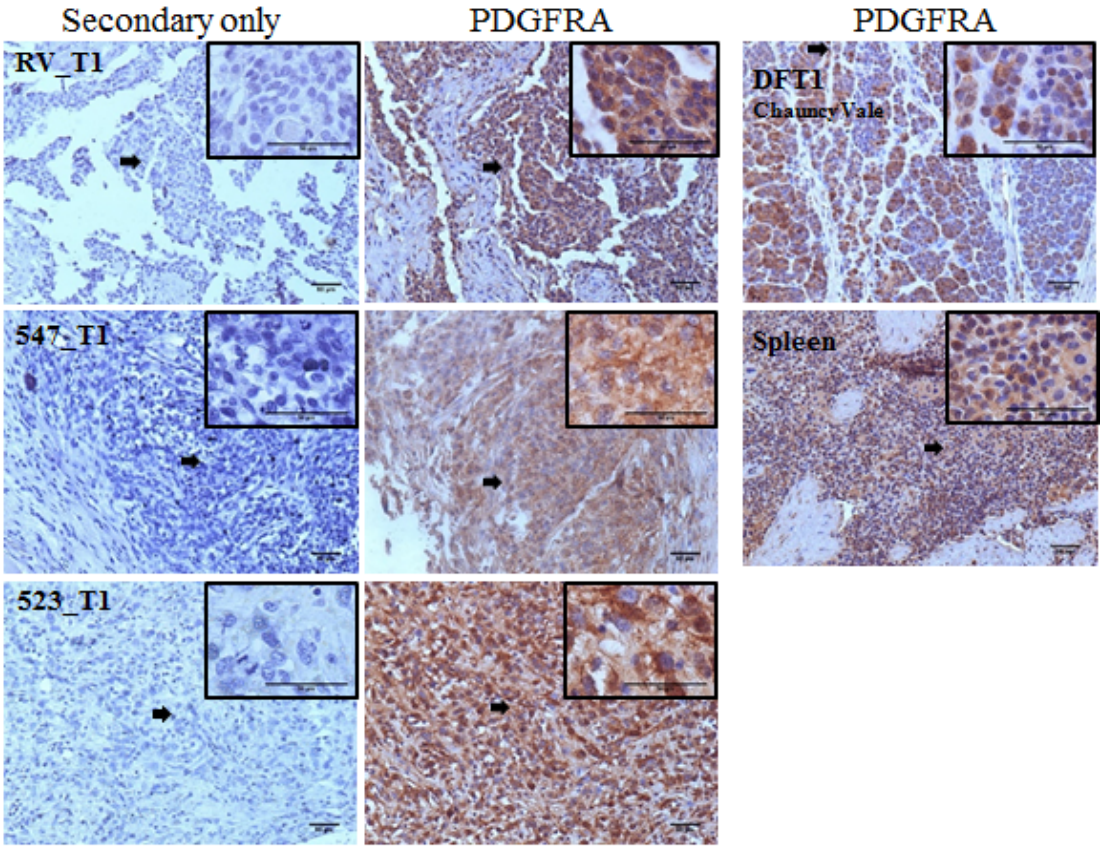


Figure 6.6: PDGFRA expression in DFT2 tumour biopsies is less variable than FABP7.

IHC staining of 3 x DFT2 tumours, one DFT1 tumour and one spleen biopsy for PDGFRA. DFT1 and spleen biopsies are labelled as such, all other biopsies are DFT2 tumours. Sections were all cut at 10 µm thickness. Main images taken at 200 x magnification. Boxed images taken at 600 x magnification. Arrows indicate the tumour cells imaged in boxed region. Positive cells are stained brown, nuclei are stained blue. All PDGFRA staining was performed by an MSci student, Callum Daniel, under supervision. All sections imaged by Rachel Owen. Scale bar indicates 50 µm.

6.4 Discussion

PRX staining of tumour biopsies has been the standard approach to diagnosing DFT1 from host derived neoplasms (Tovar *et al.* 2011; Murchison *et al.* 2010; Griner 1979). The specificity of PRX has also allowed for a detailed analysis of protein expression in tumour biopsies by reliably identifying tumour cells in serial sections which can then be stained to identify infiltrating immune cells within biopsies, or to define MHC class I heterogeneity within tumour samples (Siddle *et al.* 2013). At present there is no diagnostic biomarker for DFT2, and DFT2 tumours are negative for PRX (Pye *et al.* 2016b). Tasman-PCR exploits two chromosomal breakpoints to robustly diagnose tumours using a triplex PCR system which is useful for gross diagnosis, though carries a risk of generating false negative data and it is possible that breakpoints may emerge in newly formed tumours (Kwon *et al.* 2018). Additionally, this PCR based test cannot provide information on the identity of cells within a tumour biopsy. DFT2 tumours are highly variable and difficult to identify visually (Pye *et al.* 2016b) making it difficult to define the tumour cells in a biopsy, and a diagnostic marker is needed.

DFT2 uniquely expresses several marker proteins of myelinating glia, as well as myelin component proteins which are expressed by DFT1 but may be useful as markers for DFT2 cells once a PRX negative status is confirmed. A panel of markers were screened by RT-PCR against a panel of cell lines and tissue samples to confirm the specificity of these markers to DFT2 (Figure 6.2). Most markers demonstrated low and inconsistent expression in DFT2 cell lines, or significant expression in unrelated tissue types. Alongside previously defined glial markers FABP7 and PDGFRA, the immunosuppressive cytokine TGFB2 was also identified as specific to DFT2 cells. Whilst interesting, this cytokine was not carried forward as a potential marker as immune cytokine expression is likely to be highly variable between and within primary tumour sections (Ma *et al.* 2013) and is likely to be expressed by host immune and stromal cells in the close vicinity of the tumour (Yu *et al.* 2014). Only two markers, MBP and PMP22, showed expression in the DFT2 tumour samples, although this was accompanied by poor expression of the housekeeping gene RPL13A, indicating that the cDNA was of poor quality.

PDGFRA and FABP7, two proteins identified in the DFT2_RV proteome, were screened against whole cell lysate from three DFT2 cell lines, a DFT1 cell line and a fibroblast cell line by Western blot to confirm their specificity to DFT2 and consistency across DFT2 cell lines (Figure 6.3). Both PDGFRA and FABP7 proteins were detected in all three DFT2 cell lines and FABP7 was specific to DFT2. Blotting for PDGFRA demonstrated two bands around the expected size of PDGFRA, one which is specific to DFT2 cell lines and one which is present in DFT1 and one DFT2 cell line. While only one isoform of PDGFRA is predicted in the current devil genome assembly, it is possible that a second isoform exists. This may be an alternately spliced isoform which has not yet been identified or annotated in the devil genome, and there is evidence of several alternately spliced PDGFRA isoforms in humans according to the UniProt database (Consortium 2018; Larochelle 2016). It may

also represent a truncated or mutated form of the protein which has been differentially mutated in DFT1 and DFT2 (Wasag *et al.* 2004). The size discrepancy may also represent differential post-translational modifications on the PDGFRA protein between DFT1 and DFT2 (Parker *et al.* 2010). For example, PDGFRA is known to be heavily phosphorylated upon activation of the receptor (Matsumoto *et al.* 2000), and the higher molecular weight band in DFT2 may represent an activated form of the protein, whilst the lower band in DFT1 and DFT2_SN may represent a less heavily phosphorylated receptor. Additionally, the PDGFRA antibody used here is a polyclonal antibody against human PDGFRA (ab124392) and the antibody may be binding non-specific or additional targets, although it is unlikely that a non-specific target at the same molecular weight would be identified, and a BLAST search of the antibody immunogen against the Tasmanian devil genome only identifies PDGFRA.

A selection of glial markers including FABP7 and PDGFRA were screened by RT-PCR against a panel of DFT2 tumour samples with matched FFPE sections to assess the consistency of marker expression between tumour biopsies. PCR was extended by two cycles to increase the sensitivity of transcript detection and control for poor quality RNA which affected initial PCR screens. Marker expression across primary DFT2 tumours is strikingly variable (Figure 6.4), supporting data from Caldwell *et al.* (2018) indicating that MHC class I expression is variable, and may indicate tumour evolution during transmission events. Interestingly, the two most consistent tumours, 812_T1 and 818_T1 are the most recently sampled tumours (Caldwell *et al.* 2018) and are potentially the 'oldest' tumours in our current samples. This may indicate that DFT2 tumours are stabilising and dominant, stable clones may soon increase in frequency. Additionally RV_T1 and SN_T1 which had been circulating in the population for ~two years less than 818_T1 and 812_T1 at the time of sampling, show markedly reduced transcript expression of all markers, and little to no expression of myelin associated genes MBP, MAG and PMP22, possibly indicating that DFT2 is turning the myelination pathways back on as it spreads.

Variability is also observed at the protein level in the matched FFPE tumour sections for expression of FABP7 (Figure 6.5). IHC shows that protein expression in primary tumours does not always correlate with transcript expression, and some tumours (SN_T1) are negative for FABP7 transcripts but positive for the protein. Staining is predominantly cytoplasmic and nuclear which is where FABP7 would be expected to be detected in normal tissue (Stolte *et al.* 2014) indicating functional expression. However, the extreme variability of FABP7 in primary tumour sections coupled with the variable levels of stromal staining makes FABP7 an inappropriate marker for DFT2 tumour cells. Additionally, FABP7 is detected strongly in the tumour cells of DFT1 primary tumour biopsies and in spleen sections, indicating either wider expression of the protein than initially demonstrated or antibody cross reactivity. Blocking the antibody with DFT1 lysate prior to IHC staining did not reduce staining in DFT or spleen sections (data not shown). This indicates that the protein this antibody is binding in DFT1 tumour sections is not present in the DFT1 cell lysate. Proteomic analysis of the DFT1 cell line DFT1_4906 did not identify the presence of FABP7 (Appendix C.1,

Figure 6.3), and FABP7 transcripts were not detected in any DFT1 cell lines (Figure 4.1, Figure 6.2). This may indicate that following years in continuous culture, the DFT1 cell lines have lost expression of some proteins which are still present in tumour biopsies, and the cell lines no longer fully reflect the tumour phenotype. This is a good example of the well documented phenotypic changes in cell lines kept in continuous culture (Mouriaux *et al.* 2016). As DFT1 cells have been in culture for several years longer than DFT2 cell lines, the possibility of significantly altered tumour phenotypes particularly in DFT1 must always be considered when drawing comparisons between the two tumours, and gene and protein expression analysis should always be validated in tumour biopsies following identification in the cell lines.

PDGFRA expression was only assessed in three tumours, but while staining appears more consistent across DFT2 tumours it does not correlate with FABP7 protein expression (Figure 6.6). PDGFRA staining appears predominantly membrane localised indicating functional PDGFRA expression (Stolte *et al.* 2014), and stromal staining is weaker than seen in some FABP7 sections, indicating predominantly tumour cell specific binding. However, PDGFRA is also variably expressed in both DFT1 and spleen sections. PDGFRA appears more consistent on a transcript and protein level between primary DFT2 tumours than FABP7, and appears to stain tumour stroma less strongly, but has a broad tissue and cell expression profile (Uhlén *et al.* 2005) and is implicated in several cancers (Evans *et al.* 2017; Joglekar-Javadekar *et al.* 2017). This may mean it is present in non-DFT neoplasms in the devil, which would make PDGFRA not a suitable diagnostic marker for identifying DFT2 in primary biopsies, although it may still be useful for identifying tumour cells in known DFT2 biopsies.

This data indicates that the DFT2 tumours currently in the population are highly variable and has failed to identify a suitable marker for use on tumour biopsies. However, it is important to note that the antibodies used in this project for detection of PDGFRA and FABP7 protein were both raised against human protein. Whilst the immunogens for both antibodies are predicted to have high specificity to these proteins across multiple species, these have only been validated in human and mouse samples. A blast search of the immunogens against the devil genome in Ensembl indicates that both antibodies should have high specificity in the Tasmanian devil but more work must be done to validate the specificity of these antibodies in Tasmanian devil tissue.

6.5 Conclusions and future directions

In this chapter I have demonstrated that FABP7 and PDGFRA are specific to DFT2 cell lines on a transcript level. Additionally, FABP7 is specific to DFT2 cell lines on a protein level. PDGFRA may have a DFT2 specific isoform, although it appears at least one isoform of this protein is expressed by DFT1 cell lines. Both markers demonstrate considerable variability in DFT2 tumour biopsies on a transcript and protein level. FABP7 shows significant variability which is not correlated with transcript expression, and often shows strong staining in the tumour stroma. PDGFRA shows less variability than FABP7 on a transcript and protein level and weaker stromal staining, although variability is still present and it is possible that this protein is widely expressed in the body and by non-DFT tumours. FABP7 and PDGFRA protein is additionally detected in primary biopsies of DFT1 and spleen samples, which may indicate wider expression of these markers or may be indicative of non-specific antibody staining. Variability, stromal staining and staining of DFT1 and spleen sections make FABP7 and PDGFRA inappropriate DFT2 diagnostic markers, though PDGFRA may still be useful as a marker of DFT2 cells in pre-diagnosed tumour biopsies, and further work on defining the specificity of these antibodies may improve IHC outcomes. Protein expression of more DFT2 specific markers such as NGFR and L1CAM must also be assessed in primary DFT2 biopsies.

In addition to the assessment of PDGFRA and FABP7 as specific DFT2 markers, I have demonstrated that primary DFT2 tumours exhibit variability in the expression of all markers analysed, which supports previously generated data indicating that DFT2 tumours are highly variable. Interestingly, older DFT2 tumours which have been circulating through the population for longer are more stable in their expression of most markers and appear to upregulate myelin associated proteins which may indicate that the tumour is starting to form stable subclones, which may be redifferentiating into myelinating glial like cells. This analysis must next be expanded to a wider selection of DFT2 tumours and markers over time to understand how the expression of myelin associated markers is changing with the tumour.

Chapter 7 Generation of recombinant DFT2 cell lines expressing tagged MHC class I molecules of biological interest

7.1 Introduction

Universally low levels of cell surface MHC class I expression on DFT cells impairs the ability of the host immune system to identify the tumour as foreign during grafting (Caldwell *et al.* 2018; Siddle *et al.* 2013). In addition to this, the removal of immunogenic neoantigens and MHC class I from the surface of a tumour cell is a widely described method of immune escape (Iorgulescu *et al.* 2018; Anagnostou *et al.* 2017; Angell *et al.* 2014), and the selection pressure from genetically distinct hosts during DFT transmission is likely to result in the downregulation of both immunogenic MHC class I alleles and MHC class I alleles which bind immunogenic neoantigens from the tumour surface (Teng *et al.* 2015; Kirveskari *et al.* 1999). Evidence of MHC class I downregulation following selective immune pressure has been seen in DFT2 (Caldwell *et al.* 2018), and the remaining MHC class I on the DFT2 cell surface may present peptides which do not raise a strong immune response through the TCR, allowing DFT2 to spread despite the cell surface expression of MHC class I. Similarly, if MHC class I alleles in the Tasmanian devil bind similar peptide sequences, the ability of their immune system to recognise and raise an immune response against foreign peptides or neoantigens during DFT transmission may be impaired. Understanding the alleles expressed on the surface of DFTs and the sequence of the peptides these alleles bind may provide insight into how MHC class I positive tumours can evade immune destruction and be tolerated by the host immune system.

Conversely, has been demonstrated that MHC class I on the surface of DFT1 cells greatly increases the immunogenicity of the tumour, improving the ability of host devils to detect and raise an immune response against tumours following engraftment. This indicates that it is possible to exploit the MHC class I system in DFT1 to induce a strong and specific immune response against tumour cells. The evidence that DFT2 is losing MHC class I expression, and that MHC class I expression can be modulated in this new tumour by treatment with IFNγ implies that targeting MHC class I in DFT2 may also be an effective method of inducing anti-tumour immunity. Exploiting the MHC class I system to activate T-cell responses against DFT cells through a specific peptide vaccine targeted against tumour specific neoantigens may therefore be an effective disease management strategy moving forward.

For a peptide vaccine to be effective host antigen presenting cells must bind and present the peptide on their own MHC molecules to T-cells in the lymph node to raise an immune response, restricting peptide vaccines to certain MHC haplotypes (Harper *et al.* 2015; Bartnik *et al.* 2012; Lakkis *et al.*

2000). As demonstrated in Caldwell *et al.* (2018), DFT1 and DFT2 share several MHC class I alleles (Table 1.2). Of particular interest are two classical MHC class I alleles, *SahaI*27* and *SahaI*27-1*, two *Saha-UC* alleles which differ by one non-synonymous mutation (at position 228 in exon 3/α2 domain) which isn't predicted to interfere with peptide binding and are likely to bind the same peptide repertoire. These alleles are also found in all sequenced DFT2 host devils and are thought to be common in the general devil population (Lane *et al.* 2012). Another allele of interest is the non-classical *Saha-UK*, which is monomorphic and thus will bind the same peptides in the general population and appears to be highly expressed in the context of DFT2. Vaccine targets restricted to any of these three alleles are likely to be recognised in the broader devil population.

Understanding how devil MHC class I bind peptides is essential for improving devil immune responses to these tumours. Our lab has previously developed an antibody which recognises a native epitope on devil β_2-m molecules which can be used for immunoaffinity purification of MHC class I molecules and detection of surface MHC class I molecules by flow cytometry (Siddle *et al.* 2013, Gastaldello *et al*. in preparation). There is no similar antibody capable of recognising native epitopes on the MHC class I heavy chain and attempts to generate allele specific antibodies for devil MHC class I have thus far been unsuccessful, preventing the analysis of single-allele peptide binding properties. Similar problems in other species have been navigated by generating cell lines which overexpress tagged molecules of interest (Wynne *et al.* 2016).

Here, I will describe the generation of devil cell lines overexpressing myc-tagged, recombinant devil MHC class I, for interrogation of allele specific MHC class I peptide binding properties, an important step for studies of the antigen presentation pathway in DFT1, DFT2 and healthy devil cells.

7.2 Aims and Objectives

In this chapter I aim to generate two stable DFT2 cell lines which overexpress the MHC class I alleles *SahaI*27-1* and *Saha-UK*. These will be recombinant, myc-tagged constructs which can be purified using the myc tag for interrogation of bound peptides by mass spectrometry in the absence of allele specific antibodies.

The chapter has the following objectives:

1. Clone the full length MHC class I alleles, *SahaI*27-1* and *Saha-UK* from Tasmanian devil cDNA.

2. Generate myc-tagged MHC constructs in a mammalian expression vector.

3. Optimise transfection into a DFT2 cell line.

4. Assess the stability and functionality of the recombinant MHC class I constructs.

7.3 Results

7.3.1 The DFT2_RV cell line can be successfully transfected

This work is the first time that DFT2_RV cells have been transfected. Transfection optimisation was performed using a GFP construct over a range of Fugene:DNA ratios and final total DNA in each transfection. The fluorescence after transfection for 24 hours was measured on a flow cytometer (Figure 7.1). The percentage of GFP positive cells for each ratio is shown in Table 7.1 with the optimal ratio of 2:1 highlighted. A maximum of 19.3% GFP positive cells was achieved at this ratio.

RV cells were also transfected with a devil IFNγ-pcDNA3.0 construct which has previously been successfully transfected into CHO and DFT1_4906 cells (Siddle *et al.* 2013). Cells were harvested 48 hours post transfection and stained for surface β_2-m expression which was measured by flow cytometry. There is a significant increase in surface β_2-m expression following transfection with IFNγ, which is not demonstrated following transfection with a vector only control (Figure 7.2). This indicates a successful transfection and an increase in surface MHC class I caused by recombinant IFNγ production (Figure 5.1).

Table 7.1: Optimal Fugene:DNA ratio for transfection of DFT2_RV cells.

Red cells indicate the ratio used for all DFT2_RV transfections.

Fugene:DNA	Final µg DNA	% GFP positive cells
4:1	0.2	4.04
4:1	0.3	4.17
4:1	0.4	8.15
4:1	0.5	12.5
3:1	0.2	6.3
3:1	0.3	13.3
3:1	0.4	11.1
3:1	0.5	11.7
2:1	0.2	13
2:1	0.3	15.3
2:1	0.4	19.4
2:1	0.5	19.3
Control	0	0.32
Control	0	0

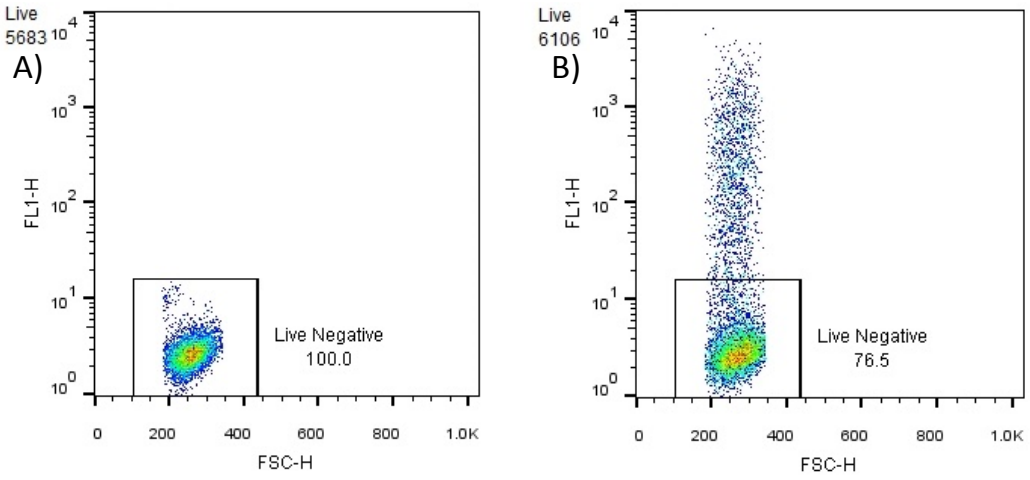
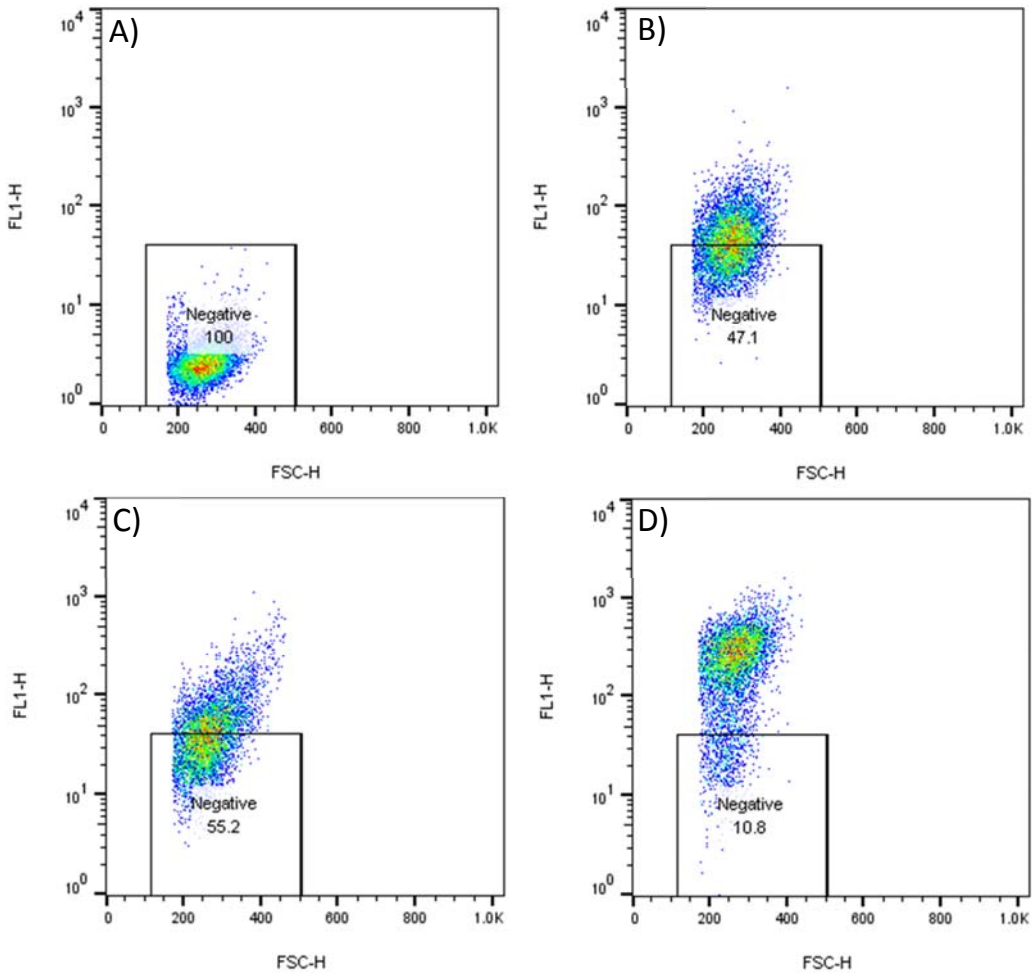


Figure 7.1: DFT2 cells fluoresce following transfection with GFP.

A) Untransfected DFT2_RV control cells B) GFP transfected DFT2 cells.

Cells were transfected using the optimal Fugene:DNA ratio indicated in Table 7.1. X-axis indicates with forward laser scatter to determine cell size. Y-axis indicates fluorescence intensity in the FL1-H channel which represents GFP fluorescence. Cells were gated to remove debris and doublets using the gating strategy detailed in Figure 2.2. Data was gathered on the BD FACSCalibur system and analysed using FlowJo.



E)

Sample	Mean fluorescence intensity (β2-m expression)
DFT2_RV secondary only control	2.8
DFT2_RV untransfected	53.3
DFT2_RV vector only control	52.4
DFT2_RV transfected with IFNγ	259

Figure 7.2: Surface β_2-m is increased following transfection of DFT2_RV with IFNγ.

Flow cytometry dot plots measuring β_2-m expression in A) Secondary antibody only control, B) Untransfected DFT2_RV, C) DFT2_RV transfected with empty pcDNA3.0, D) DFT2_RV transfected with 0.5µg IFNγ-pcDNA3.0 at 2:1 Fugene:DNA ratio. E) Table detailing the mean fluorescence intensity of the cells detected in the flow dot plots.

Y-axis represents fluorescence intensity in the FL1-H channel following staining for cell surface β_2-m. X-axis represents forward laser scatter (FSC-H) to determine cell size. Cells were gated to remove debris and doublets using the gating strategy detailed in Figure 2.2.

7.3.1 Full length *SahaI*27-1* and *Saha-UK* cloned into pcDNA3.0

Primers which amplify the full length of all devil classical MHC class I genes and add a terminal myc tag were used to amplify genes from IFNγ treated DFT1 cDNA (Primers 310/312 and universal reverse primer 317, detailed in Appendix B.2). The amplified PCR product was cloned into the blunt-ended vector pJET for sequencing. Genes were sequenced using the T7 primer within the vector and a primer in a conserved exon region of all devil class I genes (Primer 353). Almost all sequenced colonies were the UC allele *SahaI*27-1*. One colony was a *Saha-UB* allele (*SahaI*06*) which has been stored for later use (data not shown). *SahaI*27-1* was sub-cloned into the pIRES vector (SahaI*27-1-pIRES), using the EcoRI and NheI sites incorporated into the primers. The *Saha-UK* gene was previously inserted into this vector with a single myc tag (Saha-UK-pIRES) by a masters student in the lab, Rachel Coleby. An additional myc tag was inserted into both SahaI*27-1-pIRES and Saha-UK-pIRES using site-directed mutagenesis and identified by colony PCR and sequencing. Following transfection difficulties (See section 7.3.2), the *SahaI*27-1* and *Saha-UK* constructs were moved into the mammalian expression vector pcDNA3.0 by digestion with XhoI and XbaI. The translated protein sequence alignment of SahaI*27-1 in pcDNA3.0 is presented in Figure 7.3 and the genomic alignment is presented in Figure S 9 (Appendix F.1). The genomic and protein alignments for Saha-UK are presented in Figure S 10 (Appendix F.1).

Figure 7.3: Representative sequence alignment for SahaI*27-1 in pcDNA3.0.

Red box indicates 2x myc tags added during cloning process. Top sequence (labelled UC27-1) represents the protein sequence of SahaI*27-1 translated from the genomic sequence. Bottom sequence (labelled UC:27-1 in pcDNA3.0 2x myc) represents the myc tagged SahaI*27-1-pcDNA3.0 construct sequence translated to protein.

7.3.2 SahaI*27-1-pcDNA3.0 and Saha-UK-pcDNA3.0 transfect into CHO, DFT1_4906 and DFT2_RV

DFT1_4906 and DFT2_RV cells were transfected with Saha-UK-pIRES and grown under selection with G418 for one week. Cells were harvested and lysed, and whole cell lysates were assessed by Western blot for myc and Saha-UK heavy chain (UK_15-29-1, see Table 2.2) (Figure 7.4). Previous work by our laboratory has established that the Saha-UK heavy chain is detected at ~47 kDa when blotted for using the UK_15-29-1 antibody, and the myc tags are predicted to add ~3 kDa to the molecular weight, thus recombinant Saha-UK should be detected at ~50 kDa.

Blotting for myc detected no at band ~50 kDa in the transfected or untransfected cells from either cell line, indicating no myc-tagged construct was present. Blotting for Saha-UK heavy chain indicated bands at ~47 kDa in the transfected and untransfected cells from both cell lines, the Saha-UK heavy chain, but there was no increase in levels following transfection. Additional bands are detected at ~35 kDa and ~80 kDa which consistently demonstrated in myc blots and are likely to be endogenous myc protein (Appendix F.1, Figure S 11).

DFT2_RV cells were transfected with SahaI*27-1-pIRES and Saha-UK-pIRES and grown under selection with G418 for one week. Cells were harvested and stained by ICC for myc, Saha-UA/B/C (classical MHC class I) heavy chain and Saha-UK heavy chain protein to detect the presence of the recombinant constructs. DFT2_RV cells did not demonstrate an increase in staining for myc or Saha-UA/B/C heavy chain following transfection with SahaI*27-1-pIRES, or an increase in staining for Saha-UK heavy chain following transfection with Saha-UK-pIRES (Figure 7.5).

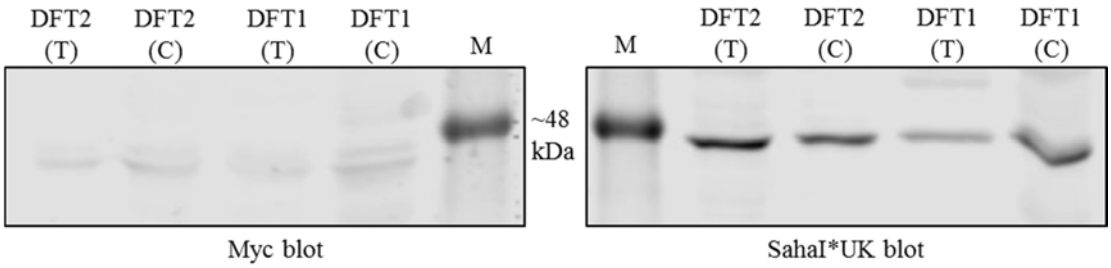


Figure 7.4: Recombinant Saha-UK construct is not detected in DFT2_RV or DFT1_4906 whole cell lysate following transfection with Saha-UK-pIRES.

Saha-UK-pIRES Transfected (T) and Untransfected (C) DFT1_4906 and DFT2_RV whole cell lysates blotted for myc and Saha-UK heavy chain. Endogenous Saha-UK is detected at ~47 kDa, recombinant Saha-UK should be detected ~50 kDa. Full blots are presented in Appendix F.2 (Figure S 11).

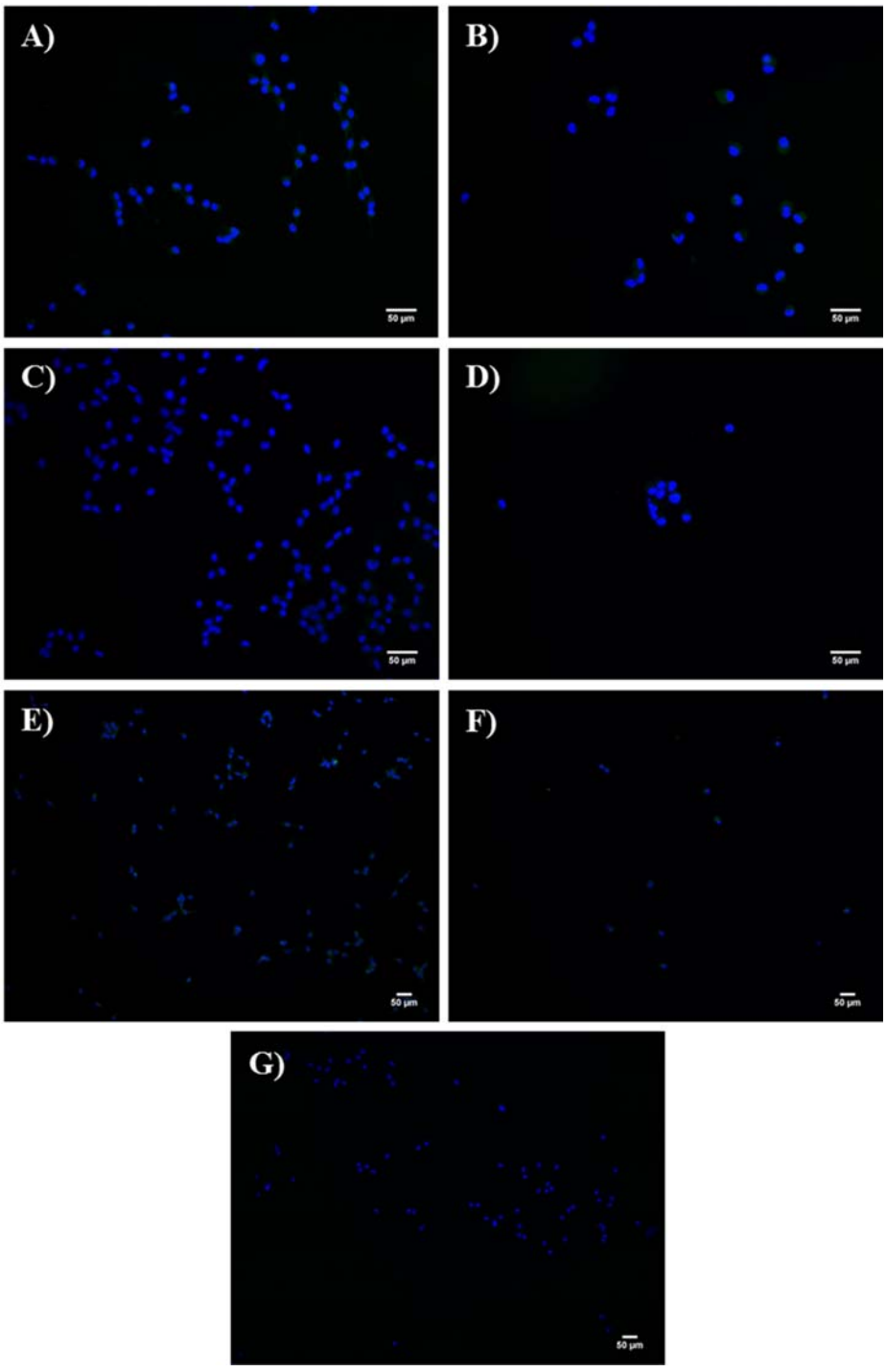


Figure 7.5: SahaI*27-1-pIRES and Saha-UK-pIRES constructs cannot be detected in DFT2_RV cells by ICC following transfection.

A) Untransfected and **B)** SahaI*27-1-pIRES transfected DFT2_RV cells stained for Saha-UA/B/C heavy chain. **C)** Untransfected and **D)** SahaI*27-1-pIRES transfected DFT2_RV cells stained for myc protein. **E)** Untransfected and **F)** Saha-UK-pIRES transfected DFT2_RV cells stained for Saha-UK heavy chain. **G)** Secondary antibody only control DFT2_RV cells.

A-D imaged at 200 X magnification. E-G images at 100 X magnification. Scale bars represent 50 µm. Images pseudo coloured in ImageJ. Blue colouring indicates DAPI stained nuclei, green colouring indicates positive staining for the protein of interest.

The MHC class I genes were moved into pcDNA3.0, a vector which has shown previous success when transfecting CHO, DFT1_4906 and DFT2_RV (Figure 7.2) cells in the lab.

CHO cells transfected with SahaI*27-1-pcDNA3.0 and selected for one week were analysed by Western blot for myc and Saha-UA/B/C heavy chain to identify the recombinant SahaI*27-1 construct. Our lab has previously demonstrated that the classical MHC class I heavy chains in devil cell lysates are typically detected at ~43 kDa and ~45 kDa using our custom pan-specific classical heavy chain antibody Saha-UA/B/C_15-25-18 (Table 2.2). The addition of myc tags in the recombinant SahaI*27-1 heavy chain is predicted to add ~3 kDa to the molecular weight, thus detection is expected to be at ~48 kDa. A band was detected at ~48 kDa in CHO cells transfected with SahaI*27-1-pcDNA3.0 when blotted for myc and for Saha-UA/B/C heavy chain, indicating successful expression of recombinant SahaI*27-1 heavy chain by the CHO cell line (Figure 7.6a).

Western blot analysis on DFT2_RV and DFT2_4906 cells transfected with Saha-UK-pcDNA3.0 detected a band at ~48 kDa when blotted for myc which was not present in untransfected control cells, which is slightly lower than the expected molecular weight but appears highly specific, indicating successful expression of recombinant Saha-UK (Figure 7.6b).

SahaI*27-1-pcDNA3.0 transfected DFT2_RV cells were harvested and stained for β_2-m for flow cytometry analysis. There was a small increase in surface β_2-m in transfected cells compared to untransfected cells, indicating an increase in surface MHC class I expression following transfection (Figure 7.7).

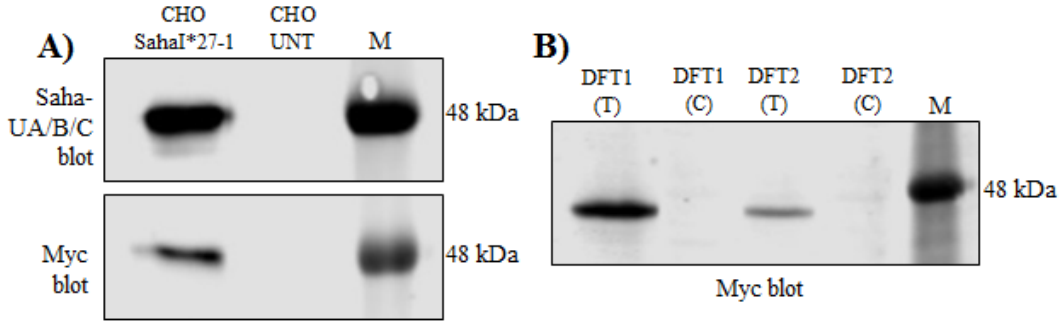
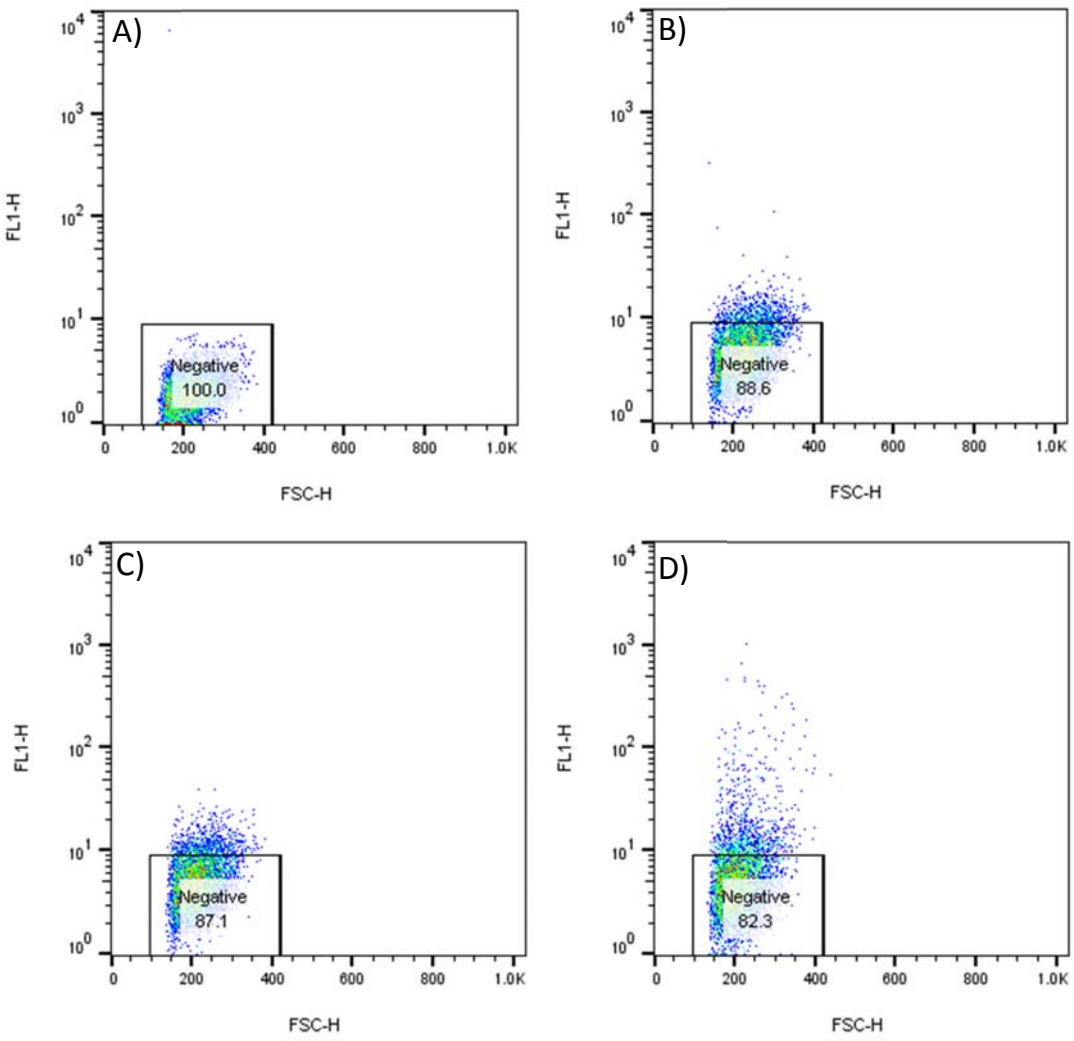


Figure 7.6: SahaI*27-1-pcDNA3.0 and Saha-UK-pcDNA3.0 successfully transfect into CHO, DFT1 and DFT2.

A) CHO transfected with SahaI*27-1-pcDNA3.0 and untransfected (UNT) blotted for Saha-UA/B/C heavy chain and myc. B) DFT1_4906 and DFT2_RV untransfected control (C) and transfected (T) with Saha-UK-pcDNA3.0. The ~48kDa constructs are detected in all cell lines following transfection. Full blots presented in Appendix F.2 (Figure S 12).



E)

Sample	Mean fluorescence intensity (β2-m expression)
DFT2_RV secondary only control	3.5
DFT2_RV untransfected	5.48
DFT2_RV vector only control	5.74
DFT2_RV transfected with SahaI*27-1-pcDNA3.0	8.63

Figure 7.7: Surface β_2-m expression is increased in DFT2_RV following transfection with SahaI*27-1-pcDNA3.0.

Flow cytometry dot plots measuring cell surface β_2-m expression in A) Secondary antibody only control DFT2_RV cells, B) Untransfected DFT2_RV cells, C) DFT2_RV transfected with empty pcDNA3.0, D) DFT2_RV transfected with 0.5µg SahaI*27-1-pcDNA3.0 at 2:1 Fugene:DNA ratio, E) Table detailing the mean fluorescence intensity of the cells detected in the flow dot plots.

Y-axis represents fluorescence intensity (FL1-H channel) following staining for cell surface β_2-m. X-axis represents forward laser scatter (FSC-H) to determine cell size. Cells were gated to remove debris and doublets using the gating strategy detailed in Figure 2.2.

7.3.3 Transfectant cell lines stably overexpress cell surface β_2-m

Saha-UK-pcDNA3.0 and SahaI*27-1-pcDNA3.0 were transfected into DFT2_RV cells. After selection with G418 for five days cells were harvested, stained for cell surface β_2-m expression (β_2-m 13-34-48, see Table 2.2) and expression levels were assessed by flow cytometry. Saha-UK-pcDNA3.0 transfected cells show increased surface β_2-m one week post-transfection (Figure 7.8a). SahaI*27-1-pcDNA3.0 transfected cells show marginally increased surface β_2-m expression one week post-transfection (Figure 7.8b).

Transfectants were selected for in G418 for a further 4 days before harvesting, live staining for β_2-m expression, and sorting using FACS. Cells were gated to sort cells with an MFI higher than the untransfected control cells. $6x10^5$ cells were sorted per construct, and 3000-4000 cells per construct were captured and sorted. Sorted cells were grown and expanded for 3 weeks before β_2-m expression was assessed by flow cytometry. DFT2_RV transfected with SahaI-UK-pcDNA3.0 (DFT2_UK) maintained overexpression of surface β_2-m expression relative to control cells three weeks post sorting (Figure 7.9a). DFT2_RV transfected with SahaI*27-1-pcDNA3.0 (DFT2_UC) showed increased expression of cell surface β_2-m relative to control cells three weeks post sorting (Figure 7.9b).

After 2 months in culture, the DFT2_UK cell line lost surface β_2-m expression which was not restored by reviving DFT2_UK from frozen storage (Figure 7.10). Additionally, myc tagged construct could not be detected in whole cell lysate (Figure 7.11). Due to this, the DFT2_UK cell line was no longer interrogated within the confines of this project and the remainder of this work focuses solely on the DFT2_UC cell line. β_2-m expression was frequently assessed by flow cytometry using the same gating strategy and parameters before every cell harvest and experiment to ensure expression was maintained.

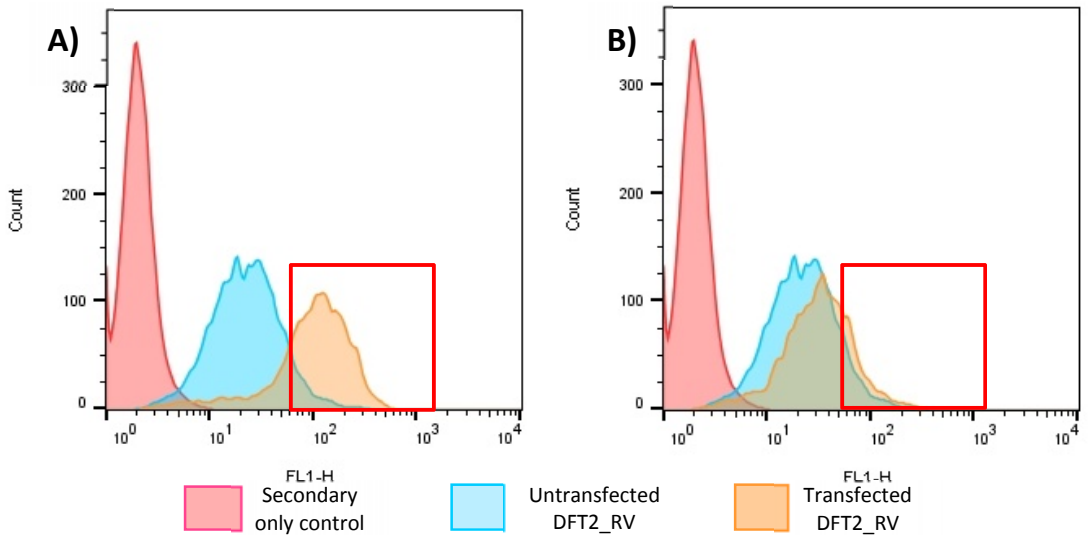


Figure 7.8: Cell surface expression of β_2-m is increased in DFT2_RV following transfection with A) Saha-UK-pcDNA3.0 and B) SahaI*27-1-pcDNA3.0.

Histograms indicating an increase in cell surface β_2-m expression in DFT2_RV cells following A) transfection with Saha-UK-pcDNA3.0 and B) SahaI*27-1-pcDNA3.0.

Red peak indicates Secondary only negative control. Cyan peak indicates untransfected DFT2_RV control cells. Orange peak indicates transfected DFT2_RV cells. Y axis indicates cell count; X axis (FL1-H) is on a log scale and indicates fluorescence intensity of the AlexaFluor488 secondary antibody following staining for cell surface β_2-m. Red boxes indicate cells captured during FACS.

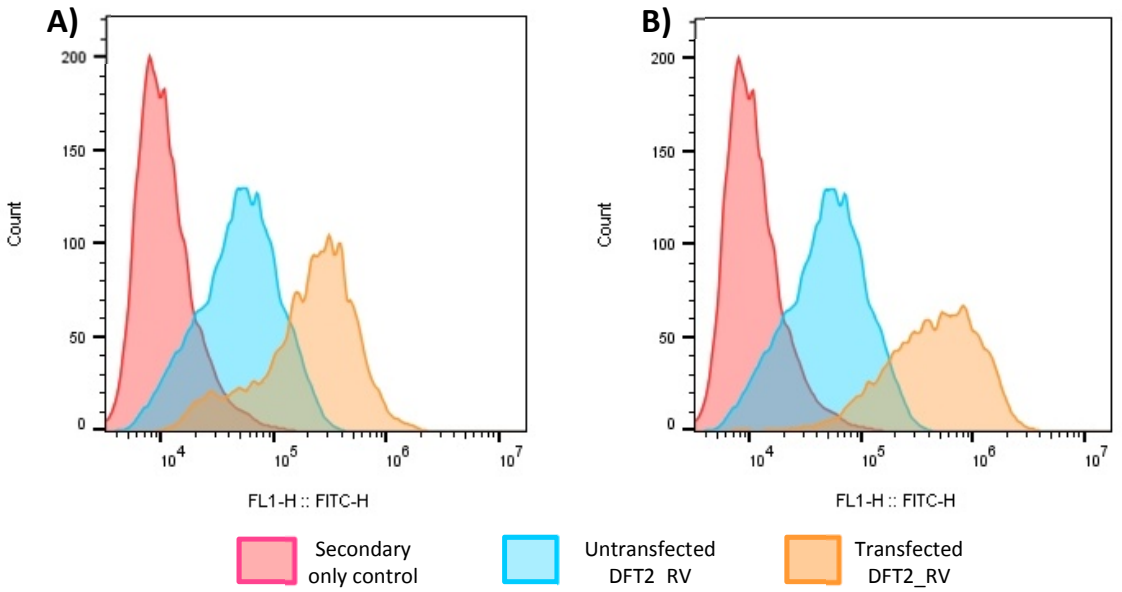
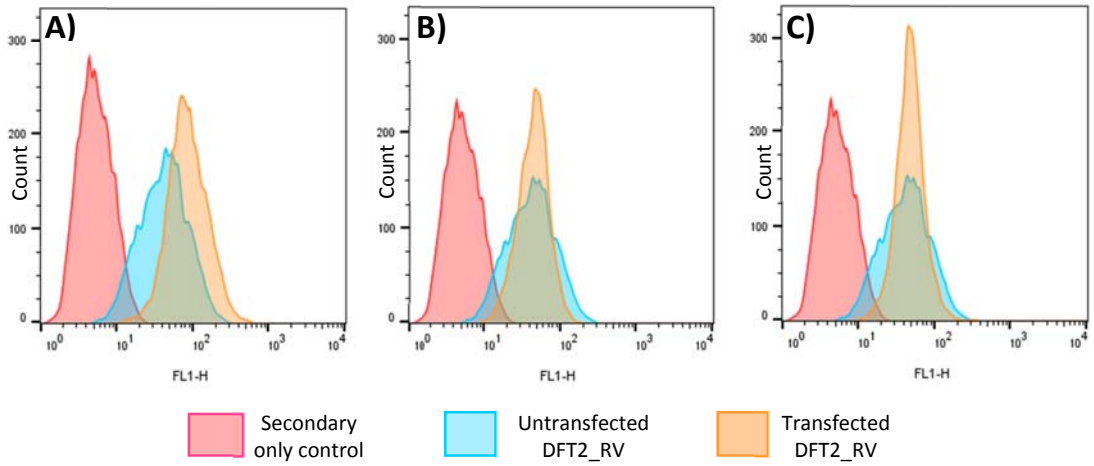


Figure 7.9: Elevated cell surface β_2-m expression is A) maintained in DFT2_UK and B) increased in DFT2_UC three weeks post sorting.

