Investigating the trafficking and interactions of the pre-T cell receptor complex

Sam Garforth

supervised by Dr. John James and Dr. Alex Jones

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Interdisciplinary Biomedical Research.

University of Warwick, Warwick Medical School.

January 2023

Contents

List of Figures

List of Tables

0.1 Acknowledgements

My sincerest thanks to my supervisor John James who has given me four years of amazing guidance, encouragement and support. Whenever this project has left me tired or confused, a meeting with John restored my enthusiasm. John has given me the freedom to try, fail and learn things myself whilst always keeping me on track. I have learned an awful lot from his mentorship.

Thanks as well to other members of the James Lab, in particular Muna. From long evenings in the lab to drinking wine at a conference in the Tuscan hills, Muna has been an incredible colleague and friend. Our PhDs were disrupted by the pandemic but we made it through together.

I'd like to thank members of the MRC DTP community for enriching my Warwick experience especially Jonathan Miller for his advice and humour. The program has been a big part of my learning and I am a convert to its interdisciplinary structure and goals.

A great many people gave me technical help, in particular Cleidi Zampronio with proteomic experiments and Karuna Sampath who supported my Zebrafish work. Thanks also to Dani, Sim, Conor, Agi and the BSU team for helping me with my fish-related confusions.

I'm grateful to everyone on floor 3 of the IBRB for creating the atmosphere that made the final year of my work particularly entertaining and enjoyable. Thanks as well to my friends, coaches and training partners in the judo club for helping me unwind.

Finally, I'd like to thank my parents for their endless support and my partner Mary for keeping me somewhat sane throughout this journey. Mary has put up with my busy schedule with nothing but patience and understanding. I was only infected with COVID once in return.

0.2 Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree apart from some background material in chapter 3 which was previously submitted for a Masters in Interdisciplinary Biomedical Research.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- Two figures in section 3.1.2 in the introduction of Chapter 3 feature data acquired by Andrei Smid with analysis by John James.
- Data in figures 3.19 and 3.25 was acquired and analysed by John James.
- Data in figures 3.24 and 3.26 was acquired by the author with additional analysis by John James.

These instances are also indicated within the text.

Some of the data presented in Chapter 3 has been submitted for publication by the author.

0.3 Abstract

T cells must pass through several checkpoints during their development in the thymus, beginning with selection of cells that have productively rearranged the TCRβ gene. Thymocytes are rescued from apoptosis at this checkpoint by transient surface expression of the pre-TCR, a complex of the expressed TCRβ chain and the invariant pTα chain. Whilst pre-TCR signalling initiates commitment to the $\alpha\beta$-lineage, signalling from the structurally similar $\gamma\delta$TCR, or premature expression of the $\alpha\beta$TCR, instead biases cells towards the $\gamma\delta$-lineage. How thymocytes cells can distinguish between the signalling initiated from these receptors is still unclear.

This work reconstitutes pre-TCR expression in a non-immune system, allowing the trafficking of the receptor to be studied in detail without the effects of receptor signalling. I find the low surface expression of the pre-TCR is a consequence of poor complex assembly and lysosomal degradation. I also present evidence that low surface expression of the pre-TCR produces a weak but distinct tonic signal. This might enable the receptor to be distinguished from the TCR at the β-selection checkpoint.

This work also investigates TMEM131, a highly conserved but poorly characterised membrane protein, previously identified as a pre-TCR interaction partner. Data presented here is consistent with TMEM131 acting as an ER chaperone. TMEM131 is directed to the ER by its signal peptide and is retained by a sequence in its C-terminal tail. The cytoplasmic tail of TMEM131 appears to be cleaved at multiple sites which I attribute to partial degradation by the proteasome. I applied a BioID assay to identify additional interaction partners and clients of TMEM131.

I have initiated work to make a TMEM131 knockout in Zebrafish in order to study the function of the protein during development.

0.4 List of Abbreviations

Ab	Antibody
APC	Allophycocyanin
APEX	Ascorbate Peroxidase
BCR	B Cell Receptor
Bcl11b	B cell lymphoma–leukaemia 11B
BioID	Biotin Identification
BP	Biotin-Phenol/Biotin-Tyramide
BS3	Bissulfosuccinimidyl suberate
cas9	CRISPR associated protein 9
CD	Cluster of Differentiation
cDC	conventional Dendritic Cell
CDR	Complementarity Determining Region
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
cTEC	cortical Thymic Epithelial Cell
CTL	Cytotoxic T Lymphocytes
DAMPs	Damage-Associated Molecular Patterns
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DN	Double Negative