Histograms indicating increased cell surface β_2-m expression in A) DFT2_UK cells three weeks post sorting, B) DFT2_UC cells three weeks post sorting.

Red peak indicates Secondary only negative control. Cyan peak indicates untransfected DFT2_RV control cells. Orange peak indicates transfected DFT2_RV cells. Y axis indicates cell count, X axis (FL1-H::FITC-H) is on a log scale and indicates fluorescence intensity of the AlexaFluor488 secondary antibody following staining for cell surface β_2-m.



D)

Sample	Mean fluorescence intensity (β2-m expression)
DFT2_RV secondary only control	5.94
DFT2_RV untransfected	52.3
DFT2_UC	101
DFT2_UK – Old	48.5
DFT2_UK - New	54.0

Figure 7.10: Elevated cell surface β_2-m expression is lost in the DFT2_UK cell line following continuous cell culture and freeze-thawing.

Histograms indicating cell surface β_2-m overexpression in A) DFT2_UC cells after 2 months in continuous culture post-sorting, and a loss of cell surface β_2-m overexpression in B) DFT2_UK cells after two months in continuous culture post-sorting (DFT2_UK – Old) and C) DFT2_UK cells after one week in culture post revival from storage at -80 °C (DFT2_UK – New). D) Table detailing the mean fluorescence intensity of the cells from each histogram.

Red peak indicates Secondary only negative control. Cyan peak indicates untransfected DFT2_RV control cells. Orange peak indicates transfected DFT2_RV cells. Y axis indicates cell count, X axis (FL1-H::FITC-H) is on a log scale and indicates fluorescence intensity of the AlexaFluor488 secondary antibody following staining for cell surface β_2-m.

7.3.4 Recombinant SahaI*27-1 associates with endogenous β_2-m

Western blotting of DFT2_UC whole cell lysate for myc and Saha-UA/B/C heavy chain detects bands at ~48 kDa in the cell line after growth in continuous culture for two weeks, indicating stable expression of recombinant SahaI*27-1 construct with intact myc tags (Figure 7.11, C). No bands are detected at ~48 kDa in whole cell lysate from DFT1_4906 (Figure 7.11, A), DFT2_RV whole cell lysate (Figure 7.11, B) or DFT2_UK whole cell lysate (Figure 7.11, D) when blotted for myc or Saha-UA/B/C. Endogenous Saha-UA/B/C heavy chain is detected at ~43 kDa and ~45 kDa in DFT2_RV, DFT2_UC and DFT2_UK when blotted for Saha-UA/B/C.

Immunoprecipitation of the recombinant SahaI*27-1 heavy chain was performed on DFT2_UC whole cell lysate using an anti-myc antibody (See Table 2.2), and assessed by Western blot. Blotting for Saha-UA/B/C following immunoprecipitation using an anti-myc antibody detects a band at ~48 kDa in DFT2_UC lysate pre-immunoprecipitation (Figure 7.12, DFT2 UC) and attached to the antibody-bead conjugate after immunoprecipitation (Figure 7.12, Myc IP), indicating recombinant SahaI*27-1 construct with myc tags in the cell lysate. No bands are detected at ~48 kDa in DFT1_4906 or DFT2_RV whole cell lysate when blotted for Saha-UA/B/C heavy chain (Figure 7.12, DFT1 4906, DFT2 RV). Endogenous Saha-UA/B/C heavy chain is detected at ~43 kDa and ~45 kDa in DFT1_4906, DFT2_RV and DFT2_UC whole cell lysate when blotted for Saha-UA/B/C, but is not detected in the antibody-bead conjugate following anti-myc immunoprecipitation (Figure 7.12).

Pan-specific immunoprecipitation of the MHC class I complex from DFT2_UC was also performed using an anti-β_2-m antibody (β_2-m_13-34-38, see Table 2.2) and analysed by Western blot. DFT1_4906 and DFT2_RV whole cell lysates are presented as negative controls and blotting for myc detects no bands at ~48 kDa (Figure 7.13) whereas blotting of DFT2_UC whole cell lysate for myc detects a band at ~48 kDa, likely the recombinant SahaI*27-1 construct. Blotting of the antibody-bead conjugate for myc detects a band at ~48 kDa, indicating that recombinant SahaI*27-1 is associating with and can be co-immunoprecipitated with endogenous β_2-m.

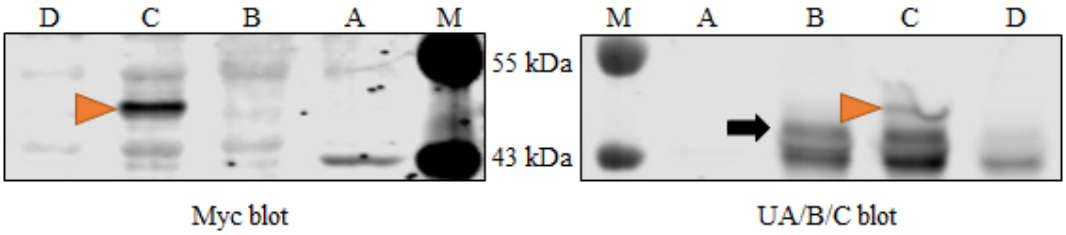


Figure 7.11: DFT2_UC expresses myc-tagged SahaI*27-1 protein.

Whole cell lysates blotted for myc and Saha-UA/B/C heavy chain.

A) DFT1_4906 whole cell lysate, B) DFT2_RV whole cell lysate, C) DFT2_RV transfected with SahaI*27-1-pcDNA3.0 (DFT2_UC) whole cell lysate, D) DFT2_RV transfected with Saha-UK-pcDNA3.0 (DFT2_UK). M indicates a protein marker of known molecular weight, indicated in kDa.

Orange arrowheads indicate the expected band for recombinant SahaI*27-1 construct at ~48 kDa. Black arrow indicates the expected band for endogenous classical MHC class I at ~45 kDa. Full blots are presented in Appendix F.2 (Figure S 13).

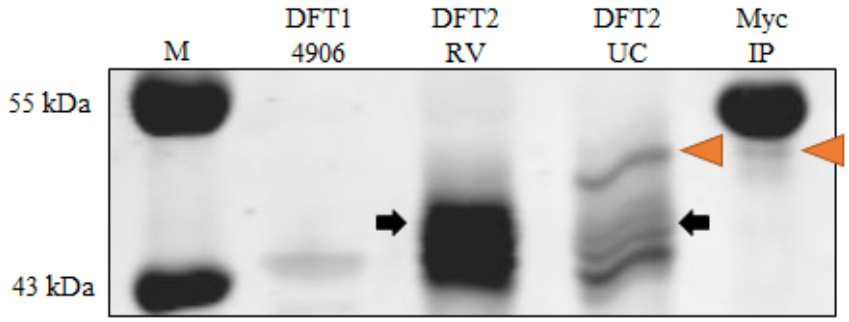


Figure 7.12: MHC class I heavy chain co-precipitates with myc.

Western blot analysis of an IP using an antibody against myc on DFT2_UC whole cell lysate blotted for Saha-UA/B/C heavy chain. Orange arrowheads indicate myc-tagged construct at ~48 kDa. Black arrows indicate endogenous Saha-UA/B/C at ~43-45 kDa. Full blots are presented in Appendix F.2 (Figure S 14).

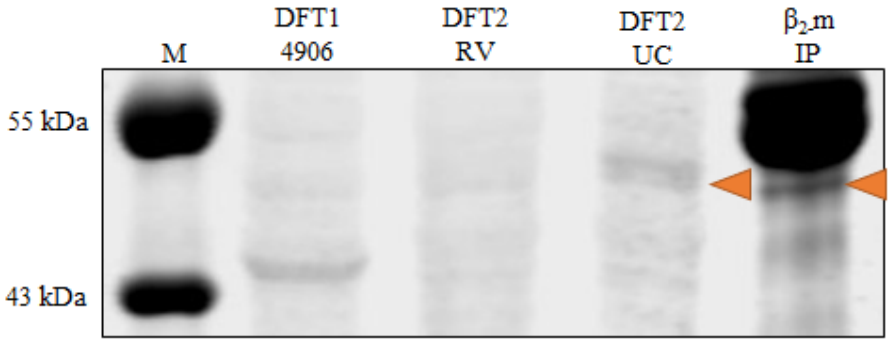


Figure 7.13: Recombinant SahaI*27-1 coprecipitates with β_2-m.

Western blot analysis of a β_2-m IP on DFT2_UC whole cell lysate blotted for myc. Orange arrowheads indicate myc-tagged construct at ~48 kDa. Full blots are presented in Appendix F.2 (Figure S 15).

7.4 Discussion

In this chapter, I have described the generation of a DFT2 cell line expressing recombinant SahaI*27-1, an MHC class I molecule thought to be common and biologically significant in the Tasmanian devil and its tumours (Caldwell *et al.* 2018; Lane *et al.* 2012).

Our lab has generated the first pan-MHC class I immunopeptidome for the Tasmanian devil (Gastaldello *et al*. in preparation), although this provides no information on specific MHC class I allele peptide binding properties. Peptide vaccines are MHC class I haplotype restricted (Kuzushima *et al.* 2001), and any large scale vaccination target must be compatible with common MHC class I alleles in the wider devil population. Therefore, in order to identify peptides which may be useful in a vaccination programme, we must define the peptide binding properties of specific MHC class I alleles which are present not only in the tumour but in the wider devil population.

GibbsCluster 2.0 is a programme which is used to cluster meaningful peptide sequences in a mixed dataset to identify binding motifs which may represent distinct MHC class I alleles (Andreatta *et al.* 2017; Andreatta *et al.* 2012). GibbsCluster 2.0 analysis on the pan-MHC class I datasets generated by Gastaldello *et al*. (in preparation) has been inconclusive, and it has not been possible to define allele specific MHC class I binding motifs in the Tasmanian devil using post-experimental analysis methods. Additionally, whilst previous collaborations between the laboratories of Dr Hannah Siddle and Prof. Karsten Skjødt have produced a number of devil specific MHC class I antibodies (which have been used in publication on several occasions, including Caldwell *et al.* (2018) and Siddle *et al.* (2013)), it has not been possible to generate antibodies which can detect native MHC class I heavy chain or distinguish between different classical MHC class I heavy chain genes and it is not possible to isolate a single MHC class I allele in our current system.

This work is the first, to our knowledge, successful stable transfection of any DFT2 cell line. I opted to generate these cell lines as DFT2 cell lines, as opposed to an MHC class I null human line such a 721.221 to ensure peptides were highly physiologically relevant to the species and the tumours. Additionally, DFT2 was chosen over DFT1 as it still expresses functional surface MHC class I (Caldwell *et al.* 2018) and should be capable of presenting recombinant constructs on the cell surface without any further treatment. This is in direct contrast to DFT1, which requires IFNγ treatment to present MHC class I on the cell surface, which directly alters the peptide repertoire and may skew any binding motifs generated (Chong *et al*. 2018). In this work I have generated a robust workflow for stably transfecting DFT2 cells with a reasonable efficiency and viability. I have also generated a workflow for identifying and subcloning highly expressing clones from a cell line which does not grow from single cell culture using FACS. These methods can be directly used with new constructs to interrogate different biological pathways in the DFT2_RV cell line. Additionally these cells can be used and adapted for further analysis of the antigen presentation pathway in the context of DFTs.

Previously, Wynne *et al.* (2016) were able to generate a single allele immunopeptidome and associated binding motifs in the Black flying fox (*Pteropus alecto*) by creating a recombinant MHC class I construct with a cytoplasmic myc tag, which was transfected into a bat cell line and isolated by anti-myc immunoaffinity. This study is proof of concept that it is possible to generate a single allele immunopeptidome in a wild species in lieu of appropriate antibodies or previously gathered peptide binding information and has been the basis for the workflow developed here. This study on bats also performed a deeper analysis of the functionality of their MHC class I allele by transfection into the human MHC class I null cell line 721.221 and IP/blotting experiments using human antibodies to identify associations between their recombinant bat MHC class I allele and the human antigen processing pathway. We have not functionally characterised our MHC class I construct to the same depth as seen in Wynne *et al.* (2016), however I have demonstrated that the recombinant SahaI*27-1 protein can be detected in whole cell lysates both immediately post transfection and following long term cell expansion, indicating stable expression of the construct (Figure 7.6, Figure 7.11). Additionally, this expression is accompanied by a detectable increase in cell surface expression of β_2-m, which is a reliable marker of construct expression and remains elevated over long term cell culture (Figure 7.8, Figure 7.9). I have also demonstrated that this construct can be immunoprecipitated using antibodies against both myc and β_2-m (Figure 7.12, Figure 7.13). The ability to coprecipitate the myc-tagged SahaI*27-1 construct with β_2-m combined with the increased cell surface β_2-m expression seen in the DFT2_UC cell line indicates that the recombinant protein is associating with endogenous β_2-m. As MHC class I constructs require all three components; heavy chain, β_2-m and peptide, to form a stable complex on the cell surface (Albring *et al.* 2004) these data indicate that the SahaI*27-1 construct is likely functional and associating with both β_2-m and peptide on the DFT2 cell surface, and it should be possible to isolate and sequence these peptides by mass spectrometry.

Despite strong initial overexpression of cell surface β_2-m, after two months in culture the DFT2_UK cell line dropped this expression and recombinant Saha-UK could no longer be detected in whole cell lysate (Figure 7.10, Figure 7.11). This is particularly unexpected as the DFT2_UK cell line expressed significantly more surface β_2-m at the point of cell sorting than the DFT2_UC cell line and was sorted with very minimal crossover with control cells (Figure 7.8). It is possible that a few low or transiently expressing transfectants outgrew the high expressing, stable clones in culture, or that the cells actively downregulated expression of the construct in response to some form of cellular stress. This cell line was recently re-established from cultures frozen very shortly after sorting by a Master of Research Student in our laboratory, Steven Turner. The newly re-established culture expressed high levels of β_2-m and was immediately re-sorted by FACS to isolate stable clones. β_2-m expression in this cell line has been maintained over a number of months indicating that initial β_2-m loss may have been due to an outgrown transiently expressing clone. Functional analysis of this cell line is currently being undertaken by Steven following the workflows developed here and in Chapter 8, with the goal of generating a Saha-UK immunopeptidome.

7.5 Conclusions and future directions

In this chapter I have generated two independent DFT2 cell lines which overexpress recombinant myc-tagged MHC class I heavy chain alleles of biological interest in the context of DFTs. I have shown that at least one of these cell lines, DFT2_UC, produces recombinant SahaI*27-1 with a cytoplasmic myc tag which can be immunoprecipitated using an antibody against the myc tag. Additionally, I have shown that this construct can also be immunoprecipitated using an antibody against β_2-m, indicating that the recombinant protein is associating with the antigen processing pathway and is likely functional. Stable transfection of these cells is accompanied by a detectable cell surface increase in β_2-m, which further supports the functionality of the complex. Increased cell surface β_2-m also acts as a transfection success indicator and as a parameter for cell sorting.

These cells are now ready to be used for the generation of an immunopeptidome specific to the SahaI*27-1 allele, an allele present in both DFT1 and DFT2 and widely expressed in the devil population, which may be the key to identifying targets for a large scale vaccination programme. The early stages of this work towards characterising the SahaI*27-1 binding motif and identifying potential vaccine targets are described in Chapter 8.

In addition to the characterisation of a biologically interesting MHC class I, this cell line is the first stable transfection of DFT2 cells, and we now have a robust workflow for transfecting and working with these difficult cells, which can be used for further detailed analysis of antigen processing in the Tasmanian devil and its tumours. There is additionally the possibility that these workflows could be further adapted for similar analyses of antigen presentation in other non-model species.

Chapter 8 Characterisation of the peptide binding properties of the MHC class I molecule SahaI*27-1

8.1 Introduction

*SahaI*27* and *SahaI*27-1* are two MHC class I alleles which differ by a single non-synonymous mutation which is not predicted to interfere with peptide binding, indicating these MHC class I heavy chains may bind the same peptide repertoire (Caldwell *et al.* 2018). Both have been identified in the DNA and mRNA of DFT1 and DFT2 cells along with DFT2 host devils (Caldwell *et al.* 2018), and previous data indicates that SahaI*27 is common in at least one devil population (Lane *et al.* 2012). Understanding the peptide binding repertoire of *SahaI*27* and *SahaI*27-1* is a crucial step to identifying vaccine targets for use in the wider devil population.

Peptide binding by MHC class I is determined by the size and electrochemical properties of the peptide binding groove, a region of the heavy chain which directly contacts peptide side chains (Leone *et al.* 2013) (Figure 1.2). The binding groove contains pockets of high electrochemical restriction which define the MHC class I "anchor residues", specific amino acid requirements at defined positions along the peptide (Sidney *et al.* 2008) (Figure 8.1). Most human and mouse MHC class I peptide restrictions are between 8-12 amino acids in length with a preference for 9mers and anchor residues at positions 2 and 9 (or c-terminal) of the peptide, though some alleles diverge from this and have been shown to bind much longer peptides (Leone *et al.* 2013; Rapin *et al.* 2008). The sequences of peptides binding a given MHC class I and their associated anchor residues can be visualised as a "binding motif". A binding motif can be used to predict whether any given peptide is likely to bind an MHC class I molecule, and with structural information prediction algorithms such as NetMHC can predict binding affinity (Andreatta and Nielsen 2016; Nielsen *et al.* 2003).

Our lab has generated pan-MHC class I immunopeptidomes for a DFT2, DFT1 (+IFNγ) and fibroblast cell line (Gastaldello *et al.* in preparation). This data has provided insights into unique features of MHC class I peptide binding in the Tasmanian devil, notably a length preference for 8mers on DFT2 and fibroblast cells and an anchor residue at position 3 (Figure 8.2). However, it currently provides no information on allele restriction, and attempts to bioinformatically define allele-specific motifs have been unsuccessful. It is currently not possible to define the allelic MHC class I binding properties in this dataset, preventing effective selection of vaccine target peptides.

In this Chapter, I will perform immunoaffinity purification of SahaI*27-1 from the recombinant cell line described in Chapter 7 in order to sequence the peptides binding this biologically interesting MHC class I allele. The determined binding motif can then be used to define allele restrictions of peptides within the larger immunopeptidome datasets across cell lines, allowing for selection of peptide vaccine targets which will be recognised within the wider Tasmanian devil population.

Additionally, very few immunopeptidome datasets exist outside of eutherian mammals, thus this work provides a foundation for studying the evolution of peptide binding specificities in different mammalian clades.

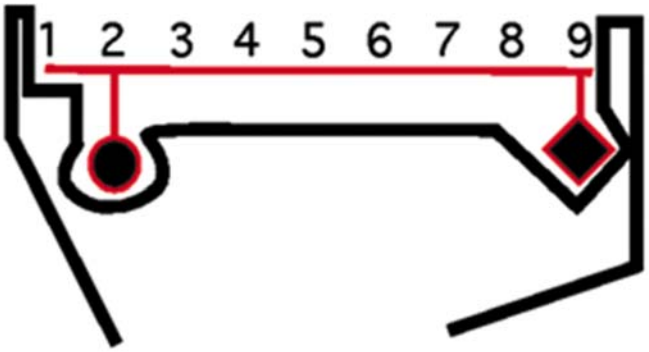


Figure 8.1: Schematic representation of the MHC class I binding groove.

Black line represents the MHC class I heavy chain peptide binding region with the associated binding groove. Red line indicates a 9mer peptide with large side chains at positions 2 and 9, which fit into the heavy chain binding pockets, conferring peptide sequence specificity (Milinski *et al*. 2005).

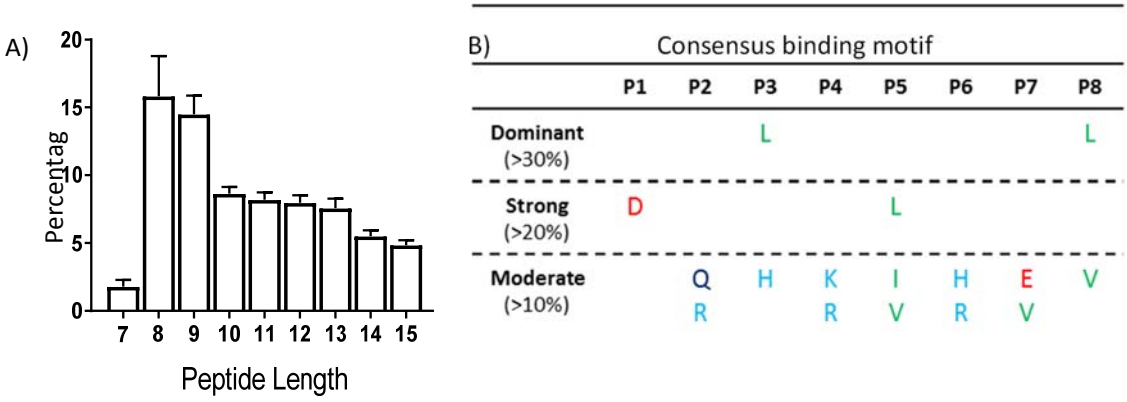


Figure 8.2: MHC class I in DFT2 preferentially binds 8mer peptides with a Leucine anchor at positions 3 and 8.

A) Peptide length distribution of experimentally determined 7-15mers in DFT2. Y axis indicates percentage of total peptides, X axis indicates peptide length. B) Consensus binding motif of 8mers in DFT2, X axis indicates position along peptide. Y axis indicates the percentage of peptides which have a given amino acid in a given position. Single letter amino acid codes are used and these are detailed in the Definitions and Abbreviations section of this thesis.

8.2 Aims and Objectives

Using the DFT2_UC cell line defined in Chapter 7, I will perform immunoaffinity purification on the SahaI*27-1 molecule using an antibody against the cytoplasmic myc tag. The purified MHC class I complex will then be fractionated by HPLC and the peptides sequenced by mass spectrometry to identify the peptide repertoire binding SahaI*27-1. The peptide sequences will be then analysed to define the peptide binding motif of SahaI*27-1, which can then be used to scan the previously generated immunopeptidomes for neoantigens in DFT1 and DFT2 which are binding SahaI*27-1 and may be useful as vaccine targets.

The chapter has the following objectives:

1. Perform immunoaffinity purification on recombinant SahaI*27-1 from DFT2_UC whole cell lysate using an anti-myc antibody.

2. Sequence the peptides binding SahaI*27-1 by mass spectrometry to generate a single allele immunopeptidome.

3. Define the binding motif of SahaI*27-1.

4. Identify neoantigens in pan-MHC class I immunopeptidome datasets which are predicted to bind the SahaI*27-1 allele.

8.3 Results

8.3.1 SahaI*27-1 successfully precipitated from whole cell DFT2_UC lysate

DFT2_UC cells (10^8/replicate, n=3) were grown in hyperflasks and harvested as described in Chapter 2.1.1. Cell pellets were lysed and incubated with anti-myc antibody (See Table 2.2) conjugated to protein A beads overnight to capture the recombinant SahaI*27-1 construct. Following myc immunoaffinity, the captured myc construct bound to the antibody-conjugated beads was separated from the remaining lysate using a glass EconoColumn, and the remaining lysate was immediately incubated with an anti-β_2-m antibody conjugated to protein A beads (β_2-m_13-34-38, Table 2.2) overnight to capture all endogenous MHC class I as an internal control dataset. A graphical representation of the workflow performed in this section is shown in Figure 8.3, and sample collection points analysed in Figure 8.4 are identified.

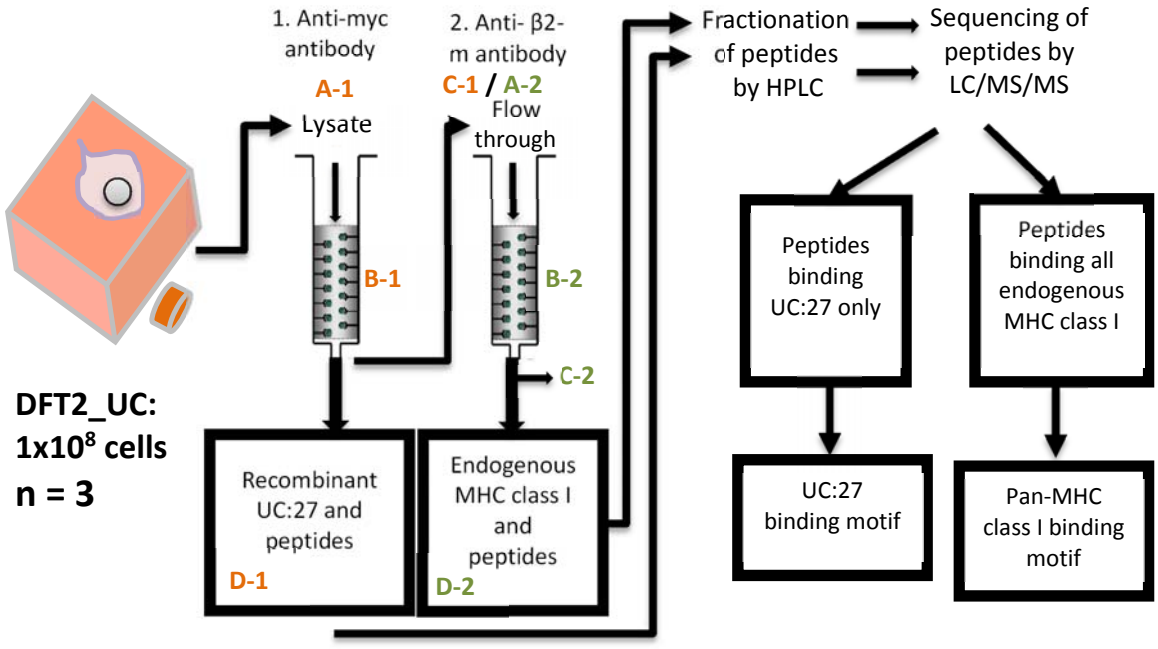


Figure 8.3: Graphical representation of the immunoaffinity purification procedure with indicated points of sample collection for analysis of the MHC class I heavy chain isolates.

Orange text indicates myc immunoaffinity purification samples and Green text indicates β_2-m immunoaffinity purification samples collected for analysis by Western blot. Labels correlate to samples analysed by Western blot in Figure 8.4.

Figure 8.4 shows western blot analysis performed on samples taken at all stages during immunoaffinity to analyse the MHC class I heavy chain isolated at each stage of the protocol (Figure 8.3). Samples were blotted with both an anti-myc antibody, to detect the myc tagged recombinant SahaI*27-1 construct, and an anti-heavy chain antibody (Saha-UA/B/C) to detect the classical MHC class I heavy chains (See Table 2.2). As previously described classical MHC class I heavy chains in devil cell lysates are detected at ~43 kDa and ~45 kDa using the Saha-UA/B/C_15-25-18 heavy chain antibody (Table 2.2). The addition of myc tags in the recombinant SahaI*27-1 heavy chain is predicted to add ~3 kDa to the molecular weight, thus detection is expected to be at ~48 kDa.

Untransfected DFT2_RV whole cell lysate is shown as a control (Figure 8.4, RV), and is positive for endogenous MHC class I heavy chain when blotted for Saha-UA/B/C (~43kDa), but negative for the recombinant SahaI*27-1 when blotted with anti-myc. Blotting of DFT2_UC whole cell lysate (Figure 8.4, A-1) for Saha-UA/B/C shows three positive bands at ~43 kDa, ~45 kDa and ~48 kDa. In contrast, blotting for myc detects a single positive band at ~48 kDa. We predict that the bands detected at ~43 kDa and ~45 kDa in the DFT2_UC and DFT2_RV whole cell lysates represent endogenous Saha-UA/B/C heavy chain, whilst the band detected at ~48 kDa in DFT2_UC whole cell lysate represents the myc tagged, recombinant SahaI*27-1 protein. The ~48kDa band is absent from the DFT2_RV whole cell lysates.

A band at ~48 kDa is detected in the antibody-bead conjugate post myc-immunoaffinity (Figure 8.4, B-1) when blotted for both Saha-UA/B/C heavy chain and for myc, indicating successful binding of the recombinant SahaI*27-1 construct to the myc antibody during immunoaffinity. An additional band at ~46 kDa is present in the myc antibody-bead conjugate when blotted for Saha-UA/B/C heavy chain, and in the post-myc immunoaffinity lysate (Figure 8.4, C-1) which may indicate cleavage of one of the myc tags during lysate preparation. Blotting of post-myc immunoaffinity DFT2_UC lysate (Figure 8.4, C-1) for Saha-UA/B/C heavy chain and myc detects no bands at ~48 kDa, indicating that the myc tagged SahaI*27-1 construct has been successfully purified from the DFT2_UC lysate. Endogenous heavy chains at ~43 kDa and ~45 kDa are still detected in the post-immunoaffinity lysate. Following elution of the MHC class I complex from the antibody-bead conjugate (Figure 8.4, D-1), a faint band is detected at ~48 kDa when blotted for Saha-UA/B/C heavy chain, but no band is detected when blotted for myc indicating that levels of recombinant SahaI*27-1 present in the final purified eluate are very low and cannot be detected by Western blot using an anti-myc antibody. These low levels of purified construct are likely to have affected downstream sample processing and sequencing of peptides.

No bands at ~48 kDa are detected in the pre β_2-m immunoaffinity lysate (Figure 8.4, A-2) when blotted for myc, though low levels of a band at ~46 kDa which may represent cleaved recombinant SahaI*27-1 are present. No bands at ~48 kDa are detected in the β_2-m antibody-bead conjugate (Figure 8.4, B-2), post β_2-m immunoaffinity lysate (Figure 8.4, C-2) or following elution of the MHC class I complex from the β_2-m antibody-bead conjugate (Figure 8.4, D-2) when blotted for myc,

indicating complete removal of recombinant SahaI*27-1 from the DFT2_UC lysate during the myc immunoaffinity purification process.

Western blot analysis of the DFT2_UC lysate pre-β_2-m immunoaffinity (post-myc immunoaffinity) (Figure 8.4, A-2) shows endogenous classical class I heavy chain at ~43 kDa and ~45 kDa when blotted for classical heavy chain. These bands are also present in the β_2-m antibody-bead conjugate (Figure 8.4, B-2), indicating the successful capture of endogenous MHC class I from the DFT2_UC lysate. Endogenous MHC class I is still detectable at comparable levels to the antibody-bead conjugate in DFT2_UC lysate post β_2-m immunoaffinity (Figure 8.4, C-2), indicating that not all endogenous MHC class I was removed from the lysate during purification. Finally, a band that could possibly represent heavy chain is present at ~43 kDa in the β_2-m immunoaffinity eluate (Figure 8.4, D-2), although this is unclear.

In summary, western blot analysis of all three replicates of immunoaffinity purification (Appendix G.1 and Figure 8.4) indicates that myc immunoaffinity purification is complete or near-complete but yields very low levels of myc-tagged construct which has likely affected downstream peptide sequencing. Additionally, this analysis indicates that whilst not all endogenous heavy chain is captured following β_2-m immunoaffinity, more endogenous construct is yielded from the purification thus more peptide is present for downstream sequencing. From here, MHC class I isolated by immunoaffinity purification using an anti-myc antibody will be referred to as “anti-myc samples”, and MHC class I isolated by immunoaffinity purification using an anti-β_2-m antibody will be referred to as “anti-β_2-m samples”.

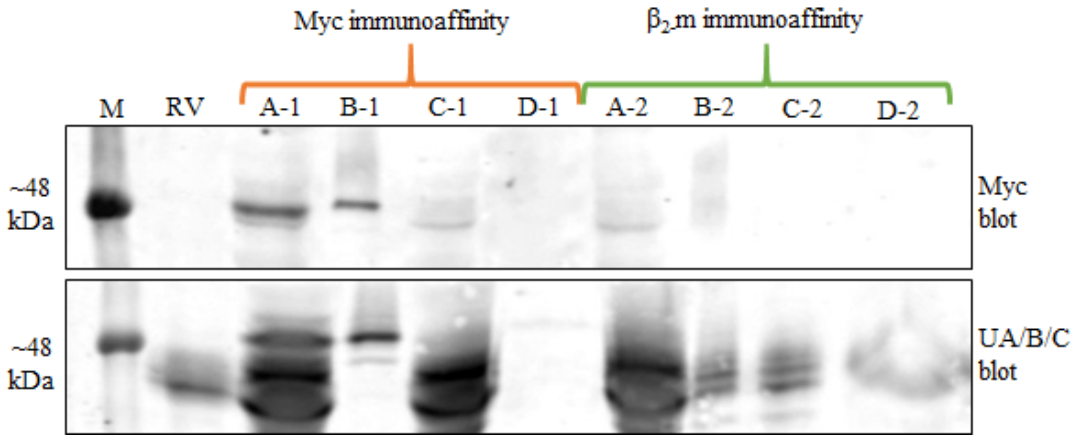


Figure 8.4: Western blot analysis of immunoaffinity indicates successful pull down of recombinant SahaI*27-1 and endogenous MHC class I.

Samples collected at various stages during A1 – D1) Myc immunoaffinity purification (anti-myc samples) and A2 - D2) β_2-m immunoaffinity purification (anti- β_2-m samples). RV) Untransfected DFT2_RV whole cell lysate, A) DFT2_UC lysate pre-immunoaffinity, B) Bead-conjugated antibody bound to protein of interest post immunoaffinity, C) DFT2_UC lysate post immunoaffinity, D) Purified MHC class I following elution from antibody-bead conjugate. Note that the post-myc immunoaffinity lysate in C-1) was immediately used in the β_2-m immunoaffinity and is the same as A-2). Full Western blot images are presented in Figure S 17, and triplicate experiments are presented in Appendix G.1 (Figure S 16, Figure S 18). Sample collection points are identified on the graphical representation of the workflow presented in Figure 8.3.

.

8.3.2 β_2-m not clearly detectable by HPLC separation of the MHC class I complex in either the anti-myc or anti-β_2-m immunoaffinity samples

Purified MHC class I complexes from immunoaffinity purification were fractionated by HPLC to isolate the peptides from β_2-m and heavy chain for mass spectrometry analysis. Previous analysis of the human MHC class I heavy chain, β_2-m and peptides on a Dionex UltiMate 3000 standard HPLC system with a Chromolith® RP-18 endcapped column of 100 mm length and 4.6 mm internal diameter with a biomodal pore structure (13 nm – 2 µm pore size) has demonstrated that human β_2-m is eluted at a retention time of 16.75 minutes (Bailey, pers com). No significant peaks were detected in the anti-myc or the anti-β_2-m immunoaffinity samples (Figure 8.5). High background signal is present around the expected β_2-m peak in both samples with no detectable peak at 16.75 in any sample. There is a peak just before 16.75 minutes in all samples (Figure 8.6). To determine whether this was species-specific size difference in β_2-m the fraction containing this peak was analysed by Western blot and 2D SDS-PAGE with Coomassie stain. Devil β_2-m should be detected at ~14 kDa but no bands are detected in any fractions, indicating no β_2-m protein is present (data not shown).

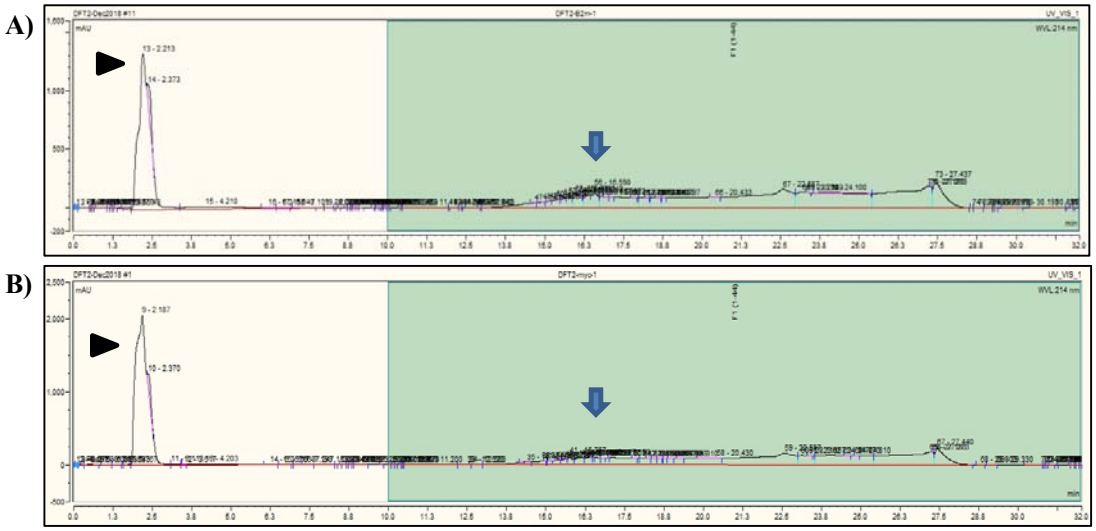


Figure 8.5: β_2-m is not detected in the HPLC traces of A) anti-myc samples or B) anti-β_2-m samples. Black line represents the HPLC trace. X axis represents retention (elution) time of the sample in minutes. Y axis represents intensity of absorbance at 280 nm in milliAbsorbance Units (mAU). Black arrowhead indicates sample injection peak. Blue arrow represents expected retention time of the β_2-m peak. Green boxed region indicates region fractions were collected from. Triplicate HPLC traces for myc and β_2-m immunoaffinity purifications are presented in Appendix G.2 (Figure S 19).

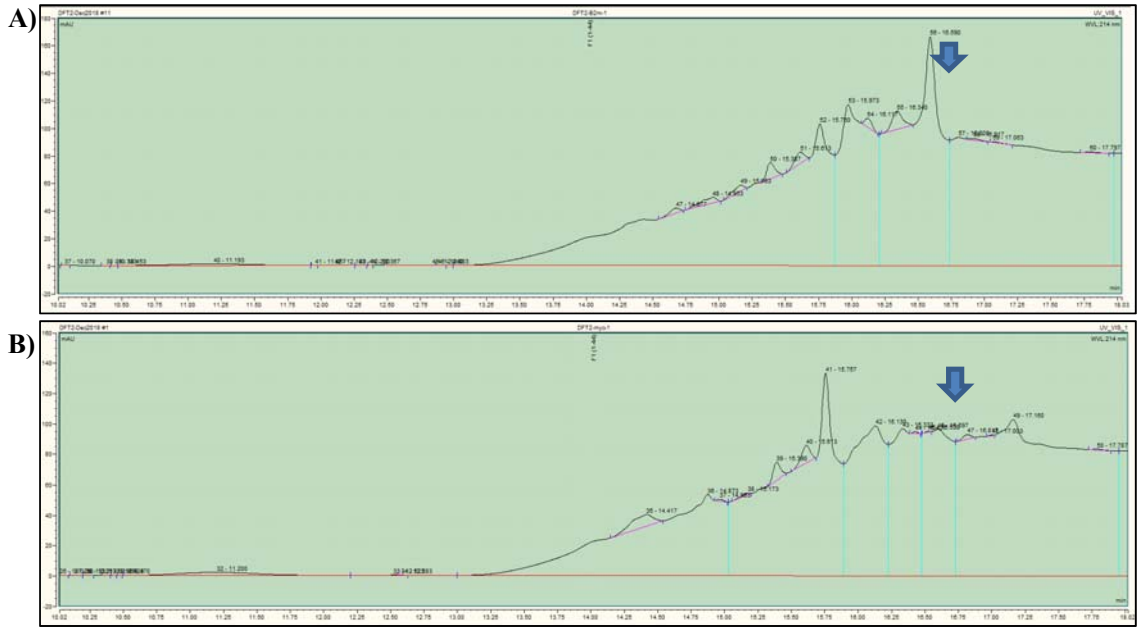


Figure 8.6: β_2-m does not elute at the expected retention time in A) myc or B) β_2-m purified samples. Black line indicates HPLC trace. X axis represents retention time of the sample. Y axis represents intensity of absorbance at 280 nm (mAU). Values indicate intensity of peaks. Blue arrow represents expected retention time of the β_2-m peak. Note these traces are the same traces as seen in Figure 8.4 but zoomed in on the collected fraction region.

8.3.3 MHC class I restricted peptide sequences were detected by mass spectrometry following β_2-m immunoaffinity, but not following myc immunoaffinity

The 10 fractions prior to the expected retention time of β_2-m were collected and analysed by mass-spectrometry. Peptide identities were defined using a custom database compiled by Alistair Bailey combining the genome with RNA-Seq data for DFT1 and DFT2 in order to capture all variants and possible mutated peptides. All identified peptides for all samples are presented in full in Appendix G.3.

A summary table of peptide numbers identified in three anti-β_2-m samples and three anti-myc samples is presented in Table 8.1. On average, the replicates of anti-myc samples produced 38% fewer peptides than replicates of anti-β_2-m samples. Over half of all peptides identified in all three anti-β_2-m samples and three anti-myc samples are between 7 and 15 amino acids in length (7-15mers). 34%, 52% and 50% of peptides in anti-β_2-m sample replicates 1, 2 and 3 respectively are between 8-12 amino acids in length (8-12mers), the standard expected length of MHC class I associated peptides. 46%, 48% and 33% of peptides in anti-myc sample replicates 1, 2 and 3 respectively, indicating a slightly lower enrichment for canonical length peptides in the myc immunoaffinity datasets.

All three anti-β_2-m samples demonstrate a strong length preference for 8mers (12%, 23% and 14% of total peptides respectively) with a consistently lower proportion of 9mers, the canonical MHC class I peptide length (7%, 12% and 11% respectively), and even lower proportions of 10mers (5%, 5% and 8% respectively). In contrast, all three anti-myc samples show reduced enrichment for 8mers (8%, 11% and 6% respectively), and no significant preference for 8mers over 9mers (9%, 10% and 5% enrichment for 9mers respectively). All three anti-myc samples show a stronger enrichment for 10mers than anti-β_2-m samples, and two of the anti-myc samples are more enriched for 10mers than 8mers (11%, 9% and 6% respectively.

Table 8.1: Anti-β_2-m samples produce more peptides and demonstrate a more canonical length preference than anti-myc samples.

Replicate	Total peptides	8-12mers	7-15mers	8mers	9mers	10mers
B_2-m 1	2085	727	1175	243	151	108
B_2-m 2	886	459	600	205	110	48
B_2-m 3	1479	742	1023	201	165	131
Myc 1	1100	511	737	88	97	118
Myc 2	432	208	288	49	45	40
Myc 3	1229	415	691	77	61	75

Combining the total peptides identified across three replicates indicates that anti-β_2-m samples are slightly more enriched for 8-12mers than anti-myc samples (43% vs 41%), and additionally are more enriched for 7mers and 13-15mers (20% vs 11%), and have a lower proportion of long peptides identified overall, indicating peptides identified in anti-β_2-m samples have a more canonical length preference than those identified in anti-myc samples (Figure 8.7).

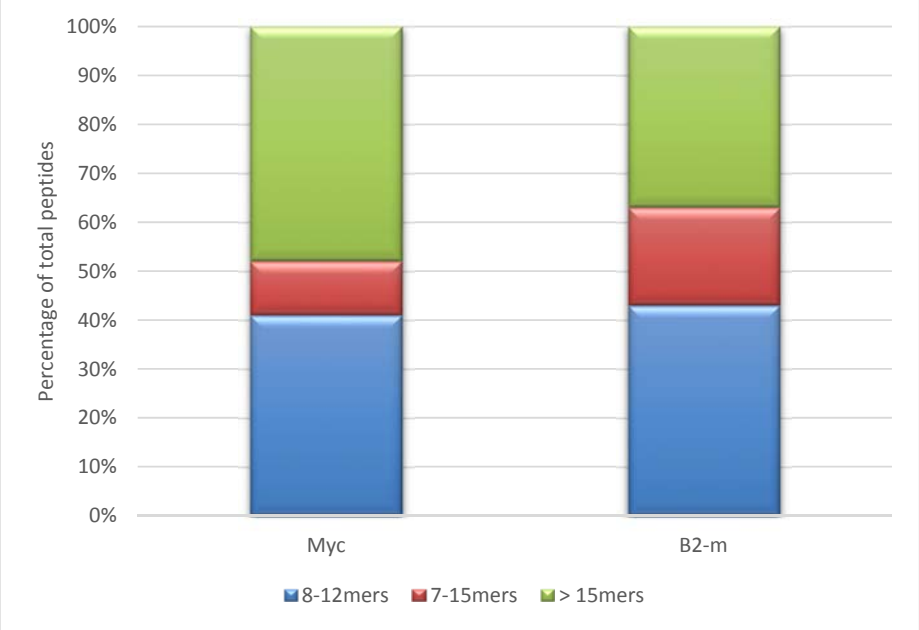


Figure 8.7: Anti-β_2-m samples are more strongly enriched for canonical length peptides than anti-myc samples.

Y axis indicates percentage of total peptides identified across all three replicates of anti-β_2-m samples and anti-myc samples. Blue sections indicate percentage of total identified peptides of 8-12 amino acids in length. Orange sections indicate percentage of peptides 7, 13, 14 and 15 amino acids in length. Green sections indicate percentage of peptides identified longer than 15 amino acids in length.

Peptide length distributions for three replicates of anti-myc and anti-β_2-m samples were combined and the standard error of mean (SEM) calculated and plotted against peptide number to determine whether the distribution was canonical for an MHC class I peptide dataset (Figure 8.8). An enrichment for peptides between 7-15 amino acids, with the highest enrichment for 8-12 amino acids is considered a canonical MHC class I peptide distribution (Trolle *et al.* 2016). Peptide length distributions in anti-β_2-m samples demonstrated higher peptide numbers, lower standard errors and averages across replicates demonstrated a more canonical length distribution than anti-myc sample peptides, with a distinct preference for 8mers and 9mers compared to other peptide lengths. Anti-myc sample peptides demonstrate lower peptide numbers and do not show any distinct length preferences. The lack of distinct length preference, unclear length distribution and generally low peptide numbers indicates that the myc immunoaffinity experiments were unsuccessful, possibly due to the low yield of purified recombinant SahaI*27-1 demonstrated in Figure 8.4. Thus, the anti-myc samples are considered to have failed, and are not analysed further in this project.

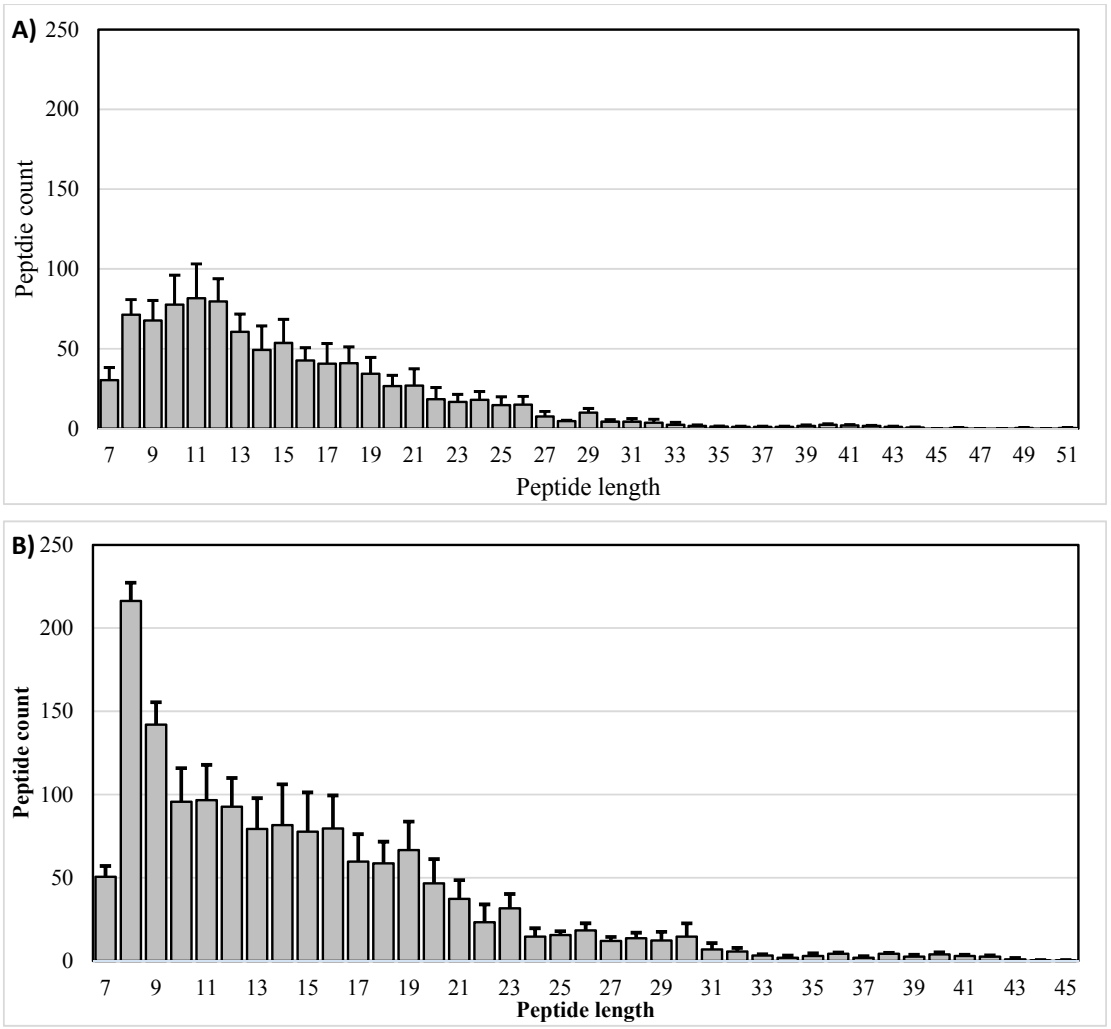


Figure 8.8: Peptides detected in anti-β_2-m samples are more reproducible and show a more canonical length distribution than peptides detected in anti-myc samples.

Composite graphs showing the average peptide length distribution across three replicates in A) anti-myc samples and B) anti-β_2-m samples. Data are mean ± SEM.

Peptide sequences in each replicate were compared to define the reproducibility of replicates within the experiment. Of the total peptides identified in anti-β_2-m samples, 13% were identified in all 3 replicates, and over two thirds of identified peptides were only identified in a single replicate. This is comparable with levels of 7-15mers shared between replicates (Figure 8.9). Peptide length distribution is canonical with a strong preference for 8mers in all individual replicates (Figure 8.10) indicating that a canonical length distribution with a preference for 8mers is conserved between replicates. Additionally, peptide length distributions are canonical with a preference for 8mers regardless of whether the peptides are identified in one, two or three replicates (Figure 8.11). Peptide length distributions of unique peptides identified in a single replicate are canonical for replicates 2 and 3 but demonstrate a small secondary peak at 14-15mers in replicate 1. However, 8mers are still the largest group of peptides in replicate 1 (Figure 8.12), further demonstrating that a canonical length distribution and a preference for 8mers is highly reproducible across replicates despite poor sequence replication.

Peptide sequences were directly compared to all peptides identified in at least one replicate of the previously generated DFT2 immunopeptidome (Gastaldello *et al.* in preparation, Appendix G.4) to determine the reproducibility of the two separate experiments (Figure 8.13). Of all peptides detected in the current experiment, 24% were also previously detected in at least one replicate by Gastaldello *et al.* Most of the peptides not detected in the Gastaldello *et al*. dataset are only detected in a single replicate of the anti-β_2-m samples. Over half of the peptides detected in all 3 anti-β_2-m sample replicates of the current experiment are also detected in the Gastaldello *et al.* dataset. A higher proportion of 7-15mers are shared with the Gastaldello dataset compared to total peptides, with 29% of 7-15mers being detected in the Gastaldello dataset. 7-15mer peptides detected in a single anti-β_2-m sample replicate are still likely to be unique to this experiment and not detectable in the Gastaldello set.

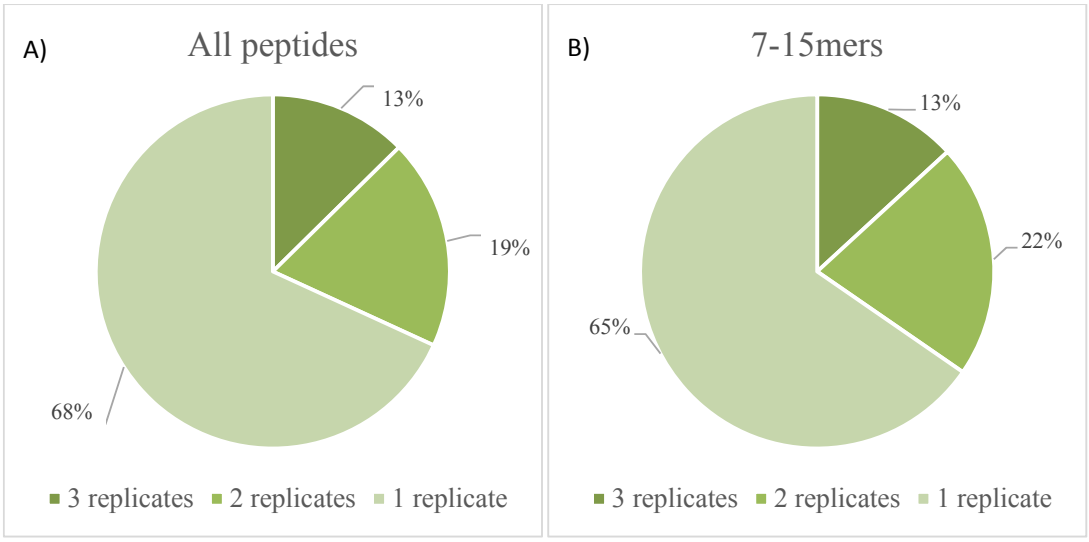


Figure 8.9: Only 13% of detected peptides are detected in all three replicates of anti-β_2-m samples.
Pie charts demonstrating percentage of peptides identified in 1, 2 or 3 replicates and anti-β_2-m samples for A) All peptides of all lengths and B) 7-15mers.

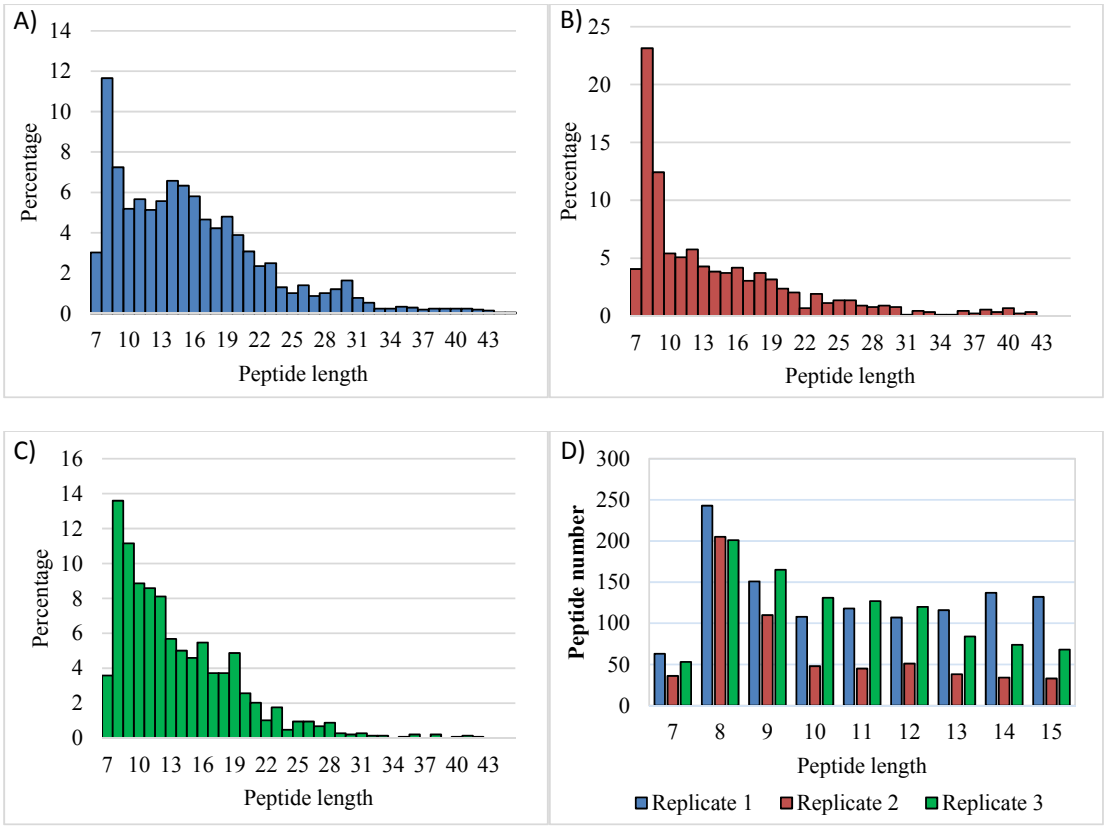


Figure 8.10: Peptide length distribution in anti-β_2-m samples indicates a canonical MHC class I peptide dataset in all three replicates.

Full peptide length distribution graphs for A) anti-β_2-m sample replicate 1, B) anti-β_2-m sample replicate 2, C) anti-β_2-m sample replicate 3. X axis indicates peptide length, Y axis indicates percentage of total identified peptides. D) All 7-15mers detected in each replicate. X axis is peptide length, Y axis is peptide number.

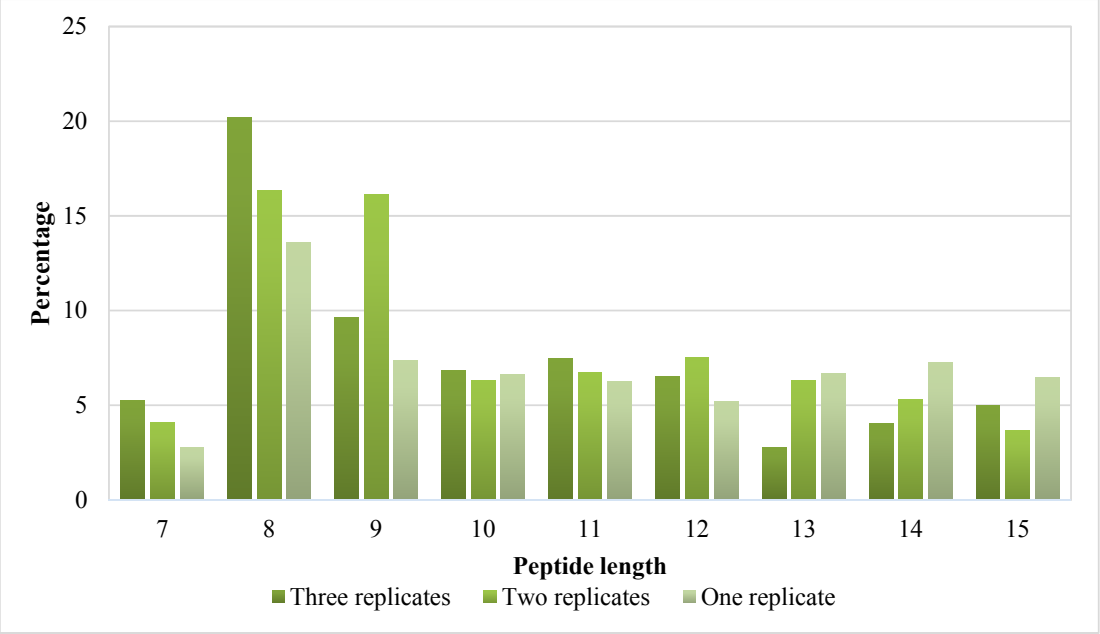


Figure 8.11: Peptides identified in anti-β_2-m samples display a canonical length distribution regardless of how many replicates they were identified in.

Bars represent peptide length distributions (between 7-15 amino acids) for peptides identified in 1, 2 or 3 replicates of anti-β_2-m samples. Y axis represents percentage of total peptides identified in 1, 2 or 3 replicates, X axis represents peptide length.

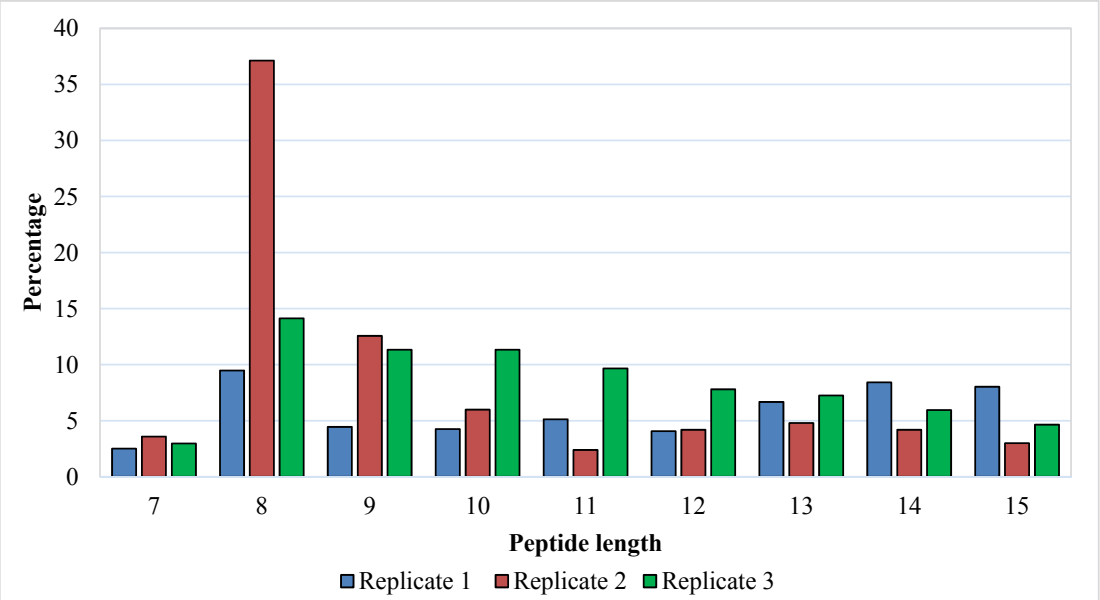


Figure 8.12: Peptides identified in a single anti-β_2-m sample replicate display a predominantly canonical length distribution.

Bars represent peptide length distributions for peptides between 7 and 15 amino acids in length which are unique to a single anti-β_2-m sample replicate. Y axis represents percentage of total peptides detected in each replicate, X axis represents peptide length.

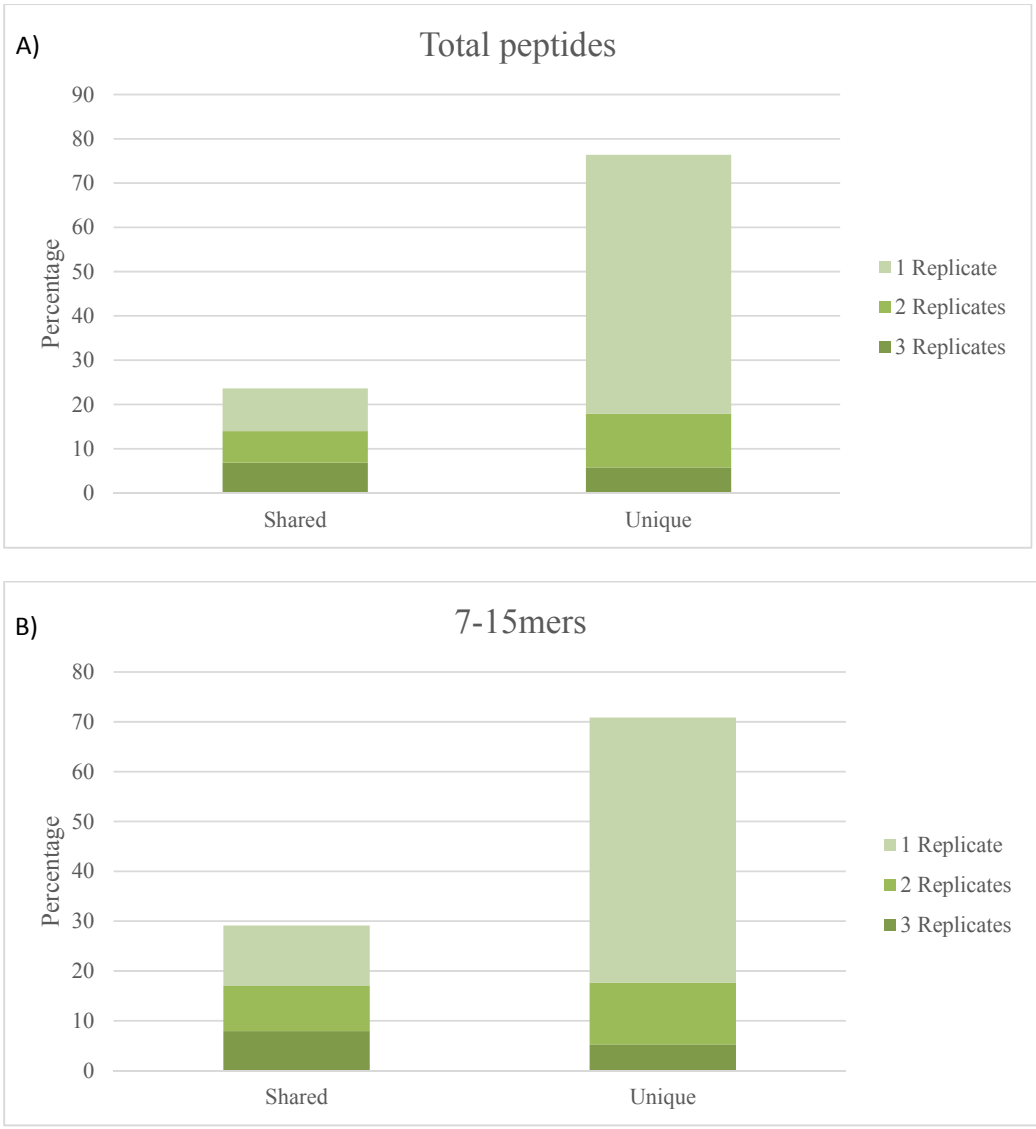


Figure 8.13: Over 20% of peptides detected in the DFT2_UC immunopeptidome have been identified in previous DFT2 immunopeptidome datasets.

Stacked column charts indicating the percentage of A) Total peptides or B) 7-15mers which are either unique to the anti-β_2-m sample datasets generated here (Unique) or shared with the DFT2 immunopeptidome dataset generated by Gastaldello *et al*. (Shared). Colours indicate how many replicates of the DFT2_UC immunopeptidome peptides were identified in.

8.3.4 Peptides identified by β_2-m immunoaffinity experiments demonstrate an unusual peptide binding motif with anchors at position 3 and the C-terminus

Sequence alignment and clustering was initially performed on all peptides identified in anti-β_2-m samples using the GibbsCluster-2.0 server (constrained for the preferred 8mer) and indicates the presence of five binding motifs within the dataset (Figure S 20). However, the suggested binding motifs are unclear (Figure S 21). GibbsCluster-2.0 uses a combination of previously generated MHC-peptide binding affinity data and immunopeptidome data from model species, and uses background amino acid frequencies that are not specific to the Tasmanian devil (Andreatta *et al.* 2012). This makes it difficult to interpret GibbsCluster data with any confidence, and thus analysis of the binding motif of the combined dataset was undertaken instead. GibbsCluster data is presented in Appendix G.5.

Peptide binding motifs were generated by calculating the frequency of each amino acid at each position along the peptide dataset. All peptides with a frequency of over 10% at a given position are plotted, with 10-19% indicating moderate overrepresentation, 20-29% indicating strong overrepresentation and >30% indicating dominant overrepresentation of an amino acid at a given peptide position. 8mer and 9mer binding motifs are reasonably well conserved across all three replicates (Figure 8.14). There is a highly conserved Leucine anchor at position 3 in both 8mers and 9mers in all three replicates. Additionally, there is a highly conserved hydrophobic anchor, predominantly Leucine, at the c-terminus of both 8mers and 9mers in all replicates. There is also a weaker but well conserved preference for Glutamate at P7 in 8mers and P8 in 9mers, a strong to moderate preference for Aspartic Acid at P1 and a moderate preference for a hydrophobic residue at P5. Due to the well conserved nature of these motifs, the three replicate datasets were combined to give a single list of every peptide identified in at least one replicate for further analysis.

A) **8mers – 219 peptides**

P1	P2	P3	P4	P5	P6	P7	P8
		L					L
D				L		E	
E N	Q R K N		I R K	I V	E R	V	V

Dominant (>30%)	Strong (>20%)	Moderate (>10%)

P1	P2	P3	P4	P5	P6	P7	P8	P9
		L						L
			P					
D S V	R Q		L K	L	I P L	Q	E	F

9mers – 136 peptides

B) **8mers – 193 peptides**

P1	P2	P3	P4	P5	P6	P7	P8
		L					L
D	Q			L			V
E S	V R N		P S I K	I V	H E	E T K	

Dominant (>30%)	Strong (>20%)	Moderate (>10%)

P1	P2	P3	P4	P5	P6	P7	P8	P9
		L						
	Q		P					L
D V	R P		L K	V P	P L I	V	E	P F

9mers – 102 peptides

C) **8mers – 179 peptides**

P1	P2	P3	P4	P5	P6	P7	P8
		L					L
D	Q V R		P I L R E	L I V	H E R	E K	V

Dominant (>30%)	Strong (>20%)	Moderate (>10%)

P1	P2	P3	P4	P5	P6	P7	P8	P9
		L						L
V	R P S		P L K	P V T	P V	Q	E	

9mers – 141 peptides

Colour legend:	Hydrophobic	Charged – Acidic	Charged - Basic	Polar neutral

Figure 8.14: Binding motifs for anti-β_2-m samples are well conserved across replicates.

A) Replicate 1, B) Replicate 2, C) Replicate 3. X axis headings indicate position along the peptide. Y axis indicates amino acid frequency. Only frequencies above 10% are plotted. Amino acid codes and characteristics are presented in full in the abbreviations section of this thesis. The legend indicates colour coding for broad amino acid properties.

To identify amino acid residues which have been positively selected for within MHC class I peptides, the amino acid frequency calculations were corrected for background frequency of each amino acid in the genome. Additionally, a Chi-squared test was performed to identify amino acids which have been significantly positively selected for within the immunopeptidome. P-values were corrected for multiple testing using Yates correction and Bonferroni correction. Only amino acid frequencies that were positively selected for in the immunopeptidome by more than 100% and were significant by both correction methods are presented here. Leucine is overrepresented in the Tasmanian devil genome and its expected frequency in any given peptide is ~10%. Both the P3 Leucine anchor and the c-terminal hydrophobic anchor remain strong to dominant following correction for background amino acid frequencies in both 8mers and 9mers (Figure 8.15). Interestingly, a dominant Aspartic Acid anchor at P1 is present in 8mers following correction for background frequencies, and a moderate Proline anchor is present at P4 in 9mers. Additionally, the weak P7/P8 Glutamate anchor remains following background frequency correction.

A)

8mers – 381 peptides

P1	P2	P3	P4	P5	P6	P7	P8		P1	P2	P3	P4	P5	P6	P7	P8	P9
		L					L	**Dominant** (>30%)			L						
D				L				**Strong** (>20%)				P					L
E	Q R		I R L P K	I V	E	E	V	**Moderate** (>10%)	V	R Q		L	L V	P L	V	E P	F

9mers – 238 peptides

B)

P1	P2	P3	P4	P5	P6	P7	P8		P1	P2	P3	P4	P5	P6	P7	P8	P9
D		L					L	**Dominant** (>300%)									
	Q	H		I	H			**Strong** (>200%)			L	P					
	N		I R	V L		E	V	**Moderate** (>100%)		Q R	H					E	L F

Colour legend:	Hydrophobic	Charged – Acidic	Charged - Basic	Polar neutral

Figure 8.15: The P3 Leucine anchor and hydrophobic c-terminal anchor remain strong to dominant after correction for background amino acid frequency.

A) 8mer and 9mer binding motifs and B) 8mer and 9mer positive selection motifs for the combined immunopeptidome of all three anti-β_2-m sample replicates. X axis headings indicate position along the peptide. Y axis indicates amino acid frequency. Only frequencies above 10% are plotted for consensus binding motif. Only selection above 100% is plotted for positive selection motifs. Amino acid codes and characteristics are presented in full in the abbreviations section of this thesis. The legend indicates colour coding for broad amino acid properties.

8.3.5 Peptides from neural specific source proteins are presented on MHC class I in the context of DFT2

The top 50 most highly expressed proteins in the DFT2 proteome (See Chapter 3.3.4 and Table 3.12) were searched against the immunopeptidome dataset to identify whether any peptides from these highly expressed proteins are presented on MHC class I on the cell surface (Table 8.2).

Overall, there is no correlation between protein expression and the number of peptides from a protein presented on MHC class I. However, 29 of the top 50 proteins were identified as source proteins for MHC class I peptides in the DFT2_UC immunopeptidome. 24 of these source proteins additionally contribute peptides to the pan-specific DFT2 immunopeptidome generated by Gastaldello *et al*. 9 proteins were not detected as peptide source proteins in the DFT2_UC immunopeptidome, but contribute peptides to the pan-specific DFT2 immunopeptidome. A further 5 proteins contribute peptides to the DFT2_UC immunopeptidome generated here, but do not contribute peptides to the pan-specific DFT2 immunopeptidome.

Interestingly, the 5th most highly expressed protein in DFT2, the neural stem cell marker protein NES, was detected as a source protein to 12 peptides across all three replicates, all of which were 8-12 amino acids in length. NES has also been identified as a source protein in the Gastaldello dataset. Of the 12 peptides contributed by NES to the DFT2_UC immunopeptidome, 11 are 8mers and 1 is a 9mer. Comparison of the peptide sequences indicates that three of these 8mer peptides are present in all replicates of the immunopeptidome (Table 8.3). One of these peptides, HNQIAQVL, is presented twice in replicate 3, with and without a post translational modification. Additionally, the 9mer peptide is an extended version of one of these replicated 8mers. These specific sequences are also identified in multiple replicates in the Gastaldello *et al*. dataset, indicating these may be stable, high affinity peptides which can be consistently identified during immunoaffinity purification. Interestingly, whilst all five individual peptide sequences demonstrate a hydrophobic c-terminal anchor residue, only one peptide which is only identified in a single replicate has a P3 leucine anchor. Understanding allele specific binding properties of MHC class I in the devil will be particularly useful for assessing the potential of these peptides as vaccine targets.

Table 8.2: Over half of the top 50 most highly expressed proteins in DFT2 are source proteins for peptides identified in the immunopeptidome.

Proteins are in order of highest to lowest expression. Red highlight indicates proteins which have a neural tissue specific expression pattern. Gastaldello dataset indicates whether a peptide from a given protein was identified as a source protein for at least one peptide in at least one replicate of the Gastaldello dataset.

Protein	Number of peptides in DFT2	8-12mers	Number of replicates	Gastaldello dataset?
VIM	40	17	3	Yes
HIST1H4A	1	0	1	Yes
ANXA2	10	10	3	Yes
PPIA	0	0	0	Yes
NES	12	12	3	Yes
TUBB	31	14	2	Yes
ANXA1	0	0	0	No
TUBB6	36	16	3	Yes
HIST1H2AA	5	4	2	No
HIST1H2AB	9	7	3	Yes
HIST1H2AD	5	4	2	No
HIST1H2AG	9	7	3	Yes
PKM	5	3	3	Yes
HIST1H1E	3	0	2	Yes
HMGB1	0	0	0	Yes
S100A11	4	0	3	Yes
SERPINH1	28	5	2	Yes
PGK1	3	1	2	Yes
CALR	2	2	1	Yes
GSN	1	1	1	No
TPI1	0	0	0	No
GPI	0	0	0	Yes
DPYSL2	28	14	3	Yes
ARHGDIA	0	0	0	No
CKB	0	0	0	Yes
MDH1	0	0	0	No
ALDOA	0	0	0	Yes
ATP5F1B	0	0	0	Yes
MAP4	57	19	3	No
CWC15	0	0	0	No
TAGLN2	0	0	0	Yes
HNRNPA1	0	0	0	Yes
CFL1	19	12	3	Yes
MSN	2	2	2	Yes
SOD1	0	0	0	No
DPYSL3	3	1	1	No
PPIB	1	0	1	Yes
RPL12	0	0	0	No
LMNA	1	1	1	Yes
RPS5	1	1	1	Yes
UBA1	0	0	0	No
FLNA	5	0	2	Yes
ANXA6	0	0	0	No
ANXA5	0	0	0	No
PHGDH	2	0	2	Yes
PRDX2	0	0	0	No
EIF4A1	10	1	3	Yes
NGFR	0	0	0	Yes
MDH2	0	0	0	Yes
VAT1	7	3	2	Yes

Table 8.3: 8mer peptides derived from the neural specific protein NES are consistently identified in immunopeptidome analysis of DFT2 cells.

Sequence of all peptides identified from NES during this analysis are presented. Owen Replicates refers to the number of replicates of the immunopeptidome presented here a peptide sequence was identified, Gastaldello replicates refers to the number of replicates of the immunopeptidome generated for DFT2 by Gastaldello *et al*. (in preparation) a peptide sequence was identified.

NES peptide sequence	Owen Replicates	Gastaldello Replicates
PQMEFREV	3	3
HNQIAQVL	3	3
DHKEIAEL	3	3
ETLQCLEV	1	0
GPQMEFREV	1	3

8.4 Discussion

Reduced MHC class I diversity has been linked with a reduced ability within a species to respond to pathogens (Maibach and Vigilant 2019) and reduced diversity of the T-cell receptor (TCR) (Migalska *et al.* 2019), both of which make a species vulnerable to infectious disease, and has been postulated to have facilitated the emergence and spread of transmissible tumours in the Tasmanian devil (Caldwell *et al.* 2018; Siddle *et al.* 2007a). DFTs evade the T-cell anti-graft immune response raised against mismatched MHC class I alleles (Takemoto *et al.* 2000; Sherman and Chattopadhyay 1993) and the slower rejection associated with recognition of polymorphisms in the bound peptide sequences known as minor histocompatibility antigens (Pierce *et al.* 1999; Wang *et al.* 1995). How DFT2 avoids major and minor histocompatibility rejection is unclear, though the suspected functional redundancy of its two major classical MHC class I alleles (*SahaI*27* and *SahaI*27-1*) may be playing a role. Understanding the peptide sequences binding these MHC class I molecules, and crucially whether they are immunogenic is important in understanding the underlying mechanisms of the emergence of tumours which transmit in the presence of MHC class I. Additionally, understanding the binding properties of MHC class I in the devil is clinically important for effective vaccine design, as peptide vaccines are haplotype restricted and any peptide target useful in a large scale vaccination programme must be capable of binding MHC class I in the wider devil population (Harper *et al.* 2015). Here, I have described the first attempts to define the specific peptide binding properties of *SahaI*27-1*, a dominant MHC class I allele in both DFT1 and DFT2 which is predicted to be functionally identical to another dominant allele in DFT1, DFT2 and the wider devil population, *SahaI*27*.

Following the optimisation and functional characterisation of a stable DFT2 cell line overexpressing recombinant, myc-tagged SahaI*27-1 (described in Chapter 7), here I have expanded this cell line and performed immunoaffinity purification of the recombinant MHC class I molecule with an anti-myc antibody. In addition to this, I also performed an immunoaffinity purification of endogenous MHC class I using an anti-β_2-m antibody, to act as an internal control for the immunoaffinity protocol, as this antibody has been previously used with success by Annalisa Gastaldello within our lab to generate immunopeptidomic datasets (in preparation). Western blot analysis of the immunoaffinity purification process indicated a successful purification of both myc-tagged and endogenous MHC class I, although the yield of recombinant SahaI*27-1 was low resulting in no β_2-m detection by HPLC and none of the myc samples generating a reliable MHC class I peptide dataset. The three β_2-m replicates were successful, and MHC class I peptides with a canonical length distribution were detected by mass-spectrometry.

Whilst both this protocol and the Gastaldello protocol used the same β_2-m antibody and the same basic process, it is worth noting a number of significant changes which are likely to have affected the outcome of these experiments, in particular with regards to the failure of the myc immunoaffinity. A crucial change in the protocol is the number of cells used for each replicate. The Gastaldello dataset

was generated using 10^9 cells per replicate, whereas the workflow presented here was optimised for 10^8 cells (Bailey *et al.* in preparation). Further, DFT2 expresses lower levels of MHC class I than fibroblast cells (Caldwell *et al.* 2018), and the recombinant construct will represent only a subset of the MHC class I on the cell surface. This is highlighted by the western blot analysis of the recombinant cell lines and the small amount of myc tagged protein detected (Figure 7.11, Figure 7.12, Figure 7.13, Figure 8.4). It has been demonstrated that high levels of MHC class I expression is a requirement for successful immunopeptidome generation (Kowalewski *et al.* 2015), and it is highly likely that there simply was not enough recombinant SahaI*27-1 for adequate size discrimination by HPLC and detection on the mass spectrometer. It is also possible that the construct is only transiently or weakly associating with β_2-m, and this association is lost during immunoaffinity. This seems unlikely as previous experiments (detailed in Chapter 7) have demonstrated that the myc tagged construct associates with β_2-m (Figure 7.13). Additionally, this experimental failure may be due to the myc antibody. The antibody used was purified from a hybridoma and there may have been a problem with the antibody batch used in this experiment. Future experiments could alter both the vector used and the cell line for transfection, possibly transfecting into an MHC class I heavy chain null human cell line such as 721.221.

Despite the difficulties with immunoaffinity purification of the myc tagged MHC class I construct, I have generated triplicate pan-MHC class I immunopeptidome data by performing immunoaffinity purification with β_2-m. We have generated this dataset using fewer cells than are standardly used in such experiments (Caron *et al.* 2015b), and indeed fewer than used in the previously generated DFT2 immunopeptidome from Gastaldello *et al.* (in preparation). The peptides in this dataset overlap with the previously generated DFT2 immunopeptidome, with 24% of the peptides identified found in both datasets (Figure 8.13). These experiments were performed on slightly different cell types (Transfected vs Untransfected DFT2_RV), using different cell numbers with different batches of antibody. Further, the immunoaffinity purification was performed using different protocols in different labs, using different HPLC systems and mass spectrometers. It is not uncommon in immunopeptidomics for low reproducibility between experiments, even when resampling the same biological and technical replicates (Caron *et al.* 2015b; Michalski *et al.* 2011; Tabb *et al.* 2010). Thus, 24% of peptides shared despites these differences in conditions is promising when considering the work is being undertaken in a non-model species. Additionally, characteristics of peptide binding, including length preference and binding motif are conserved, indicating functional similarities between the two immunopeptidomes.

Despite the significantly lower amount of starting material, the peptides detected by mass spectrometry exhibit a similar length distribution to the Gastaldello dataset (Figure 8.2, Figure 8.10). The length distribution indicates a significant preference for 8mers over 9mers, a finding contrasting with widely demonstrated MHC class I binding preference for 9mers in multiple species (Nielsen *et al.* 2018; Trolle *et al.* 2016; Nielsen and Andreatta 2016; Rapin *et al.* 2008). This preference is also noted in the Gastaldello *et al*. immunopeptidome for DFT2 (Figure 8.2), although to a lesser degree

than demonstrated here, indicating that this is a species specific characteristic of devil MHC class I. This data has demonstrated a strong preference for hydrophobic amino acids at the C-terminus of bound peptides in DFT2. This has been previously demonstrated in DFT2 by the Gastaldello dataset, and additionally in DFT1 and devil fibroblasts (Gastaldello *et al.* in preparation). A hydrophobic C-terminal anchor residue is a common binding preference for MHC class I molecules, thus the presence of this anchor in the Tasmanian devil, even in a pan-specific dataset, is not particularly unusual (Rapin *et al.* 2008; Janeway *et al.* 2001; Falk *et al.* 1991). In addition, the data presented here indicates a strong to dominant preference for Leucine at position 3 in both 8mers and 9mers, and a strong to dominant preference for Leucine or another hydrophobic amino acid at the C-terminus of the peptide (Figure 8.14). Leucine is the most common amino acid in the devil genome, but correction for this still demonstrates dominant positive selection for Leucine at positions 3 and 8 in 8mers, and moderate to strong selection for Leucine at position 3 and hydrophobic amino acids at position 9 in 9mers. Leucine is not uncommon as an MHC class I anchor residue (Jurtz *et al.* 2017; Rapin *et al.* 2008) and is most notably the predominant amino acid in the P2 anchor of the human MHC class I allele HLA-A*02:01 (Drijfhout *et al.* 1995) and an orthologous allele in chimps (Rapin *et al.* 2008). Although leucine anchors exist and leucine is seen in many of the known MHC class I binding motifs (Rapin *et al.* 2008), it is rarely as dominant as demonstrated at position 3 in this dataset. This may indicate that the antigen presentation pathway and peptide loading systems in the Tasmanian devil have a preference for leucine compared to other species. The position of the P3 leucine anchor in this dataset is also striking, as most MHC class I molecules studied demonstrate strong position 2 anchor residues and rarely demonstrate significant anchoring at position 3 (Rapin *et al.* 2008). Interestingly, the mouse MHC class I allele, H-2K^{b}, demonstrates similar binding properties, with a preference for 8mers and a secondary anchor at position 3, with an additional anchor at position 5 (Garstka *et al.* 2015; Rapin *et al.* 2008; Fremont *et al.* 1995). Additionally, some human MHC class I, particularly HLA-C alleles such as HLA-C04:01 and HLA-C06:02, show preferences for 8mer binding and P3 anchors, which has been postulated to relate to the role of HLA-C in humans as a primary NK cell receptor (Gfeller *et al.* 2018; Bassani-Sternberg *et al.* 2017). Additionally, the 8mer positive selection motif indicates at dominant Aspartic acid residue at position 1, and the 9mer positive selection motif indicates a strong secondary Proline anchor at position 4. The data generated in this chapter confirms the unique and species specific MHC class I binding properties in the Tasmanian devil which have been previously identified by Gastaldello *et al.*, indicating that these specificities are likely reliable and reflect the true peptide repertoire of MHC class I in the Tasmanian devil.

Analysis of the peptides being presented from highly expressed proteins in DFT2 has identified NES as a source protein of interest. NES, or Nestin, is an undifferentiated neural stem cell marker protein with restricted expression in the adult body (Silvestroff *et al.* 2013). Three peptide sequences from the Nestin source protein were identified in all three replicates of the immunopeptidome generated here and in all three replicates of the immunopeptidome generated by Gastaldello *et al.* (Table 8.2,

Table 8.3), thus are likely to be highly expressed on the cell surface and are potential high affinity binders. These peptides are all 8mers and an additional commonly detected 9mer sequence contains a nested 8mer within its sequence. Interestingly, whilst these peptides adhere to the clear length preference of MHC class I in the context of DFT2 and all demonstrate a c-terminal hydrophobic anchor residue, none of the peptides identified in all three replicates have a P3 leucine residue. This may indicate that the P3 anchor is less important for peptide stability. Additionally, it may be that these peptides bind distinct MHC class I alleles which do not have a P3 leucine anchor residue. Further analysis of these peptides can be carried out using refolding assays with MHC class I heavy chain and β_2-m to determine binding specificity.

The immunopeptidome here is a pan-MHC class I dataset, and thus we would expect it to contain peptides from a number of MHC class I alleles, each with a unique binding specificity and motif. In that respect we would not expect to see such a strong binding motif in a mixed dataset. The analysis of the immunopeptidome as a whole presented in section 8.3.4 reveals clear binding specificities and anchor residues. The clarity of a single motif in a mixed dataset may be due to a number of reasons, some of which could provide insight into how and why the species has developed two contagious tumours. Firstly, SahaI*27 and Saha-UK are both expressed by DFT2 cells at a higher level than other alleles (Caldwell *et al.* 2018), and it may be that a large proportion of the identified peptides are binding one of these molecules, resulting in this motif effectively "swamping" the immunopeptidome and hiding the motifs from less dominant alleles. Additionally, whilst human MHC class I consists of three genes which are expressed in all cells, it is not uncommon for other species to display region configuration polymorphism of the MHC class I locus, expressing only some of the MHC class I genes in the genome or expressing a dominant MHC class I gene (Codner *et al.* 2012; Doxiadis *et al.* 2011). Whilst this is usually associated with an increase in MHC class I variation, it is possible that DFT2 has exploited this system to reduce cell surface MHC class I to a single, common allele which is not as readily identified as foreign between hosts. It is also possible that the single binding motif reflects functional redundancy of the MHC class I system in Tasmanian devils. If devil MHC class I alleles all bind similar peptide repertoires, the minor histocompatibility antigen graft response is likely to be impaired, which may explain how both DFT1 and DFT2 have emerged and are able to spread throughout the devil population. Defining the specific binding motifs of the dominant MHC class I molecules SahaI*27, SahaI*27-1 and Saha-UK is an essential next step for vaccine design, and also for understanding the fundamental immunology underlying the emergence and spread of transmissible tumours in a wild species.

8.5 Conclusions and future directions

The data generated in this Chapter has given some unique insights into the peptide binding specificities of MHC class I in the context of DFT2, demonstrating a unique peptide length preference for 8mers over 9mers and a non-canonical P3 anchor residue alongside a seemingly single dominant motif from a multi-allele dataset. These characteristics have also previously been identified in a devil fibroblast immunopeptidome (Gastaldello *et al*. in preparation), indicating that these are species specific MHC class I binding characteristic rather than tumour specific. These data may indicate some of the mechanisms behind the emergence and spread of transmissible tumours in an endangered species, something which is still not entirely understood.

The next steps will be to recharacterise and reoptimise the DFT2_UC cell line in order to generate an allele specific immunopeptidome for this biologically interesting MHC class I molecule. This will allow us to define the possible reasons for a single dominant allele in pan-MHC class I datasets and is also essential for moving forward with vaccine target selection. Any appropriate vaccine targets identified from this will then need to be screened *in vitro* by tetramer assays to determine whether they are immunogenic and likely to raise an immune response in wild devils.

Chapter 9 General discussion

9.1 DFT2 emerged from a similar cell type to DFT1

Prior to the turn of the 21st century, there was only a single tumour known to be able to spread as an allograft between genetically distinct and MHC discordant hosts. Canine Transmissible Venereal Tumour (CTVT) is an ancient parasitic tumour lineage which emerged in an ancient dog breed or wolf around 11000 years ago from a myeloid cell (Murchison *et al.* 2014). In the last 20 years, a further seven independent tumour lineages with the ability to spread as allografts have been identified; DFT1, DFT2 (Pye *et al.* 2016b; Pearse and Swift 2006) and five lineages of neoplasia circulating in four species of marine bivalves (Metzger *et al.* 2016; Metzger *et al.* 2015). The major difference between the tumours identified in bivalves and DFTs is that whilst the tumours circulating in bivalves have been understood for a number of years and were only identified as transmissible recently, there is no evidence for DFT1, DFT2 or any other transmissible tumour in the Tasmanian devil prior to the initial discovery of these tumours (Hawkins *et al.* 2006). Tasmanian devils are the only mammalian species to harbour two transmissible tumour lineages, and all evidence indicates that these tumours have emerged independently with no detectable exogenous cause and had likely not existed in the population long prior to their first discovery (Stammnitz *et al.* 2018; Pye *et al.* 2016b). The independent emergence of two transmissible tumours in a single mammalian species within 20 years of each other was an unprecedented event and has raised concerns that the devil may be prone to forming transmissible tumours. Recent long-term modelling of the disease has indicated that DFT1 is unlikely to result in the extinction of the devil (Wells *et al.* 2019), but these models do not consider the emergence of DFT2, and a second widespread tumour may prove an unmanageable disease burden for the remaining wild population.

DFT2 has been shown to differ from DFT1 in several ways, including being PRX negative, possessing a detectable Y chromosome (Pye *et al.* 2016b) and the expression of functional cell surface MHC class I (Caldwell *et al.* 2018), along with somatic mutations which have been described in detail by Stammnitz *et al.* (2018). Despite these differences, the two tumours share an indistinguishable gross phenotype, spread between animals in the same way and have broadly similar expression of some neuroectodermal markers (Stammnitz *et al.* 2018; Pye *et al.* 2016b), resulting in speculation that DFT2 may have emerged from a similar cell type to DFT1, a Schwann cell (Murchison *et al.* 2010) though this had not been further investigated.

In this project, I have performed a deep analysis of protein expression by a DFT2 cell line, making direct comparisons of protein expression profiles of DFT2 and DFT1, using a fibroblast cell line as a 'healthy' control. This analysis has shown that DFT2 is functionally more similar to DFT1 than it is to fibroblasts. Crucially, I have identified a significant enrichment in the DFT2 proteome for proteins which are overexpressed in or specific to the nervous system. These include component

proteins of the myelin sheath, a highly specialised cellular compartment which found only in two cell types; Schwann cells in the PNS and oligodendrocytes in the CNS (Snaidero and Simons 2014). The expression of myelin components such as MPZ, MAG and MBP has been identified at a protein and transcript level in both DFT2 cell lines and primary tumour biopsies. I have also identified the expression of specific markers of both oligodendrocyte and Schwann cell lineages in the DFT2 proteome including PDGFRA, CSPG4, FABP7 and L1CAM, three of which have been additionally validated on a transcript level in three independent DFT2 cell lines and several primary tumour biopsies. This analysis has strongly indicated that DFT2 has emerged from a myelinating glial cell type, similar to DFT1.

Schwann cells and oligodendrocytes are similar in protein expression and biological function, but are two distinct cell types which derive from different regions of the developing neuroectoderm (Liu and Cheung 2016; Danesin *et al.* 2006) and myelinate axons in distinct manners (Salzer and Zalc 2016) (Figure 9.1). Discriminating between them with regards to the DFT2 progenitor is a difficult process relying on the expression of a few markers. DFT2 expresses markers such as MPZ that indicate a Schwann cell expression profile (Kilfoyle *et al.* 2006) and markers such as PDGFRA and CSPG4 which indicate an oligodendrocyte lineage (Tripathi *et al.* 2010). Schwann cell specific PRX (Gillespie *et al.* 1994) and oligodendrocyte marker Olig2 (Wegener *et al.* 2015) were not detected in the DFT2 proteome, and primers against central myelin genes MOG and PLP1 were inconclusive (Peschl *et al.* 2017; Morell and Quarles 1999). The lack of devil oligodendrocyte samples to compare DFT2 to has also hindered the definition of the DFT2 progenitor state.

Thus far, the data generated in this project indicates that DFT2 likely emerged from a less differentiated cell type to its DFT1 counterpart, demonstrated by the reduced expression of myelin specific proteins and genes (MBP, MAG, MPZ) in comparison to DFT1, and an increased enrichment for early glial markers such as FABP7, L1CAM and PDGFRA. DFT2 appears more enriched at a proteomic and transcriptomic level for markers of oligodendroglial lineage rather than Schwann cell, demonstrating expression in the proteome for specific oligodendrocyte markers CSPG4 and QKI. In addition, the response of DFT2 to IFNγ may indicate a differentiation signature rather than a dedifferentiation signature, a response more indicative of an oligodendrocyte precursor cell (OPC) than a Schwann cell. There is mounting evidence that OPCs are able to differentiate into myelinating Schwann cells following spinal cord injury or disease induced demyelination (Assinck *et al.* 2017; Zawadzka *et al.* 2010), and the expression of Schwann cell specific myelin protein MPZ by DFT2 combined with a predominantly OPC phenotype may be indicative of this differentiation occurring within the tumour as the phenotype is affected by tumourigenesis. The DFT2 origins require further definition, and in the future this may be possible through the comparison of glial marker expression across a larger panel of devil tissues, including brain and oligodendrocyte samples, on an mRNA and protein level, and additionally through co-culture of the tumour cells with neurons in order to induce myelin production, where it may be possible to identify the distinct myelination patterns of Schwann

cells and oligodendrocytes by the tumours, as well as the composition of any functional myelin produced (Hyung *et al.* 2015; Pang *et al.* 2012) (Figure 9.1).

Whilst further definition of the cellular origins of DFT2 must be undertaken, this project has shown evidence that DFT2 has emerged from a myelinating glial cell type, and at present an oligodendrocyte precursor appears most likely. The independent emergence of two transmissible tumours from two highly similar cell types may indicate a propensity for devil myelinating glial cells to form tumours which can evade immune detection, a concerning prospect for the future of the species. Indeed, the work presented here indicates that Tasmanian devil Schwann cells are negative for classical MHC class I, and may have an immune evasive MHC class I phenotype which is primed for the formation of transmissible tumours. However, it may also indicate that these tumours have undergone similar mutagenic processes (as has already been indicated in Stammnitz *et al.* (2018)) and may be expressing the same Tumour Specific or Tumour Associated antigens (TSA/TAA). If these two similar tumours do share any TSAs or TAAs, it may be possible to develop a single vaccine that can raise an effective immune response against both tumours. This would be significant step forward for conservation of the remaining wild population and would greatly reduce the costs and manual labour required for a large scale vaccination programme.

Broad vaccination strategies in a wild species against a pathogen of this nature carry some risks and difficulties both for vaccine development and ultimately vaccine delivery. There is significant evidence that peptide vaccines against targets which are not highly specific can induce severe autoimmune responses in patients (Sultan *et al.* 2017), and it will be critical to the success of the vaccine programme that any TSA or TAA identified as a potential target is validated for safety to ensure no severe autoimmune reactions occur in vaccinated devils. Additionally, vaccination has been shown to exert a strong selective pressure on pathogens resulting in the fast evolution of resistant strains which may downregulate the expression of the target antigen (Chong and Ikematsu 2018; Xu *et al.* 2015), and we would expect that DFT1 and DFT2 may undergo the same evolution in response to a wide vaccination programme. Thus, it is important to consider targeting the vaccine against multiple antigens to reduce the ability of the tumour to evolve resistance. Previously identified potential therapeutic targets such as PD-L1 inhibitors and heat shock proteins (Tovar *et al.* 2018; Flies *et al.* 2017; Flies *et al.* 2016) could be combined with suitable peptide targets to ensure a robust, multi-epitope approach to prevent the continued spread of the tumours whilst reducing the selective pressure which results in resistant and likely more aggressive subclones. Understanding the cellular origins of these two tumours will now allow for direct comparison of antigens between the tumour cells and their healthy progenitor cell types to allow to robust selection of neoantigens as vaccine candidates, which will hopefully reduce the risk of off-target autoimmune effects and allow for the selection of multiple candidates.

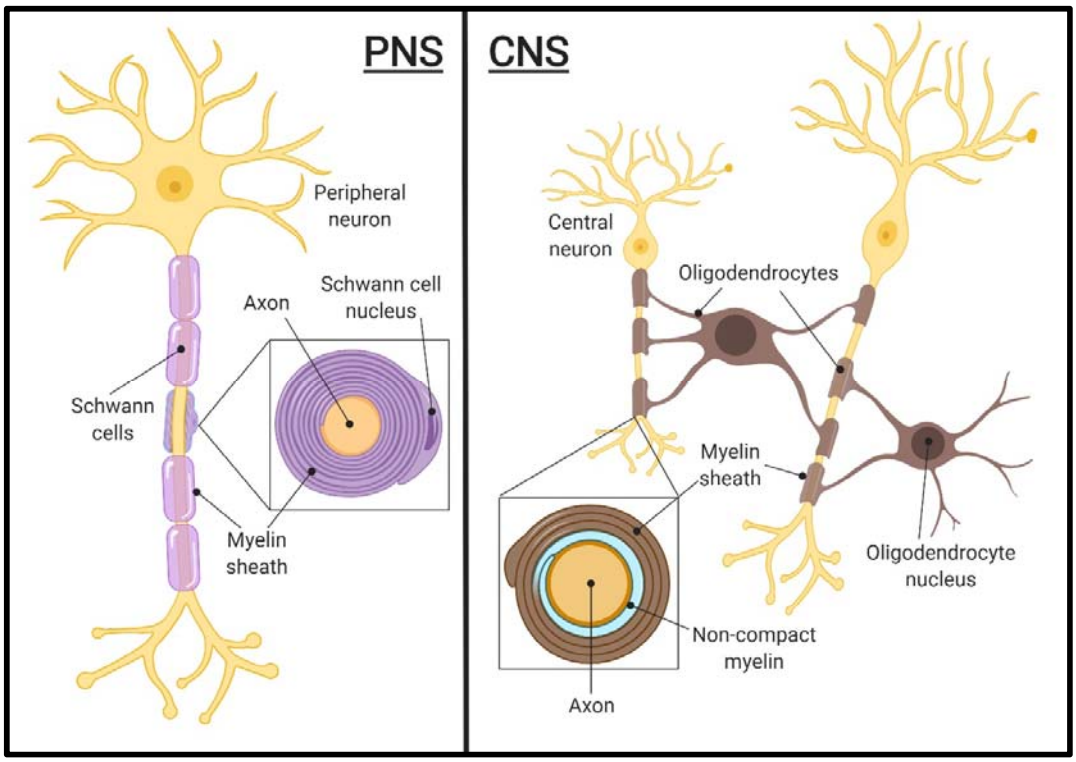


Figure 9.1: Schwann cells and oligodendrocytes myelinate axons in distinct manners.