DP	Double Positive
DTT	Dithiothreitol
ER	Endoplasmic Reticulum
ERAD	Endoplasmic reticulum-associated protein degradation
FBS	Fetal Bovine Serum
FKBP	FK506 binding protein
FRB	FKBP-Rapamycin binding
GFP	Green fluorescent protein
Gt	Goat
GTA	GFP-trap agarose
HA	Haemagglutinin
HEK	Human Embryonic Kidney
HRP	Horseradish Peroxidase
IFNγ	Interferon γ
IL	Interleukin
ITAM	Immunoreceptor tyrosine-based activation motif
LAT	Linker for Activation of T cells
LB	Lysogeny broth
Lck	Lymphocyte-specific protein tyrosine kinase
LPS	Lipopolysaccharide
MHC	Major Histocompatibility Complex
Ms	Mouse
mScar	mScarlet fluorescent protein
mTCR	mature T Cell Receptor
mTEC	medullary Thymic Epithelial Cell
Nb	Nanobody
NK	Natural Killer cell
PAMPs	Pathogen-Associated Molecular Patterns
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
pMHC	peptide MHC molecule
pre-TCR or pTCR	pre-T cell receptor
PRRs	Pattern Recognition Receptors
pTα	pre-T alpha
RAG	Recombination-Activating Genes
Rb	Rabbit
RUSH	Retention Using Selective Hooks
SBP	Streptavidin Binding Peptide
SNPs	Single-nucleotide polymorphisms
SOC	Super Optimum Broth
SP	Single Positive
SPPLAT	Selective Proteomic Proximity Labeling Assay Using Tyramide
T7E1	T7 Endonuclease I
TCR	T Cell Receptor
TMEM131	Transmembrane protein 131
TMEM131L	Transmembrane protein 131-Like
TRA	Tissue Restricted Antigen

TSS	Transcription Start Site
Ub	Ubiquitin
V(D)J	Variable (Diversity) Joining
WT	Wild-type
ZAP70	ζ-associated protein kinase of 70kD

Chapter 1

General Introduction

1.1 Divisions of the mammalian immune system

The immune system covers the entirety of host mechanisms to prevent, contain and eliminate pathogens: entities that induce disease in the host to reproduce and spread. These pathogens range from viruses, extracellular and intracellular bacteria, fungi, protozoan endoparasites and larger parasites such as helminths. This is a diverse group of potential threats made worse by the ability of pathogens, particularly viruses, to evolve rapidly to evade host responses. As initiating any immune response requires recognising the invading pathogen, two complementary strategies have evolved to identify pathogenic infection with a limited set of receptor genes: the innate and adaptive immune systems.

The innate immune system includes non-cellular mechanisms, such as secreted anti-microbial compounds, and cellular responses using germline-encoded pattern recognition receptors (PRRs). PRRs recognise conserved features of many pathogens (known as pathogen-associated molecular patterns or PAMPs) as well as inflammatory molecules produced by host cells (damage associated molecular patterns or DAMPs). Innate immune systems of varying complexity are present in all multicellular organisms and are fast to initiate a response to infection. Though the innate immune system prevents or eliminates the majority of pathogenic infections, it is limited by a lack of specificity to individual pathogens. This means pathogens have evolved ways to evade or suppress the innate immune response and little long-term immunity can be created through innate mechanisms to reduce the risk of subsequent infections.

The adaptive immune system is unique to vertebrates and evolved separately in jawless fish and jawed fish from which mammals are descended.[1] This system uses the somatic recombination of a limited set of genetic elements to create immune receptors with enormous sequence diversity in a pool of specialised immune cells. In mammals, the adaptive immune system features two recombined immune receptors: the T cell receptors (TCRs) of T lymphocytes and the B cell receptors (BCRs) of B lymphocytes which are precursors to secreted antibodies. During development of these cells, these receptors must be extensively tested to remove receptors that are nonfunctional or worse autoreactive to the host. Cells with receptors failing this process are killed or prevented from developing further. Of the remaining cells, only a few may ever encounter a ligand recognised by their recombined immune receptor and actually initiate an adaptive immune response. These cells are then activated and initiate humoral and cellular responses specific to the invading pathogen. The time needed for the correct cells to be activated and amplify means adaptive immune responses are typically slow. Once active however, the response is very specific

to the pathogen encountered. A proportion of the activated cells differentiate into memory cells that are able to respond more quickly if the pathogen is encountered again. However, the adaptive immune system has a cost. It is inherently inefficient due to the low proportion of cells generated that go on to be used and errors in the recombination and selection process lead to oncogenesis and autoimmune diseases.

In this chapter, the receptors and effector cells of the innate immune system will be summarised followed by an explanation of how the innate immune system initiates an adaptive immune response. The T cell and B cell responses of the adaptive immune system will then be covered before an in-depth discussion of the function and development of the TCR via the pre-T cell receptor (pre-TCR). The cell biology of the pre-TCR and one of its interaction partners, Transmembrane protein 131, will be the main focus of this thesis.

1.2 Components of the innate immune system

1.2.1 Non-cellular components

The first defence against pathogens is physical barriers including skin and mucus membranes at body orifices such as the genitourinary tract. Pathogens entering the body through gaps in this epithelial barrier encounter the host complement system, a cascade of proteolysis reactions involving over 30 soluble blood proteins and membrane proteins on host cells. These reactions are triggered near the surface of pathogen membranes through three pathways known as the classical, lectin and alternative pathways. These pathways can have three impacts: the release of small molecules that attract innate immune system cells, the deposition of complement proteins on the pathogen's surface marking it for phagocytosis (opsonisation), and the formation of a protein complex called the membrane attack complex that is able to cause lysis of target membranes.[2] In practice, the membrane attack complex appears to be most effective against gram negative *Neisseria* species and the opsonising and inflammatory effector components of pathway are more vital for host immunity.

1.2.2 Effector cells of the innate immune system

The effector cells of the immune system are white blood cells, otherwise known as leukocytes. These ultimately derive from pluripotent haematopoietic stem cells in the bone marrow that initially differentiate to produce precursors cells of the common lymphoid and myeloid lineages. The common lymphoid lineage produces the T and B lymphocytes of the adaptive immune system whilst the common myeloid lineage produces the effector cells of the innate immune namely granulocytes, macrophages, dendritic cells, and mast cells.[3]

Leukocytes signal to each other through peptide chemical messengers called cytokines[4] allowing both the innate and adaptive immune systems to coordinate a response targeted to a subset of pathogens. For instance, cytokine IL-12 is produced in response to infection by intracellular pathogens such as viruses or intracellular bacteria and induces a cell-mediated response from macrophages, T cells and NK cells. IL-4 meanwhile, coordinates a humoral immune response with a greater focus on antibody production more suited to defence from extracellular pathogens such as helminths. Some cytokines, called chemokines, act as chemical messengers to attract immune cells to specific locations.

Neutrophils, a type of granulocyte distinctive for their polymorphic nuclei, are by far the most numerous type of leukocyte and accumulate quickly at sites of tissue damage. These are mobile phagocytic cells that can engulf many pathogens and kill them with a burst of radical oxygen species. However, their lifespan is very short and they are unable to resolve many infections. Later in the course of an infection, circulating cells called monocytes arrive at the site of tissue damage and differentiate into a specialised phagocytic cell called a macrophage. Initially these macrophages adopt an aggressive M1 phenotype involving the phagocytosis of both pathogens and dying neutrophils and the secretion of pro-inflammatory cytokines such as IL-12 and TNFα. Later in the course of infection, IL-4 signalling from T_H2 lymphocytes encourages macrophages to adopt an anti-inflammatory M2 phenotype and contribute to the process of wound healing.[5] Resident macrophages are also present in tissues such as the lungs, liver, intestine and central nervous system and adopt a heterogenous range of functions. Along with dendritic cells discussed later, macrophages and monocytes are specialised antigen presenting cells that enable the innate immune system to initiate an adaptive immune response.

Eosinophils, basophils, and mast cells are granulocytes making up a more minor fraction of innate immune cells. All are characterised by their polymorphic nuclei and cytoplasmic granules of antimicrobial compounds and enzymes. They play a key role in inducing an inflammatory response and wound healing as well as the immune response to parasitic worms.[6]

1.2.3 Recognition of conserved pathogenic features by innate immune receptors

The aforementioned effector cells of the innate immune system use a variety of genome-encoded pattern recognition receptors (PRRs) that recognise conserved features of groups of pathogens. Ideal targets are features of pathogens that are not expressed by host cells (i.e. PAMPs) such as bacterial LPS and viral genomes. A key family of PRRs is the Toll-like receptors (TLRs) discovered first in *Drosophila* but widely expressed in both vertebrates and invertebrates.[7] TLRs expressed on the cell surface of macrophages recognise conserved bacterial features such LPS. Dendritic cells and B cells express TLRs on internal endosomal membranes than can identify nucleic acid components of viral genomes. TLR signalling leads to the production of cytokines such as IFNγ through the NF-κB pathway.[8] Extracellular PAMPs can also be sensed by members of the C-type lectin receptor (CLR) family. Many of these receptors recognise carbohydrate ligands such as mannose. Mannose residues are normally hidden by subsequent glycosylations in human glycoproteins but are exposed in the cell walls of fungi and mycobacteria.[9]