Axons are presented in orange. Schwann cells form the peripheral (PNS) myelin sheath and are presented in purple. Oligodendrocytes form the central (CNS) myelin sheath and are presented in brown. Light blue indicates non-compact myelin which forms at the growth cone as the oligodendrocyte cytoplasm extends. Each Schwann cell wraps its entire cytoplasm around a single axon, whilst oligodendrocytes extend processes to myelinate multiple axons at multiple points. Adapted from Salzer and Zalc (2016). Created with BioRender.com.

9.2 DFT2 is a rapidly evolving tumour

DFT1 was first described in 1996 but was not sampled until 2001, and regular sampling and analysis of the tumour was not initiated until 2004 (Hawkins *et al.* 2006). This has left at least an 8 year black hole in sampling of the early years of DFT1, and by the time analysis was regularly occurring dominant subclones of the tumour had likely already emerged and stabilised. In contrast, DFT2 has been regularly sampled since its first description, which genetic analysis indicates is likely to be very close to the initial emergence of the tumour (Stammnitz *et al.* 2018; Pye *et al.* 2016b). This means that we can track the early evolutionary changes associated with a tumour that is under extreme immune pressure as it spreads between hosts. This is not only useful information for planning disease management strategies in the wild species, but also informs the fundamental biology behind the emergence and spread of transmissible tumours and the mechanisms tumours use to avoid the immune system.

It is possible that DFT2 is at present a rapidly evolving tumour. There is evidence that subclones in female hosts can lose the Y chromosome in response to immune pressure (Stammnitz *et al.* 2018), and evidence of highly variable expression of MHC class I across primary DFT2 tumours which may be reducing as the tumour encounters more genetically distinct host devils (Caldwell *et al.* 2018). In this project, I have further demonstrated the variability between primary DFT2 tumours on a panel of genes and proteins associated with the myelinating glial origins of the tumour. On both a transcript and protein level, different DFT2 tumours express variable levels of these markers, and there is also heterogeneity within tumour samples. This data has demonstrated that the most stable expression of all markers assessed appears to be in the tumour samples which have been circulating in the population for the longest, 812_T1 and 818_T1. Interestingly. 812_T1 is a tumour previously identified as being negative for classical MHC class I protein (Caldwell *et al.* 2018), and this may indicate that this tumour is becoming a dominant subclone.

9.3 DFT2 creates an immunosuppressive microenvironment to evade immune destruction

Whilst initially thought to be due to low genetic diversity at the MHC class I locus, DFT1 has been shown to avoid immune detection between hosts by epigenetic downregulation of surface MHC class I expression (Siddle *et al.* 2013; Siddle *et al.* 2007a). This downregulation is reversible by treatment of DFT1 cells with the inflammatory cytokine IFNγ (Siddle *et al.* 2013). It has also been shown that increasing MHC class I on the cell surface of DFT1 increases the ability of the host devil immune system to detect and clear DFT1 cells (Tovar *et al.* 2017). In contrast, DFT2 expresses cell surface MHC class I, and currently appears to circulate in devils with similar MHC class I genotypes (Caldwell *et al.* 2018). Tasmanian devils are able to reject skin grafts even in the case of few to no MHC class I mismatches, indicating an ability to respond to minor histocompatibility antigens (Kreiss *et al.* 2011), and evidence indicates that CD3+ lymphocytes infiltrate into some DFT2 tumours (Caldwell *et al.* 2018), indicative of an immune response to the tumour, though it does not appear to prevent tumour growth.

It is clear that DFT2 is employing distinct methods of immune evasion to DFT1, allowing it to spread between individuals in the presence of cell surface MHC class I. In this project I have assessed how DFT2 responds to the inflammatory cytokine IFNγ, defining how MHC class I expression is modulated by inflammation in the context of DFT2 and assessing the transcription of a panel of genes associated with tumour immune evasion both pre and post IFNγ to identify potential mechanisms of immune evasion that are unique to DFT2.

This work has demonstrated that DFT2 cells upregulate surface MHC class I in response to IFNγ significantly more than DFT1 cells, indicating this may be a useful mechanism to exploit during vaccine design and development. Transcript analysis by RT-PCR also demonstrated that DFT2 upregulates both classical and non-classical MHC class I in response to IFNγ, similar to DFT1. Assessment of a small panel of immune associated genes has demonstrated some striking differences between DFT1 and DFT2 which may highlight how DFT2 evades the host immune response. DFT2 demonstrates significantly higher expression of the immunosuppressive cytokines TGFB1 and TGFB2 than DFT1. The expression of both cytokines is unaffected in DFT2 by IFNγ treatment, whilst TGFB2 expression is increased in DFT1 following IFNγ treatment. TGFB signalling has been implicated in tumour immune evasion and impaired cytotoxic T-cell responses (Oh and Li 2013; Ouhtit *et al.* 2013; Donkor *et al.* 2012), and the constitutive expression of these cytokines by DFT2 is indicative of an immunosuppressive profile. Additionally, DFT2 shows significantly reduces levels of the IL-1 receptor IL1R1 both pre and post IFNγ treatment compared to DFT1. This may indicate that IL-1 signalling in DFT2 is attenuated, thus the ability of the host immune system to induce proinflammatory cytotoxic signals through the tumour cells is impeded (Lachmann *et al.* 2009). DFT2 also strongly upregulates PD-L1 in response to IFNγ in a comparable manner to DFT1, a

widely described mechanism of tumour immune evasion which prevents inflammation induced death of tumour cells (Gato-Cañas *et al.* 2017).

Combined, this data paints a picture of a tumour which has impeded pro-inflammatory signalling and is producing an immunosuppressive environment to avoid destruction by host immune systems. The constitutive expression of TGFB1 and 2 is particularly striking, given the well described role of TGFB signalling in the prevention of antigen specific T-cell responses (Tinoco *et al.* 2009; Bruijnzeel-Koomen *et al.* 2003; Gilbert *et al.* 1997), and may reflect a key mechanism behind DFT2s ability to evade host immune responses despite expressing surface MHC class I. Further analysis of immune signalling in DFT2, in particular through the TGFB and IL-1 pathways needs to be carried out to fully understand how this tumour is effectively avoiding host immune responses as it spreads, and to identify mechanisms which can be exploited in the production of an effective vaccine against DFTs. Additionally, further comparison of the immune cytokines and signalling pathways active in DFT1 and DFT2 will help to identify the shared and unique mechanisms of immune evasion which have contributed to the spread of these tumours.

9.4 Defining the peptide binding properties of SahaI*27-1

*SahaI*27* and *SahaI*27-1* are two classical MHC class I alleles which are present in both DFT1 and DFT2, have been identified in DFT2 host devils and are common in at least one devil population (Caldwell *et al.* 2018; Lane *et al.* 2012) and are predicted to bind a highly similar peptide repertoire. The prevalence of these two alleles is of biological interest in the context of DFTs. Reduced MHC class I diversity has been linked with a reduced ability within a species to respond to pathogens (Maibach and Vigilant 2019) and reduced diversity of the T-cell receptor (TCR) (Cheng *et al.* 2019; Migalska *et al.* 2019), both of which make a species vulnerable to infectious disease and has been postulated to have facilitated the emergence and spread of transmissible tumours in the Tasmanian devil (Cheng *et al.* 2019; Caldwell *et al.* 2018; Siddle *et al.* 2007a). DFTs evade both the strong T-cell anti-graft immune response raised against mismatched MHC class I alleles (Takemoto *et al.* 2000; Sherman and Chattopadhyay 1993) and the slower rejection associated with recognition of polymorphisms in the bound peptide sequences known as minor histocompatibility antigens (Pierce *et al.* 1999; Wang *et al.* 1995). The functional redundancy of these two seemingly dominant major classical MHC class I alleles may also be playing a role in the spread of DFT2.

Characterising the peptide sequences binding these MHC class I molecules and identifying whether they are immunogenic is important in understanding the underlying mechanisms of the emergence of tumours which transmit in the presence of MHC class I. Additionally, the binding properties of MHC class I in the devil are important for effective vaccine design, as peptide vaccines are haplotype restricted and any peptide target useful in a large scale vaccination programme must be capable of binding MHC class I in the wider devil population (Harper *et al.* 2015) (Figure 9.2). The presence of these alleles in both tumours combined with their similar cellular origins raises the possibility that these alleles may be binding TSAs or TAAs which are common to both tumours, which could allow for the production of a single prophylactic vaccine against both tumours.

In this project, I have successfully generated a DFT2 cell line expressing recombinant *SahaI*27-1* with a cytoplasmic myc tag which can be used for immunoaffinity and analysis of the peptide binding properties of this biologically interesting MHC class I allele. I have also demonstrated that this construct interacts with endogenous β_2-m and successful expression of the construct is accompanied by a detectable increase in cell surface β_2-m, indicating the recombinant construct is folding and being transported to the cell surface. A second cell line expressing the monomorphic, widely expressed non-classical allele *Saha-UK* has also been generated, and is in the process of being analysed by a Masters student. This has been the first transfection of DFT2 cells to my knowledge, and the robust transfection pipeline developed here can be used for further studies on antigen presentation in DFT2 cells. This could include interaction and localisation studies of components of the antigen presentation pathway in Tasmanian devils, something which has never been fully defined.

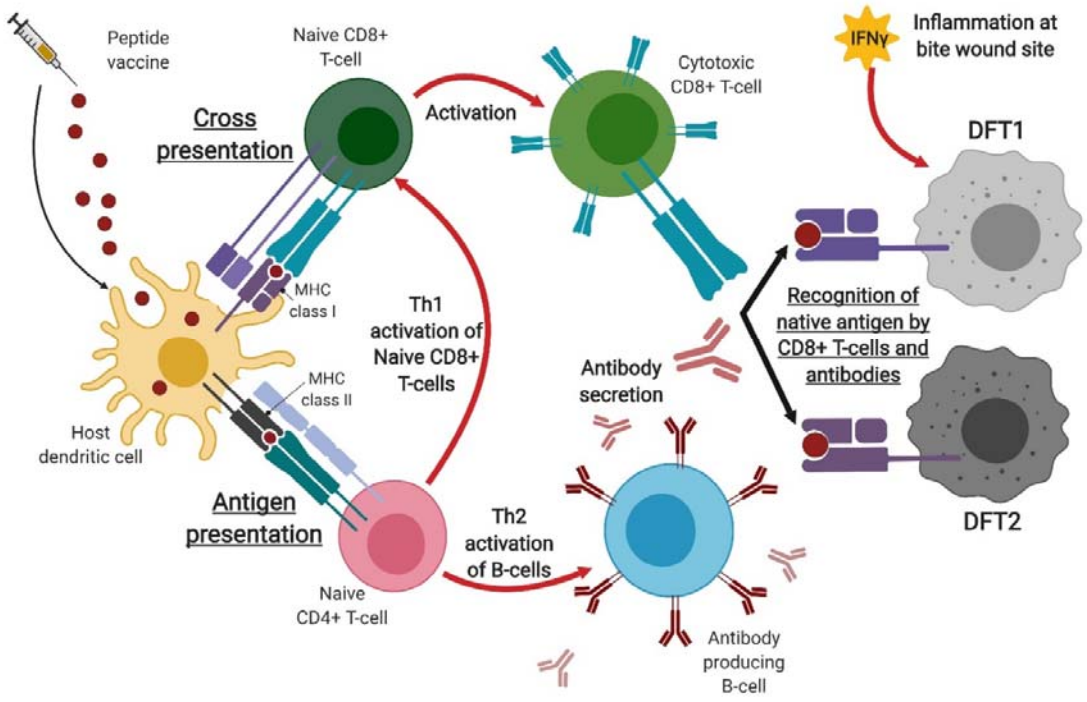


Figure 9.2: Schematic of peptide vaccine induced immunity against DFTs.

Inoculation with an identified TSA or TAA can induce immune responses against the antigen through both cell-mediated and humoral mechanisms. Target antigens are taken up by host dendritic cells where they are presented on MHC class II to naive CD4+ T-cells, resulting in activation. Activated CD4+ T-cells will proliferate and undergo differentiation into Th1 or Th2 type T-cells depending on the cytokines in the extracellular environment. Th1 type T-cells will subsequently activate naïve CD8+ T-cells against the target antigen. This triggers a cell-mediated cytotoxic immune response against DFTs through mature cytotoxic CD8+ T-cells with TCRs which recognise the target antigen. Th2 type T-cells activate B-cells against the target antigen, resulting in the production and release of antibodies which can neutralise cells presenting the antigen. Additionally, the inoculated antigen can be cross presented directly onto MHC class I following endocytosis by the host dendritic cell, directly activating naïve CD8+ T-cells and eliciting a cell-mediated cytotoxic response against tumour cells. Adapted from Skwarczynski and Toth (2016). Created with BioRender.com.

The recombinant cell line developed in this project is for the purpose of generating an allele specific immunopeptidome for SahaI*27/SahaI*27-1. From this, we can generate a binding motif which can be used to scan previously generated immunopeptidomes for DFT2, DFT1 and fibroblasts (Gastaldello *et al*. in preparation) to predict which experimentally determined peptides are able to bind SahaI*27-1, in turn informing the selection of appropriate potential vaccine targets for further study. Unfortunately, the replicates generated during the confines of this project for the SahaI*27-1 allele were unsuccessful. This was determined based on a combination of poor peptide yield and poor length distributions that did not have a strong enrichment for peptides of 8-12 amino acids in length, the canonical size restriction on MHC class I bound peptides (Trolle *et al.* 2016) which has been demonstrated by Gastaldello *et al.* (in preparation) in DFT2 pan-specific MHC class I immunopeptidome datasets. This experimental failure may be down to; lack of starting material, low levels of recombinant MHC class I, cytoplasmic retention of recombinant MHC class I or weak or otherwise unstable association with β_2-m and peptide. Recharacterisation of the recombinant SahaI*27-1 construct within the cell line needs to be carried out to identify the underlying problems behind this experimental failure.

Despite the limitations described above, this project has generated a pan-specific MHC class I immunopeptidome for the DFT2_UC cell line. The ability to generate an immunopeptidome using far fewer cells than the standard experimental procedure (Schuster *et al.* 2018; Wynne *et al.* 2016; Caron *et al.* 2015b) will be invaluable in further studies of peptide binding in the devil, significantly reducing the costs and time associated with the cell culture aspect of these experiments. The immunopeptidome generated in this project shares several characteristics with the DFT2_RV immunopeptidome generated by Gastaldello et al. (in preparation) including 24% of identified peptides, a high level of reproducibility given the use of different protocols and different equipment and the low reproducibility in immunopeptidomics data (Caron *et al.* 2015a; Michalski *et al.* 2011; Tabb *et al.* 2010). The immunopeptidome generated in this project demonstrates a canonical length distribution with a preference for 8mers over 9mers. The majority of MHC class I molecules in experimentally assessed in several species, including humans, mice and cows tend towards a length preference for 9mers (Nielsen *et al.* 2018; Trolle *et al.* 2016; Hansen *et al.* 2014; Rapin *et al.* 2008), indicating that the strong preference for 8mers in the Tasmanian devil is a species specific feature of peptide binding. Additionally, the immunopeptidome generated here has confirmed the presence of a P3 Leucine anchor residue in DFT2 MHC class I and a hydrophobic C-terminal anchor. This P3 anchor is an unusual feature of Tasmanian devil MHC class I peptide binding, as most experimentally determined MHC class I binding motifs have primary anchors at P2 and P9, and primary P3 anchor residues are a rare occurrence (Leone *et al.* 2013; Rapin *et al.* 2008). There are several human and mouse MHC class I alleles that have demonstrated an 8mer length preference and a P3 anchor residue, although the P3 anchor is rarely dominant (Gfeller *et al.* 2018; Bassani-Sternberg *et al.* 2017; Garstka *et al.* 2015). Interestingly, these binding preferences have been speculated in human MHC class I to relate to the role of HLA-C as an NK cell receptor (Gfeller *et al.* 2018; Bassani-Sternberg

et al. 2017), indicating that MHC class I in the context of DFT2 may be primarily geared towards interactions with NK cells. The presence of such strong anchors in a pan-specific MHC class I dataset indicates that either a single allele is dominant on the surface of devil cells, despite the presence of multiple alleles in the genome, or high similarity between MHC class I binding motifs in the devil, which may indicate a level of functional redundancy which could aid in the emergence and spread of the tumours.

This work has also identified three 8mer peptides derived from the neural specific protein NES, one of the most highly expressed proteins in the DFT2 proteome, which are detected in all replicates of the immunopeptidome generated here and in all replicates of the immunopeptidome generated by Gastaldello et al. (in preparation), indicating that these peptides are likely high affinity and stable MHC class I binders. Interestingly, none of these peptides contain a P3 Leucine residue, indicating that they may bind a less dominant MHC class I allele or that a P3 Leucine is not crucial for peptide binding stability. Further work must be done to characterise the binding affinity and specificity of these peptides, and their immunogenicity must be assessed, but they may present useful vaccine targets further down the line.

It is important to consider the fact that DFT2 is a fast evolving tumour whilst interpreting this work. Evidence presented here and by Caldwell *et al.* (2018) has demonstrated that DFT2 tumours are highly variable in protein and gene expression, and may be losing cell surface expression of MHC class I, indicating that the immunopeptidome of DFT2 may also vary dramatically between cell lines and tumour samples. A fibroblast immunopeptidome by Gastaldello *et al.* (in preparation) demonstrates the same 8mer length preference and P3 leucine anchor as DFT2, indicating that MHC class I binding properties and motifs are likely stable and consistent. However, it is highly likely that exact peptides identified during this work as potential targets may not be expressed by all DFT2 tumours, thus it is essential that candidate peptides are validated in different cell lines and tumours, and peptides should be selected from proteins with stable expression across DFT2 tumours in order to ensure the target peptides are likely to be widely expressed.

The immunopeptidome dataset generated here confirms unusual traits about MHC class I peptide binding in the devil and may also point to some of the underlying mechanisms which have resulted in the emergence of two independent lineages of transmissible tumour in the species in just 20 years.

9.5 Future directions: Towards a dual DFT vaccine

This PhD project has been part of a wider study working towards the identification of MHC class I peptides which could be useful targets for a vaccine against DFT1 and DFT2. The major finding of this PhD project is that DFT1 and DFT2 have emerged from a similar cell type, indicating that these two tumours may share neoantigens which could be targeted by a single peptide vaccine. Implementing a vaccination programme in a wild species is an expensive and labour intensive task, and a single vaccine that protects against both tumours is an attractive prospect to keep costs down. Additionally, our lab has demonstrated that an MHC class I allele *SahaI*27*, and the closely related *SahaI*27-1*, are highly expressed in both DFT1 and DFT2 and shared by host devils (Caldwell *et al.* 2018), and evidence indicates that *SahaI*27* is common in at least one wild devil population (Lane *et al.* 2012). The similar cellular origins of DFT1 and DFT2, combined with the shared expression of the *SahaI*27* and *SahaI*27*-1 alleles, indicate that non-self or altered peptides that bind to these alleles are promising targets for vaccine design. As *SahaI*27* and *SahaI*27-1* differ by a single non-synonymous mutation not predicted to affect peptide binding (Caldwell *et al*. 2018), we would predict that these alleles bind the same peptide sequences. Thus, I hypothesise that SahaI*27 and SahaI*27-1 may bind and present peptides that derive from mutated or over-expressed proteins on the surface of both DFT2 and DFT1+IFNγ cells. Indeed, this work has shown that at least three peptides presented on DFT2 cells derive from the neural specific protein NES which is highly expressed in DFT2 compared to both DFT1 and fibroblasts. Due to the shared cellular origins of DFT1 and DFT2, neoantigens may be shared by both cancers, and their binding to SahaI*27/SahaI*27-1 indicates they will be recognised by the immune systems of a large number of devils in the wider population. The next steps from this project will be working to identify potential vaccine targets and assessing their immunogenicity.

In the first instance, an immunopeptidome for SahaI*27-1 must be generated using the cell line whose generation is described in Chapter 7. The recombinant construct being expressed by this cell line must be recharacterise to ensure the construct is associating with β_2-m by performing immunoprecipitation experiments using an antibody against β_2-m and blotting for myc. Additionally, the cell line can be re-sorted by FACS to remove low β_2-m expressing clones from the heterogenous cell population, something which has been successful at increasing cell surface β_2-m in the cell line expressing recombinant Saha-UK (data not shown). DFT2_RV can also be re-transfected with the recombinant SahaI*27-1 construct if the recombinant heavy chain has stopped associating with β_2-m over time. Immunoaffinity purification using an anti-myc antibody can then be performed on a higher number of cells than performed here account for low expression of the construct. Additionally, if further problems arise with immunoaffinity and mass spectrometry in the DFT2_UC cell line, the construct can instead be transfected into an MHC class I null cell line, either DFT2_RV following CRISPR knockout of endogenous MHC class I genes (Saidulu *et al.* 2018) or into an MHC class I deficient non-devil cell line, such as LCL 721.221 (Kennedy *et al.* 2019; Shimizu and DeMars 1989)

from which the single allele of interest could be pulled down using an anti-β_2-m antibody. Whilst immunoaffinity purification from a non-devil cell line would not yield specific peptide targets of interest, it can still be used to generate a binding motif for SahaI*27-1 (and we would expect also SahaI*27). However, generation of an immunopeptidome from a devil cell line, and in particular a DFT cell line, is preferable due to the high physiological relevance of the specific peptide sequences detected, thus these options should be followed up first before transfection into a non-devil cell line.

Once an immunopeptidome is generated for SahaI*27-1 and a binding motif defined, the motif can be used in combination with pan-specific MHC immunopeptidome datasets generated in this project and previously by Gastaldello *et al*. (in preparation) to identify which peptides in DFT2, DFT1 (+IFNγ) and fibroblasts are predicted to be binding SahaI*27-1 and SahaI*27 (Jurtz *et al.* 2017; Nielsen and Andreatta 2016). The fibroblast immunopeptidome will be used as a control dataset to identify SahaI*27-1 peptides unique to DFT1 and DFT2, which are potentially neoantigens. In addition, using the custom peptide identification database created by Alistair Bailey (in preparation) which combines fibroblast genomic data with transcriptomic data from DFT1 and DFT2, we can identify novel peptides which originate from mutated proteins in the tumour cells, with a specific focus on peptides presented by the cancers which are derived from proteins unique to or overexpressed in DFT1 and DFT2 thus are less likely to be presented in other body tissues. Using the mass spectrometry data, we can also identify peptides which have unique post-translational modifications in DFT1 and DFT2 which are either not modified or have different modifications in fibroblasts, which can be targeted by a vaccine (Silva *et al.* 2013). By scanning the previously generated immunopeptidomes for peptides binding SahaI*27-1 in DFT2, DFT1 (+IFNγ) and fibroblasts, we can select candidate peptides which are unique to one or both of the tumours, by sequence or by modification which may be useful as vaccine targets.

Following candidate peptide selection, the validity of these peptides must be confirmed by re-folding the MHC class I complex with β_2-m and the peptide target of interest in a cell free system. This validates that the peptide identified experimentally is a real MHC class I restricted peptide and can also be used to experimentally assess the binding affinity of the peptide to the MHC class I molecule of interest, and the stability of the final MHC class I-peptide complex (van Hateren *et al.* 2017; van Hateren *et al.* 2013). Peptide binding affinity and stability data can not only be used to inform bioinformatic analysis and improve future peptide binding predictions (Bonsack *et al.* 2019) but is also an important factor in the immunogenicity of a given peptide, with high binding affinity and stability being a requirement for peptide immunogenicity, though high binding affinity alone does not guarantee high immunogenicity (Wang *et al.* 2016; Sette *et al.* 1994). Different amino acid and peptide identities activate T-cells to differing extents and ultimately some peptides may raise large immune responses whilst others effectively raise none, regardless of whether they are “foreign” (Meister *et al.* 2019; Rasmussen *et al.* 2016; Calis *et al.* 2013). Whilst an immune response to a peptide vaccine can be accentuated using adjuvants (Awate *et al.* 2013), an effective peptide vaccine candidate must also be capable of raising an immune response on its own.

Peptide candidates which have been validated and demonstrated to have a high binding affinity to SahaI*27-1 in a cell-free system can be screened for immunogenicity using MHC class I tetramers (Figure 9.3). This is an *in vitro* method of detecting the antigen specific cytotoxic T-cell responses which are difficult to detect *in vivo,* which can be used to directly visualise and quantify antigen-specific effector T-cells in blood samples (Altman *et al.* 1996). Briefly; MHC-peptide complexes are refolded with β_2-m and biotinylated. Fluorophore tagged streptavidin is added, resulting in four biotinylated MHC class I complexes bound by a single fluorescent streptavidin through their biotin tag. Tetramers can be made for any MHC-peptide complex of interest. Tetramers are then mixed with the cell population of interest, which in this case will be devil blood samples, where antigen-specific T-cells will bind to fluorescent tetramers through their TCR. Samples can then be analysed by flow cytometry, and the number of antigen-specific T-cells in a blood sample can be quantified (Sims *et al.* 2010). Tetramer staining is a powerful method of assessing antigen specific immune responses, which is becoming more suitable for large scale analysis of candidate peptides (Luimstra *et al.* 2018) and has more recently been demonstrated as effective at defining immune responses in tissues, something which could be useful for looking at immune cells in DFT tumour samples (Fehlings *et al.* 2018; Li *et al.* 2017). Peptides which are identified as high affinity and immunogenic will eventually be screened *in vivo* by immunisation of devils and detection of antibody responses. We would hope that any immunogenic candidates identified from this would could be fed directly into the current vaccination studies (Reviewed in Section 1.4.7) and combined with other current potential therapeutic targets in DFT1 such as PD-L1 (Flies *et al.* 2016) and heat shock proteins (Tovar *et al.* 2018) to induce strong and specific anti-tumour immune responses post-vaccination. We would hope that any high immunogenicity candidates we find will aid in the development of a specific, prophylactic vaccine against both DFT1 and DFT2, which could be used to prevent the widespread transmission of DFT2 and to stablise local populations.

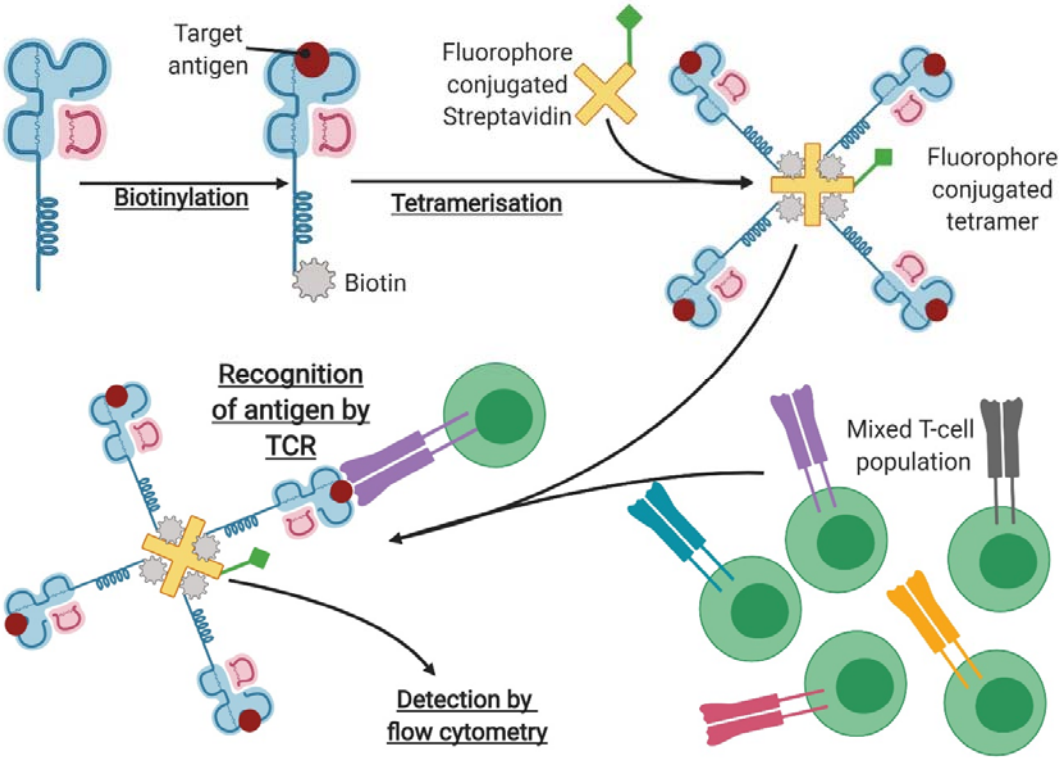


Figure 9.3: Schematic diagram of an MHC class I tetramer assay to assess peptide-MHC immunogenicity.

MHC class I-peptide complexes are biotinylated and conjugated into tetramers using streptavidin conjugated to a fluorophore. Tetramers are incubated with mixed T-cell populations (whole blood or mixed lymphocytes) and TCRs which recognise the target antigen bind to the tetramers. Interaction of T-cells with tetramers can be assessed and quantified by flow cytometry using the conjugated fluorophore for detection. Adapted from Sims *et al.* (2010). Created with BioRender.com.

9.6 Conclusions

Despite recent improvements in predicted DFT1 outcomes, the Tasmanian devil remains vulnerable to the effects of two transmissible tumours circulating within the wild population, and it is currently unknown what effect the widespread transmission of the newly emerged DFT2 could have on the remaining species. A prophylactic vaccine to support local population survival is an attractive prospect for conserving the species in the wild and would be particularly crucial along the DFT2 disease front. Additionally, the emergence of two transmissible tumours in a single species is unprecedented and is a unique opportunity to study the fundamental biology of tumour immune evasion and metastasis.

The data presented in this thesis has demonstrated that DFT1 and DFT2 have arisen from a similar cell type. I have demonstrated that DFT2 is a fast evolving, highly variable tumour, but that it expresses components of the myelin sheath and specific myelinating glial markers both *in vitro* and *in vivo*, indicating the tumour has arisen from either a Schwann cell or an oligodendrocyte, similar to the Schwann cell derived DFT1. Additionally, I have demonstrated that DFT2 upregulates MHC class I in response to IFNγ in a similar manner to DFT1, indicating that it is possible to modulate the immunogenicity of this new tumour, a crucial aspect of vaccine design. This data has also indicated that DFT2 may be employing distinct methods of immune evasion to DFT1, and the newer tumour shows an immunosuppressive cytokine profile at the transcript level which may indicate a mechanism for how the tumour spreads in the presence of MHC class I. Further, I have demonstrated that devil sciatic nerve and likely Schwann cells have a potentially immune evasive MHC class I phenotype, something which may have aided in the emergence of transmissible tumours.

The data presented here demonstrates an unusual peptide length preference for 8mers and a P3 Leucine anchor in devil MHC class I, which appear to be species specific features. This data also indicates either a functional redundancy of MHC class I in the Tasmanian devil, or the presence of a single highly dominant allele on the cell surface, which may be key to the emergence and spread of two transmissible tumours in the species. A neural specific protein highly expressed in DFT2 has been indicated as a source protein for peptides which are highly reproducible across immunopeptidome datasets, which may be interesting candidate peptides for further analysis. Additionally, this work is the first transfection of DFT2 cells to our knowledge and has generated two independent and stable DFT2 cell lines overexpressing recombinant MHC class I alleles of biological interest, which can be used for allele specific immunopeptidome studies as well as deeper analysis of the antigen processing and presentation pathway in Tasmanian devils.

The data presented in this thesis has provided some of the required knowledge to move the DFT vaccination programme forward, including cellular origins and peptide binding characteristics, and has generated the tools needed to identify specific MHC class I allele binding properties. Additionally, it has identified some potential underlying mechanisms for how these two tumours, in particular DFT2, have emerged and are able to spread throughout the wild devil population.

Appendix A List of supplementary figures and tables

A.1 List of supplementary figures

A.2 List of supplementary tables

Appendix B Chapter 2

B.1 RStudio code for data normalisation, imputation and visualisation

This example code will use "Unmodified_proteins.csv" (Appendix C.2), a dataset containing average protein expression in DFT1, DFT2 and fibroblasts. Protein expression and ontology was analysed in the same manner, and datasets used for analysis are presented in Appendix C.2 and Appendix C.3. Text in **bold** indicates functional R code, text in *italics* indicates notes explaining the code which are not functional. Code was generated based on code written by Tony Lin (available at https://datascienceplus.com/proteomics-data-analysis-2-3-data-filtering-and-missing-value-imputation/)

1. Load data frame into RStudio for analysis and view the first few rows to check data is correct

> protein_dat <- read.csv("Unmodified_proteins.csv")

> head(protein_dat)

2. Use protein names as row names

> rownames(protein_dat) <- protein_dat$Protein

3. Create new columns in the data frame containing Log2 transformed data for protein expression in each cell line *(NB. for functional analysis data was -log10 transformed by replacing log2 in the code with -log10).*

> protein_dat$Log2_DFT2 <- log2(protein_dat$DFT2.Average)

> protein_dat$Log2_DFT1 <- log2(protein_dat$DFT1.Average)

> protein_dat$Log2_Fibs <- log2(protein_dat$Salem.Average)

4. Select only transformed data columns for analysis and visualisation

> protein_df <- protein_dat[,5:7]

5. Create a data filtering function to filter data by how many cell types a protein is expressed in.

> filter_valids = function(protein_df, celltype, min_count, at_least_one = TRUE) {

 # celltype = vector which dictates grouping

 # min_count = vector indicating the minimum number of valid values in each cell type for retention

 # at_least_one = TRUE indicates the row is kept if min_count is met for at least one cell type (The protein is expressed in at least one cell type)

 # FALSE = means min_count must be met across all conditions for retention (The protein is expressed in all three cell types)

> log2.names = grep("^Log2", names(protein_df), value = TRUE) *# Extract columns with Log2 in the name from the dataframe*

> cell.names = lapply("^Log2", *# Group column names by cell type*

> function(x) grep(x, log2.names, value = TRUE, perl = TRUE))

> cell.filter = sapply(1:length(cell.names), function(i) {

> prot_df1 = protein_df[cell.names[[i]]] *# Extract data columns*

> prot_df1 = as.matrix(prot_df1) *# Convert data frame to matrix for filtering command*

> sums = rowSums(is.finite(prot_df1)) *#Count how many valid values are present in each cell line (How many proteins are expressed in each cell line)*

> sums >= min_count[i] *# Calculate if min_count has been met*

> })

> if(at_least_one) {

> protein_df$KEEP = apply(cell.filter, 1, any)} else{

> protein_df$KEEP = apply(cell.filter, 1, all)}

> return(protein_df)} *# Prevents omission of data following filtering, filter results are instead added as a new column labelled "KEEP" which indicates whether proteins are to be kept in further analysis "TRUE" or removed "FALSE"*

6. Apply filtering to either keep all proteins for analysis (and imputation) or to only keep proteins detected in all three cell lines (no imputation needed)

> protein_df.F = filter_valids(protein_df, *#Create a new filtered dataframe called protein_df.F with only the proteins for further analysis (in this case, all proteins)*

> celltype = c("DFT2", "DFT1", "Fibroblasts"),

> min_count = c(1, 1), *# Set to c(3, 3), to keep proteins detected in all three cell lines*

> at_least_one = TRUE) *#Set to FALSE to keep proteins detected in all three cell lines*

7. Create a median centred data normalisation function

> median_centering = function(protein_df) {

> log2.names = grep("^Log2", names(protein_df), value = TRUE)

> protein_df[, log2.names] = lapply(log2.names,

> function(x) {

> LOG2 = protein_df[[x]]

> LOG2[!is.finite(LOG2)] = NA *# Excludes any missing values from calculation of the median*

> gMedian = median(LOG2, na.rm = TRUE)

> LOG2 - gMedian})

>return(protein_df)}

8. Apply normalisation function to the filtered dataset and check normalisation by visualising the data as a distribution

>prot_df.FN = median_centering(protein_df.F*) #Create a new object containing filtered, normalised data.*

#Create histograms for the data in each cell line to check the distribution

>hist(prot_df.FN$Log2_DFT2*)*

>hist(prot_df.FN$Log2_DFT1*)*

>hist(prot_df.FN$Log2_Fibs*)*

9. Define the data imputation function

> impute_data = function(protein_df.F, width = 0.3, downshift = 1.8) {

 # Imputation assumes that missing data following a normal distribution which is narrowed and downshifted from the main dataset (Proteins not detected in a cell line are likely not highly expressed)

> log2.names = grep("^Log2", names(protein_df.F), value = TRUE)

> impute.names = sub("^Log2", "impute", log2.names) *# Add a new column to the dataframe which indicates whether each value in each column is imputed or not*

> protein_df.F[impute.names] = lapply(log2.names, function(x)

!is.finite(protein_df.F[, x]))

> set.seed(1)

> protein_df.F[log2.names] = lapply(log2.names,

> function(x) {temp = protein_df.F[[x]]

> temp[!is.finite(temp)] = NA

> temp.sd = width * sd(temp[protein_df.F$KEEP], na.rm = TRUE) *# reduces the width of the standard deviation under the assumption missing values fit a narrowed distribution*

> temp.mean = mean(temp[protein_df.F$KEEP], na.rm = TRUE) -

> downshift * sd(temp[protein_df.F$KEEP], na.rm = TRUE) *#shifts the mean of imputed values down from the main values under the assumption that missing values fit a downshifted distribution*

> n.missing = sum(is.na(temp))

> temp[is.na(temp)] = rnorm(n.missing, mean = temp.mean, sd = temp.sd)

> return(temp)})

> return(protein_df.F)}

10. Impute missing values to the dataset so proteins missing from a cell line are given a value for cluster analysis

> prot_df.FNI = impute_data(prot_df.FN) *#Creates a new object containing filtered data which has been normalised and missing values imputed*

11. Hierarchical clustering of data and visualisation using the heatmap function

> prot_heatmap <- prot_df.FNI[,1:3] *#Select only columns with data in for cluster analysis*

> prot_heatmap_matrix <- data.matrix(prot_heatmap) *#Convert imputed data to a matrix for cluster analysis*

#Hierarchical clustering of rows and columns occurs within the heatmap function

> heatmap(prot_heatmap_matrix, Rowv = NULL, Colv = NULL, distfun = dist,

> hclustfun = hclust, col = heat.colors(100))

#Note that clustering of rows has no biological significance and is not included in interpretation of analysis

B.2 Primers used in the generation of this thesis

Table S 1: Details of all primers used during the generation of this thesis.

Primer pairs and expected fragment size are detailed in the notes and optimised temperatures columns. Primers marked with + were designed by Hannah Siddle. Primers marked with * were designed by Alison Caldwell. All other primers were design by Rachel Owen unless otherwise specified.

Primer	Sequence 5' - 3'	Direction	Tm	GC%	Notes	Optimised temperatures
Devmen F+	ATGGAGAATGTGGACCGGGAC	F	63	57	Saha-UD	60 w/ Primer Devmen R
Devmen R+	TGAGTTCACTGCCTCATTCACT	R	60	45	Saha-UD – 275 bp w/ Primer Devmen F	60 w/ Primer Devmen F
22+	TGTGCATCCTTCCCTACCTGGAGG	F	61	58	β_2-m	60 w/ Primer 23
23+	CATTGTTGAAAGACAGATCGGACCGC	R	60	50	β_2-m – 300 bp w/ Primer 22.	60 w/ Primer 22
85+	CCGTGGGCTACGTGGACGATCAGC	F	64	67	MHC class I (amplifies all loci in conserved region of exon2)	60 w/ Primer 86
86+	GCCCTCGTGCTGAACTCGGCAGG	R	64	70	MHC class I (amplifies all loci in conserved region of exon2)	60 w/ Primer 85
91+	GTCGTAGGCGAACTGAAG	R			Classical Class I - 296 bp w/ Primer 85.	60 w/ Primer 85
138a+	TGGTTGGACAAGAGTAA	F	48	41	Saha-UK	60 w/ Primer 140
140+	CCTCAGGAAGATCCAGTCGTAAGTC	R	67	52	Saha-UK - 190 bp w/ Primer 138a	60 w/ Primer 138a
192+	TCCACCACATTGCCCATCTTACTCCG	F	61	54	IL10	No success
193+	ACCACGCTGTTCTTGATCTCCG	R	55	57	IL10 – 229 bp w/ Primer 192	No success
256*	CACTATGGACAACTACCCCAAGC	F	62	52	EGR1 (KNOX-24)	62 w/ Primer 257
257*	TTGGGTAGCTTGTCTCCGATAGC	R	62	52	EGR1 (KNOX-24) – 265 bp w/ Primer 256	62 w/ Primer 256
258*	GAGCAGTTTGCTAAACCAACTGC	F	61	48	EGR3	No success
259*	AAGTAGGTCACCGTCTTGTTGC	R	60	50	EGR3 – 218 bp w/ Primer 258	No success
260*	CATTTTGGACCTGCAAATCGC	F	58	48	ID2	No success
261*	TAGCCACAGAGTGCTTTGC	R	57	53	ID2 – 191 bp w/ Primer 260	No success
262*	AAGATGACTGGAAGGTGGTCAGC	F	62	52	AXL	58 w/ Primer 263
263*	TTGGCAGCTGAGGTTGAACG	R	59	55	AXL – 205 bp w/ Primer 262	58 w/ Primer 262
264*	CTGGGAATCCAAACACAAGACATGC	F	63	48	MERTK	63 w/ Primer 265
265*	GCTTCTCAGTAACACGACTGC	R	60	52	MERTK – 518 bp w/ Primer 264	63 w/ Primer 264
268 (Murchison *et al*. 2010)	GGATCAAAGTACCTGGCAACTGCA	F	63	50	MBP	60 w/ Primer 269
269	CCTTGCCTGAGGCCCTCCTGGGCA	R	71	71	MBP – 144 bp w/ Primer 268	60 w/ Primer 268

(Murchison *et al.* 2010)						
270*	TCT TAA CTT GAC TGA GGC CG	F	57	50	S100B	No success
271*	AATACTGGTGGAAC GTCTCG	R	57	50	S100B – 97 bp w/ Primer 270	No success
272*	CGATTCGATTTTCCC GATGAGC	F	60	50	MAG	59 w/ Primer 273
273*	GTATTTGATGTCCA AGCTGGCG	R	60	50	MAG – 587 bp w/ Primer 272	59 w/ Primer 272
274 (Murchison *et al.* 2010)	CAATGTCCAGCACT GTTTCTCATCCTCTA	F	65	45	PMP22	65 w/ Primer 275
275 (Murchison *et al.* 2010)	GAACAGGAACAAG GAAAGGACGCTGAA	R	65	48	PMP22 – 105 bp w/ Primer 274	65 w/ Primer 274
276*	AACGTATGGGATAC AGCTGGC	F	60	52	RAN	58 w/ Primer 277
277*	TAGCAAGCCAAAGG AAGGGC	R	60	55	RAN – 312 bp w/ Primer 276	58 w/ Primer 276
278*	CTGGGCAAATGCAA GAGC	F	56	56	IL6	60 w/ Primer 279
279*	TATGGTGCTGCTAT TGTCGC	R	57	50	IL6 – 104 bp w/ Primer 278	60 w/ Primer 278
280 (Murchison *et al.* 2010)	CAACTACAAGACCC TGAAAAGTCAAGGA GA	F	65	43	STAT3	65 w/ Primer 281
281 (Murchison *et al.* 2010)	AGCATCTGTTCCTA GCTGCTGCATCTT	R	65	48	STAT3 – 100 bp w/ Primer 280	65 w/ Primer 280
282*	GTATTATTGGTTAC GGGGGAGC	F	60	50	CD274/PDL1	62 w/ Primer 283
283*	TGTGGCATTGACCC TGAGAGTGC	R	64	57	CD274/PDL1 – 259 bp w/ Primer 282	62 w/ Primer 282
284*	AAATTTGGCAACGG CTCCTGGCG	F	64	57	TGFB1/1	63 w/ Primer 285
285*	CGAAAATCGATGTA GAGCTGCCG	R	62	52	TGFB1/1 – 388 bp w/ Primer 284	63 w/ Primer 284
286*	CGACATGGATCAGT TCATGC	F	57	50	TGFB2	57 w/ Primer 287
287*	CCTTCACCAAGTTG GAAGC	R	57	53	TGFB2 – 352 bp w/ Primer 286	57 w/ Primer 286
288*	TTGCTGTTCGTCAC AGTAGC	F	57	50	IL1R1	57 w/ Primer 289
289*	ATTGCCAAGGAGAG GTTTGC	R	57	50	IL1R1 – 391 bp w/ Primer 288	57 w/ Primer 288
292*	CCTGGATTTCAGGT GTCTTTGC	F	62	50	TP53	64 w/ Primer 293
293*	ATCTTTCATGGTGG GGACAGC	R	61	52	TP53 – 307 bp w/ Primer 292	64 w/ Primer 292
296*	CGAGGTTTTCGGTG TTCACAAC	F	62	50	NOTCH1	No success
297*	ACATCGGCACTTGA AGTCAGTG	R	62	50	NOTCH1 – 331 bp w/ Primer 296	No success
307*	AGTTAGCTAGCGCC ACCATGGAACCTTT GAAGCTC	F	79	51	Full length SahaI*UK	60 w/ Primer 309 and Primer 317
309*	GAAATAAGTTTTTG TTCAGCTTTGGCTGT CAGAGAG	R	74	39	Full length SahaI*UK with NheI restriction site – 1000 bp w/ Primer 307	60 w/ Primer 307

310*	AGTTAGCTAGCGCCACCATGGGCTCTCCGGCGCGCGC	F	91	70	Full length Classical MHC class I	60 w/ Primer 312 and Primer 317
312*	GAAATAAGTTTTTGTTCTTTGGCTGTCAGAGAGAC	R	72	37	Full length Classical MHC class I with NheI restriction site – 1000 bp w/ Primer 310	60 w/ Primer 310
313*	AGTTAGCTAGCGCCACCATGGGCTCTTATGCCTGCTCTC	F	83	56	Full length SahaI*UD	59 w/ Primer 315 and Primer 317
315*	CTCTCTGACAGCCAAAGCTGAACAAAAACTTATTTC	R	71	37	Full length SahaI*UD with NheI restriction site – 1000 bp w/ Primer 310	59 w/ Primer 313
317*	ATAGCGAATTCTCTAGATCAGATCTTCTTCAGAAATAAGTTTTTG	R	73	31	Universal full length MHC class I reverse primer with c-terminal myc tag	60 w/ Primer 307 and Primer 310 59 w/ Primer 315
318	AACTGCCCACCATCAAGATGCC	F	56	55	PRX	56 w/ Primer 319
319	TTCTGGGATGGACACTTCAGGC	R	56	55	PRX – 300 bp w/ Primer 318	56 w/ Primer 318
320	CCAGCCTCAAACTGAGACAGC	R	56	57	Devil GAP43	59 w/ Primer 321
321	TGACTTGGGATCTTTCCTGC	R	52	50	Devil GAP43 – 347 bp w/ Primer 320	59 w/ Primer 320
322	CCTACGGATACTACCAAGACG	F	54	52	Devil NGFR	58 w/ Primer 323
323	GCCCAATAAAGGTGTGGTCCG	R	56	57	Devil NGFR – 286 bp w/ Primer 322	58 w/ Primer 322
324	GAGAGTGGACAAGAGTGATGAAGC	F	57	50	Devil NCAM1	59 w/ Primer 325
325	CGGTGCAGATGTACTCTCCG	R	56	60	Devil NCAM1 – 337 bp w/ Primer 324	59 w/ Primer 324
326	TCGCCTGAGCCAAATATCAGC	F	54	52	Devil NCAM2	No success
327	AGGAATGGGTTCTCCTTCTGC	R	54	52	Devil NCAM2 – 282 bp w/ Primer 326	No success
347	agcgaagaagatctgATCTAGAGAATTCGCTATATCTTTC	F	63	38	Forward primer for myc tag site directed mutagenesis	60 w/ Primer 348
348	aatcagtttctgttcCAGATCTTCTTCAGAAATAAGTTTTTG	R	62	31	Reverse primer for myc tag site directed mutagenesis	60 w/ Primer 347
353	CAGCACCAGGACACAGAGTTC	f	56	57	Primer for sequencing of myc tag in pIRES	N/A
354	GTGGTTTACACGGACAGGGAG	F	56	57	MPZ	57 w/ Primer 355
355	CTTTCCACCGAGGATCCCCAA	R	56	57	MPZ – 234 bp / Primer 354	57 w/ Primer 354
356	AGAAGTTGCAGAAGAGGAGG	F	52	50	APP	58 w/ Primer 357
357	TTGACAAGGACCAGTCTCAGC	R	54	52	APP – 222 bp w/ Primer 356	58 w/ Primer 356
358	ACGGTCAATACCCACTTTGG	F	52	50	NLGN3	59 w/ Primer 359
359	GGATATAGACCATGACTGGC	R	52	50	NLGN3 – 442 bp w/ Primer 358	59 w/ Primer 358
360	ACTGGCTACCCTCAACCTCAG	F	56	57	TPBG	60 w/ Primer 361
361	TCCAGGTGCTTGAGTTTAGGC	R	54	52	TPBG – 380 bp w/ Primer 360	60 w/ Primer 360
362	ATTCTAAACCTCAGGGCCAG	F	52	50	FABP7	57 w/ Primer 363

363	TCCTGATCACCACT TTGTCC	R	52	50	FABP7 – 243 bp w/ Primer 362	57 w/ Primer 362
364	ACAAGACCAAGCGA GGCCAGG	F	60	62	RPL13A	60 w/ Primer 365
365	GCCTGGTATTTCCA GCCAACCTCA	R	60	54	RPL13A – 183 bp w/ Primer 364	60 w/ Primer 364
368	TGGGCCACAACTTT GGAATGAG	F	55	50	ADAM19	No success
369	GTCACACTCTTCTCC ATCTTCC	R	55	50	ADAM19 – 253 bp w/ Primer 368	No success
370	TATGTCCCCTCTAC ACCATGG	F	54	52	RRN3	No success
371	CCTTTCAGCTGTATC TGTCCC	R	54	52	RRN3 – 269 bp w/ Primer 370	No success
372	ACACTACGGCCAAT TCACCTG	F	54	52	NR2F1	No success
373	TGGGTTGGAGGCAT TCTTCCTC	R	57	55	NR2F1 – 206 bp w/ Primer 372	No success
374	TTCCCCAGCAAGAC CTCAGC	F	56	60	PLP1	58 w/ Primer 375
375	GCCCATGAGTTTCA GGACAGC	R	56	57	PLP1 – 260 bp w/ Primer 374	58 w/ Primer 374
376	TTGTGCAACCCCGA ATCTACTG	F	60	50	L1CAM	62 w/ Primer 377
377	AAACCTTCAGGCAG GAGCAAGC	R	62	55	L1CAM – 252 bp w/ Primer 376	62 w/ Primer 376
378	GCATGTGCTGAACT CCACCATTG	F	62	52	THY1	58 w/ Primer 379
379	CTCATGACCCCACA GGTT TG	R	59	55	THY1 – 221 bp w/ Primer 378	58 w/ Primer 378
380	TACCCAGGATTACG GCAAGG	F	59	55	MRC2	No success
381	TTCTCCTCACTGGG GTTGTC	R	59	55	MRC2 – 366 bp w/ Primer 380	No success
382	CTCCAGAGCTTGAA GTTGAG	F	57	50	PDGFRA	58 w/ Primer 383
383	GTTTCCTTGGTAGC ATGACG	R	57	50	PDGFRA – 264 bp w/ Primer 382	58 w/ Primer 382
384	GGATCAAGATGGAG AGCAAG	F	57	50	MOG	61 w/ Primer 385
385	TTCCAGCAAGGCAC TCTCAG	R	59	55	MOG – 358 bp w/ Primer 384	61 w/ Primer 384
386	ACAGATATGGCGGA ATGTGC	F	57	50	NF1	63 w/ Primer 387
387	TGATGTAGGTGCTG GCTTTGC	R	60	52	NF1 – 285 bp w/ Primer 386	63 w/ Primer 386
392	ATGGGCTCCGTGGT CAAGTC	F	61	60	SOX2	58 w/ Primer 393
393	ACATGTGAAGTCTG CTGGGC	R	59	55	SOX2 – 165 bp w/ Primer 392	58 w/ Primer 392
394	TGAACGAGAGCGAC AAGAGG	F	59	55	SOX10	No success
395	GACTCTGCCTTGCTT GATTGC	R	60	53	SOX10 – 330 bp w/ Primer 394	No success

Appendix C Chapter 3

C.1 Non-quantified whole cell proteomes

AVAILABLE AT: https://doi.org/10.5258/SOTON/D0982

Table S 2: Raw, non-quantified whole cell proteome data for three replicates each of DFT2_RV, DFT1_4906 (+IFNγ) and fibroblast (Salem) cell lines.

Columns "ensembl_peptide_id", "ensembl_gene_id" and "name" contain the information on peptide (and hence, protein) identity. DFT1 replicate data is presented in the sheets titled "DFT1 Replicate 1", "DFT1 Replicate 2" and "DFT1 Replicate 3", DFT2 replicate data is presented in the sheets titled "DFT2 Replicate 1", "DFT2 Replicate 2" and "DFT2 Replicate 3", fibroblast replicate data is presented in the sheets titled "Salem Replicate 1", "Salem Replicate 2" and "Salem Replicate 3".

C.2 Quantified whole cell proteomes

AVAILABLE AT: https://doi.org/10.5258/SOTON/D0982

Table S 3: Raw data for the quantification of proteins from the whole cell proteomes of DFT2_RV, DFT1_4906 and fibroblast (Salem) cell lines.

All quantified proteins were detected in all three replicates of the given cell line of the proteome presented in Appendix 3A. Columns "x_1_intensity", "x_2_intensity" and "x_3_intensity" represent the raw expression values for a given protein, and "Average" represents the average expression values used for functional analysis. Columns "ensembl_peptide_id", "ensembl_gene_id" and "name" contain the information on peptide (and hence, protein) identity. The sheets titled "Protein_replicates.csv" and "Unmodified_proteins.csv" represent the data used to generate the heatmaps in Figure 3.5 and Appendix B.1. Triplicate data is presented for DFT1, DFT2 and devil fibroblasts in the sheets titled "Quantified and annotated DFT1", "Quantified and annotated DFT2" and "Quantified and annotated Salem" respectively. The sheet titled "Fibroblasts_Karbiener2017" contains proteomic data on sheep oral mucosa fibroblasts previously published in Karbiener *et al*. 2017 which was used as a control dataset during functional analysis. Proteins were ranked based on an average of the values presented in the columns titled "LFQintensity.OMF.BR1", "LFQintensity.OMF.BR2", "LFQintensity.OMF.BR3" and "LFQintensity.OMF.BR4". Sheets titled "Unmodified_proteins.csv" and "Protein_replicates.csv" represent the csv files used to generate the heatmaps in Figure 3.5 as described in Chapter 2.3.4 and Appendix B.1.

C.3 Functional analysis of full quantified whole cell proteomes

AVAILABLE AT: https://doi.org/10.5258/SOTON/D0982

Table S 4: Full results for functional analysis of the full, quantified proteomes of each cell line against the Gene Ontology: Biological process, Gene Ontology: Cellular compartment, Reactome and Human Protein Atlas databases.

Proteins were ranked by expression from highest to lowest using the average values presented in Appendix C.2. The "source" column indicates the database each enriched functional term is derived from, where GO:BP = Gene Ontology: Biological process, GO:CC = Gene Ontology: Cellular compartment, REAC = Reactome and HPA = Human protein atlas. "term_name" and "term_id" are the identifiers for each enriched process. "adjusted_p_value" is the significance of enrichment of each term, adjusted for multiple testing using the g:SCS algorithm (Chapter 2.3.4). "term_size" is the number of proteins in the database associated with a given process. "query_size" is the number of recognised proteins submitted for analysis. "intersection_size" is the number of proteins in the query which are associated with a given process, and these proteins are listed in "intersections". Functional analysis for DFT1, DFT2, devil fibroblasts and sheep fibroblasts (Karbiener *et al*. 2017) is presented in the sheets titled "DFT1 full functional analysis", "DFT2 full functional analysis", "Salem full functional analysis" and "SheepFibs_Karbiener2017" respectively. The sheets titled "Biological_process_full.csv", "Cell_compartment_full.csv", "Reactome_full.csv" and "Human_protein_ontology_full.csv" correspond to the data used to generate the heatmaps in Figure 3.6 as described in Chapter 2.3.4 and Appendix B.1.

C.4 Categorisation and comparison of full quantified whole cell proteome functional enrichment

AVAILABLE AT: https://doi.org/10.5258/SOTON/D0982

Table S 5: Cross cell line comparisons of the functional analysis presented in Appendix C.3.
Adjusted p-values from Appendix C.3 are used to compare functional analysis between cell lines. NA indicates a functional term was not identified as enriched in that cell line. GO:BP and GO:CC analysis is further annotated to identify terms which give cell-type specific information in order to identify DFT2s cell type of origin, as discussed in Chapter 3.3.2. Yellow cell highlight indicates functional terms identified as providing cell-type specific information in the context of this project. These terms have been further categorised according to the key. Sheet titles indicate the functional database used in analysis.

C.5 Functional analysis of proteins unique to a single cell line

AVAILABLE AT: https://doi.org/10.5258/SOTON/D0982

Table S 6: Full results for functional analysis of proteins unique to and quantified in a single cell line against the Gene Ontology: Biological process, Gene Ontology: Cellular compartment, Reactome and Human Protein Atlas databases.

Proteins were ranked by expression from highest to lowest using the average values presented in Appendix C.2. The "source" column indicates the database each enriched functional term is derived from, where GO:BP = Gene Ontology: Biological process, GO:CC = Gene Ontology: Cellular compartment, REAC = Reactome and HPA = Human protein atlas. "term_name" and "term_id" are the identifiers for each enriched process. "adjusted_p_value" is the significance of enrichment of each term, adjusted for multiple testing using the g:SCS algorithm (Chapter 2.3.4). "term_size" is the number of proteins in the database associated with a given process. "query_size" is the number of recognised proteins submitted for analysis. "intersection_size" is the number of proteins in the query which are associated with a given process, and these proteins are listed in "intersections". Functional analysis of proteins unique to DFT1, DFT2 and fibroblasts are presented in sheets titled "DFT1 Unique Funct. Analysis", "DFT2 Unique Funct. Analysis" and "Salem Unique Funct. Analysis" respectively.

C.6 Categorisation and comparison of functional enrichment in proteins unique to a single cell line

AVAILABLE AT: https://doi.org/10.5258/SOTON/D0982

Table S 7: Cross cell line comparisons of the functional analysis presented in Appendix C.5.
Adjusted p-values from Appendix C.5 are used to compare functional analysis between cell lines. NA indicates a functional term was not identified as enriched in that cell line. GO:BP, GO:CC and Reactome analysis is further annotated to identify terms which give cell-type specific information in order to identify DFT2s cell type of origin, as discussed in Chapter 3.1.2. Yellow cell highlight indicates functional terms identified as providing cell-type specific information in the context of this project. These terms have been further categorised according to the key. Sheet titles indicate the functional database used in analysis.

C.7 Functional analysis of tumour specific proteins

AVAILABLE AT: https://doi.org/10.5258/SOTON/D0982

Table S 8: Functional analysis of tumour specific and ubiquitous proteins.

A) **Proteins unique to DFT2, plus proteins quantified in DFT1 and DFT2 but not in fibroblasts, presented in sheet labelled "DFT2 not Salem Funct. Analysis". Referred to in text as DFT2 + tumour specific proteins.**
B) **Proteins unique to DFT1, plus proteins quantified in DFT1 and DFT2 but not in fibroblasts, presented in sheet labelled "DFT1 not Salem Funct. Analysis". Referred to in text as DFT1 + tumour specific proteins.**
C) **Proteins unique to fibroblasts, plus proteins quantified in all three cell lines, presented in sheet titled "Salem Unique All Cells Funct.". Referred to in text as Fibroblasts + ubiquitous proteins.**

against the Gene Ontology: Biological process, Gene Ontology: Cellular compartment, Reactome and Human Protein Atlas databases.

Proteins were ranked by expression from highest to lowest using the average values presented in Appendix C.2. The "source" column indicates the database each enriched functional term is derived from, where GO:BP = Gene Ontology: Biological process, GO:CC = Gene Ontology: Cellular compartment, REAC = Reactome and HPA = Human protein atlas. "term_name" and "term_id" are the identifiers for each enriched process. "adjusted_p_value" is the significance of enrichment of each term, adjusted for multiple testing using the g:SCS algorithm (Chapter 2.3.4). "term_size" is the number of proteins in the database associated with a given process. "query_size" is the number of recognised proteins submitted for analysis. "intersection_size" is the number of proteins in the query which are associated with a given process, and these proteins are listed in "intersections".

Appendix D Chapter 4

D.1 Schwann cell isolation protocol

Adapted from Stevens *et al.* (1998)

Slide and plate coating

A 3.5 cm cell culture dish or 8 well chamber slide was coated with poly-L-lysine (Sigma-Aldrich, Cat no. P8920-100ML) for two hours at 37 ºC. Plates/wells were washed with MilliQ water and coated overnight at 37 ºC with 1.5 µg/cm2 of Laminin (Sigma-Aldrich, Cat no. L2020) which was removed immediately prior to plating.

Dissection

Young adult mice between 1-9 months were killed by cervical dislocation by a trained and licensed individual. Sciatic nerves were immediately removed and placed into cold L-15 media (Sigma-Aldrich, Cat no. L1518) on ice. The perineurium of the nerves was removed under a dissection microscope to separate the axons within the nerve fascicle and expose the myelin sheaths. The clean nerves were placed into fresh L-15 media on ice.

Isolation

Nerves were mechanically disrupted using sterile tweezers in 0.5 mg/mL collagenase (Sigma-Aldrich, Cat no. C0130-100MG) in serum free DMEM and incubated at 37 ºC 5 % CO2 for 30 min. Nerves were triturated using a blue Gilson pipette tip and incubated for a further 10 min at 37 ºC, 5 % CO2. Nerves were triturated again with a yellow tip before addition of isolation media (DMEM (Gibco) containing 5 % heat inactivated horse serum (Gibco), 50 µg/mL penicillin, 50 µg/mL streptomycin (Gibco), 1 % N2 supplement (Sigma-Aldrich, Cat. no. 17502048), 5 mM forskolin (Sigma-Aldrich, Cat. no. F6886, originated from *Coleus forskohlii*)) to inactivate collagenase. Cells were centrifuged for 10 min at 300 g. Cells were resuspended in fresh isolation media and plated. Cells were incubated at 37 ºC and 5 % CO2 for 72 hours without being disturbed.

Culture conditions

After 72 hours, cells were gently washed with dPBS (no longer than 1 min contact time). Media replaced with growth media (isolation media + 100 µM dbCAMP (Sigma-Aldrich, Cat. no. D0627)). Cells could be maintained and expanded for up to 3 weeks. Populations were enriched for Schwann cells during splitting. Schwann cells and fibroblasts were visually identified by light microscopy (Figure S 1). Schwann cell detachment in TrypLE was visualised and the detachment was stopped by growth media addition when Schwann cells were detached but fibroblasts were not. Cells were centrifuged at 300 g for 10 min. Cells were resuspended in growth media and replated onto a laminin/poly-l-lysine coated plate.

Cells were preserved at -80 $_{o}$C by resuspending in freezing media (90 % heat inactivated horse serum, 10 % DMSO) and storage in a screwcap cryotube. Cryotube placed in polystyrene box at -80 oC. Cells recovered by the gradual addition of warmed media to the frozen cells, before centrifugation, media replacement and replating. Schwann cells could be stained by ICC for Schwann cell markers MBP and S100 (Figure S 2).

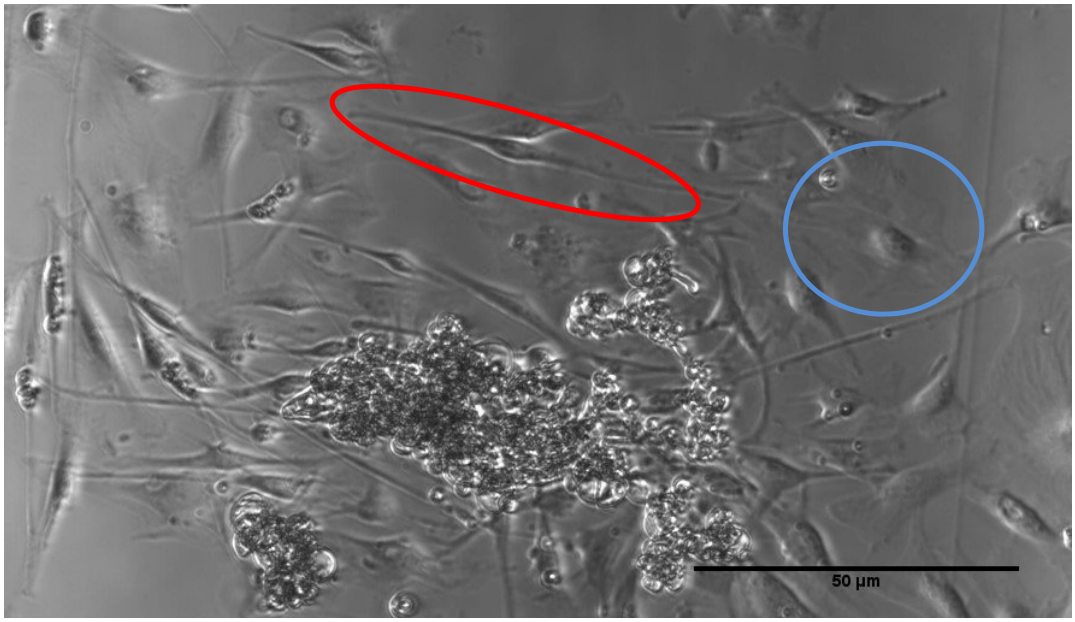


Figure S 1: Schwann cells and contaminating fibroblasts can be visually identified in culture.
The red circle identifies a Schwann cell, with a large central nucleus and two long processes. The blue circle identifies a fibroblast with a large central nucleus and a flat, egg shaped cytoplasm. (400X magnification). Scale bar represents 50 µm.

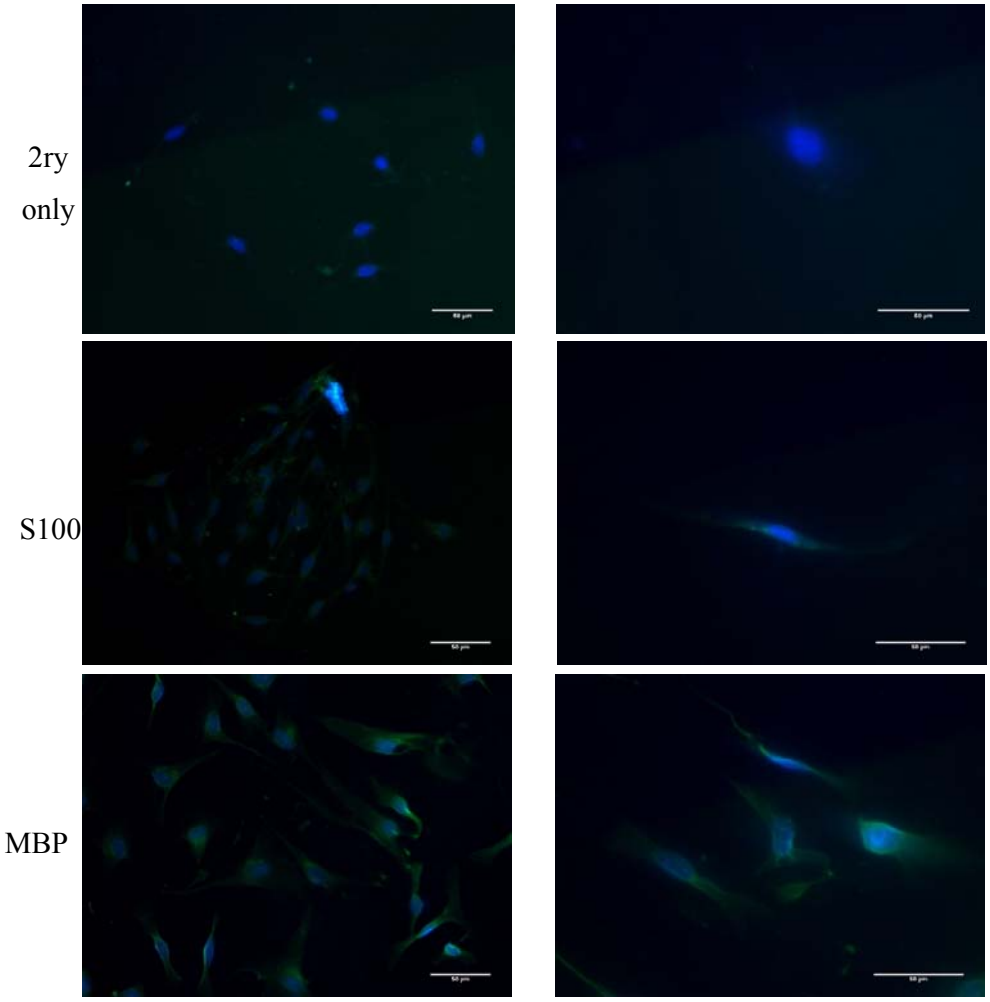
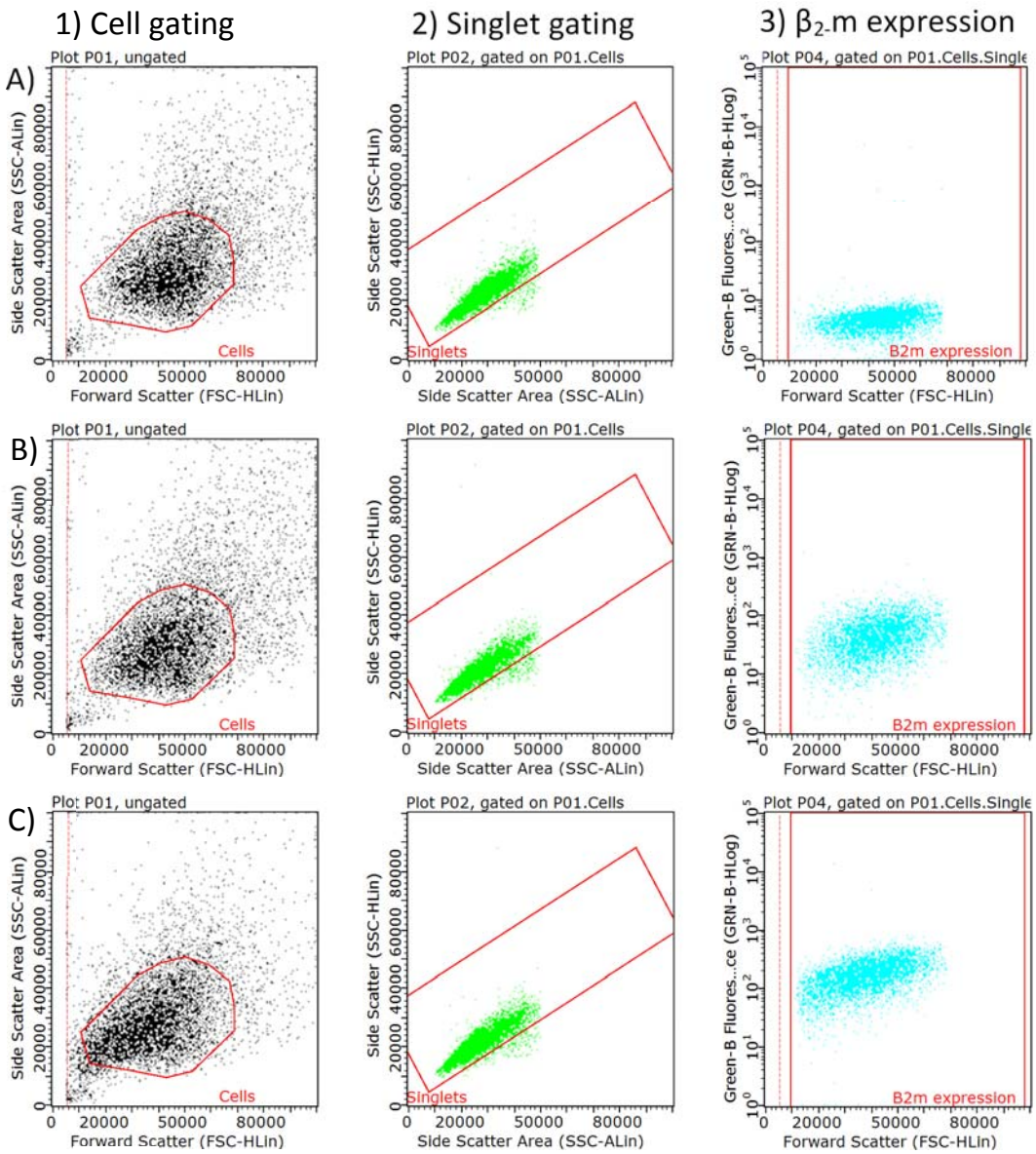


Figure S 2: Schwann cells isolated and cultured from adult mice express Schwann cell markers MBP and S100.

Fluorescence microscopy images of Schwann cell markers on cultured primary mouse Schwann cells. Images taken at 400x (left) and 600x (right) magnification. Antibody details can be found in Table 2.2. Green colour indicates positive staining for the protein being interrogated. Blue staining indicates DAPI stained nuclei. Scale bar represents 50 μm.

Appendix E Chapter 5

E.1 Flow cytometry results for triplicate IFNγ treatments on DFT2

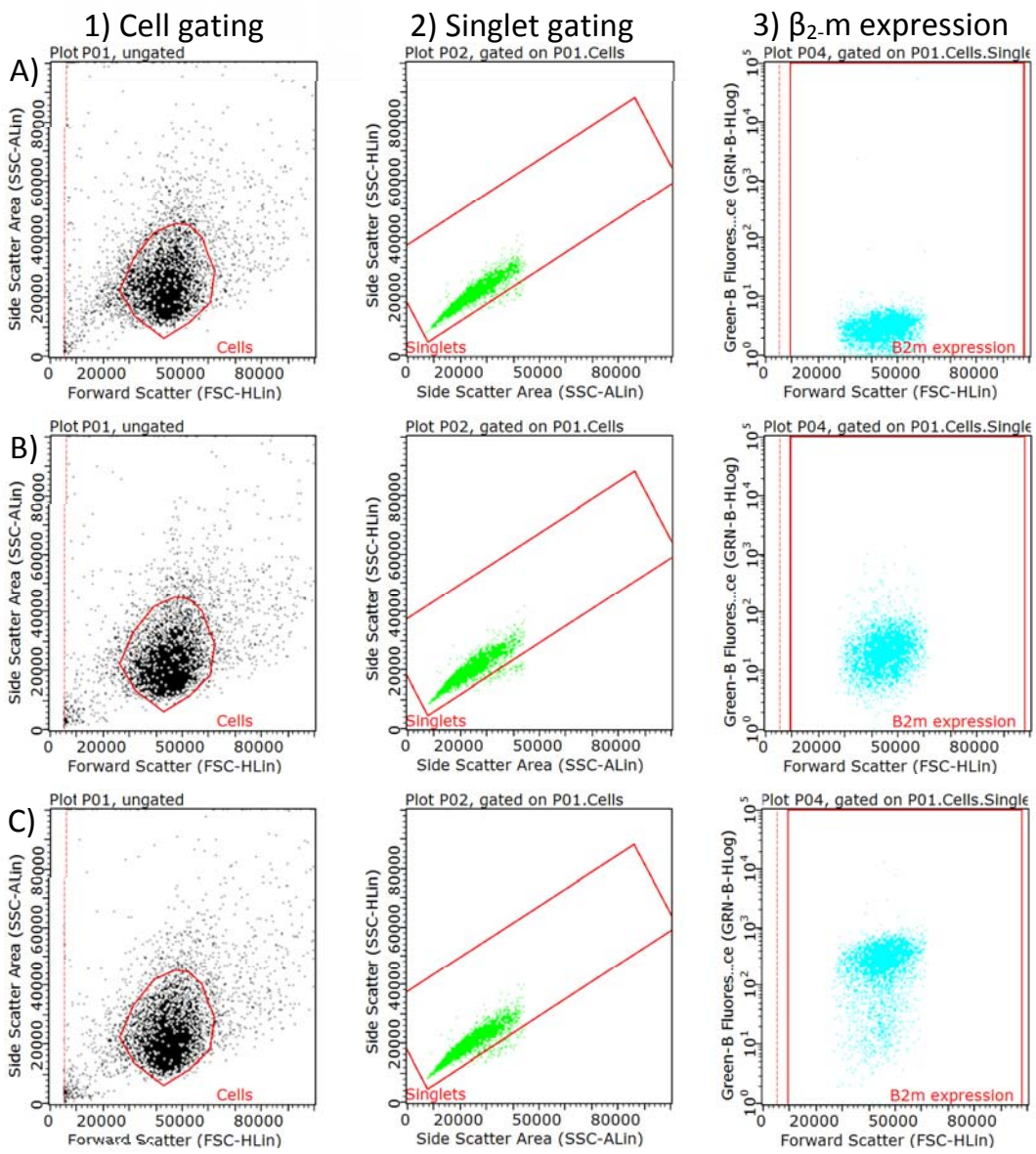


D)

DFT2_RV Sample	Mean fluorescence intensity	Median fluorescence intensity
Untreated 2ry only	4.47	4.58
Untreated β_2-m	39.30	41.08
IFNγ treated 2ry only	4.35	4.55
IFNγ treated β_2-m replicate 1	140.22	157.42
IFNγ treated β_2-m replicate 2	148.98	163.57
IFNγ treated β_2-m replicate 3	151.23	162.52

Figure S 3: Gating strategy and replicate fluorescence values for β_2-m expression in DFT2_RV following treatment with IFNγ.

Flow cytometry dot plots for A) Untreated DFT2_RV cells, 2ry only control cells. B) Untreated DFT2_RV cells stained for β_2-m expression. C) DFT2_RV cells treated with IFNγ and stained for β_2-m expression.
Dot plots with black dots in the column "1) Cell Gating" indicate the removal of dead or abnormal cells and debris from the dataset based on forward scatter (X axis) and side scatter (Y axis) properties of the cells. Only cells within red gate are analysed further. Dot plots with green dots in the 2) Singlet gating" column indicate the removal of doublets. Y axis indicates side scatter, X axis indicates area under the side scatter curve which is proportional to time taken for the droplet to pass through the detection laser. Plot is gated on column 1). Only cells within the red gate are carried forward for further analysis. Dot plots with cyan dots in the third column ("β_2-m expression") indicate fluorescence in the Green- B channel of the 488 fluorophore and represents β_2-m expression. Plots gated on 2). D) Table showing mean and median fluorescence intensity for all samples assessed, including biological replicates of IFNγ treatment in the DFT2_RV cell line.



D)

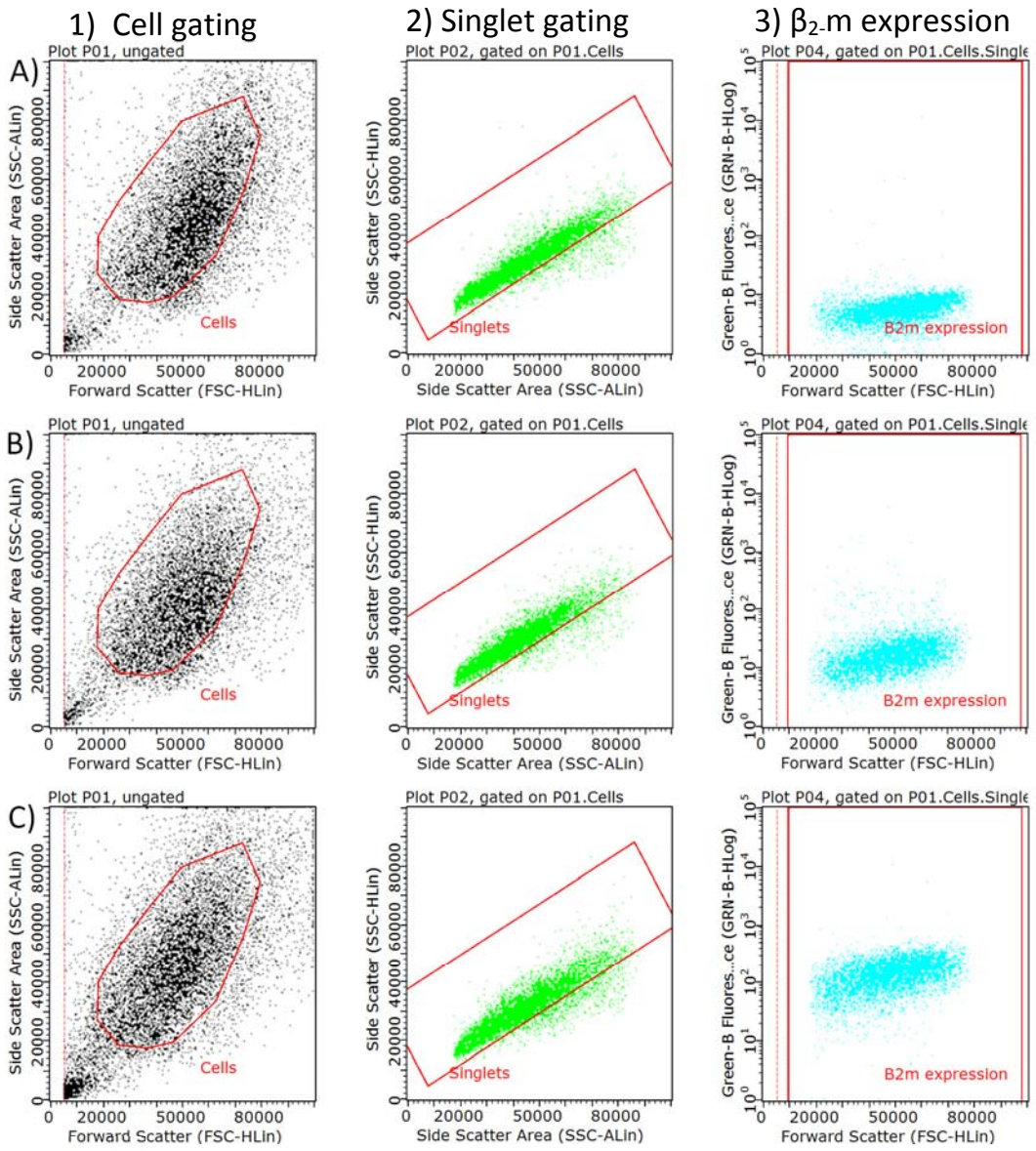
DFT2_SN Sample	Mean fluorescence intensity	Median fluorescence intensity
Untreated 2ry only	2.78	2.85
Untreated β_2-m	19.54	19.70
IFNγ treated 2ry only	3.25	3.33
IFNγ treated β_2-m replicate 1	202.81	275.72
IFNγ treated β_2-m replicate 2	189.29	263.42
IFNγ treated β_2-m replicate 3	229.55	310.34

Figure S 4: Gating strategy and replicate fluorescence values for β_2-m expression in DFT2_SN following treatment with IFNγ.

Flow cytometry dot plots for A) Untreated DFT2_SN cells, 2ry only control cells. B) Untreated DFT2_SN cells stained for β_2-m expression. C) DFT2_SN cells treated with IFNγ and stained for β_2-m expression.

Dot plots with black dots in the column "1) Cell Gating" indicate the removal of dead or abnormal cells and debris from the dataset based on forward scatter (X axis) and side scatter (Y axis) properties of the cells. Dot plots with green dots in the 2) Singlet gating" column indicate the removal of doublets. Y axis indicates side

(Fig. S 4, cont.) scatter, X axis indicates area under the side scatter curve which is proportional to time taken for the droplet to pass through the detection laser. Plot is gated on column 1). Only cells within the red gate are carried forward for further analysis. Dot plots with cyan dots in the third column ("β_2-m expression") indicate fluorescence in the Green-B channel of the 488 fluorophore and represents β_2-m expression. Plots gated on 2). D) Table showing mean and median fluorescence intensity for all samples assessed, including biological replicates of IFNγ treatment in the DFT2_SN cell line.



D)

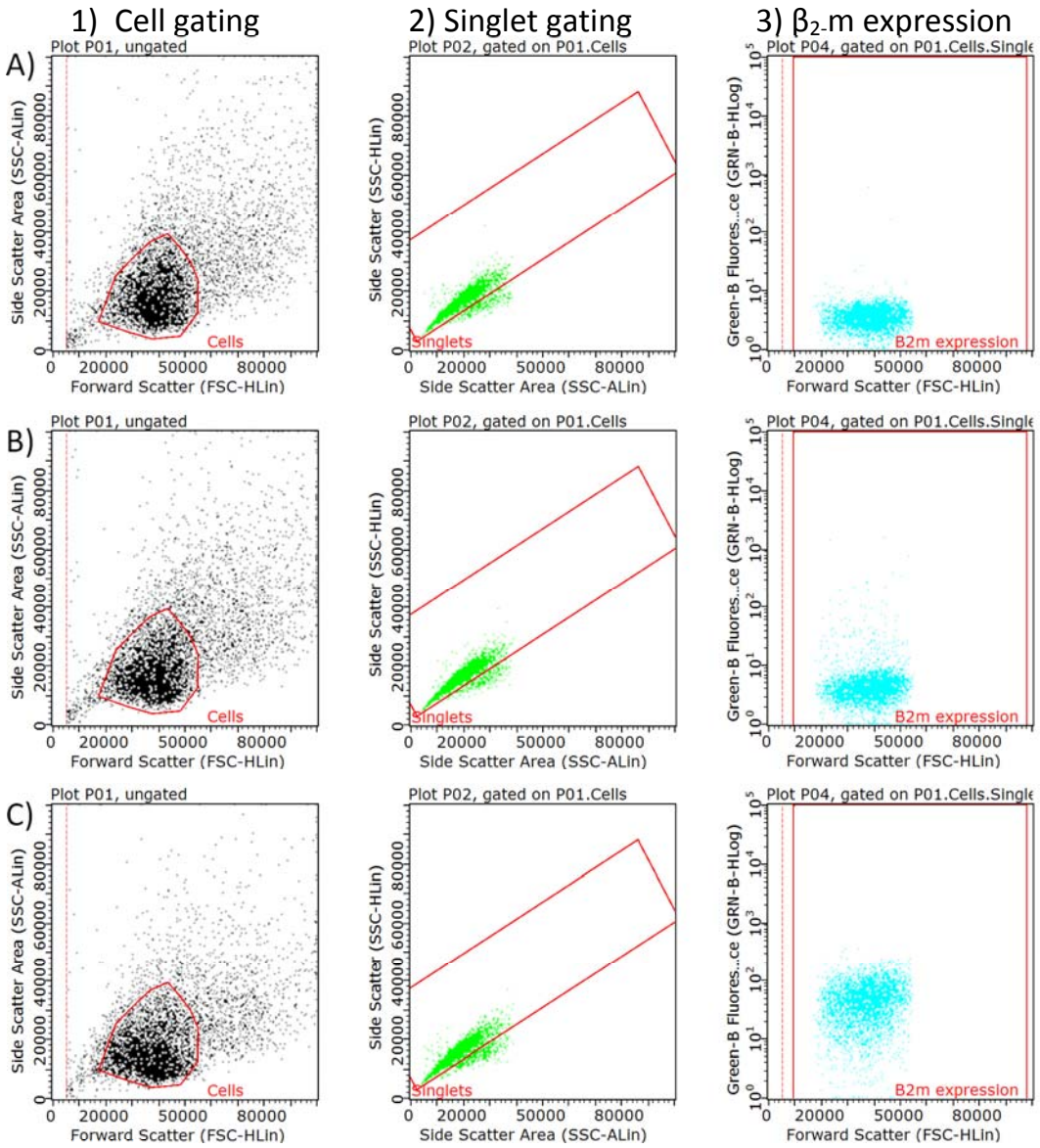
DFT2_549 Sample	Mean fluorescence intensity	Median fluorescence intensity
Untreated 2ry only	5.46	5.75
Untreated β_2-m	15.50	15.19
IFNγ treated 2ry only	5.42	5.63
IFNγ treated β_2-m replicate 1	129.95	136.06
IFNγ treated β_2-m replicate 2	130.38	136.15
IFNγ treated β_2-m replicate 3	134.62	145.01

Figure S 5: Gating strategy and replicate fluorescence values for β_2-m expression in DFT2_549 following treatment with IFNγ.

Flow cytometry dot plots for A) Untreated DFT2_549 cells, 2ry only control cells. B) Untreated DFT2_549 cells stained for β_2-m expression. C) DFT2_549 cells treated with IFNγ and stained for β_2-m expression.

Dot plots with black dots in the column "1) Cell Gating" indicate the removal of dead or abnormal cells and debris from the dataset based on forward scatter (X axis) and side scatter (Y axis) properties of the cells. Dot plots with green dots in the 2) Singlet gating" column indicate the removal of doublets. Y axis indicates side

(Fig. S 5, cont.) scatter, X axis indicates area under the side scatter curve which is proportional to time taken for the droplet to pass through the detection laser. Plot is gated on column 1). Only cells within the red gate are carried forward for further analysis. Dot plots with cyan dots in the third column ("β_2-m expression") indicate fluorescence in the Green-B channel of the 488 fluorophore and represents β_2-m expression. Plots gated on 2). D) Table showing mean and median fluorescence intensity for all samples assessed, including biological replicates of IFNγ treatment in the DFT2_549 cell line.



D)

DFT1_4906 Sample	Mean fluorescence intensity	Median fluorescence intensity
Untreated 2ry only	3.27	3.35
Untreated β_2-m	4.12	3.96
IFNγ treated 2ry only	3.15	3.24
IFNγ treated β_2-m replicate 1	34.77	40.23
IFNγ treated β_2-m replicate 2	36.15	42.25
IFNγ treated β_2-m replicate 3	38.09	43.54

Figure S 6: Gating strategy and replicate fluorescence values for β_2-m expression in DFT1_4906 following treatment with IFNγ.

Flow cytometry dot plots for A) Untreated DFT1_4906 cells, 2ry only control cells. B) Untreated DFT1_4906 cells stained for β_2-m expression. C) DFT1_4906 cells treated with IFNγ and stained for β_2-m expression. Dot plots with black dots in the column "1) Cell Gating" indicate the removal of dead or abnormal cells and debris from the dataset based on forward scatter (X axis) and side scatter (Y axis) properties of the cells. Dot plots with green dots in the 2) Singlet gating" column indicate the removal of doublets. Y axis indicates side

(Fig. S 6, cont.) scatter, X axis indicates area under the side scatter curve which is proportional to time taken for the droplet to pass through the detection laser. Plot is gated on column 1). Only cells within the red gate are carried forward for further analysis. Dot plots with cyan dots in the third column ("β_2-m expression") indicate fluorescence in the Green-B channel of the 488 fluorophore and represents β_2-m expression. Plots gated on 2). D) Table showing mean and median fluorescence intensity for all samples assessed, including biological replicates of IFNγ treatment in the DFT1_4906 cell line.

A)

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	18187	3	6062	F (3, 8) = 99.97	P<0.0001
Row Factor (Treatment)	73777	1	73777	F (1, 8) = 1217	P<0.0001
Column Factor (Cell line)	28836	3	9612	F (3, 8) = 171.6	P<0.0001
Subject	448.2	8	56.03	F (8, 8) = 0.9239	P=0.5432
Residual	485.2	8	60.64		

B)

	DFT2_RV	DFT2_SN	DFT2_549	DFT1_4906
DFT2_RV		0.0466	0.0054	0.0003
DFT2_SN	<0.0001		0.7076	0.0893
DFT2_549	0.0833	<0.0001		0.4750
DFT1_4906	<0.0001	<0.0001	<0.0001	

Untreated

IFNγ Treated

Figure S 7: Numerical results of Two-way ANOVA and post-hoc Tukey's t-test on IFNγ mediated upregulation of cell surface β_2-m between cell lines.

A) ANOVA table indicating significant differences in β_2-m upregulation between treated and untreated cells (Row factor), between cell lines (Column factor), and a significant interaction between the two (Interaction). B) P-values for post-hoc Tukey's t-test demonstrating significant differences between β_2-m expression in untreated cells and in the upregulation of β_2-m between cell lines.

E.2 RT-PCR results for triplicate IFNγ treatments

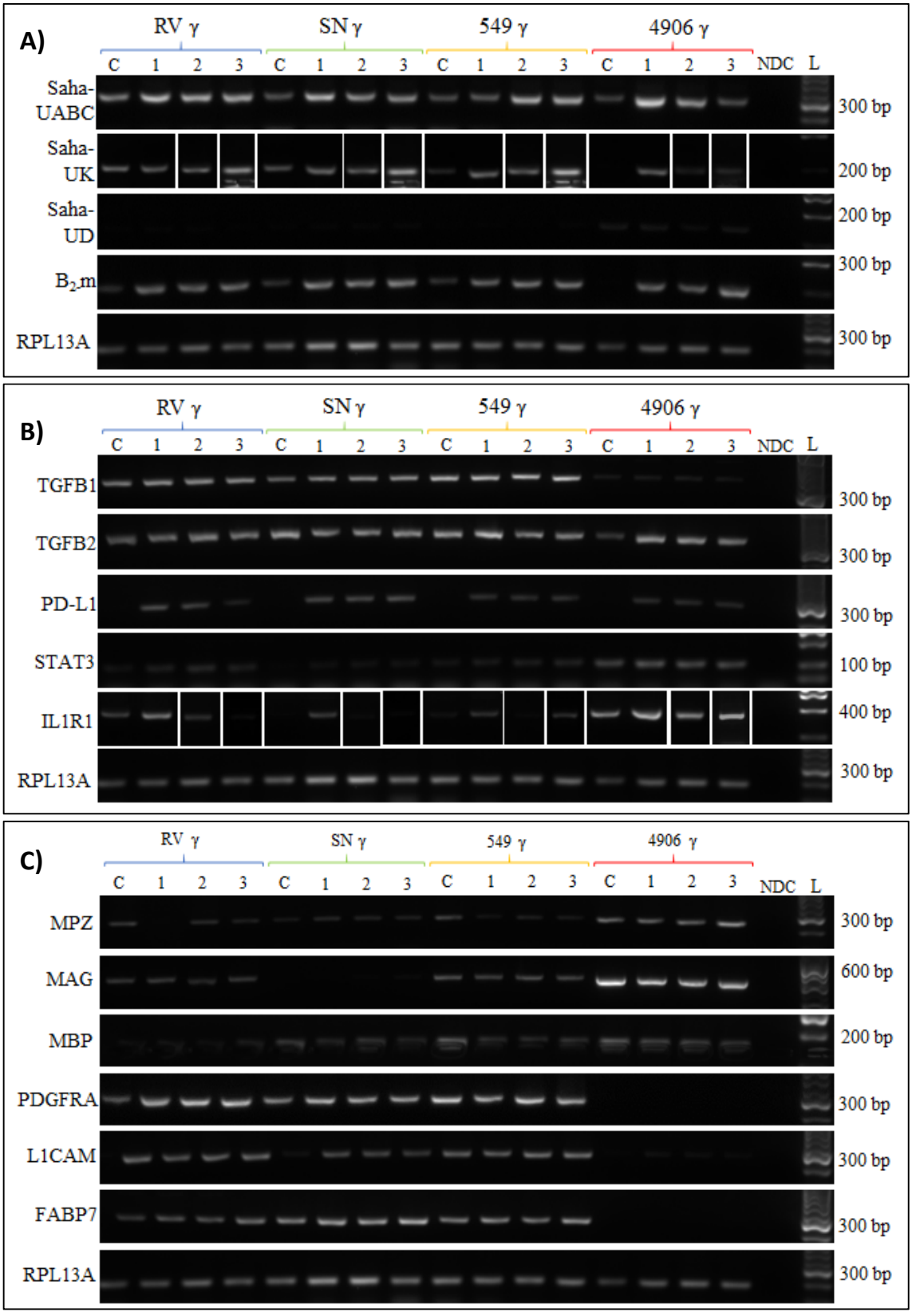


Figure S 8: Triplicate RT-PCR results for IFNγ treatments in DFT2 and DFT1.

(Fig. S 8, cont.) Transcriptional regulation of A) MHC class I associated genes, B) Immune associated genes, C) Myelin and glial associated genes by DFT2 and DFT1 cell lines in response to IFNγ treatment. Cell lines are denoted by coloured brackets. C denotes untreated control cDNA. 1, 2 and 3 indicate three independent biological replicates of IFNγ treatment. RPL13A is a housekeeping control gene.

Appendix F Chapter 7

F.1 Genomic sequence alignments for SahaI*27-1-pcDNA3.0 and genomic and translated protein sequence alignments for SahaI*UK in pcDNA3.0

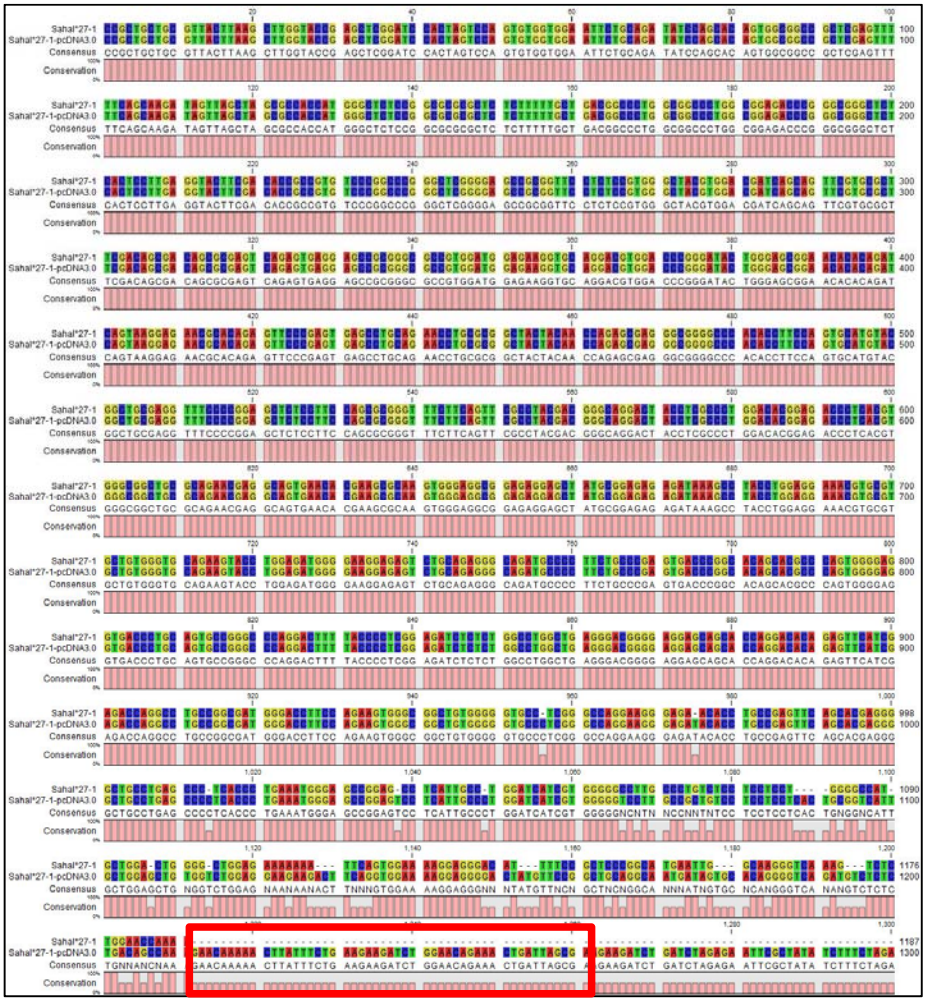


Figure S 9: Genomic sequence alignment of SahaI*27-1 with recombinant SahaI*27-1-pcDNA3.0.
Red box indicates 2x myc tags added during cloning process. Top sequence (labelled SahaI*27-1) represents the gemonic sequence of SahaI*27-1. Bottom sequence (labelled SahaI*27-1-pcDNA3.0) represents the myc tagged SahaI*27-1-pcDNA3.0 construct sequence.

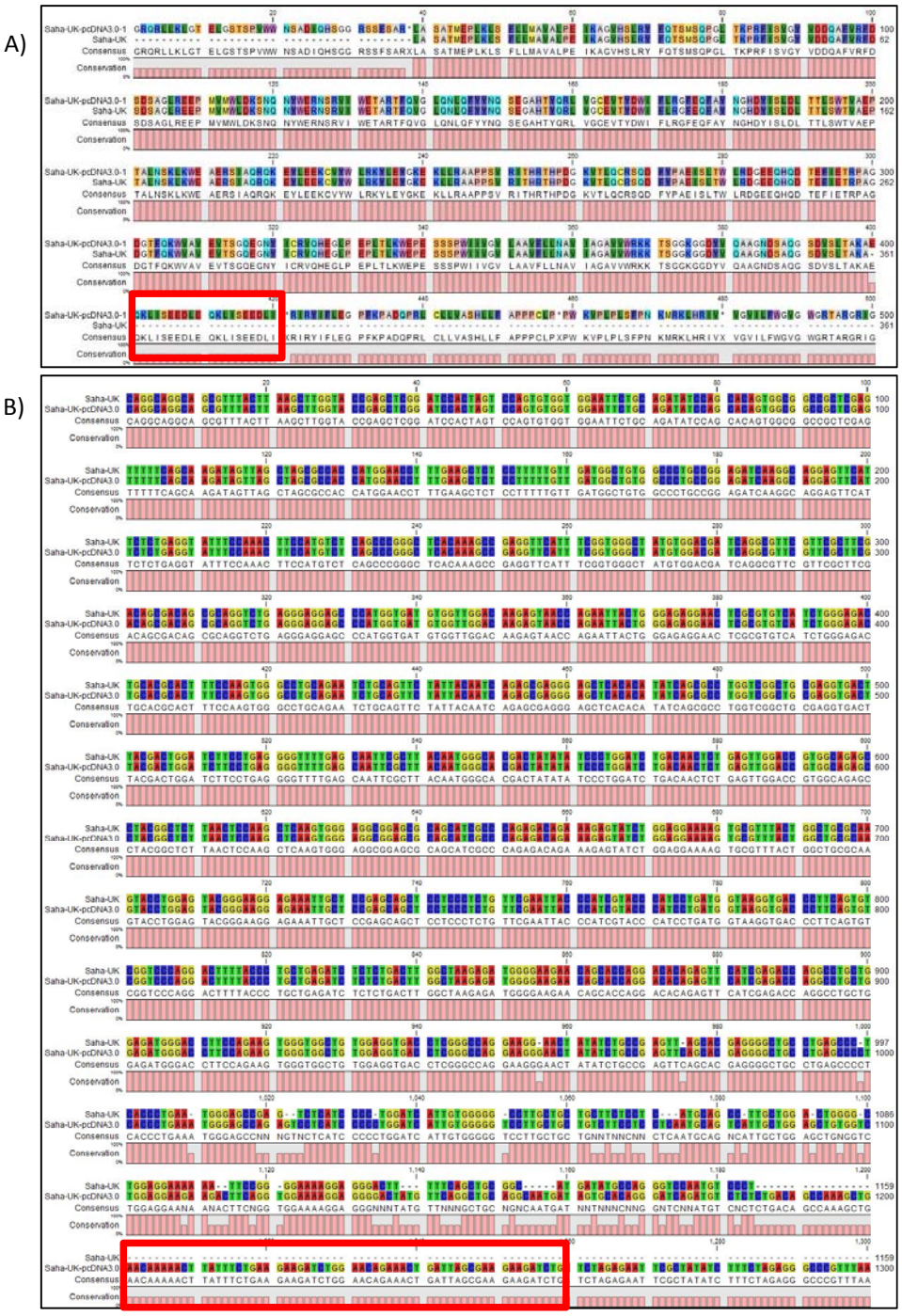


Figure S 10: Sequence alignments of A) Protein and B) Genomic sequence of Saha-UK and Saha-UK-pcDNA3.0.

A) Protein alignment of recombinant Saha-UK-pcDNA3.0 (top sequence labelled Saha-UK-pcDNA3.0-1) against genomic Saha-UK (bottom sequence labelled Saha-UK) translated from DNA sequences. B) Genomic sequence alignment of genomic Saha-UK (top sequence labelled Saha-UK) and recombinant Saha-UK-pcDNA3.0 construct (bottom sequence labelled Saha-UK-pcDNA3.0). Red box indicates 2x myc tags added during cloning process.

F.2 Full western blot images from Chapter 7

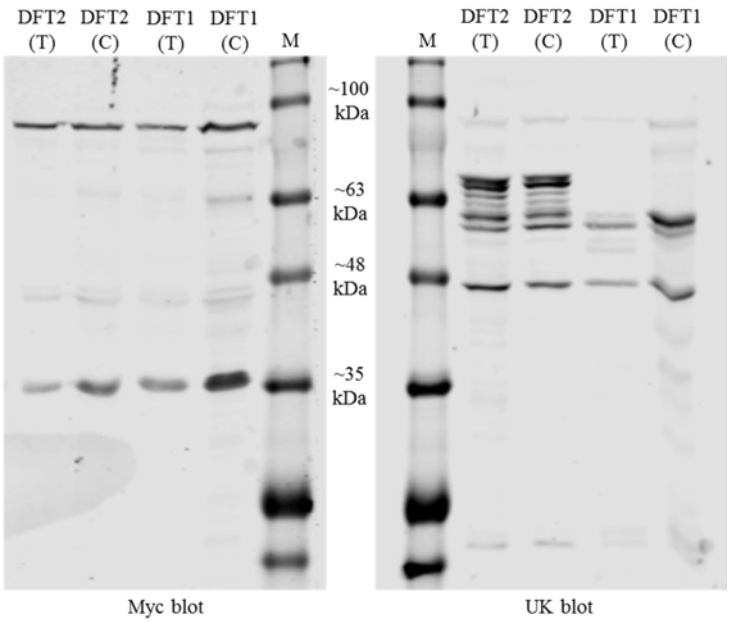


Figure S 11: Full Western blot images from Figure 7.4.

SahaI*UK-pIRES Transfected (T) and Untransfected (C) DFT1_4906 and DFT2_RV whole cell lysates blotted for myc and SahaI*UK protein. SahaI*UK is detected at ~48 kDa, myc tagged construct should be detected ~51 kDa. Endogenous myc protein is detected at ~35 kDa and ~80 kDa.

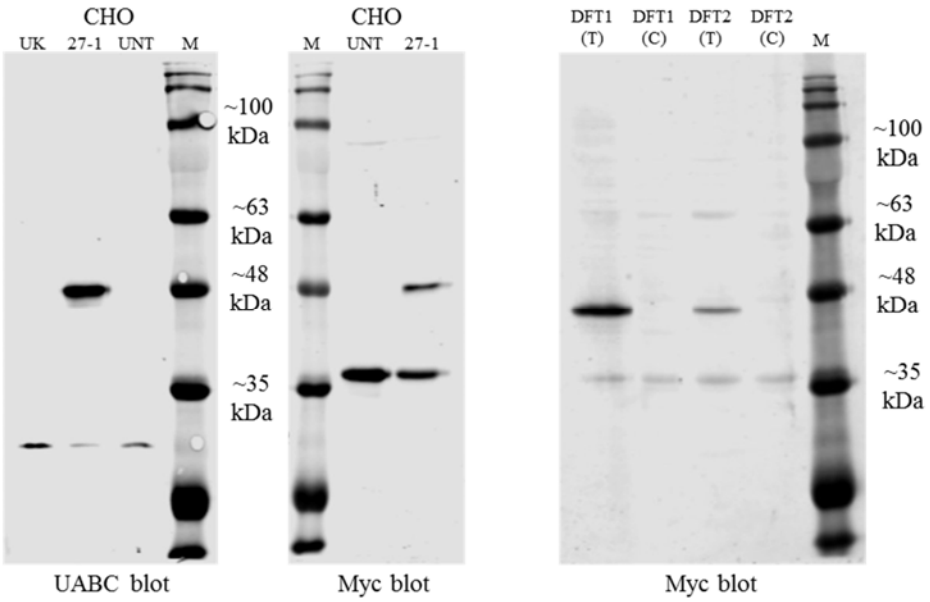


Figure S 12: Full Western blot images from Figure 7.6.

A) CHO transfected with SahaI*27-1-pcDNA3.0 and SahaI*UK and untransfected (UNT) blotted for classical MHC class I (Saha-UA/B/C). B) CHO transfected with SahaI*27-1-pcDNA3.0 and untransfected (UNT) blotted for myc. C) DFT1_4906 and DFT2_RV untransfected control (C) and transfected (T) with SahaI*UK-pcDNA3.0.

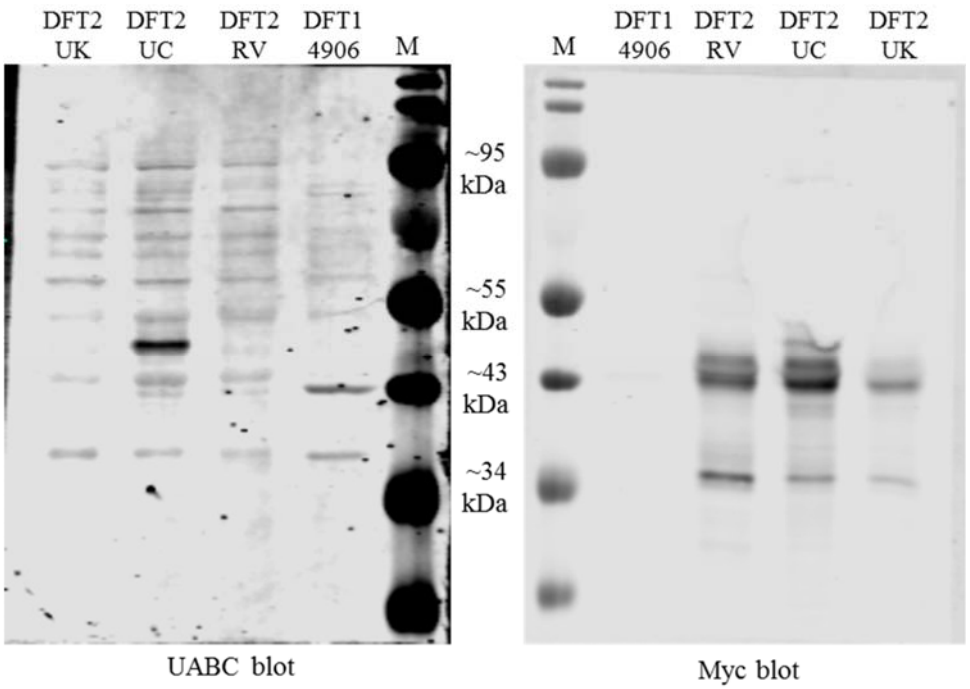


Figure S 13: Full Western blot images from Figure 7.11.

DFT1_4906 whole cell lysate, DFT2_RV whole cell lysate, DFT2_RV transfected with SahaI*27-1-pcDNA3.0 (DFT2_UC) whole cell lysate and DFT2_RV transfected with Saha-UK-pcDNA3.0 (DFT2_UK) whole cell lysate blotted for myc and classical MHC class I (UA/B/C). M indicates a protein marker of known molecular weight, indicated in kDa.

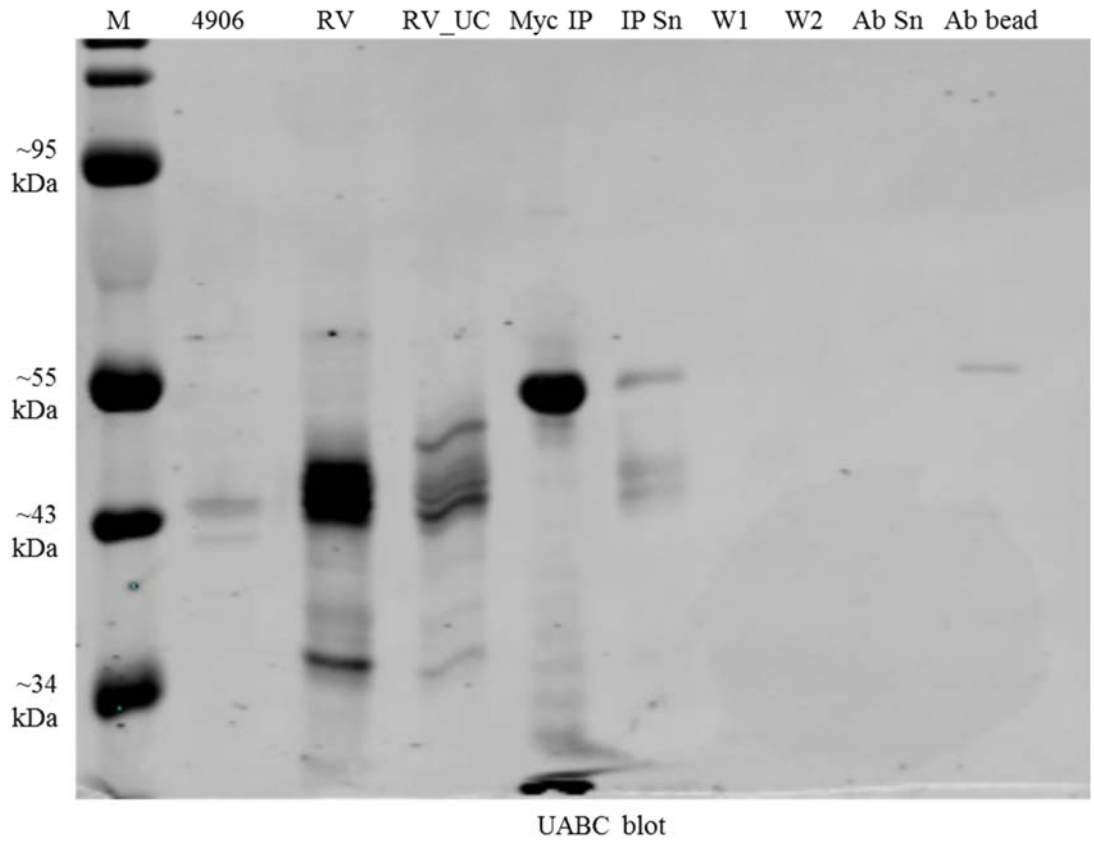


Figure S 14: Full Western blot image from Figure 7.12.

Western blot analysis of a myc IP on DFT2_UC whole cell lysate blotted for Saha-UA/B/C heavy chain. From left to right the blot reads as follows:

Protein marker (M), DFT1_4906 whole cell lysate (4906), DFT2_RV whole cell lysate (RV), DFT2_UC whole cell lysate (RV_UC), Immunoprecipitation of recombinant SahaI*27-1 construct from DFT2_UC lysate using anti-myc antibody (Myc IP), Post immunoaffinity DFT2_UC lysate (IP Sn), Supernatant from washes of the immunoaffinity beads post purification (W1, W2), Supernatant from conjugation of myc antibody to protein A beads (Ab Sn), Protein A bead bound to anti-myc antibody (Ab bead).

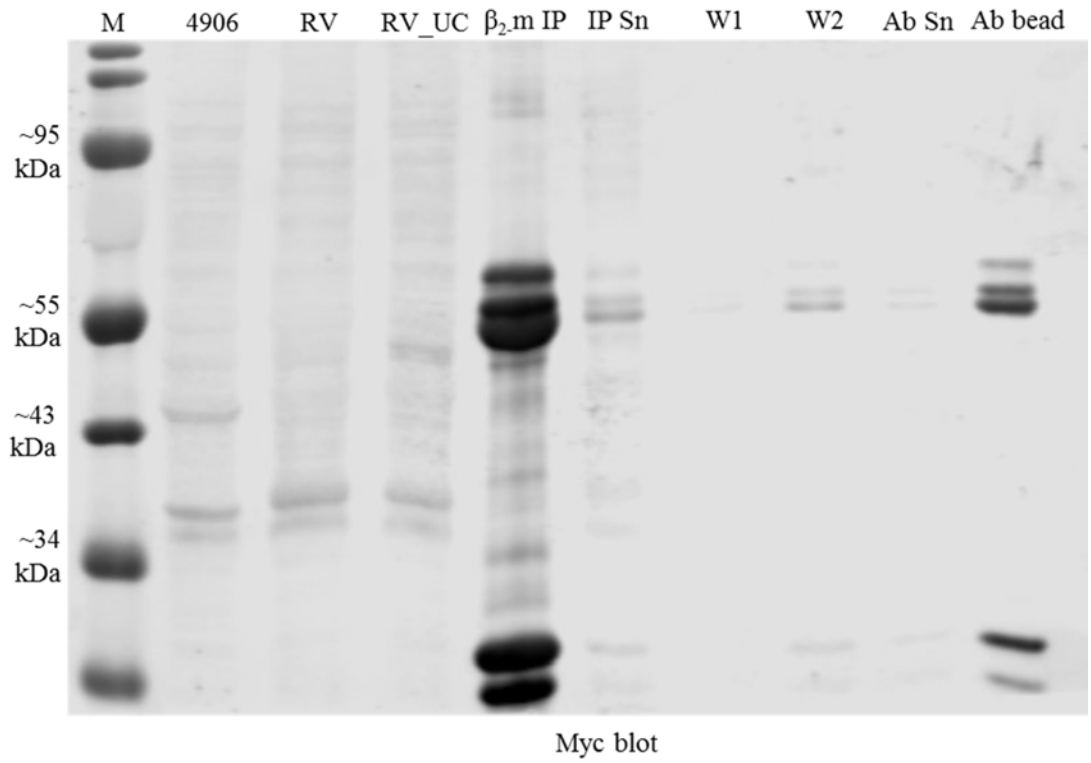


Figure S 15: Full Western blot image from Figure 7.13.

Western blot analysis of a β_2-m IP on DFT2_UC whole cell lysate blotted for myc. From left to right the blot reads as follows:

Protein marker (M), DFT1_4906 whole cell lysate (4906), DFT2_RV whole cell lysate (RV), DFT2_UC whole cell lysate (RV_UC), Immunoprecipitation of recombinant SahaI*27-1 construct from DFT2_UC lysate using anti- β_2-m antibody (β_2-m IP), Post immunoaffinity DFT2_UC lysate (IP Sn), Supernatant from washes of the immunoaffinity beads post purification (W1, W2), Supernatant from conjugation of β_2-m antibody to protein A beads (Ab Sn), Protein A bead bound to anti- β_2-m antibody (Ab bead).

Appendix G Chapter 8

G.1 Western blot analysis of triplicate immunoaffinity purification experiments

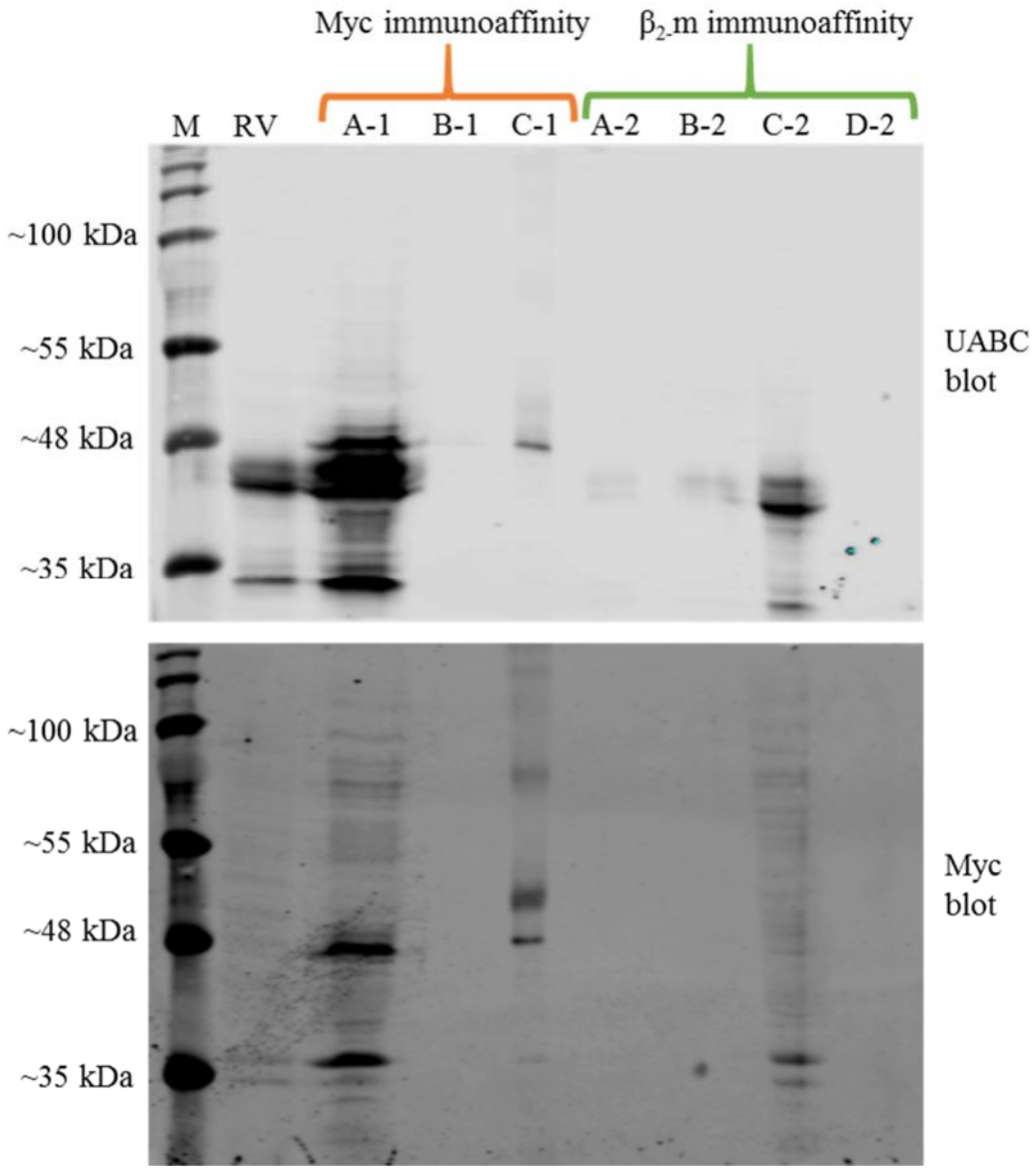


Figure S 16: Full Western blot images for Replicate 1 of the immunoaffinity experiments described in Chapter 8.3.1.

M denotes a protein marker of known size which is indicated in kDa. Samples collected at various stages during 1) Myc immunoaffinity purification and 2) β_2-m immunoaffinity purification. A) DFT2_UC lysate pre-immunoaffinity, B) Eluted protein from bead-conjugated antibodies, C) Bead-conjugated antibody bound to protein of interest post immunoaffinity. Note that the pre-β_2-m immunoaffinity lysate in A-2) was used immediately post-myc immunoaffinity and should not contain any myc-tagged construct.

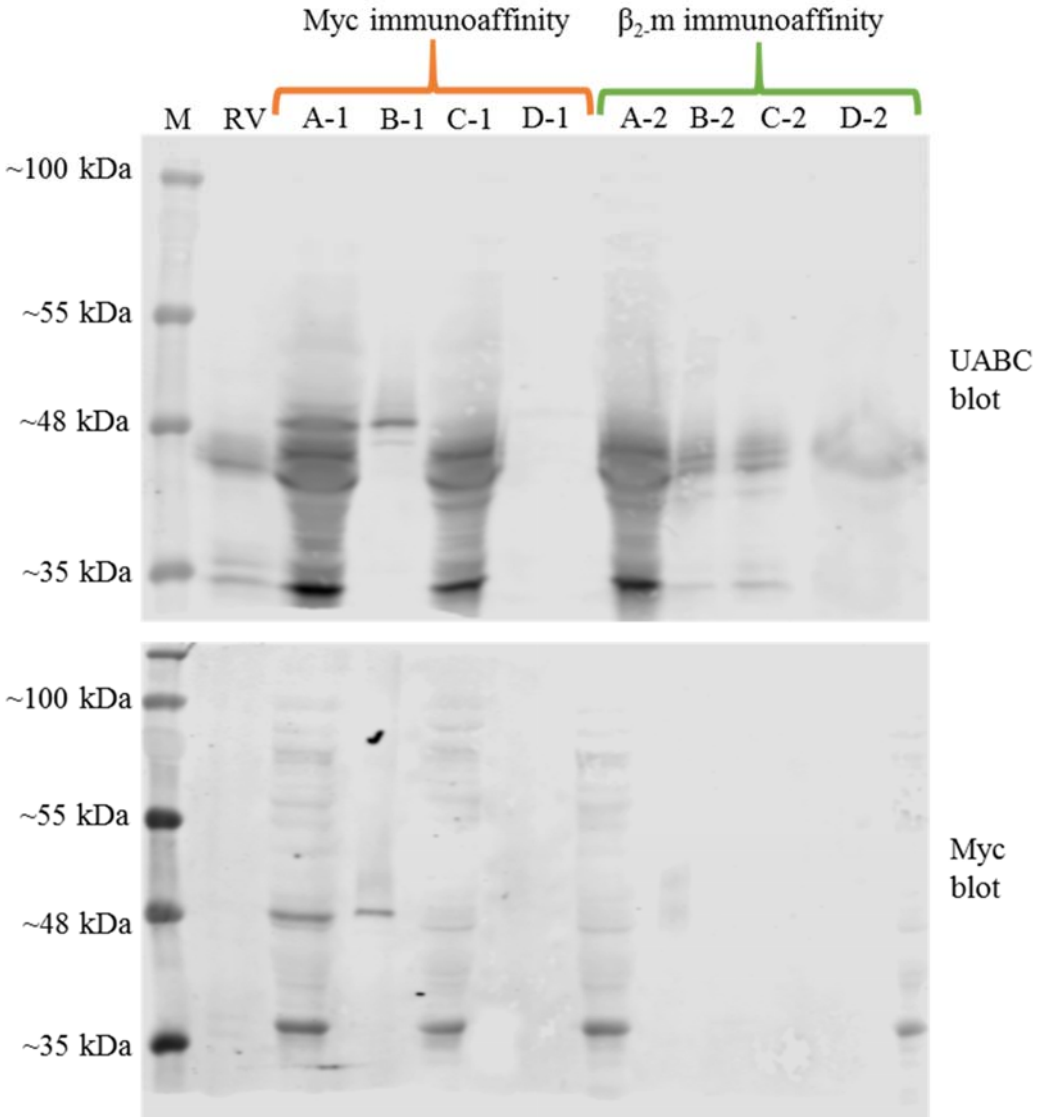


Figure S 17: Full Western blot images for Replicate 2 of the immunoaffinity experiments described in Chapter 8.3.1 and presented in Figure 8.3.

Samples collected at various stages during 1) Myc immunoaffinity purification and 2) β_2-m immunoaffinity purification. A) DFT2_UC lysate pre-immunoaffinity, B) Bead-conjugated antibody bound to protein of interest post immunoaffinity, C) DFT2_UC lysate post immunoaffinity, D) Eluted protein from bead-conjugated antibodies. Note that the post-myc immunoaffinity lysate in C-1) was immediately used in the β_2-m immunoaffinity and is the same as β_2-m A-2).

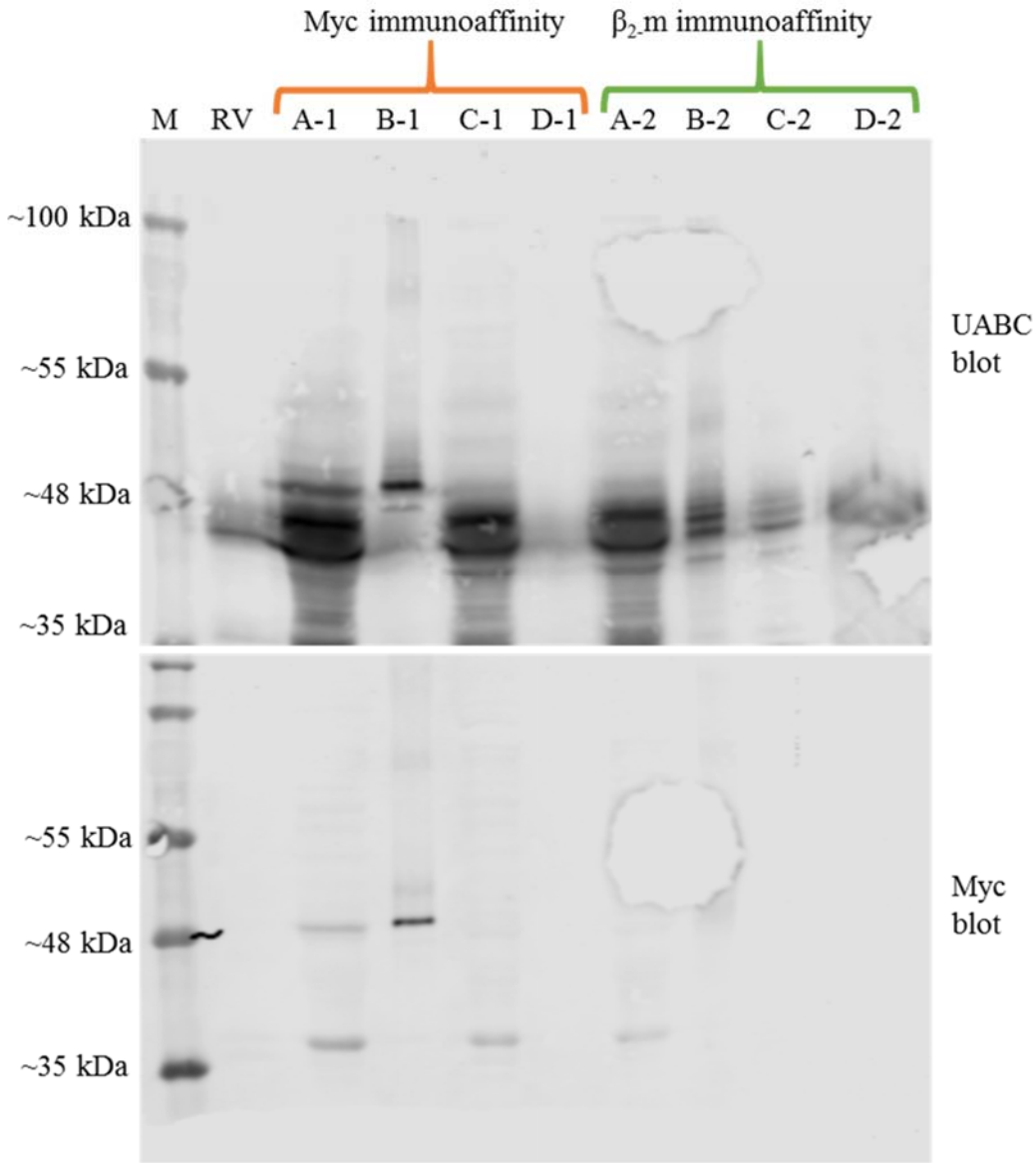


Figure S 18: Full Western blot images for Replicate 3 of the immunoaffinity experiments detailed in Chapter 8.3.1.

Samples collected at various stages during 1) Myc immunoaffinity purification and 2) β_2-m immunoaffinity purification. A) DFT2_UC lysate pre-immunoaffinity, B) Bead-conjugated antibody bound to protein of interest post immunoaffinity, C) DFT2_UC lysate post immunoaffinity, D) Eluted protein from bead-conjugated antibodies. Note that the post-myc immunoaffinity lysate in C-1) was immediately used in the β_2-m immunoaffinity and is the same as β_2-m A-2).

G.2 HPLC traces for triplicate immunoaffinity purifications

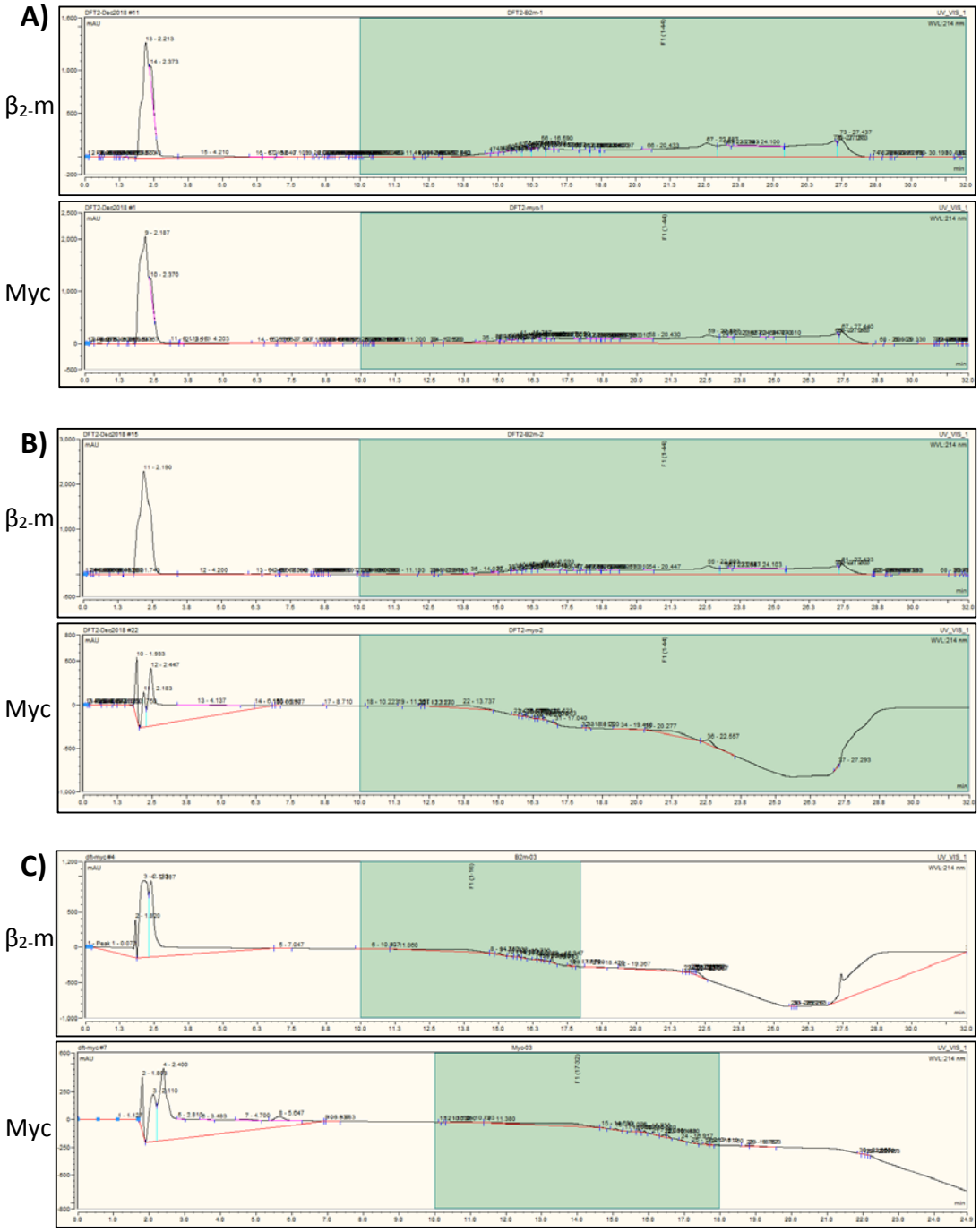


Figure S 19: Triplicate HPLC traces for biological replicates of β2-m and myc immunoaffinity experiments.

β2-m and myc immunoaffinity HPLC traces for A) Replicate 1, B) Replicate 2, C) Replicate 3. Black line represents the HPLC trace. X axis represents retention (elution) time of the sample in minutes. Y axis represents intensity of detection. Green boxed region indicates region fractions were collected from. Replicate 1 is presented in the main text as Figure 8.4. Note the traces for Replicate 3 and for myc immunoaffinity in Replicate 2 are inverted, however the peaks remain at the expected positions.

G.3 Peptides identified by mass spectrometry following immunoaffinity purification of the MHC class I complex

AVAILABLE AT: https://doi.org/10.5258/SOTON/D0982

Table S 9: Raw mass spectrometry data and sequenced peptide information following immunoaffinity purification of the MHC class I complex.

Peptide sequence is given in the column labelled "peptide_cln", peptide length is given in the column "length". "exp" refers to the replicate each peptide was detected in, MYC_1, MYC_2 and MYC_3 are the three replicates of myc immunoaffinity purification, B2M_1, B2M_2 and B2M_3 are the corresponding β_2-m immunoaffinity replicates. Peptide and gene Ensembl identifiers for the source protein of detected peptides are presented in columns "accession", "ensembl_peptide_id" and "ensembl_gene_id" where available. Official gene symbol and protein description are presented in "name" and "description" where available. RefSeq protein ID is given in "IDs" where available. Genome scaffold, chromosome and start and end of protein sequences in the genome of the source protein are presented in the columns "seqnames", "chr", "start" and "end" where available.

G.4 Peptides identified by mass spectrometry following immunoaffinity purification of the MHC class I complex by Annalisa Gastaldello

AVAILABLE AT: https://doi.org/10.5258/SOTON/D0982

Table S 10: Raw mass spectrometry data and sequenced peptide information following immunoaffinity purification of the MHC class I complex performed by Annalisa Gastaldello *et al.* (in preparation).
Peptide sequence is given in the column labelled "peptide_cln", peptide length is given in the column "length". "exp" refers to the replicate each peptide was detected in, B2M_4, B2M_5 and B2M_7 are the corresponding β_2-m immunoaffinity replicates. B2M_6 was a failed experiment and is not included in this analysis. Peptide and gene Ensembl identifiers for the source protein of detected peptides are presented in columns "accession", "ensembl_peptide_id" and "ensembl_gene_id" where available. Official gene symbol and protein description are presented in "name" and "description" where available. RefSeq protein ID is given in "IDs" where available. Genome scaffold, chromosome and start and end of protein sequences in the genome of the source protein are presented in the columns "seqnames", "chr", "start" and "end" where available.

G.5 GibbsCluster2.0 results for MHC class I peptide binding in DFT2

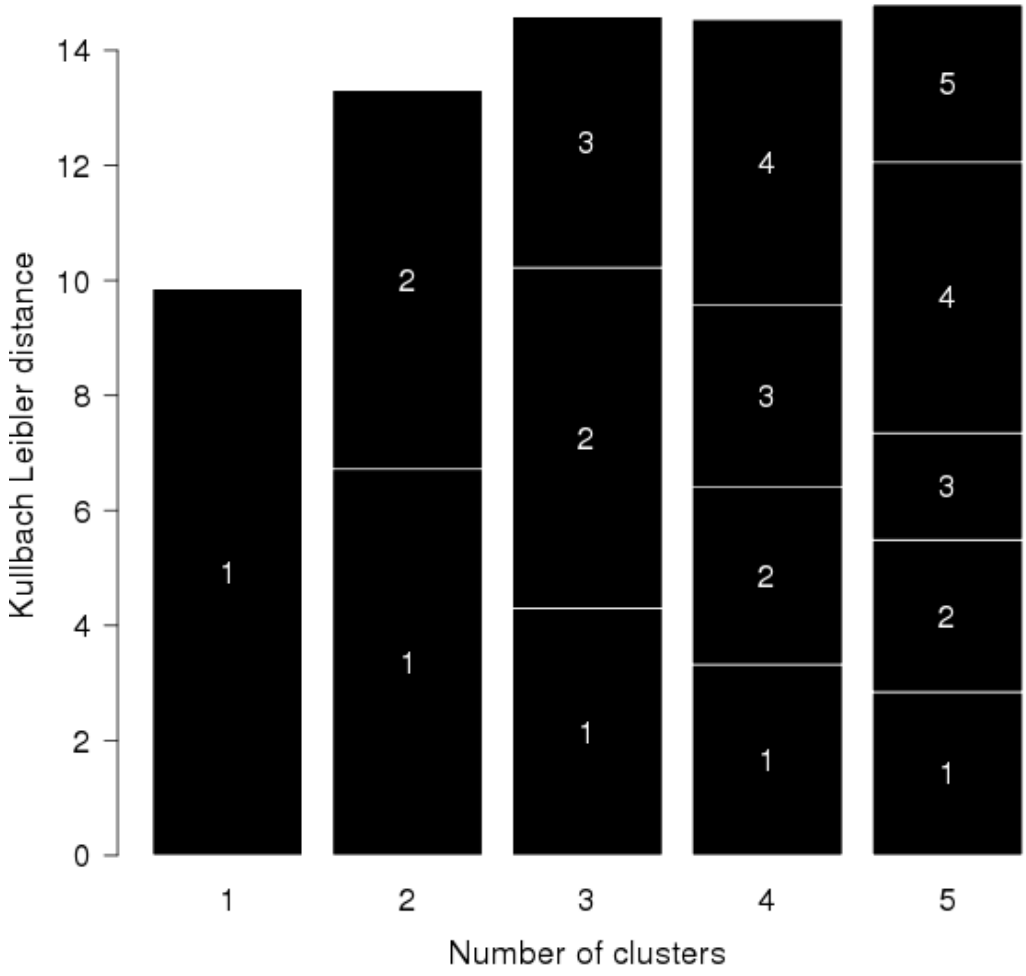


Figure S 20: Kullbach Leibler distance vs number of clusters for all peptides identified from β_2-m immunoaffinity purification from DFT2_UC as calculated by the GibbsCluster2.0 server.

A higher Kullbach Leibler distance as calculated by the GibbsCluster2.0 server indicates a stronger clustering of the data, and the size of each block within each bar indicates the proportion of peptides calculated to be within each cluster. E.g for 5 clusters, more peptides align to cluster 4 than to cluster 3. Number of clusters indicates the number of predicted MHC class I binding motifs, hence the number of predicted MHC class I alleles in the peptide dataset. GibbsCluster indicates that 5 alleles are present, although the distances are close and the clustering data is largely inconclusive. Suggested binding motifs for these clusters are presented in Figure S20.

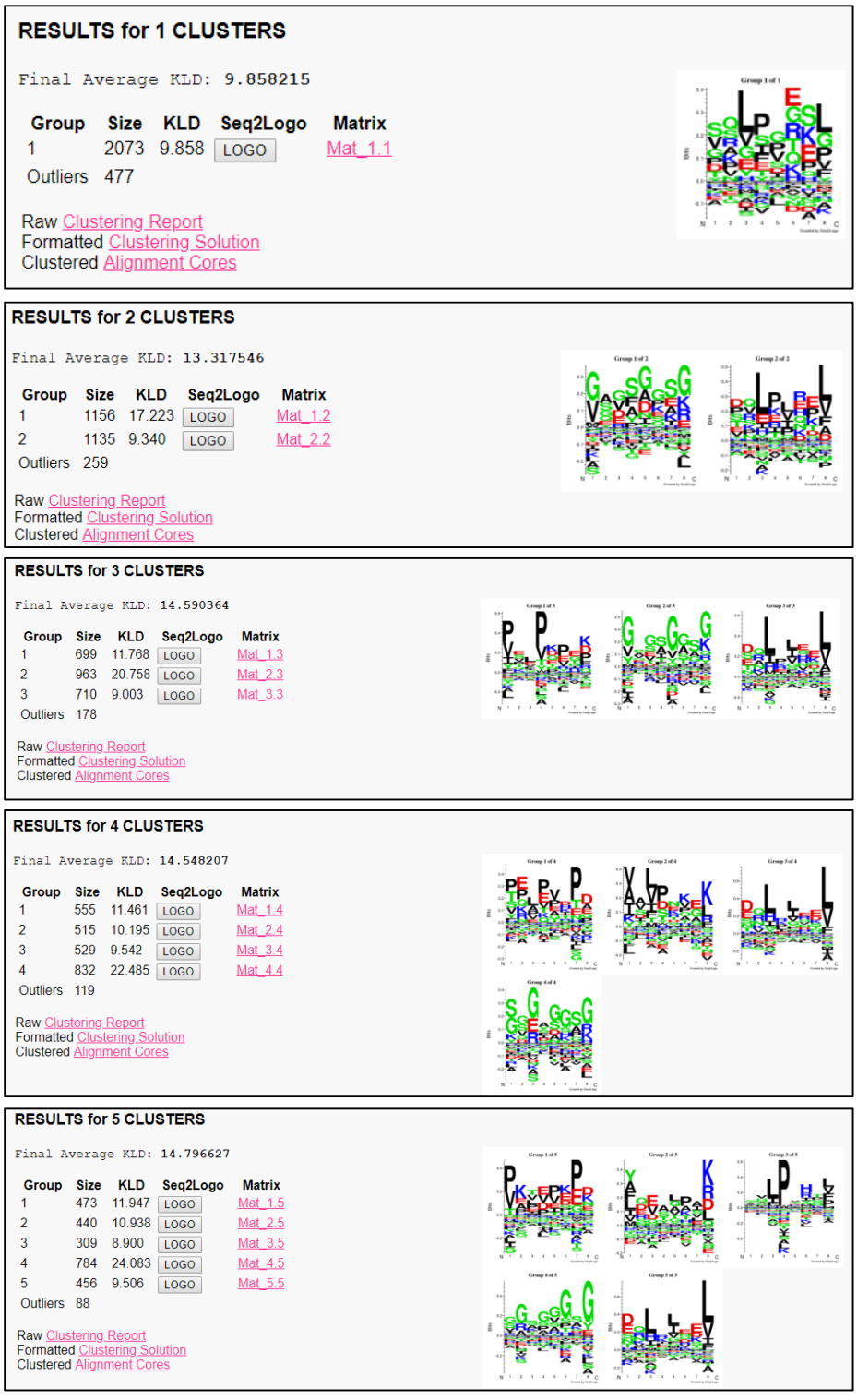


Figure S 21: Summary reports and predicted binding motifs for GibbsCluster2.0 clustering of peptides. Predicted binding motifs for all suggested MHC class I alleles (clusters) in the β_2-m immunoaffinity peptide dataset have a high background and no clear binding motif or preferences. Binding motifs do not clarify the highly similar cluster analysis presented in Figure S19. GibbsCluster2.0 analysis of this dataset is inconclusive and does not indicate the presence of multiple, clear binding motifs from different MHC class I alleles.

List of References

Adams, E.J. and Parham, P. (2001). Species-specific evolution of MHC class I genes in the higher primates. *Immunological Reviews* **183**:41–64.

Ahmad, S., Ranaghan, K.E. and Azam, S.S. (2019). Combating tigecycline resistant Acinetobacter baumannii: A leap forward towards multi-epitope based vaccine discovery. *European Journal of Pharmaceutical Sciences* **132**:1–17.

Albanese, F., Poli, A., Millanta, F. and Abramo, F. (2002). Primary cutaneous extragenital canine transmissible venereal tumour with Leishmania-laden neoplastic cells: a further suggestion of histiocytic origin? *Veterinary Dermatology* **13**:243–246.

Albrecht, V., Zweiniger, C., Surendranath, V., Lang, K., Schöfl, G., Dahl, A., Winkler, S., *et al.* (2017). Dual redundant sequencing strategy: Full-length gene characterisation of 1056 novel and confirmatory HLA alleles. *HLA* **90**:79–87.

Albring, J., Koopmann, J.-O., Hämmerling, G.J. and Momburg, F. (2004). Retrotranslocation of MHC class I heavy chain from the endoplasmic reticulum to the cytosol is dependent on ATP supply to the ER lumen. *Molecular Immunology* **40**:733–741.

Allen, R.L. and Hogan, L. (2001). Non-Classical MHC Class I Molecules (MHC-Ib). In: *ELS*. John Wiley & Sons, Ltd.

Altman, J.D., Moss, P.A.H., Goulder, P.J.R., Barouch, D.H., McHeyzer-Williams, M.G., Bell, J.I., McMichael, A.J., *et al.* (1996). Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**:94–96.

Anagnostou, V., Smith, K.N., Forde, P.M., Niknafs, N., Bhattacharya, R., White, J., Zhang, T., *et al.* (2017). Evolution of neoantigen landscape during immune checkpoint blockade in non-small cell lung cancer. *Cancer discovery* **7**:264–276.

Andersen, N.D., Srinivas, S., Piñero, G. and Monje, P. V (2016). A rapid and versatile method for the isolation, purification and cryogenic storage of Schwann cells from adult rodent nerves. *Scientific Reports* **6**:31781.

Anderson, C.K., Reilly, E.C., Lee, A.Y. and Brossay, L. (2019). Qa-1-restricted CD8(+) T cells can compensate for the absence of conventional T cells during viral infection. *Cell Reports* **27**:537–548.

Andreatta, M., Alvarez, B. and Nielsen, M. (2017). GibbsCluster: unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Research* **45**:458–463.

Andreatta, M., Lund, O. and Nielsen, M. (2012). Simultaneous alignment and clustering of peptide

data using a Gibbs sampling approach. *Bioinformatics* **29**:8–14.

Andreatta, M. and Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32**:511–517.

Angell, T.E., Lechner, M.G., Jang, J.K., LoPresti, J.S. and Epstein, A.L. (2014). MHC class I loss is a frequent mechanism of immune escape in papillary thyroid cancer that is reversed by interferon and selumetinib treatment in vitro. *Clinical Cancer Research* **20**:6034–6044.

Antoniou, A.N., Powis, S.J. and Elliott, T. (2003). Assembly and export of MHC class I peptide ligands. *Current Opinion in Immunology* **15**:75–81.

Armbruster, D.A. and Pry, T. (2008). Limit of blank, limit of detection and limit of quantitation. *The Clinical biochemist. Reviews* **29 Suppl 1**:S49–S52.

Arnett, H.A., Wang, Y., Matsushima, G.K., Suzuki, K. and Ting, J.P.-Y. (2003). Functional genomic analysis of remyelination reveals importance of inflammation in oligodendrocyte regeneration. *The Journal of Neuroscience* **23**:9824–9832.

Asada, K., Nakajima, T., Shimazu, T., Yamamichi, N., Maekita, T., Yokoi, C., Oda, I., *et al.* (2015). Demonstration of the usefulness of epigenetic cancer risk prediction by a multicentre prospective cohort study. *Gut* **64**:388 LP – 396.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., *et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* **25**:25–29.

Assinck, P., Duncan, G.J., Plemel, J.R., Lee, M.J., Stratton, J.A., Manesh, S.B., Liu, J., *et al.* (2017). Myelinogenic plasticity of oligodendrocyte precursor cells following spinal cord contusion injury. *The Journal of Neuroscience* **37**:8635–8654.

Awate, S., Babiuk, L.A. and Mutwiri, G. (2013). Mechanisms of action of adjuvants. *Frontiers in Immunology* **4**:114.

Ayala Garcia, M.A., Gonzalez Yebra, B., Lopez Flores, A.L. and Guani Guerra, E. (2012). The major histocompatibility complex in transplantation. *Journal of Transplantation* **2012**:842141.

Bacallao, K. and Monje, P. V (2015). Requirement of cAMP signaling for Schwann cell differentiation restricts the onset of myelination. *PLoS ONE* **10**:e0116948.

Balabanov, R., Strand, K., Goswami, R., McMahon, E., Begolka, W., Miller, S.D. and Popko, B. (2007). Interferon-γ-oligodendrocyte interactions in the regulation of experimental autoimmune encephalomyelitis. *The Journal of Neuroscience* **27**:2013–2024.

Barbin, G., Aigrot, M.S., Charles, P., Foucher, A., Grumet, M., Schachner, M., Zalc, B., *et al.* (2004).

Axonal cell-adhesion molecule L1 in CNS myelination. *Neuron Glia Biology* **1**:65–72.

Bartnik, A., Nirmal, A.J. and Yang, S.-Y. (2012). Peptide vaccine therapy in colorectal cancer. *Vaccines* **1**:1–16.

Bassani-Sternberg, M., Chong, C., Guillaume, P., Solleder, M., Pak, H., Gannon, P.O., Kandalaft, L.E., *et al.* (2017). Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity. *PLoS Computational Biology* **13**:e1005725.

Bautz, D.J., Sherpa, A.T. and Threadgill, D.W. (2017). Prophylactic vaccination targeting ERBB3 decreases polyp burden in a mouse model of human colorectal cancer. *OncoImmunology* **6**:e1255395.

Baxevanis, C.N. (2019). T-cell recognition of non-mutated tumor antigens in healthy individuals: connecting endogenous immunity and tumor dormancy. *Cancer Immunology, Immunotherapy* **68**:705–707.

Berd, D., Sato, T., Maguire, H.C., Kairys, J. and Mastrangelo, M.J. (2004). Immunopharmacologic analysis of an autologous, hapten-modified human melanoma vaccine. *Journal of Clinical Oncology* **22**:403–415.

Bhatheja, K. and Field, J. (2006). Schwann cells: Origins and role in axonal maintenance and regeneration. *The International Journal of Biochemistry & Cell Biology* **38**:1995–1999.

Binkowski, T.A., Marino, S.R. and Joachimiak, A. (2012). Predicting HLA class I non-permissive amino acid residues substitutions. *PLoS ONE* **7**:e41710.

Birkeland, S.A., Storm, H.H., Lamm, L.U., Barlow, L., Blohmé, I., Forsberg, B., Eklund, B., *et al.* (1995). Cancer risk after renal transplantation in the nordic countries, 1964–1986. *International Journal of Cancer* **60**:183–189.

Biswas, D., Marnane, C.N., Mal, R. and Baldwin, D. (2007). Extracranial head and neck schwannomas—A 10-year review. *Auris Nasus Larynx* **34**:353–359.

Blaine, D.P. (1810). *A Domestic Treatise on the Diseases of Horses and Dogs ...* . 4th ed., with very large additions. London,: T. Boosey,.

Blaj, C., Schmidt, E.M., Lamprecht, S., Hermeking, H., Jung, A., Kirchner, T. and Horst, D. (2017). Oncogenic effects of high MAPK activity in colorectal cancer mark progenitor cells and persist irrespective of RAS mutations. *Cancer Research* **77**:1763–1774.

Blanpain, C. (2013). Tracing the cellular origin of cancer. *Nature Cell Biology* **15**:126–134.

Blockus, H. and Chédotal, A. (2016). Slit-Robo signaling. *Development* **143**:3037–3044.

Bonsack, M., Hoppe, S., Winter, J., Tichy, D., Zeller, C., Küpper, M., Schitter, E.C., *et al.* (2019). Performance evaluation of MHC class-I binding prediction tools based on an experimentally validated MHC-peptide binding dataset. *Cancer Immunology Research* **18**:0584.

Bouallegui, Y. (2019). Immunity in mussels: An overview of molecular components and mechanisms with a focus on the functional defenses. *Fish & Shellfish Immunology* **89**:158–169.

Braud, V.M., Allan, D.S.J. and McMichael, A.J. (1999). Functions of nonclassical MHC and non-MHC-encoded class I molecules. *Current Opinion in Immunology* **11**:100–108.

Briske-Anderson, M.J., Finley, J.W. and Newman, S.M. (1997). The influence of culture time and passage number on the morphological and physiological development of caco-2 cells. *Proceedings of the Society for Experimental Biology and Medicine* **214**:248–257.

Brose, K., Bland, K.S., Wang, K.H., Arnott, D., Henzel, W., Goodman, C.S., Tessier-Lavigne, M., *et al.* (1999). Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive axon guidance. *Cell* **96**:795–806.

Brown, G.K., Kreiss, A., Lyons, A.B. and Woods, G.M. (2011). Natural killer cell mediated cytotoxic responses in the Tasmanian Devil. *PLoS ONE* **6**:e24475.

Brown, G.K., Tovar, C., Cooray, A.A., Kreiss, A., Darby, J., Murphy, J.M., Corcoran, L.M., *et al.* (2016). Mitogen-activated Tasmanian devil blood mononuclear cells kill devil facial tumour disease cells. *Immunology & Cell Biology* **94**:673–679.

Brown, R.S., Wolke, R.E., Saila, S.B. and Brown, C.W. (1978). Prevalence of neoplasia in 10 New England populations of the soft-shell clam (Mya arenaria). *Annals of the New York Academy of Sciences* **298**:522–534.

Bruijnzeel-Koomen, C.A.F.M., Schmidt-Weber, C.B., Knol, E.F., Van Hoffen, E., Garssen, J., Tiemessen, M.M. and Kunzmann, S. (2003). Transforming growth factor-β inhibits human antigen-specific CD4+ T cell proliferation without modulating the cytokine response. *International Immunology* **15**:1495–1504.

Butte, M.J., Keir, M.E., Phamduy, T.B., Sharpe, A.H. and Freeman, G.J. (2007). Programmed Death-1 Ligand 1 interacts specifically with the B7-1 costimulatory molecule to inhibit T cell responses. *Immunity* **27**:111–122.

Caldwell, A., Coleby, R., Tovar, C., Stammnitz, M.R., Kwon, Y.M., Owen, R.S., Tringides, M., *et al.* (2018). The newly-arisen Devil facial tumour disease 2 (DFT2) reveals a mechanism for the emergence of a contagious cancer. *eLife* **7**:e35314.

Calis, J.J.A., Maybeno, M., Greenbaum, J.A., Weiskopf, D., De Silva, A.D., Sette, A., Ke?mir, C., *et al.* (2013). Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS*

Computational Biology **9**:e1003266.

Cancer Research UK (2019). *Oligodendroglioma* [Online]. Available at: https://www.cancerresearchuk.org/about-cancer/brain-tumours/types/oligodendroglioma.

Caron, E., Espona, L., Kowalewski, D.J., Schuster, H., Ternette, N., Alpízar, A., Schittenhelm, R.B., *et al.* (2015). An open-source computational and data resource to analyze digital maps of immunopeptidomes. *eLife* **4**:e07661.

Caron, E., Kowalewski, D.J., Chiek Koh, C., Sturm, T., Schuster, H. and Aebersold, R. (2015). Analysis of Major Histocompatibility Complex (MHC) immunopeptidomes using mass spectrometry. *Molecular & Cellular Proteomics* **14**:3105–3117.

Castelnovo, L.F., Bonalume, V., Melfi, S., Ballabio, M., Colleoni, D. and Magnaghi, V. (2017). Schwann cell development, maturation and regeneration: a focus on classic and emerging intracellular signaling pathways. *Neural Regeneration Research* **12**:1013–1023.

Chauhan, V., Rungta, T., Goyal, K. and Singh, M.P. (2019). Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Scientific Reports* **9**:2517.

Chen, Y., Tian, D., Ku, L., Osterhout, D.J. and Feng, Y. (2007). The selective RNA-binding protein Quaking I (QKI) is necessary and sufficient for promoting oligodendroglia differentiation. *Journal of Biological Chemistry* **282**:23553–23560.

Cheng, J., Torkamani, A., Grover, R.K., Jones, T.M., Ruiz, D.I., Schork, N.J., Quigley, M.M., *et al.* (2011). Ectopic B-cell clusters that infiltrate transplanted human kidneys are clonal. *Proceedings of the National Academy of Sciences of the United States of America* **108**:5560–5565.

Cheng, Y. and Belov, K. (2014). Characterisation of non-classical MHC class I genes in the Tasmanian devil (Sarcophilus harrisii). *Immunogenetics* **66**:727–735.

Cheng, Y. and Belov, K. (2012). Isolation and characterisation of 11 MHC-linked microsatellite loci in the Tasmanian devil (Sarcophilus harrisii). *Conservation Genetics Resources* **4**:463–465.

Cheng, Y., Makara, M., Peel, E., Fox, S., Papenfuss, A.T. and Belov, K. (2019). Tasmanian devils with contagious cancer exhibit a constricted T-cell repertoire diversity. *Communications Biology* **2**:99.

Cheng, Y., Stuart, A., Morris, K., Taylor, R., Siddle, H., Deakin, J., Jones, M., *et al.* (2012). Antigen-presenting genes and genomic copy number variations in the Tasmanian devil MHC. *BMC Genomics* **13**:1–14.

Chiang, E.Y. and Stroynowski, I. (2004). A nonclassical MHC class I molecule restricts CTL-mediated rejection of a syngeneic melanoma tumor. *The Journal of Immunology* **173**:4394–4401.

Chimal-Ramírez, G.K., Espinoza-Sánchez, N.A. and Fuentes-Pananá, E.M. (2013). Protumor

activities of the immune response: insights in the mechanisms of immunological shift, oncotraining, and oncopromotion. *Journal of oncology* **2013**:835956.

Chong, Y. and Ikematsu, H. (2018). Is seasonal vaccination a contributing factor to the selection of influenza epidemic variants? *Human vaccines & immunotherapeutics* **14**:518–522.

Clark, A.J., Kaller, M.S., Galino, J., Willison, H.J., Rinaldi, S. and Bennett, D.L.H. (2017). Co-cultures with stem cell-derived human sensory neurons reveal regulators of peripheral myelination. *Brain* **140**:898–913.

Codner, G.F., Birch, J., Hammond, J.A. and Ellis, S.A. (2012). Constraints on haplotype structure and variable gene frequencies suggest a functional hierarchy within cattle MHC class I. *Immunogenetics* **64**:435–445.

Collins, B.C., Hunter, C.L., Liu, Y., Schilling, B., Rosenberger, G., Bader, S.L., Chan, D.W., *et al.* (2017). Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nature Communications* **8**:291.

Collinson, I.R., Skehel, J.M., Fearnley, I.M., Runswick, M.J. and Walker, J.E. (1996). The F1Fo-ATPase complex from bovine heart mitochondria: the molar ratio of the subunits in the stalk region linking the F1 and Fo domains. *Biochemistry* **35**:12640–12646.

Consortium, T.U. (2018). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**:D506–D515.

Corfas, G., Velardez, M.O., Ko, C.-P., Ratner, N. and Peles, E. (2004). Mechanisms and roles of axon-Schwann cell interactions. *The Journal of Neuroscience* **24**:9250–9260.

Corthay, A., Skovseth, D.K., Lundin, K.U., Røsjø, E., Omholt, H., Hofgaard, P.O., Haraldsen, G., *et al.* (2005). Primary antitumor immune response mediated by CD4+ T cells. *Immunity* **22**:371–383.

Cruz, F.M., Colbert, J.D., Merino, E., Kriegsman, B.A. and Rock, K.L. (2017). The biology and underlying mechanisms of cross-presentation of exogenous antigens on MHC-I molecules. *Annual Review of Immunology* **35**:149–176.

Danesin, C., Agius, E., Escalas, N., Ai, X., Emerson, C., Cochard, P. and Soula, C. (2006). Ventral neural progenitors switch toward an oligodendroglial fate in response to increased Sonic Hedgehog (Shh) activity: involvement of sulfatase 1 in modulating Shh signaling in the ventral spinal cord. *The Journal of Neuroscience* **26**:5037–5048.

Dash, P., Fiore-Gartland, A.J., Hertz, T., Wang, G.C., Sharma, S., Souquette, A., Crawford, J.C., *et al.* (2017). Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**:89–93.

Deakin, J.E., Bender, H.S., Pearse, A.-M., Rens, W., O'Brien, P.C.M., Ferguson-Smith, M.A., Cheng,

Y., *et al.* (2012). Genomic restructuring in the Tasmanian Devil facial tumour: chromosome painting and gene mapping provide clues to evolution of a transmissible tumour. *PLoS Genetics* **8**:e1002483.

Dent, K.A., Christie, K.J., Bye, N., Basrai, H.S., Turbic, A., Habgood, M., Cate, H.S., *et al.* (2015). Oligodendrocyte birth and death following traumatic brain injury in adult mice. *PLoS ONE* **10**:e0121541.

Deteix, C., Attuil-Audenis, V., Duthey, A., Patey, N., McGregor, B., Dubois, V., Caligiuri, G., *et al.* (2010). Intragraft Th17 Infiltrate Promotes Lymphoid Neogenesis and Hastens Clinical Chronic Rejection. *The Journal of Immunology* **184**:5344 LP – 5351.

Devine, L., Sun, J., Barr, M.R. and Kavathas, P.B. (1999). Orientation of the Ig domains of CD8αβ relative to MHC class I. *The Journal of Immunology* **162**:846–851.

Doddrell, R.D.S., Dun, X.-P., Moate, R.M., Jessen, K.R., Mirsky, R. and Parkinson, D.B. (2012). Regulation of Schwann cell differentiation and proliferation by the Pax-3 transcription factor. *Glia* **60**:1269–1278.

Donkor, M.K., Sarkar, A. and Li, M.O. (2012). Tgf-β1 produced by activated CD4(+) T cells antagonizes T cell surveillance of tumor development. *OncoImmunology* **1**:162–171.

Doxiadis, G.G.M., de Groot, N., Otting, N., Blokhuis, J.H. and Bontrop, R.E. (2011). Genomic plasticity of the MHC class I A region in rhesus macaques: extensive haplotype diversity at the population level as revealed by microsatellites. *Immunogenetics* **63**:73–83.

Drijfhout, J.W., Brandt, R.M.P., D'Amaro, J., Kast, W.M. and Melief, C.J.M. (1995). Detailed motifs for peptide binding to HLA-A∗0201 derived from large random sets of peptides using a cellular binding assay. *Human Immunology* **43**:1–12.

Epstein, B., Jones, M., Hamede, R., Hendricks, S., McCallum, H., Murchison, E.P., Schönfeld, B., *et al.* (2016). Rapid evolutionary response to a transmissible cancer in Tasmanian devils. *Nature Communications* **7**:12684.

Evans, E.K., Gardino, A.K., Kim, J.L., Hodous, B.L., Shutes, A., Davis, A., Zhu, X.J., *et al.* (2017). A precision therapy against cancers driven by KIT/PDGFRA mutations. *Science Translational Medicine* **9**:eaao1690.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., *et al.* (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Research* **46**:649–655.

Falcão, A.M., van Bruggen, D., Marques, S., Meijer, M., Jäkel, S., Agirre, E., Samudyata, *et al.* (2018). Disease-specific oligodendrocyte lineage cells arise in multiple sclerosis. *Nature Medicine* **24**:1837–1844.

Falk, K., Rötzschke, O., Stevanović, S., Jung, G. and Rammensee, H.-G. (1991). Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**:290–296.

Fancy, S.P.J., Zhao, C. and Franklin, R.J.M. (2004). Increased expression of Nkx2.2 and Olig2 identifies reactive oligodendrocyte progenitor cells responding to demyelination in the adult CNS. *Molecular and Cellular Neuroscience* **27**:247–254.

Fassati, A. and Mitchison, N.A. (2010). Testing the theory of immune selection in cancers that break the rules of transplantation. *Cancer Immunology and Immunotherapy* **59**:643–651.

Fehlings, M., Chakarov, S., Simoni, Y., Sivasankar, B., Ginhoux, F. and Newell, E.W. (2018). Multiplex peptide-MHC tetramer staining using mass cytometry for deep analysis of the influenza-specific T-cell response in mice. *Journal of Immunological Methods* **453**:30–36.

Finley, D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annual Review of Biochemistry* **78**:477–513.

Finzsch, M., Schreiner, S., Kichko, T., Reeh, P., Tamm, E.R., Bösl, M.R., Meijer, D., *et al.* (2010). Sox10 is required for Schwann cell identity and progression beyond the immature Schwann cell stage. *The Journal of Cell Biology* **189**:701–712.

Flajnik, M.F. and Kasahara, M. (2001). Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* **15**:351–362.

Flies, A.S., Blackburn, N.B., Lyons, A.B., Hayball, J.D. and Woods, G.M. (2017). Comparative Analysis of Immune Checkpoint Molecules and Their Potential Role in the Transmissible Tasmanian Devil Facial Tumor Disease. *Frontiers in Immunology* **8**.

Flies, A.S., Lyons, A.B., Corcoran, L.M., Papenfuss, A.T., Murphy, J.M., Knowles, G.W., Woods, G.M., *et al.* (2016). PD-L1 is not constitutively expressed on Tasmanian Devil facial tumor cells but is strongly upregulated in response to IFN-γ and can be expressed in the tumor microenvironment. *Frontiers in Immunology* **7**:581.

Focosi, D., De Donno, M., Barbuti, S., Davini, S., Fornaciari, S., Curcio, M., Mariotti, M.L., *et al.* (2011). Cancer transmissibility across HLA barriers between immunocompetent individuals: rare but not impossible. *Human Immunology* **72**:1–4.

Fox, I.J. and Chowdhury, J.R. (2004). Hepatocyte transplantation. *American Journal of Transplantation* **4**:7–13.

Frampton, D., Schwenzer, H., Marino, G., Butcher, L.M., Pollara, G., Kriston-Vizi, J., Venturini, C., *et al.* (2018). Molecular signatures of regression of the Canine Transmissible Venereal Tumor. *Cancer cell* **33**:620-633.e6.

Fremont, D.H., Stura, E.A., Matsumura, M., Peterson, P.A. and Wilson, I.A. (1995). Crystal

structure of an H-2Kb-ovalbumin peptide complex reveals the interplay of primary and secondary anchor positions in the major histocompatibility complex binding groove. *Proceedings of the National Academy of Sciences* **92**:2479.

Fu, R., Zhang, Y.-W., Li, H.-M., Lv, W.-C., Zhao, L., Guo, Q.-L., Lu, T., *et al.* (2018). LW106, a novel indoleamine 2,3-dioxygenase 1 inhibitor, suppresses tumour progression by limiting stroma-immune crosstalk and cancer stem cell enrichment in tumour micro-environment. *British journal of pharmacology* **175**:3034–3049.

Fujiwara, S., Hoshikawa, S., Ueno, T., Hirata, M., Saito, T., Ikeda, T., Kawaguchi, H., *et al.* (2014). SOX10 transactivates S100B to suppress Schwann cell proliferation and to promote myelination. *PLoS ONE* **9**:e115400.

Fukami, N., Ramachandran, S., Saini, D., Walter, M., Chapman, W., Patterson, G.A. and Mohanakumar, T. (2009). Antibodies to MHC class I induce autoimmunity: role in the pathogenesis of chronic rejection. *Journal of immunology (Baltimore, Md. : 1950)* **182**:309–318.

Furuhashi, M. and Hotamisligil, G.S. (2008). Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. *Nature Reviews Drug discovery* **7**:489–503.

Furuta, E. and Yamaguchi, K. (2011). Transplant rejection in terrestrial molluscs. *Invertebrate Survival Journal* **8**:15–20.

Galluzzi, L., Vacchelli, E., Eggermont, A., Fridman, W.H., Galon, J., Sautès-Fridman, C., Tartour, E., *et al.* (2012). Trial Watch. *OncoImmunology* **1**:699–739.

Gao, Y., Tao, J., Li, M.O., Zhang, D., Chi, H., Henegariu, O., Kaech, S.M., *et al.* (2005). JNK1 is essential for CD8+ T cell-mediated tumor immune surveillance. *The Journal of Immunology* **175**:5783–5789.

Garrido, C., Garrido, F., Berruguilla, E., Romero, I., Paco, L., Garcia-Lora, A.M., Stefansky, J., *et al.* (2012). MHC class I molecules act as tumor suppressor genes regulating the cell cycle gene expression, invasion and intrinsic tumorigenicity of melanoma cells. *Carcinogenesis* **33**:687–693.

Garrido, F., Perea, F., Bernal, M., Sánchez-Palencia, A., Aptsiauri, N. and Ruiz-Cabello, F. (2017). The escape of cancer from T cell-mediated immune surveillance: HLA class I loss and tumor tissue architecture. *Vaccines* **5**:7.

Garstka, M.A., Fish, A., Celie, P.H.N., Joosten, R.P., Janssen, G.M.C., Berlin, I., Hoppes, R., *et al.* (2015). The first step of peptide selection in antigen presentation by MHC class I molecules. *Proceedings of the National Academy of Sciences* **112**:1505–1510.

Gato-Cañas, M., Zuazo, M., Arasanz, H., Ibañez-Vea, M., Lorenzo, L., Fernandez-Hinojal, G., Vera, R., *et al.* (2017). PDL1 signals through conserved sequence motifs to overcome interferon-

mediated cytotoxicity. *Cell Reports* **20**:1818–1829.

Gestal, C., Roch, P., Renault, T., Pallavicini, A., Paillard, C., Novoa, B., Oubella, R., *et al.* (2008). Study of diseases and the immune system of bivalves using molecular biology and genomics. *Reviews in Fisheries Science* **16**:133–156.

Gfeller, D., Guillaume, P., Michaux, J., Pak, H.-S., Daniel, R.T., Racle, J., Coukos, G., *et al.* (2018). The length distribution and multiple specificity of naturally presented HLA-I ligands. *The Journal of Immunology* **201**:3705–3716.

Gilbert, K.M., Thoman, M., Bauche, K., Pham, T. and Weigle, W.O. (1997). Transforming growth factor-β1 induces antigen-specific unresponsiveness in naive t cells. *Immunological Investigations* **26**:459–472.

Gillespie, C.S., Sherman, D.L., Blair, G.E. and Brophy, P.J. (1994). Periaxin, a novel protein of myelinating Schwann cells with a possible role in axonal ensheathment. *Neuron* **12**:497–508.

Glezer, I., Lapointe, A. and Rivest, S. (2006). Innate immunity triggers oligodendrocyte progenitor reactivity and confines damages to brain injuries. *The FASEB Journal* **20**:750–752.

Goertz, O., Langer, S., Uthoff, D., Ring, A., Stricker, I., Tannapfel, A. and Steinau, H.U. (2014). Diagnosis, treatment and survival of 65 patients with malignant peripheral nerve sheath tumors. *Anticancer Res* **34**:777–783.

Goldstein, A.S. and Witte, O.N. (2013). Does the microenvironment influence the cell types of origin for prostate cancer? *Genes & Development* **27**:1539–1544.

Grey, H.M., Kubo, R.T., Colon, S.M., Poulik, M.D., Cresswell, P., Springer, T., Turner, M., *et al.* (1973). The small subunit of HL-A antigens is beta 2-microglobulin. *The Journal of Experimental Medicine* **138**:1608–1612.

Griffin, J.W. and Thompson, W.J. (2008). Biology and pathology of nonmyelinating Schwann cells. *Glia* **56**:1518–1531.

Griner, L.A. (1979). Neoplasms in Tasmanian Devils (Sarcophilus harrisii). *Journal of the National Cancer Institute* **62**:589–595.

Hachem, S., Aguirre, A., Vives, V., Marks, A., Gallo, V. and Legraverend, C. (2005). Spatial and temporal expression of S100B in cells of oligodendrocyte lineage. *Glia* **51**:81–97.

Hamede, R., Lachish, S., Belov, K., Woods, G., Kreiss, A., Pearse, A.-. M., Lazenby, B., *et al.* (2011). Reduced effect of Tasmanian Devil facial tumor disease at the disease front. *Conservation Biology* **26**:124–134.

Hanahan, D. and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* **144**:646–

674.

Hanahan, D. and Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* **100**:57–70.

Hannoun, Z., Lin, Z., Brackenridge, S., Kuse, N., Akahoshi, T., Borthwick, N., McMichael, A., *et al.* (2018). Identification of novel HIV-1-derived HLA-E-binding peptides. *Immunology Letters* **202**:65–72.

Hansen, A.M., Rasmussen, M., Svitek, N., Harndahl, M., Golde, W.T., Barlow, J., Nene, V., *et al.* (2014). Characterization of binding specificities of bovine leucocyte class I molecules: impacts for rational epitope discovery. *Immunogenetics* **66**:705–718.

Haririan, A., Nogueira, J., Kukuruga, D., Schweitzer, E., Hess, J., Gurk-Turner, C., Jacobs, S., *et al.* (2009). Positive cross-match living donor kidney transplantation: longer-term outcomes. *American Journal of Transplantation* **9**:536–542.

Harper, S.J.F., Ali, J.M., Wlodek, E., Negus, M.C., Harper, I.G., Chhabra, M., Qureshi, M.S., *et al.* (2015). CD8 T-cell recognition of acquired alloantigen promotes acute allograft rejection. *Proceedings of the National Academy of Sciences* **112**:12788–12793.

van Hateren, A., Bailey, A. and Elliott, T. (2017). Recent advances in Major Histocompatibility Complex (MHC) class I antigen presentation: Plastic MHC molecules and TAPBPR-mediated quality control. *F1000Research* **6**:158.

van Hateren, A., Carter, R., Bailey, A., Kontouli, N., Williams, A.P., Kaufman, J. and Elliott, T. (2013). A mechanistic basis for the co-evolution of chicken tapasin and Major Histocompatibility Complex class I (MHC I) proteins. *Journal of Biological Chemistry* **288**:32797–32808.

Hawkins, C.E., Baars, C., Hesterman, H., Hocking, G.J., Jones, M.E., Lazenby, B., Mann, D., *et al.* (2006). Emerging disease and population decline of an island endemic, the Tasmanian devil Sarcophilus harrisii. *Biological Conservation* **131**:307–324.

Hawkins, C.E., McCallum, H., Mooney, N., Jones, M. and Holdsworth, M. (2008). *Sarcophilus Harrisii.*

Haworth, K.B., Leddon, J.L., Chen, C.-Y., Horwitz, E.M., Mackall, C.L. and Cripe, T.P. (2015). Going back to class I: MHC and immunotherapies for childhood cancer. *Pediatric Blood & Cancer* **62**:571–576.

Haynes-Gilmore, N., Banach, M., Edholm, E.-S., Lord, E. and Robert, J. (2014). A critical role of non-classical MHC in tumor immune evasion in the amphibian Xenopus model. *Carcinogenesis* **35**:1807–1813.

Hennecke, J. and Wiley, D.C. (2001). T cell receptor - MHC interactions up close. *Cell* **104**:1–4.

Hirsch, M.R., Wietzerbin, J., Pierres, M. and Goridis, C. (1983). Expression of Ia antigens by cultured astrocytes treated with gamma-interferon. *Neuroscience Letters* **41**:199–204.

Höftberger, R., Aboul-Enein, F., Brueck, W., Lucchinetti, C., Rodriguez, M., Schmidbauer, M., Jellinger, K., *et al.* (2004). Expression of Major Histocompatibility Complex class I molecules on the different cell types in multiple sclerosis lesions. *Brain Pathology* **14**:43–50.

Hörig, H., Young, A.C.M., Papadopoulos, N.J., DiLorenzo, T.P. and Nathenson, S.G. (1999). Binding of longer peptides to the H-2Kb heterodimer is restricted to peptides extended at their C terminus: refinement of the inherent MHC class I peptide binding criteria. *The Journal of Immunology* **163**:4434–4441.

Horiuchi, M., Lindsten, T., Pleasure, D. and Itoh, T. (2010). Differing in vitro survival dependency of mouse and rat NG2+ oligodendroglial progenitor cells. *Journal of Neuroscience Research* **88**:957–970.

Howcroft, T.K. and Singer, D.S. (2003). Expression of nonclassical MHC class Ib genes. *Immunologic Research* **27**:1–29.

Howe, C., Garstka, M., Al-Balushi, M., Ghanem, E., Antoniou, A.N., Fritzsche, S., Jankevicius, G., *et al.* (2009). Calreticulin-dependent recycling in the early secretory pathway mediates optimal peptide loading of MHC class I molecules. *The EMBO Journal* **28**:3730–3744.

Hsiao, Y.W., Liao, K.W., Hung, S.W. and Chu, R.M. (2002). Effect of tumor infiltrating lymphocytes on the expression of MHC molecules in canine transmissible venereal tumor cells. *Veterinary Immunology and Immunopathology* **87**:19–27.

Hunt, J.S., Petroff, M.G., Morales, P., Sedlmayr, P., Geraghty, D.E. and Ober, C. (2000). HLA-G in reproduction: studies on the maternal–fetal interface. *Human Immunology* **61**:1113–1117.

Hunt, S.E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., Parton, A., *et al.* (2018). Ensembl variation resources. *Database* **2018**:119.

Hyung, S., Im, S.-K., Lee, B.Y., Shin, J., Park, J.-C., Lee, C., Suh, J.-K.F., *et al.* (2019). Dedifferentiated Schwann cells secrete progranulin that enhances the survival and axon growth of motor neurons. *Glia* **67**:360–375.

Hyung, S., Yoon Lee, B., Park, J.-C., Kim, J., Hur, E.-M. and Francis Suh, J.-K. (2015). Coculture of primary motor neurons and Schwann cells as a model for in vitro myelination. *Scientific Reports* **5**:15122.

Imai, D., Yoshizumi, T., Okano, S., Uchiyama, H., Ikegami, T., Harimoto, N., Itoh, S., *et al.* (2017). The prognostic impact of programmed cell death ligand 1 and human leukocyte antigen class I in pancreatic cancer. *Cancer Medicine* **6**:1614–1626.

Ince, T.A., Richardson, A.L., Bell, G.W., Saitoh, M., Godar, S., Karnoub, A.E., Iglehart, J.D., *et al.* (2007). Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. *Cancer Cell* **12**:160–170.

Iorgulescu, J.B., Braun, D., Oliveira, G., Keskin, D.B. and Wu, C.J. (2018). Acquired mechanisms of immune escape in cancer following immunotherapy. *Genome medicine* **10**:87.

Ishitani, A., Sageshima, N., Lee, N., Dorofeeva, N., Hatake, K., Marquardt, H. and Geraghty, D.E. (2003). Protein expression and peptide binding suggest unique and interacting functional roles for HLA-E, F, and G in maternal-placental immune recognition. *The Journal of Immunology* **171**:1376–1384.

Jain, S. and Baranwal, M. (2019). Conserved peptide vaccine candidates containing multiple Ebola nucleoprotein epitopes display interactions with diverse HLA molecules. *Medical Microbiology and Immunology*.

James, E., Bailey, I., Sugiyarto, G. and Elliott, T. (2013). Induction of protective antitumor immunity through attenuation of ERAAP function. *The Journal of Immunology* **190**:5839–5846.

James, S., Jennings, G., Kwon, Y.M., Stammnitz, M., Fraik, A., Storfer, A., Comte, S., *et al.* (2019). Tracing the rise of malignant cell lines: Distribution, epidemiology and evolutionary interactions of two transmissible cancers in Tasmanian devils. *Evolutionary Applications* **0**.

Janeway, C., Travers, P., Walport, M. and C., S.M. (2001). The major histocompatibility complex and its functions. In: *Immunobiology: The Immune System in Health and Disease*. 5th ed. Garland Science.

Jang, B., Jeon, Y.-C., Shin, H.-Y., Lee, Y.-J., Kim, H., Kondo, Y., Ishigami, A., *et al.* (2017). Myelin Basic Protein citrullination, a hallmark of central nervous system demyelination, assessed by novel monoclonal antibodies in prion diseases. *Molecular Neurobiology* **55**:3172–3184.

Jang, S.Y., Yoon, B.-A., Shin, Y.K., Yun, S.H., Jo, Y.R., Choi, Y.Y., Ahn, M., *et al.* (2017). Schwann cell dedifferentiation-associated demyelination leads to exocytotic myelin clearance in inflammatory segmental demyelination. *Glia* **65**:1848–1862.

Jessen, K.R., Brennan, A., Morgan, L., Mirsky, R., Kent, A., Hashimoto, Y. and Gavrilovic, J. (1994). The schwann cell precursor and its fate: A study of cell death and differentiation during gliogenesis in rat embryonic nerves. *Neuron* **12**:509–527.

Jessen, K.R. and Mirsky, R. (2005). The origin and development of glial cells in peripheral nerves. *Nature Reviews Neuroscience* **6**:671–682.

Jessen, K.R. and Mirsky, R. (2016). The repair Schwann cell and its function in regenerating nerves. *The Journal of physiology* **594**:3521–3531.

Jessen, K.R., Morgan, L., Stewart, H.J. and Mirsky, R. (1990). Three markers of adult non-myelin-forming Schwann cells, 217c(Ran-1), A5E3 and GFAP: development and regulation by neuron-Schwann cell interactions. *Development* **109**:91–103.

Jiang, H., Canfield, S.M., Gallagher, M.P., Jiang, H.H., Jiang, Y., Zheng, Z. and Chess, L. (2010). HLA-E-restricted regulatory CD8(+) T cells are involved in development and control of human autoimmune type 1 diabetes. *The Journal of clinical investigation* **120**:3641–3650.

Joglekar-Javadekar, M., Van Laere, S., Bourne, M., Moalwi, M., Finetti, P., Vermeulen, P.B., Birnbaum, D., *et al.* (2017). Characterization and targeting of Platelet-Derived Growth Factor Receptor alpha (PDGFRA) in Inflammatory Breast Cancer (IBC). *Neoplasia* **19**:564–573.

Jones, M.E., Paetkau, D., Geffen, E. and Moritz, C. (2004). Genetic diversity and population structure of Tasmanian devils, the largest marsupial carnivore. *Molecular Ecology* **13**:2197–2209.

Joyce, J.A. and Fearon, D.T. (2015). T cell exclusion, immune privilege, and the tumor microenvironment. *Science* **348**:74 LP – 80.

Jurewicz, A., Biddison, W.E. and Antel, J.P. (1998). MHC class I-restricted lysis of human oligodendrocytes by Myelin Basic Protein peptide-specific CD8 T lymphocytes. *The Journal of Immunology* **160**:3056–3059.

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B. and Nielsen, M. (2017). NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *The Journal of Immunology* **199**:3360–3368.

Kaewkhaw, R., Scutt, A.M. and Haycock, J.W. (2012). Integrated culture and purification of rat Schwann cells from freshly isolated adult tissue. *Nature Protocols* **7**:1996–2004.

Karbiener, M., Darnhofer, B., Frisch, M.-T., Rinner, B., Birner-Gruenberger, R. and Gugatschka, M. (2017). Comparative proteomics of paired vocal fold and oral mucosa fibroblasts. *Journal of Proteomics* **155**:11–21.

Karlsson, F.A., Groth, T., Sege, K., Wibell, L. and Peterson, P.E.R.A. (1980). Turnover in humans of β2-microglobulin: the constant chain of HLA-antigens. *European Journal of Clinical Investigation* **10**:293–300.

Kasahara, T., Hooks, J.J., Dougherty, S.F. and Oppenheim, J.J. (1983). Interleukin 2-mediated immune interferon (IFN-gamma) production by human T cells and T cell subsets. *The Journal of Immunology* **130**:1784–1789.

Kaur, G. and Dufour, J.M. (2012). Cell lines: Valuable tools or useless artifacts. *Spermatogenesis* **2**:1–5.

Kawamura, K., Kawamura, N., Okamoto, N. and Manabe, M. (2013). Suppression of

choriocarcinoma invasion and metastasis following blockade of BDNF/TrkB signaling. *Cancer Medicine* **2**:849–861.

Keenan, B.P., Saenger, Y., Kafrouni, M.I., Leubner, A., Lauer, P., Maitra, A., Rucki, A.A., *et al.* (2014). A listeria vaccine and depletion of T-regulatory cells activate immunity against early stage pancreatic intraepithelial neoplasms and prolong survival of mice. *Gastroenterology* **146**:1784–1794.

Kelley, J., Walter, L. and Trowsdale, J. (2005). Comparative genomics of major histocompatibility complexes. *Immunogenetics* **56**:683–695.

Kennedy, P.R., Barthen, C., Williamson, D.J. and Davis, D.M. (2019). HLA-B and HLA-C differ in their nanoscale organization at cell surfaces. *Frontiers in Immunology* **10**:61.

Keppel, M.P., Saucier, N., Mah, A.Y., Vogel, T.P. and Cooper, M.A. (2015). Activation-specific metabolic requirements for NK cell IFN-γ production. *The Journal of Immunology* **194**:1954–1962.

Kibbe, W.A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research* **35**:43–46.

Kilfoyle, D.H., Dyck, P.J., Wu, Y., Litchy, W.J., Klein, D.M., Dyck, P.J.B., Kumar, N., *et al.* (2006). Myelin protein zero mutation His39Pro: hereditary motor and sensory neuropathy with variable onset, hearing loss, restless legs and multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry* **77**:963–966.

Kim, H.-S., Lee, J., Lee, D.Y., Kim, Y.-D., Kim, J.Y., Lim, H.J., Lim, S., *et al.* (2017). Schwann cell precursors from human pluripotent stem cells as a potential therapeutic target for myelin repair. *Stem Cell Reports* **8**:1714–1726.

Kim, J.K., Lee, H.J. and Park, H.T. (2014). Two faces of Schwann cell dedifferentiation in peripheral neurodegenerative diseases: pro-demyelinating and axon-preservative functions. *Neural Regeneration Research* **9**:1952–1954.

Kimman, T.G., Vandebriel, R.J. and Hoebee, B. (2007). Genetic variation in the response to vaccination. *Community Genetics* **10**:201–217.

Kipp, M., Gingele, S., Pott, F., Clarner, T., van der Valk, P., Denecke, B., Gan, L., *et al.* (2011). BLBP-expression in astrocytes during experimental demyelination and in human multiple sclerosis lesions. *Brain, Behavior, and Immunity* **25**:1554–1568.

Kirveskari, J., He, Q., Leirisalo-Repo, M., Mäki-Ikola, O., Wuorela, M., Putto-Laurila, A. and Granfors, K. (1999). Enterobacterial infection modulates major histocompatibility complex class I expression on mononuclear cells. *Immunology* **97**:420–428.

Klein, D., Groh, J., Wettmarshausen, J. and Martini, R. (2014). Nonuniform molecular features of

myelinating Schwann cells in models for CMT1: Distinct disease patterns are associated with NCAM and c-Jun upregulation. *Glia* **62**:736–750.

Kobayashi, M., Ishibashi, S., Tomimitsu, H., Yokota, T. and Mizusawa, H. (2012). Proliferating immature Schwann cells contribute to nerve regeneration after ischemic peripheral nerve injury. *Journal of Neuropathology & Experimental Neurology* **71**:511–519.

Kochan, G., Escors, D., Breckpot, K. and Guerrero-Setas, D. (2013). Role of non-classical MHC class I molecules in cancer immunosuppression. *OncoImmunology* **2**:e26491.

Korkolopoulou, P., Kaklamanis, L., Pezzella, F., Harris, A.L. and Gatter, K.C. (1996). Loss of antigen-presenting molecules (MHC class I and TAP-1) in lung cancer. *British Journal of Cancer* **73**:148–153.

Kosack, L., Wingelhofer, B., Popa, A., Orlova, A., Agerer, B., Vilagos, B., Majek, P., *et al.* (2019). The ERBB-STAT3 axis drives Tasmanian Devil facial tumor disease. *Cancer Cell* **35**:125–139.

Koschmann, C., Zamler, D., MacKay, A., Robinson, D., Wu, Y.-M., Doherty, R., Marini, B., *et al.* (2016). Characterizing and targeting PDGFRA alterations in pediatric high-grade glioma. *Oncotarget* **7**:65696–65706.

Kowalewski, D.J., Schuster, H., Backert, L., Berlin, C., Kahn, S., Kanz, L., Salih, H.R., *et al.* (2015). HLA ligandome analysis identifies the underlying specificities of spontaneous antileukemia immune responses in chronic lymphocytic leukemia (CLL). *Proceedings of the National Academy of Sciences* **112**:166–175.

Kreiss, A., Brown, G.K., Tovar, C., Lyons, A.B. and Woods, G.M. (2015). Evidence for induction of humoral and cytotoxic immune responses against devil facial tumor disease cells in Tasmanian devils (Sarcophilus harrisii) immunized with killed cell preparations. *Vaccine* **33**:3016–3025.

Kreiss, A., Cheng, Y., Kimble, F., Wells, B., Donovan, S., Belov, K. and Woods, G.M. (2011). Allorecognition in the Tasmanian devil (Sarcophilus harrisii), an endangered marsupial species with limited genetic diversity. *PLoS ONE* **6**:e22402.

Kreiss, A., Fox, N., Bergfeld, J., Quinn, S.J., Pyecroft, S. and Woods, G.M. (2008). Assessment of cellular immune responses of healthy and diseased Tasmanian devils (Sarcophilus harrisii). *Developmental & Comparative Immunology* **32**:544–553.

Kulis, M. and Esteller, M. (2010). 2 - DNA Methylation and Cancer. In: Herceg, Z. and Ushijima, T. B. T.-A. in G. (eds.) *Epigenetics and Cancer, Part A*. Academic Press, pp. 27–56.

Kuzushima, K., Hayashi, N., Kimura, H. and Tsurumi, T. (2001). Efficient identification of HLA-A*2402–restricted cytomegalovirus-specific CD8+ T-cell epitopes by a computer algorithm and an enzyme-linked immunospot assay. *Blood* **98**:1872–1881.

Kwa, M.S., van Schaik, I.N., De Jonge, R.R., Brand, A., Kalaydjieva, L., van Belzen, N., Vermeulen, M., *et al.* (2003). Autoimmunoreactivity to Schwann cells in patients with inflammatory neuropathies. *Brain* **126**:361–375.

Kwon, Y.M., R, S.M., Jinhong, W., Kate, S., W, K.G., J, P.R., Alexandre, K., *et al.* (2018). Tasman-PCR: a genetic diagnostic assay for Tasmanian devil facial tumour diseases. *Royal Society Open Science* **5**:180870.

Lachmann, H.J., Lowe, P., Felix, S.D., Rordorf, C., Leslie, K., Madhoo, S., Wittkowski, H., *et al.* (2009). In vivo regulation of interleukin 1beta in patients with cryopyrin-associated periodic syndromes. *The Journal of Experimental Medicine* **206**:1029–1036.

Lakkis, F.G., Arakelov, A., Konieczny, B.T. and Inoue, Y. (2000). Immunologic 'ignorance' of vascularized organ transplants in the absence of secondary lymphoid tissue. *Nature Medicine* **6**:686–688.

Lakkis, F.G. and Lechler, R.I. (2013). Origin and biology of the allogeneic response. *Cold Spring Harbor Perspectives in Medicine* **3**:a014993.

Lane, A., Cheng, Y., Wright, B., Hamede, R., Levan, L., Jones, M., Ujvari, B., *et al.* (2012). New insights into the role of MHC diversity in Devil Facial Tumour Disease. *PLoS ONE* **7**:e36955.

Larochelle, S. (2016). Protein isoforms: more than meets the eye. *Nature Methods* **13**:291.

Lazenby, B.T., Tobler, M.W., Brown, W.E., Hawkins, C.E., Hocking, G.J., Hume, F., Huxtable, S., *et al.* (2018). Density trends and demographic signals uncover the long-term impact of transmissible cancer in Tasmanian devils. *Journal of Applied Ecology* **55**:1368–1379.

Le, T.M., Le, Q.V.C., Truong, D.M., Lee, H.-J., Choi, M.-K., Cho, H., Chung, H.-J., *et al.* (2017). β2-microglobulin gene duplication in cetartiodactyla remains intact only in pigs and possibly confers selective advantage to the species. *PLoS ONE* **12**:e0182322.

Lee, P.R., Cohen, J.E. and Fields, R.D. (2006). Immune system evasion by peripheral nerve sheath tumor. *Neuroscience Letters* **397**:126–129.

Lee, S.-J., Jang, B.-C., Lee, S.-W., Yang, Y.-I., Suh, S.-I., Park, Y.-M., Oh, S., *et al.* (2006). Interferon regulatory factor-1 is prerequisite to the constitutive expression and IFN-γ-induced upregulation of B7-H1 (CD274). *FEBS Letters* **580**:755–762.

Lee, Y., Morrison, B.M., Li, Y., Lengacher, S., Farah, M.H., Hoffman, P.N., Liu, Y., *et al.* (2012). Oligodendroglia metabolically support axons and contribute to neurodegeneration. *Nature* **487**:443–448.

Leone, P., Shin, E.-C., Perosa, F., Vacca, A., Dammacco, F. and Racanelli, V. (2013). MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells.

Journal of the National Cancer Institute **105**:1172–1187.

Li, M.O., Wan, Y.Y. and Flavell, R.A. (2007). T cell-produced Transforming Growth Factor-β1 controls T cell tolerance and regulates Th1- and Th17-cell differentiation. *Immunity* **26**:579–591.

Li, S., Mwakalundwa, G. and Skinner, P.J. (2017). In situ MHC-tetramer staining and quantitative analysis to determine the location, abundance, and phenotype of antigen-specific CD8 T cells in tissues. *Journal of Visualized Experiments*:56130.

Li, X.-J., Zhang, X., Lin, A., Ruan, Y.-Y. and Yan, W.-H. (2012). Human leukocyte antigen-G (HLA-G) expression in cervical cancer lesions is associated with disease progression. *Human Immunology* **73**:946–949.

Lilje, O. and Armati, P.J. (1997). The distribution and abundance of MHC and ICAM-1 on Schwann cells in vitro. *Journal of Neuroimmunology* **77**:75–84.

Lin, C.M. and Gill, R.G. (2016). Direct and indirect allograft recognition: pathways dictating graft rejection mechanisms. *Current opinion in organ transplantation* **21**:40–44.

Linette, G.P., Stadtmauer, E.A., Maus, M. V, Rapoport, A.P., Levine, B.L., Emery, L., Litzky, L., *et al.* (2013). Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. *Blood* **122**:863–871.

Lisak, R.P., Bealmear, B., Benjamins, J. and Skoff, A. (1998). Inflammatory cytokines inhibit upregulation of glycolipid expression by Schwann cells in vitro. *Neurology* **51**:1661–1665.

Lisak, R.P., Bealmear, B., Benjamins, J.A. and Skoff, A.M. (2001). Interferon-γ, tumor necrosis factor-α, and transforming growth factor-β inhibit cyclic AMP-induced Schwann cell differentiation. *Glia* **36**:354–363.

Little, C.C. and Tyzzer, E.E. (1916). Further experimental studies on the inheritance of susceptibility to a Transplantable tumor, Carcinoma (J. W. A.) of the Japanese waltzing Mouse . *The Journal of Medical Research* **33**:393–453.

Liu, J.A. and Cheung, M. (2016). Neural crest stem cells and their potential therapeutic applications. *Developmental Biology* **419**:199–216.

Liu, Y., Beyer, A. and Aebersold, R. (2016). On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**:535–550.

Liu, Z., Hu, X., Cai, J., Liu, B., Peng, X., Wegner, M. and Qiu, M. (2007). Induction of oligodendrocyte differentiation by Olig2 and Sox10: Evidence for reciprocal interactions and dosage-dependent mechanisms. *Developmental Biology* **302**:683–693.

Liu, Z., Jin, Y.-Q., Chen, L., Wang, Y., Yang, X., Cheng, J., Wu, W., *et al.* (2015). Specific marker

expression and cell state of Schwann cells during culture in vitro. *PLoS ONE* **10**:e0123278.

Loh, R., Bergfeld, J., Hayes, D., O'Hara, A., Pyecroft, S., Raidal, S. and Sharpe, R. (2006). The pathology of Devil Facial Tumor Disease (DFTD) in Tasmanian Devils (Sarcophilus harrisii). *Veterinary Pathology Online* **43**:890–895.

Luimstra, J.J., Garstka, M.A., Roex, M.C.J., Redeker, A., Janssen, G.M.C., van Veelen, P.A., Arens, R., *et al.* (2018). A flexible MHC class I multimer loading system for large-scale detection of antigen-specific T cells. *The Journal of Experimental Medicine* **215**:1493–1504.

Ma, G.-F., Miao, Q., Zeng, X.-Q., Luo, T.-C., Ma, L.-L., Liu, Y.-M., Lian, J.-J., *et al.* (2013). Transforming Growth Factor-β1 and -β2 in gastric precancer and cancer and roles in tumor-cell interactions with peripheral blood mononuclear cells in vitro. *PLoS ONE* **8**:e54249.

Maekawa, N., Konnai, S., Ikebuchi, R., Okagawa, T., Adachi, M., Takagi, S., Kagawa, Y., *et al.* (2014). Expression of PD-L1 on canine tumor cells and enhancement of IFN-γ production from tumor-infiltrating cells by PD-L1 blockade. *PloS ONE* **9**:e98415.

Maibach, V. and Vigilant, L. (2019). Reduced bonobo MHC class I diversity predicts a reduced viral peptide binding ability compared to chimpanzees. *BMC Evolutionary Biology* **19**:14.

Mailliard, R.B., Egawa, S., Cai, Q., Kalinska, A., Bykovskaya, S.N., Lotze, M.T., Kapsenberg, M.L., *et al.* (2002). Complementary dendritic cell-activating function of CD8+ and CD4+ T cells: helper role of CD8+ T cells in the development of T helper type 1 responses. *The Journal of experimental medicine* **195**:473–483.

Malmberg, K.J., Carlsten, M., Bjorklund, A., Sohlberg, E., Bryceson, Y.T. and Ljunggren, H.G. (2017). Natural killer cell-mediated immunosurveillance of human cancer. *Seminars in Immunology* **31**:20–29.

Marijt, K.A., Blijleven, L., Verdegaal, E.M.E., Kester, M.G., Kowalewski, D.J., Rammensee, H.-G., Stevanović, S., *et al.* (2018). Identification of non-mutated neoantigens presented by TAP-deficient tumors. *The Journal of Experimental Medicine* **215**:2325–2337.

Martini, M., Testi, M.G., Pasetto, M., Picchio, M.C., Innamorati, G., Mazzocco, M., Ugel, S., *et al.* (2010). IFN-γ-mediated upmodulation of MHC class I expression activates tumor-specific immune response in a mouse model of prostate cancer. *Vaccine* **28**:3548–3557.

Mason, D. (1998). A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunology Today* **19**:395–404.

Mata, M., Alessi, D. and Fink, D.J. (1990). S100 is preferentially distributed in myelin-forming Schwann cells. *Journal of Neurocytology* **19**:432–442.

Matsumoto, T., Yokote, K., Take, A., Takemoto, M., Asaumi, S., Hashimoto, Y., Matsuda, M., *et al.*

(2000). Differential interaction of CrkII adaptor protein with platelet-derived growth factor α- and β-receptors is determined by its internal tyrosine phosphorylation. *Biochemical and Biophysical Research Communications* **270**:28–33.

Matsushita, H., Hosoi, A., Ueha, S., Abe, J., Fujieda, N., Tomura, M., Maekawa, R., *et al.* (2015). Cytotoxic T lymphocytes block tumor growth both by lytic activity and IFNγ-dependent cell-cycle arrest. *Cancer Immunology Research* **3**:26–36.

Matthews, M.A. and Duncan, D. (1971). A quantitative study of morphological changes accompanying the initiation and progress of myelin production in the dorsal funiculus of the rat spinal cord. *Journal of Comparative Neurology* **142**:1–22.

McCallum, H., Jones, M., Hawkins, C., Hamede, R., Lachish, S., Sinn, D.L., Beeton, N., *et al.* (2009). Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. *Ecology* **90**:3379–3392.

McCallum, H., Tompkins, D., Jones, M., Lachish, S., Marvanek, S., Lazenby, B., Hocking, G., *et al.* (2007). Distribution and Impacts of Tasmanian Devil Facial Tumor Disease. *EcoHealth* **4**:318–325.

McCarthy, B.J., Rankin, K.M., Aldape, K., Bondy, M.L., Brännström, T., Broholm, H., Feychting, M., *et al.* (2011). Risk factors for oligodendroglial tumors: a pooled international study. *Neuro-oncology* **13**:242–250.

McKnight, M. (2008). *Thylacinus Cynocephalus*.

Meister, D., Taimoory, S.M. and Trant, J.F. (2019). Unnatural amino acids improve affinity and modulate immunogenicity: Developing peptides to treat MHC type II autoimmune disorders. *Peptide Science* **111**:e24058.

Melief, C.J.M. and Kessler, J.H. (2017). Novel insights into the HLA class I immunopeptidome and T-cell immunosurveillance. *Genome Medicine* **9**:44.

Mengel, M., Sis, B., Haas, M., Colvin, R.B., Halloran, P.F., Racusen, L.C., Solez, K., *et al.* (2012). Banff 2011 Meeting report: new concepts in antibody-mediated rejection. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* **12**:563–570.

Messick, T.E., Smith, G.R., Soldan, S.S., McDonnell, M.E., Deakyne, J.S., Malecka, K.A., Tolvinski, L., *et al.* (2019). Structure-based design of small-molecule inhibitors of EBNA1 DNA binding blocks Epstein-Barr virus latent infection and tumor growth. *Science Translational Medicine* **11**:5612.

Metzger, M.J., Reinisch, C., Sherry, J. and Goff, S.P. (2015). Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams. *Cell* **161**:255–263.

Metzger, M.J., Villalba, A., Carballal, M.J., Iglesias, D., Sherry, J., Reinisch, C., Muttray, A.F., *et al.*

(2016). Widespread transmission of independent cancer lineages within multiple bivalve species. *Nature* **534**:705–709.

Meydan, C., Otu, H.H. and Sezerman, O.U. (2013). Prediction of peptides binding to MHC class I and II alleles by temporal motif mining. *BMC Bioinformatics* **14**:13.

Meyer zu Horste, G., Heidenreich, H., Mausberg, A.K., Lehmann, H.C., ten Asbroek, A.L., Saavedra, J.T., Baas, F., *et al.* (2010). Mouse Schwann cells activate MHC class I and II restricted T-cell responses, but require external peptide processing for MHC class II presentation. *Neurobiology of Disease* **37**:483–490.

Meyer Zu Hörste, G., Hu, W., Hartung, H.-P., Lehmann, H.C. and Kieseier, B.C. (2008). The immunocompetence of Schwann cells. *Muscle & Nerve* **37**:3–13.

Michalski, A., Cox, J. and Mann, M. (2011). More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC−MS/MS. *Journal of Proteome Research* **10**:1785–1793.

Migalska, M., Sebastian, A. and Radwan, J. (2019). Major histocompatibility complex class I diversity limits the repertoire of T cell receptors. *Proceedings of the National Academy of Sciences* **116**:5021–5026.

Mikulak, J., Negrini, S., Klajn, A., D'Alessandro, R., Mavilio, D. and Meldolesi, J. (2012). Dual REST-dependence of L1CAM: from gene expression to alternative splicing governed by Nova2 in neural cells. *Journal of Neurochemistry* **120**:699–709.

Milinski, M., Griffiths, S., Wegner, K.M., Reusch, T.B.H., Haas-Assenbaum, A. and Boehm, T. (2005). Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proceedings of the National Academy of Sciences* **102**:4414–4418.

Miller, S.J., Li, H., Rizvi, T.A., Huang, Y., Johansson, G., Bowersock, J., Sidani, A., *et al.* (2003). Brain lipid binding protein in axon-Schwann cell interactions and peripheral nerve tumorigenesis. *Molecular and Cellular Biology* **23**:2213–2224.

Mimura, K., Teh, J.L., Okayama, H., Shiraishi, K., Kua, L.-F., Koh, V., Smoot, D.T., *et al.* (2018). PD-L1 expression is mainly regulated by interferon gamma associated with JAK-STAT pathway in gastric cancer. *Cancer Science* **109**:43–53.

Mina, M. (2018). Live-attenuated viral vaccines may alter within- and between-host dynamics of non-vaccine targeted bacterial pathogens. *American Journal of Clinical Pathology* **149**:S163–S164.

Mirsky, R., Woodhoo, A., Parkinson, D.B., Arthur-Farraj, P., Bhaskaran, A. and Jessen, K.R. (2008). Novel signals controlling embryonic Schwann cell development, myelination and dedifferentiation. *Journal of the Peripheral Nervous System* **13**:122–135.

Monje, P. V, Soto, J., Bacallao, K. and Wood, P.M. (2010). Schwann cell dedifferentiation is independent of mitogenic signaling and uncoupled to proliferation: role of cAMP and JNK in the maintenance of the differentiated state. *Journal of Biological Chemistry* **285**:31024–31036.

Moreau, A., Varey, E., Anegon, I. and Cuturi, M.-C. (2013). Effector mechanisms of rejection. *Cold Spring Harbor Perspectives in Medicine* **3**:a015461.

Morell, P. and Quarles, R.H. (1999). Characteristic composition of myelin. In: Siegel, G. J., Agranoff, B. W., Albers, R. W. and et al. (eds.) *Basic Neurochemistry: Molecular, Cellular and Medical Aspects.* 6th ed. Philadelphia: Lippincott-Raven.

Morgan, R.A., Chinnasamy, N., Abate-Daga, D., Gros, A., Robbins, P.F., Zheng, Z., Dudley, M.E., *et al.* (2013). Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy. *Journal of Immunotherapy* **36**:133–151.

Morihiro, Y., Yasumoto, Y., Vaidyan, L.K., Sadahiro, H., Uchida, T., Inamura, A., Sharifi, K., *et al.* (2013). Fatty acid binding protein 7 as a marker of glioma stem cells. *Pathology International* **63**:546–553.

Mougiakakos, D., Johansson, C.C., Trocme, E., All-Ericsson, C., Economou, M.A., Larsson, O., Seregard, S., *et al.* (2010). Intratumoral forkhead box P3-positive regulatory T cells predict poor survival in cyclooxygenase-2–positive uveal melanoma. *Cancer* **116**:2224–2233.

Mouriaux, F., Zaniolo, K., Bergeron, M.-A., Weidmann, C., De La Fouchardière, A., Fournier, F., Droit, A., *et al.* (2016). Effects of long-term serial passaging on the characteristics and properties of cell lines derived from uveal melanoma primary tumors. *Investigative Ophthalmology & Visual Science* **57**:5288–5301.

Moyon, S., Dubessy, A.L., Aigrot, M.S., Trotter, M., Huang, J.K., Dauphinot, L., Potier, M.C., *et al.* (2015). Demyelination causes adult CNS progenitors to revert to an immature state and express immune cues that support their migration. *Journal of Neuroscience* **35**:4–20.

Muehlenbachs, A., Bhatnagar, J., Agudelo, C.A., Hidron, A., Eberhard, M.L., Mathison, B.A., Frace, M.A., *et al.* (2015). Malignant transformation of Hymenolepis nana in a human host. *New England Journal of Medicine* **373**:1845–1852.

Mukaratirwa, S., Chimonyo, M., Obwolo, M., Gruys, E. and Nederbragt, H. (2004). Stromal cells and extracellular matrix components in spontaneous canine transmissible venereal tumour at different stages of growth. *Histology and histopathology* **19**:1117–1123.

Murat, P. and Tellam, J. (2015). Effects of messenger RNA structure and other translational control mechanisms on major histocompatibility complex-I mediated antigen presentation. *Wiley Interdisciplinary Reviews RNA* **6**:157–171.

Murchison, E.P., Schulz-Trieglaff, O.B., Ning, Z., Alexandrov, L.B., Bauer, M.J., Fu, B., Hims, M., *et al.* (2012). Genome sequencing and analysis of the Tasmanian Devil and its transmissible cancer. *Cell* **148**:780–791.

Murchison, E.P., Tovar, C., Hsu, A., Bender, H.S., Kheradpour, P., Rebbeck, C.A., Obendorf, D., *et al.* (2010). The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. *Science* **327**:84–87.

Murchison, E.P., Wedge, D.C., Alexandrov, L.B., Fu, B., Martincorena, I., Ning, Z., Tubio, J.M.C., *et al.* (2014). Transmissible dog cancer genome reveals the origin and history of an ancient cell lineage. *Science* **343**:437–440.

Murgia, C., Pritchard, J.K., Kim, S.Y., Fassati, A. and Weiss, R.A. (2006). Clonal origin and evolution of a transmissible cancer. *Cell* **126**:477–487.

Myers, J.P., Santiago-Medina, M. and Gomez, T.M. (2011). Regulation of axonal outgrowth and pathfinding by integrin-ECM interactions. *Developmental Neurobiology* **71**:901–923.

Najafi, M., Farhood, B. and Mortezaee, K. (2019). Contribution of regulatory T cells to cancer: A review. *Journal of Cellular Physiology* **234**:7983–7993.

Nakajima, C., Uekusa, Y., Iwasaki, M., Yamaguchi, N., Mukai, T., Gao, P., Tomura, M., *et al.* (2001). A Role of Interferon-γ (IFN-γ) in Tumor Immunity. *Cancer Research* **61**:3399–3405.

Nave, K.-A. (2010). Myelination and support of axonal integrity by glia. *Nature* **468**:244–252.

Nave, K.-A. and Werner, H.B. (2014). Myelination of the nervous system: mechanisms and functions. *Annual Review of Cell and Developmental Biology* **30**:503–533.

Neefjes, J. and Ovaa, H. (2013). A peptide's perspective on antigen presentation to the immune system. *Nature Chemical Biology* **9**:769–775.

Neerincx, A., Hermann, C., Antrobus, R., van Hateren, A., Cao, H., Trautwein, N., Stevanović, S., *et al.* (2017). TAPBPR bridges UDP-glucose:glycoprotein glucosyltransferase 1 onto MHC class I to provide quality control in the antigen presentation pathway. *eLife* **6**:e23049.

Nguyen, A. V, Wu, Y.-Y., Liu, Q., Wang, D., Nguyen, S., Loh, R., Pang, J., *et al.* (2013). STAT3 in epithelial cells regulates inflammation and tumor progression to malignant state in colon. *Neoplasia* **15**:998–1008.

Nicotra, M.L., Powell, A.E., Rosengarten, R.D., Moreno, M., Grimwood, J., Lakkis, F.G., Dellaporta, S.L., *et al.* (2009). A hypervariable invertebrate allodeterminant. *Current Biology* **19**:583–589.

Nielsen, M. and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets.

Genome Medicine **8**:33.

Nielsen, M., Connelley, T. and Ternette, N. (2018). Improved prediction of Bovine Leucocyte Antigens (BoLA) presented ligands by use of mass-spectrometry-determined ligand and in vitro binding data. *Journal of Proteome Research* **17**:559–567.

Nielsen, M., Lundegaard, C., Worning, P., Lauemøller, S.L., Lamberth, K., Buus, S., Brunak, S., *et al.* (2003). Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Science* **12**:1007–1017.

O'Neil, D., Swanton, C., Jones, A., Medd, P.G., Rayment, N. and Chain, B. (1999). IFN-γ down-regulates MHC expression and antigen processing in a human B cell line. *The Journal of Immunology* **162**:791–798.

Ochando, J. and Braza, M.S. (2017). Nanoparticle-based modulation and monitoring of antigen-presenting cells in organ transplantation. *Frontiers in Immunology* **8**:1888.

Offringa, R. (2009). Antigen choice in adoptive T-cell therapy of cancer. *Current Opinion in Immunology* **21**:190–199.

Oh, S.A. and Li, M.O. (2013). TGF-β: guardian of T cell function. *The Journal of Immunology* **191**:3973–3979.

Ohteki, T., Fukao, T., Suzue, K., Maki, C., Ito, M., Nakamura, M. and Koyasu, S. (1999). Interleukin 12–dependent interferon γ production by CD8α+ lymphoid dendritic cells. *The Journal of Experimental Medicine* **189**:1981–1986.

Ortonne, N., Wolkenstein, P., Blakeley, J.O., Korf, B., Plotkin, S.R., Riccardi, V.M., Miller, D.C., *et al.* (2018). Cutaneous neurofibromas. *Neurology* **91**:S5 LP-S13.

Ostroumov, D., Fekete-Drimusz, N., Saborowski, M., Kühnel, F. and Woller, N. (2018). CD4 and CD8 T lymphocyte interplay in controlling tumor growth. *Cellular and Molecular Life Sciences* **75**:689–713.

Ou, W., Thapa, R.K., Jiang, L., Soe, Z.C., Gautam, M., Chang, J.-H., Jeong, J.-H., *et al.* (2018). Regulatory T cell-targeted hybrid nanoparticles combined with immuno-checkpoint blockage for cancer immunotherapy. *Journal of Controlled Release* **281**:84–96.

Ouhtit, A., Madani, S., Gupta, I., Shanmuganathan, S., Abdraboh, M.E., Al-Riyami, H., Al-Farsi, Y.M., *et al.* (2013). TGF-β2: a novel target of CD44-promoted breast cancer invasion. *Journal of Cancer* **4**:566–572.

Owen, D. (2005). *Tasmanian Devil : A Unique and Threatened Animal*. Pemberton, D. (ed.). London: Natural History Museum.

Ozawa, T., Brennan, C.W., Wang, L., Squatrito, M., Sasayama, T., Nakada, M., Huse, J.T., *et al.* (2010). PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas. *Genes & Development* **24**:2205–2218.

Pang, Y., Zheng, B., Kimberly, S.L., Cai, Z., Rhodes, P.G. and Lin, R.C.S. (2012). Neuron-oligodendrocyte myelination co-culture derived from embryonic rat spinal cord and cerebral cortex. *Brain and Behavior* **2**:53–67.

Parham, P. (2009). *The Immune System*. 3rd ed. New York: Garland Science.

Parker, C.E., Mocanu, V., Mocanu, M., Dicheva, N. and Warren, M.R. (2010). Mass spectrometry for post-translational modifications. In: Alzate, O. (ed.) *Neuroproteomics*. Boca Raton (FL): CRC press/Taylor & Francis.

Parkinson, D.B., Bhaskaran, A., Arthur-Farraj, P., Noon, L.A., Woodhoo, A., Lloyd, A.C., Feltri, M.L., *et al.* (2008). c-Jun is a negative regulator of myelination. *Journal of Cell Biology* **181**:625–637.

Patchett, A.L., Tovar, C., Corcoran, L.M., Lyons, A.B. and Woods, G.M. (2017). The toll-like receptor ligands Hiltonol® (polyICLC) and imiquimod effectively activate antigen-specific immune responses in Tasmanian devils (Sarcophilus harrisii). *Developmental & Comparative Immunology* **76**:352–360.

Pearse, A.-M., Swift, K., Hodson, P., Hua, B., McCallum, H., Pyecroft, S., Taylor, R., *et al.* (2012). Evolution in a transmissible cancer: a study of the chromosomal changes in devil facial tumor (DFT) as it spreads through the wild Tasmanian devil population. *Cancer Genetics* **205**:101–112.

Pearse, A.M. and Swift, K. (2006). Allograft theory: Transmission of devil facial-tumour disease. *Nature* **439**:549.

Pérez, J., J. Day, M. and Mozos, E. (1998). Immunohistochemical study of the local inflammatory infiltrate in spontaneous canine transmissible venereal tumour at different stages of growth. *Veterinary Immunology and Immunopathology* **64**:133–147.

Perry, A., Roth, K.A., Banerjee, R., Fuller, C.E. and Gutmann, D.H. (2001). NF1 deletions in S-100 protein-positive and negative cells of sporadic and neurofibromatosis 1 (NF1)-associated plexiform neurofibromas and malignant peripheral nerve sheath tumors. *The American journal of pathology* **159**:57–61.

Persson, G., Melsted, W.N., Nilsson, L.L. and Hviid, T.V.F. (2017). HLA class Ib in pregnancy and pregnancy-related disorders. *Immunogenetics* **69**:581–595.

Peschl, P., Bradl, M., Höftberger, R., Berger, T. and Reindl, M. (2017). Myelin Oligodendrocyte Glycoprotein: deciphering a target in inflammatory demyelinating diseases. *Frontiers in Immunology* **8**:529.

Pierce, R.A., Field, E.D., den Haan, J.M.M., Caldwell, J.A., White, F.M., Marto, J.A., Wang, W., *et al.* (1999). Cutting edge: the HLA-A*0101-restricted HY minor histocompatibility antigen originates from DFFRY and contains a cysteinylated cysteine residue as identified by a novel mass spectrometric technique. *The Journal of Immunology* **163**:6360–6364.

Pinfold, T.L., Brown, G.K., Bettiol, S.S. and Woods, G.M. (2014). Mouse model of Devil Facial Tumour Disease establishes that an effective immune response can be generated against the cancer cells. *Frontiers in Immunology* **5**:251.

Potter, J.F. and Schoeneman, M. (1970). Metastasis of maternal cancer to the placenta and fetus. *Cancer* **25**:380–388.

Pye, R., Patchett, A., McLennan, E., Thomson, R., Carver, S., Fox, S., Pemberton, D., *et al.* (2018). Immunization strategies producing a humoral IgG immune response against Devil Facial Tumor Disease in the majority of Tasmanian Devils destined for wild release. *Frontiers in Immunology* **9**:259.

Pye, R.J., Hamede, R., Siddle, H. V, Caldwell, A., Knowles, G.W., Swift, K., Kreiss, A., *et al.* (2016). Demonstration of immune responses against devil facial tumour disease in wild Tasmanian devils. *Biology Letters* **12**:0553.

Pye, R.J., Pemberton, D., Tovar, C., Tubio, J.M.C., Dun, K.A., Fox, S., Darby, J., *et al.* (2016). A second transmissible cancer in Tasmanian devils. *Proceedings of the National Academy of Sciences* **113**:374–379.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* [Online]. Available at: http://www.r-project.org/.

Rapin, N., Hoof, I., Lund, O. and Nielsen, M. (2008). MHC motif viewer. *Immunogenetics* **60**:759.

Rasmussen, M., Fenoy, E., Harndahl, M., Kristensen, A.B., Nielsen, I.K., Nielsen, M. and Buus, S. (2016). Pan-specific prediction of peptide–MHC class I complex stability, a correlate of T cell immunogenicity. *The Journal of Immunology* **197**:1517–1524.

Ravera, S., Panfoli, I., Aluigi, M., Calzia, D. and Morelli, A. (2011). Characterization of Myelin Sheath FoF1-ATP Synthase and its Regulation by IF1. *Cell Biochemistry and Biophysics* **59**:63–70.

Reddy, S., Jia, S., Geoffrey, R., Lorier, R., Suchi, M., Broeckel, U., Hessner, M.J., *et al.* (2009). An autoinflammatory disease due to homozygous deletion of the IL1RN locus. *The New England Journal of Medicine* **360**:2438–2444.

Redwine, J.M., Buchmeier, M.J. and Evans, C.F. (2001). In vivo expression of major histocompatibility complex molecules on oligodendrocytes and neurons during viral infection. *The American Journal of Pathology* **159**:1219–1224.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H. and Vilo, J. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Research* **44**:83–89.

Reimand, J., Kull, M., Peterson, H., Hansen, J. and Vilo, J. (2007). g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research* **35**:193–200.

Rivers, L.E., Young, K.M., Rizzi, M., Jamen, F., Psachoulia, K., Wade, A., Kessaris, N., *et al.* (2008). PDGFRA/NG2 glia generate myelinating oligodendrocytes and piriform projection neurons in adult mice. *Nature Neuroscience* **11**:1392.

Rodriguez, F.J., Folpe, A.L., Giannini, C. and Perry, A. (2012). Pathology of peripheral nerve sheath tumors: diagnostic overview and update on selected diagnostic problems. *Acta Neuropathologica* **123**:295–319.

Rogel, A., Willoughby, J.E., Buchan, S.L., Leonard, H.J., Thirdborough, S.M. and Al-Shamkhani, A. (2017). Akt signaling is critical for memory CD8+ T-cell development and tumor immune surveillance. *Proceedings of the National Academy of Sciences* **114**:1178–1187.

Rogers, J.H. (1985). Mouse histocompatibility-related genes are not conserved in other mammals. *The EMBO journal* **4**:749–753.

De Rosa, A., Pellegatta, S., Rossi, M., Tunici, P., Magnoni, L., Speranza, M.C., Malusa, F., *et al.* (2012). A radial glia gene marker, Fatty Acid Binding Protein 7 (FABP7), is involved in proliferation and invasion of glioblastoma cells. *PLoS ONE* **7**:e52113.

Rossjohn, J., Gras, S., Miles, J.J., Turner, S.J., Godfrey, D.I. and McCluskey, J. (2015). T cell antigen receptor recognition of antigen-presenting molecules. *Annual Review of Immunology* **33**:169–200.

Ruiz-Aravena, M., Jones, M.E., Carver, S., Estay, S., Espejo, C., Storfer, A. and Hamede, R.K. (2018). Sex bias in ability to cope with cancer: Tasmanian devils and facial tumour disease. *Proceedings of the Royal Society B: Biological Sciences* **285**:20182239.

Saadipour, K., MacLean, M., Pirkle, S., Ali, S., Lopez-Redondo, M.L., Stokes, D.L. and Chao, M. V (2017). The transmembrane domain of the p75 neurotrophin receptor stimulates phosphorylation of the TrkB tyrosine kinase receptor. *Journal of Biological Chemistry* **292**:16594–16604.

Saidulu, M., M, P.K., G, F.V., Esther, Z., E, L.F., D, R.T., M, T.T., *et al.* (2018). Human Leukocyte Antigen class I and II knockout human induced pluripotent stem cell–derived cells: universal donor for cell therapy. *Journal of the American Heart Association* **7**:e010239.

Sainsbury, J.R., Anderson, T.J. and Morgan, D.A. (2000). ABC of breast diseases: breast cancer. *BMJ* **321**:745–750.

Saitoh, F., Wakatsuki, S., Tokunaga, S., Fujieda, H. and Araki, T. (2016). Glutamate signals through mGluR2 to control Schwann cell differentiation and proliferation. *Scientific Reports* **6**:29856.

Sakaguchi, S., Yamaguchi, T., Nomura, T. and Ono, M. (2008). Regulatory T cells and immune tolerance. *Cell* **133**:775–787.

Salter, R.D., Benjamin, R.J., Wesley, P.K., Buxton, S.E., Garrett, T.P.J., Clayberger, C., Krensky, A.M., *et al.* (1990). A binding site for the T-cell co-receptor CD8 on the α3 domain of HLA-A2. *Nature* **345**:41–46.

Salzer, J.L. and Zalc, B. (2016). Myelination. *Current Biology* **26**:971–975.

Samuel, N.M., Mirsky, R., Grange, J.M. and Jessen, K.R. (1987). Expression of major histocompatibility complex class I and class II antigens in human Schwann cell cultures and effects of infection with Mycobacterium leprae. *Clinical and Experimental Immunology* **68**:500–509.

Sánchez-Paulete, A.R., Teijeira, A., Cueto, F.J., Garasa, S., Pérez-Gracia, J.L., Sánchez-Arráez, A., Sancho, D., *et al.* (2017). Antigen cross-presentation and T-cell cross-priming in cancer immunology and immunotherapy. *Annals of Oncology* **28**:xii44–xii55.

Sandusky, G.E., Carlton, W.W. and Wightman, K.A. (1987). Diagnostic immunohistochemistry of canine round cell tumors. *Veterinary Pathology* **24**:495–499.

Sarma, N.J., Tiriveedhi, V., Angaswamy, N. and Mohanakumar, T. (2012). Role of antibodies to self-antigens in chronic allograft rejection: potential mechanism and therapeutic implications. *Human immunology* **73**:1275–1281.

Schaefer, M.K.E. and Frotscher, M. (2012). Role of L1CAM in axon sprouting and branching. *Cell and Tissue Research* **349**:39–48.

Schiffman, M., Doorbar, J., Wentzensen, N., de Sanjosé, S., Fakhry, C., Monk, B.J., Stanley, M.A., *et al.* (2016). Carcinogenic human papillomavirus infection. *Nature Reviews Disease Primers* **2**:16086.

Schmid, B. V, Keşmir, C. and de Boer, R.J. (2008). The specificity and polymorphism of the MHC class I prevents the global adaptation of HIV-1 to the monomorphic proteasome and TAP. *PLoS ONE* **3**:e3525.

Schneider-Schaulies, J., Kirchhoff, F., Archelos, J. and Schachner, M. (1991). Down-regulation of myelin-associated glycoprotein on Schwann cells by interferon-gamma; and tumor necrosis factor-alpha affects neurite outgrowth. *Neuron* **7**:995–1005.

Schuster, H., Shao, W., Weiss, T., Pedrioli, P.G.A., Roth, P., Weller, M., Campbell, D.S., *et al.* (2018). A tissue-based draft map of the murine MHC class I immunopeptidome. *Scientific Data* **5**:180157.

Sebille, F., Gagne, K., Guillet, M., Degauque, N., Pallier, A., Brouard, S., Vanhove, B., *et al.* (2001).

Direct recognition of foreign MHC determinants by naive T cells mobilizes specific Vβ families without skewing of the Complementarity-Determining Region 3 length distribution. *The Journal of Immunology* **167**:3082–3088.

Sedlacek, A.L., Younker, T.P., Zhou, Y.J., Borghesi, L., Shcheglova, T., Mandoiu, I.I. and Binder, R.J. (2019). CD91 on dendritic cells governs immunosurveillance of nascent, emerging tumors. *JCI Insight* **4**:e127239.

Sette, A., Vitiello, A., Reherman, B., Fowler, P., Nayersina, R., Kast, W.M., Melief, C.J., *et al.* (1994). The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *The Journal of Immunology* **153**:5586–5592.

Shabaneh, T.B., Molodtsov, A.K., Steinberg, S.M., Zhang, P., Torres, G.M., Mohamed, G.A., Boni, A., *et al.* (2018). Oncogenic BRAFV600E governs regulatory T-cell recruitment during melanoma tumorigenesis. *Cancer Research* **78**:5038 LP – 5049.

Shankaran, V., Ikeda, H., Bruce, A.T., White, J.M., Swanson, P.E., Old, L.J. and Schreiber, R.D. (2001). IFNγ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410**:1107–1111.

Sharifi, K., Ebrahimi, M., Kagawa, Y., Islam, A., Tuerxun, T., Yasumoto, Y., Hara, T., *et al.* (2013). Differential expression and regulatory roles of FABP5 and FABP7 in oligodendrocyte lineage cells. *Cell and Tissue Research* **354**:683–695.

Sharifi, K., Morihiro, Y., Maekawa, M., Yasumoto, Y., Hoshi, H., Adachi, Y., Sawada, T., *et al.* (2011). FABP7 expression in normal and stab-injured brain cortex and its role in astrocyte proliferation. *Histochemistry and Cell Biology* **136**:501–513.

Shen, X. and Zhao, B. (2018). Efficacy of PD-1 or PD-L1 inhibitors and PD-L1 expression status in cancer: meta-analysis. *BMJ* **362**:3529.

Sherman, L.A. and Chattopadhyay, S. (1993). The molecular basis of allorecognition. *Annual Review of Immunology* **11**:385–402.

Shimizu, M., Ueno, K. and Yachie, A. (2012). Relapse of systemic juvenile idiopathic arthritis after influenza vaccination in a patient receiving tocilizumab. *Clinical and vaccine immunology : CVI* **19**:1700–1702.

Shimizu, Y. and DeMars, R. (1989). Production of human cells expressing individual transferred HLA-A,-B,-C genes using an HLA-A,-B,-C null human cell line. *The Journal of Immunology* **142**:3320–3328.

Sia, D., Villanueva, A., Friedman, S.L. and Llovet, J.M. (2017). Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterology* **152**:745–761.

Siddle, H. V, Kreiss, A., Eldridge, M.D., Noonan, E., Clarke, C.J., Pyecroft, S., Woods, G.M., *et al.* (2007). Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proceedings of the National Academy of Sciences* **104**:16221–16226.

Siddle, H. V, Kreiss, A., Tovar, C., Yuen, C.K., Cheng, Y., Belov, K., Swift, K., *et al.* (2013). Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. *Proceedings of the National Academy of Sciences* **110**:5103–5108.

Siddle, H. V, Marzec, J., Cheng, Y., Jones, M. and Belov, K. (2010). MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. *Proceedings of the Royal Society B* **277**.

Siddle, H. V, Sanderson, C. and Belov, K. (2007). Characterization of major histocompatibility complex class I and class II genes from the Tasmanian devil (Sarcophilus harrisi). *Immunogenetics* **59**:753–760.

Sidney, J., Assarsson, E., Moore, C., Ngo, S., Pinilla, C., Sette, A. and Peters, B. (2008). Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Research* **4**:2.

Silva, A., Mount, A., Krstevska, K., Pejoski, D., Hardy, M.P., Owczarek, C., Scotney, P., *et al.* (2015). The combination of ISCOMATRIX adjuvant and TLR agonists induces regression of established solid tumors in vivo. *The Journal of Immunology* **194**:2199–2207.

Silva, A.M.N., Vitorino, R., Domingues, M.R.M., Spickett, C.M. and Domingues, P. (2013). Post-translational modifications and mass spectrometry detection. *Free Radical Biology and Medicine* **65**:925–941.

da Silva, I.L., Montero-Montero, L., Ferreira, E. and Quintanilla, M. (2018). New insights into the role of Qa-2 and HLA-G non-classical MHC-I complexes in malignancy. *Frontiers in Immunology* **9**:2894.

Silvestroff, L., Franco, P.G. and Pasquini, J.M. (2013). Neural and oligodendrocyte progenitor cells: transferrin effects on cell proliferation. *ASN Neuro* **5**:e00107.

Simpson, J.A.D., Al-Attar, A., Watson, N.F.S., Scholefield, J.H., Ilyas, M. and Durrant, L.G. (2010). Intratumoral T cell infiltration, MHC class I and STAT1 as biomarkers of good prognosis in colorectal cancer. *Gut* **59**:926–933.

Sims, S., Willberg, C. and Klenerman, P. (2010). MHC–peptide tetramers for the analysis of antigen-specific T cells. *Expert Review of Vaccines* **9**:765–774.

Skwarczynski, M. and Toth, I. (2016). Peptide-based synthetic vaccines. *Chemical Science* **7**:842–

854.

Slukvin, I.I., Lunn, D.P., Watkins, D.I. and Golos, T.G. (2000). Placental expression of the nonclassical MHC class I molecule Mamu-AG at implantation in the rhesus monkey. *Proceedings of the National Academy of Sciences* **97**:9104–9109.

Small, E.J., Sacks, N., Nemunaitis, J., Urba, W.J., Dula, E., Centeno, A.S., Nelson, W.G., *et al.* (2007). Granulocyte macrophage colony-stimulating factor–secreting allogeneic cellular immunotherapy for hormone-refractory prostate cancer. *Clinical Cancer Research* **13**:3883–3891.

Smyth, M.J., Sullivan, L.C., Brooks, A.G. and Andrews, D.M. (2013). Non-classical MHC Class I molecules regulating natural killer cell function. *OncoImmunology* **2**:e23336.

Snaidero, N. and Simons, M. (2014). Myelination at a glance. *Journal of Cell Science* **127**:2999–3004.

Sokolovska, A., Hem, S.L. and HogenEsch, H. (2007). Activation of dendritic cells and induction of CD4+ T cell differentiation by aluminum-containing adjuvants. *Vaccine* **25**:4575–4585.

Spiegel, I., Adamsky, K., Eshed, Y., Milo, R., Sabanay, H., Sarig-Nadir, O., Horresh, I., *et al.* (2007). A central role for Necl4 (SynCAM4) in Schwann cell-axon interaction and myelination. *Nature Neuroscience* **10**:861–869.

Spierings, E. (2014). Minor histocompatibility antigens: past, present, and future. *Tissue Antigens* **84**:360–374.

Spyra, M., Kluwe, L., Hagel, C., Nguyen, R., Panse, J., Kurtz, A., Mautner, V.F., *et al.* (2011). Cancer stem cell-like cells derived from malignant peripheral nerve sheath tumors. *PloS ONE* **6**:e21099.

Stammnitz, M.R., Coorens, T.H.H., Gori, K.C., Hayes, D., Fu, B., Wang, J., Martin-Herranz, D.E., *et al.* (2018). The origins and vulnerabilities of two transmissible cancers in Tasmanian Devils. *Cancer Cell* **33**:607–619.

Steinman, R.M. and Swanson, J. (1995). The endocytic activity of dendritic cells. *The Journal of Experimental Medicine* **182**:283 LP – 288.

Stevens, B., Tanner, S. and Fields, R.D. (1998). Control of myelination by specific patterns of neural impulses. *The Journal of Neuroscience* **18**:9303–9311.

Stierli, S., Napoli, I., White, I.J., Cattin, A.-L., Monteza Cabrejos, A., Garcia Calavia, N., Malong, L., *et al.* (2018). The regulation of the homeostasis and regeneration of peripheral nerve is distinct from the CNS and independent of a stem cell population. *Development* **145**:170316.

Stolte, C., Binder, J.X., Tsafou, K., Schneider, R., O'Donoghue, S.I., Pletscher-Frankild, S. and Jensen, L.J. (2014). COMPARTMENTS: unification and visualization of protein subcellular

localization evidence. *Database* **2014**:012.

Stone, S., Wu, S., Jamison, S., Durose, W., Pallais, J.P. and Lin, W. (2018). Activating transcription factor 6α deficiency exacerbates oligodendrocyte death and myelin damage in immune-mediated demyelinating diseases. *Glia* **66**:1331–1345.

Strakova, A. and Murchison, E.P. (2015). The cancer which survived: insights from the genome of an 11000 year-old cancer. *Current Opinion in Genetics & Development* **30**:49–55.

Street, S.E.A., Trapani, J.A., MacGregor, D. and Smyth, M.J. (2002). Suppression of lymphoma and epithelial malignancies effected by interferon γ. *The Journal of Experimental Medicine* **196**:129–134.

Sugiura, Y. and Lin, W. (2011). Neuron–glia interactions: the roles of Schwann cells in neuromuscular synapse formation and function. *Bioscience Reports* **31**:295–302.

Sullivan, L.C., Hoare, H.L., McCluskey, J., Rossjohn, J. and Brooks, A.G. (2006). A structural perspective on MHC class Ib molecules in adaptive immunity. *Trends in Immunology* **27**:413–420.

Sultan, H., Trillo-Tinoco, J., Rodriguez, P. and Celis, E. (2017). Effective antitumor peptide vaccines can induce severe autoimmune pathology. *Oncotarget* **8**:70317–70331.

Swann, J.B. and Smyth, M.J. (2007). Immune surveillance of tumors. *Journal of Clinical Investigation* **117**:1137–1146.

Tabb, D.L., Vega-Montoto, L., Rudnick, P.A., Variyath, A.M., Ham, A.-J.L., Bunk, D.M., Kilpatrick, L.E., *et al.* (2010). Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *Journal of Proteome Research* **9**:761–776.

Tagliavacca, L., Colombo, F., Racchetti, G. and Meldolesi, J. (2013). L1CAM and its cell-surface mutants: new mechanisms and effects relevant to the physiology and pathology of neural cells. *Journal of Neurochemistry* **124**:397–409.

Takeda, K., Kaisho, T., Yoshida, N., Takeda, J., Kishimoto, T. and Akira, S. (1998). Stat3 activation is responsible for IL-6-dependent T cell proliferation through preventing apoptosis: generation and characterization of T cell-specific Stat3-deficient mice. *The Journal of Immunology* **161**:4652–4660.

Takemoto, S.K., Terasaki, P.I., Gjertson, D.W. and Cecka, J.M. (2000). Twelve years' experience with national sharing of HLA-matched cadaveric kidneys for transplantation. *New England Journal of Medicine* **343**:1078–1084.

Takeuchi, M., Alard, P. and Streilein, J.W. (1998). TGF-β promotes immune deviation by altering accessory signals of antigen-presenting cells. *The Journal of Immunology* **160**:1589–1597.

Tanner, D.C., Cherry, J.D. and Mayer-Pröschel, M. (2011). Oligodendrocyte progenitors reversibly exit the cell cycle and give rise to astrocytes in response to interferon-γ. *The Journal of Neuroscience* **31**:6235–6246.

Teng, M.W.L., Galon, J., Fridman, W.-H. and Smyth, M.J. (2015). From mice to humans: developments in cancer immunoediting. *The Journal of clinical investigation* **125**:3338–3346.

Terabe, M., Robertson, F.C., Clark, K., De Ravin, E., Bloom, A., Venzon, D.J., Kato, S., *et al.* (2017). Blockade of only TGF-β 1 and 2 is sufficient to enhance the efficacy of vaccine and PD-1 checkpoint blockade immunotherapy. *OncoImmunology* **6**:e1308616.

Thaunat, O., Field, A.-C., Dai, J., Louedec, L., Patey, N., Bloch, M.-F., Mandet, C., *et al.* (2005). Lymphoid neogenesis in chronic rejection: Evidence for a local humoral alloimmune response. *Proceedings of the National Academy of Sciences* **102**:14723–14728.

The Gene Ontology Consortium (2018). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* **47**:330–338.

Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., *et al.* (2017). A subcellular map of the human proteome. *Science* **356**:e3321.

Tinoco, R., Alcalde, V., Yang, Y., Sauer, K. and Zuniga, E.I. (2009). Cell-intrinsic transforming growth factor-beta signaling mediates virus-specific CD8+ T cell deletion and viral persistence in vivo. *Immunity* **31**:145–157.

Tomita, K., Kubo, T., Matsuda, K., Fujiwara, T., Yano, K., Winograd, J.M., Tohyama, M., *et al.* (2007). The neurotrophin receptor p75NTR in Schwann cells is implicated in remyelination and motor recovery after peripheral nerve injury. *Glia* **55**:1199–1208.

Torii, T., Miyamoto, Y., Yamamoto, M., Ohbuchi, K., Tsumura, H., Kawahara, K., Tanoue, A., *et al.* (2015). Arf6 mediates Schwann cell differentiation and myelination. *Biochemical and Biophysical Research Communications* **465**:450–457.

Tovar, C., Obendorf, D., Murchison, E.P., Papenfuss, A.T., Kreiss, A. and Woods, G.M. (2011). Tumor-specific diagnostic marker for transmissible facial tumors of Tasmanian devils: immunohistochemistry studies. *Veterinary Pathology* **48**:1195–1203.

Tovar, C., Patchett, A.L., Kim, V., Wilson, R., Darby, J., Lyons, A.B. and Woods, G.M. (2018). Heat shock proteins expressed in the marsupial Tasmanian devil are potential antigenic candidates in a vaccine against devil facial tumour disease. *PLoS ONE* **13**:e0196469.

Tovar, C., Pye, R.J., Kreiss, A., Cheng, Y., Brown, G.K., Darby, J., Malley, R.C., *et al.* (2017). Regression of devil facial tumour disease following immunotherapy in immunised Tasmanian devils. *Scientific Reports* **7**:43827.

Tripathi, R.B., Rivers, L.E., Young, K.M., Jamen, F. and Richardson, W.D. (2010). NG2 glia generate new oligodendrocytes but few astrocytes in a murine experimental autoimmune encephalomyelitis model of demyelinating disease. *The Journal of Neuroscience* **30**:16383–16390.

Trolle, T., McMurtrey, C.P., Sidney, J., Bardet, W., Osborn, S.C., Kaever, T., Sette, A., *et al.* (2016). The length distribution of class I-restricted T cell epitopes is determined by both peptide supply and MHC allele-specific binding preference. *The Journal of Immunology* **196**:1480–1487.

Uhlén, M., Björling, E., Agaton, C., Szigyarto, C.A.-K., Amini, B., Andersen, E., Andersson, A.-C., *et al.* (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. *Molecular & Cellular Proteomics* **4**:1920–1932.

Ujvari, B. and Belov, K. (2011). Major Histocompatibility Complex (MHC) markers in conservation biology. *International Journal of Molecular Sciences* **12**:5168–5186.

Ujvari, B., Pearse, A.-M., Peck, S., Harmsen, C., Taylor, R., Pyecroft, S., Madsen, T., *et al.* (2013). Evolution of a contagious cancer: epigenetic variation in Devil Facial Tumour Disease. *Proceedings. Biological sciences* **280**:20121720.

Vallières, N., Barrette, B., Wang, L.X., Bélanger, E., Thiry, L., Schneider, M.R., Filali, M., *et al.* (2017). Betacellulin regulates schwann cell proliferation and myelin formation in the injured mouse peripheral nerve. *Glia* **65**:657–669.

de Veer, M.J., Holko, M., Frevel, M., Walker, E., Der, S., Paranjape, J.M., Silverman, R.H., *et al.* (2001). Functional classification of interferon-stimulated genes identified using microarrays. *Journal of Leukocyte Biology* **69**:912–920.

Velghe, A.I., Van Cauwenberghe, S., Polyansky, A.A., Chand, D., Montano-Almendras, C.P., Charni, S., Hallberg, B., *et al.* (2013). PDGFRA alterations in cancer: characterization of a gain-of-function V536E transmembrane mutant as well as loss-of-function and passenger mutations. *Oncogene* **33**:2568–2576.

Vella, L.A., Yu, M., Fuhrmann, S.R., El-Amine, M., Epperson, D.E. and Finn, O.J. (2009). Healthy individuals have T-cell and antibody responses to the tumor antigen cyclin B1 that when elicited in mice protect from cancer. *Proceedings of the National Academy of Sciences* **106**:14010–14015.

Visvader, J.E. (2011). Cells of origin in cancer. *Nature* **469**:314–322.

Vitale, I., Sistigu, A., Manic, G., Rudqvist, N.-P., Trajanoski, Z. and Galluzzi, L. (2019). Mutational and antigenic landscape in tumor progression and cancer immunotherapy. *Trends in Cell Biology* **29**:396–419.

Waldhauer, I. and Steinle, A. (2008). NK cells and cancer immunosurveillance. *Oncogene* **27**:5932–5943.

Wang, D., Quiros, J., Mahuron, K., Pai, C.-C., Ranzani, V., Young, A., Silveria, S., *et al.* (2018). Targeting EZH2 reprograms intratumoral regulatory T cells to enhance cancer immunity. *Cell Reports* **23**:3262–3274.

Wang, H.B., Wang, X.P., Zhong, S.Z. and Shen, Z.L. (2013). Novel method for culturing Schwann cells from adult mouse sciatic nerve in vitro. *Molecular Medicine Reports* **7**:449–453.

Wang, L., Wang, Y., Song, Z., Chu, J. and Qu, X. (2014). Deficiency of interferon-gamma or its receptor promotes colorectal cancer development. *Journal of Interferon & Cytokine Research* **35**:273–280.

Wang, R., Natarajan, K. and Margulies, D.H. (2009). Structural basis of the CD8αβ/MHCI interaction: focused recognition orients CD8β to a T cell proximal position. *The Journal of Immunology* **183**:2554–2564.

Wang, S., Li, J., Chen, X., Wang, L., Liu, W. and Wu, Y. (2016). Analyzing the effect of peptide-HLA-binding ability on the immunogenicity of potential CD8+ and CD4+ T cell epitopes in a large dataset. *Immunologic Research* **64**:908–918.

Wang, W., Meadows, L.R., den Haan, J.M., Sherman, N.E., Chen, Y., Blokland, E., Shabanowitz, J., *et al.* (1995). Human H-Y: a male-specific histocompatibility antigen derived from the SMCY protein. *Science* **269**:1588–1590.

Wasag, B., Debiec-Rychter, M., Pauwels, P., Stul, M., Vranckx, H., Oosterom, A. Van, Hagemeijer, A., *et al.* (2004). Differential expression of KIT/PDGFRA mutant isoforms in epithelioid and mixed variants of gastrointestinal stromal tumors depends predominantly on the tumor site. *Modern Pathology* **17**:889–894.

Watson, N.F.S., Ramage, J.M., Madjd, Z., Spendlove, I., Ellis, I.O., Scholefield, J.H. and Durrant, L.G. (2006). Immunosurveillance is active in colorectal cancer as downregulation but not complete loss of MHC class I expression correlates with a poor prognosis. *International Journal of Cancer* **118**:6–10.

Wegener, A., Deboux, C., Bachelin, C., Frah, M., Kerninon, C., Seilhean, D., Weider, M., *et al.* (2015). Gain of Olig2 function in oligodendrocyte progenitors promotes remyelination. *Brain* **138**:120–135.

Wei, X.X., Fong, L. and Small, E.J. (2015). Prostate cancer immunotherapy with Sipuleucel-T: current standards and future directions. *Expert Review of Vaccines* **14**:1529–1541.

Wells, K., Hamede, R.K., Jones, M.E., Hohenlohe, P.A., Storfer, A. and McCallum, H.I. (2019). Individual and temporal variation in pathogen load predicts long-term impacts of an emerging infectious disease. *Ecology* **100**:e02613.

Wenger, S.L., Senft, J.R., Sargent, L.M., Bamezai, R., Bairwa, N. and Grant, S.G. (2004). Comparison of established cell lines at different passages by karyotype and comparative genomic hybridization. *Bioscience Reports* **24**:631 LP – 639.

Wieczorek, M., Abualrous, E.T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F. and Freund, C. (2017). Major Histocompatibility Complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. *Frontiers in Immunology* **8**:292.

Wilson, D.B., Wilson, D.H., Schroder, K., Pinilla, C., Blondelle, S., Houghten, R.A. and Garcia, K.C. (2004). Specificity and degeneracy of T cells. *Molecular Immunology* **40**:1047–1055.

Wood, C.G. and Mulders, P. (2009). Vitespen: a preclinical and clinical review. *Future Oncology* **5**:763–774.

Woodhoo, A., Alonso, M.B., Droggiti, A., Turmaine, M., D'Antonio, M., Parkinson, D.B., Wilton, D.K., *et al.* (2009). Notch controls embryonic Schwann cell differentiation, postnatal myelination and adult plasticity. *Nature Neuroscience* **12**:839–847.

Woods, G., Kreiss, A., Belov, K., Siddle, H., Obendorf, D. and Muller, H. (2007). The immune response of the Tasmanian Devil (Sarcophilus harrisi) and Devil Facial Tumour Disease. *EcoHealth* **4**:338–345.

Wynne, J.W., Woon, A.P., Dudek, N.L., Croft, N.P., Ng, J.H.J., Baker, M.L., Wang, L.-F., *et al.* (2016). Characterization of the antigen processing machinery and endogenous peptide presentation of a bat MHC class I molecule. *The Journal of Immunology* **196**:4468–4476.

Xiong, J., Su, T., Zhu, P., Ao, Q., Ruan, Q. and Wang, G. (2017). Malignant rhabdoid tumor in the renal allograft of an adult transplant recipient: a unique case of a rare tumor. *Diagnostic Pathology* **12**:86.

Xu, Y., Liu, B., Gröndahl-Yli-Hannuksila, K., Tan, Y., Feng, L., Kallonen, T., Wang, Lichan, *et al.* (2015). Whole-genome sequencing reveals the effect of vaccination on the evolution of Bordetella pertussis. *Scientific Reports* **5**:12888.

Yadav, M., Jhunjhunwala, S., Phung, Q.T., Lupardus, P., Tanguay, J., Bumbaca, S., Franci, C., *et al.* (2014). Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515**:572–576.

Yagasaki, H., Ohashi, H., Ito, M., Kobayashi, S., Kato, M., Shichino, H., Chin, M., *et al.* (2011). A novel mechanism of transplacental cancer transmission: natural killer/T-cell lymphoma in the paratesticular region is of maternal origin. *Blood* **117**:6046–6047.

Yang, T.J., Chandler, J.P. and Dunne-Anway, S. (1987). Growth stage dependent expression of MHC antigens on the canine transmissible venereal sarcoma. *British Journal of Cancer* **55**:131–

134.

Yang, Y., Xiang, Z., Ertl, H.C. and Wilson, J.M. (1995). Upregulation of class I major histocompatibility complex antigens by interferon gamma is necessary for T-cell-mediated elimination of recombinant adenovirus-infected hepatocytes in vivo. *Proceedings of the National Academy of Sciences* **92**:7257–7261.

Yu, Y., Xiao, C.-H., Tan, L.-D., Wang, Q.-S., Li, X.-Q. and Feng, Y.-M. (2014). Cancer-associated fibroblasts induce epithelial-mesenchymal transition of breast cancer cells through paracrine TGF-β signalling. *British Journal of Cancer* **110**:724–732.

Zawadzka, M., Rivers, L.E., Fancy, S.P.J., Zhao, C., Tripathi, R., Jamen, F., Young, K., *et al.* (2010). CNS-resident glial progenitor/stem cells produce Schwann cells as well as oligodendrocytes during repair of CNS demyelination. *Cell Stem Cell* **6**:578–590.

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., *et al.* (2012). PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics* **11**:010587.

Zhang, S.-C., Ge, B. and Duncan, I.D. (1999). Adult brain retains the potential to generate oligodendroglial progenitors with extensive myelination capacity. *Proceedings of the National Academy of Sciences* **96**:4089.

Zhao, L., Zhang, M. and Cong, H. (2013). Advances in the study of HLA-restricted epitope vaccines. *Human Vaccines & Immunotherapeutics* **9**:2566–2577.

Zhao, T., Bernstein, K.E., Fang, J. and Shen, X.Z. (2017). Angiotensin-converting enzyme affects the presentation of MHC class II antigens. *Laboratory Investigation* **97**:764.

Zheng, H., Chang, L., Patel, N., Yang, J., Lowe, L., Burns, D.K. and Zhu, Y. (2008). Induction of Abnormal Proliferation by Nonmyelinating Schwann Cells Triggers Neurofibroma Formation. *Cancer Cell* **13**:117–128.

Zhou, Y. and Reddy, V. (2014). Metastatic oligodendroglioma in a patient with no surgical treatment. *American Journal of Clinical Pathology* **142**:A003–A003.

Zilberberg, J., Feinman, R. and Korngold, R. (2015). Strategies for the identification of T cell-recognized tumor antigens in hematological malignancies for improved graft-versus-tumor responses after allogeneic blood and marrow transplantation. *Biology of Blood and Marrow Transplantation* **21**:1000–1007.

Zuliani, E., Stegagno, C., Moretto, G., Mazzucco, S., Bonetti, B., Zanusso, G., Valdo, P., *et al.* (2002). Enhanced expression of NGF receptors in multiple sclerosis lesions. *Journal of Neuropathology & Experimental Neurology* **61**:91–98.