Mammals have a number of cytoplasmic PRRs to identify viral infection, particularly through detecting viral RNA or DNA genomes. These PRRs include RIG-I that recognises short RNA sequences with 5'triphosphate caps and MDA5 that recognises long genomic RNA. Cytoplasmic DNA can be detected via the cGAS-STING pathway.[10,11]

1.3 The central role of $\alpha\beta$T cells in the adaptive immune response

1.3.1 Antigen presentation using Major Histocompatibility Complex (MHC) molecules

Antigens are molecules, usually proteins, that can generate an adaptive immune response, classically but not exclusively the production of antibodies by B cells. Whilst most proteins within a cell are host proteins performing their normal function, these are not easily distinguished from proteins that are pathogenic in origin. Antigen presentation is the process by which cells display a sample of peptides from their protein contents to T lymphocytes of the adaptive immune system so that foreign peptides can be identified. Mechanistically, peptides are presented by surface proteins called Major Histocompatibility Complex (MHC) molecules and peptides tested by the T cell receptor (TCR). The strength of TCR signalling in response to presented peptides is fundamental to the effector functions of the adaptive immune system.

There are two types of MHC molecule which differ in their structure, expression, genetics and the origin of peptides they present. Each type also interacts with different subsets of T lymphocytes: the CD8+ cytotoxic T lymphocytes (CTLs) which are directly involved in cell-mediated immunity and CD4+ T helper cells (T_H) which orchestrate an immune response including production of antibodies.

MHC-I molecules present peptides from the degradation of cytoplasmic proteins

MHC class I molecules are expressed by all nucleated, somatic cells and present peptides from the intracellular digestion of cytoplasmic proteins by the proteasome. This includes both host proteins and proteins from intracellular pathogens such as viruses. Although typically just representing a random sampling of the cell's proteome, the peptides presented can be changed by downstream processes and cellular stress. The rate of protein degradation is thought to be increased by the formation of defective ribosomal products (DRiPs) from misfolded or mistranslated proteins.[12] These DRiPs are produced in greater amounts during high protein synthesis such as during viral infection. Some cells such as dendritic cells express proteasome subunit variants forming the immunoproteasome that produces different peptide specificities. These subunits can be expressed in other cell types due to interferon signalling accompanying viral infection. Whilst a large majority of peptides are degraded completely, some are imported into the ER by the ABC-family transporter TAP and trimmed by ER aminopeptidase 1 before being loaded onto MHC class I molecules by the peptide loading complex (tapasin, calreticulin and protein disulphide isomerase ERp57).[13] Peptide-MHC I complexes are trafficked to the cell surface and are sampled by the T cell receptors of CD8+ cytotoxic T cells.

MHC-I molecules are heterodimers of a variable, single pass transmembrane α chain and an invariant, single domain β chain called β_2-microglobin. The two N-terminal domains of the α chain form a deep peptide-binding groove, consisting of two flanking alpha-helices and a base of eight anti-parallel beta strands. The ends of this groove are closed off by conserved tyrosine residues limiting most bound peptides to 8-10 amino acids in length. The most important binding interactions between the peptide and the MHC molecule occur between conserved residues of the groove and the peptide termini.[14] This permits the central sequence of the peptide that interacts with the TCR to vary, though different alleles have different peptides specificities due to additional interactions between amino acids in this region and

residues on the base and sides of the groove.[15]

MHC-II molecules present peptides from the degradation of engulfed material

MHC class II molecules are constitutively expressed by specialised antigen presenting cells such as B cells, dendritic cells and macrophages, but can also be induced on epithelial cells and endothelial cells at sites of inflammation by IFNγ. These molecules present peptides originating primarily from engulfed material, such as extracellular pathogens that have been digested in late endosomal compartments. MHC II molecules are synthesised in the ER but can only leave in complex with an invariant chain (Ii). The cytoplasmic sequence of this invariant chain has sorting motifs that directs the complex to late endosomes, during which a portion of the chain known as the class-II associated Ii peptide (CLIP) occupies the peptide binding groove of the complex. In the endosomes most of the invariant chain is digested leaving only the CLIP that is removed by the non-classical MHC II molecule HLA-DM to allow binding of endosomal peptides. When bound to a peptide, MHC-II molecules are transported to the cell surface and can be sampled by the T cell receptors of CD4+ T_H cells. Peptides from cytoplasmic proteins can also be presented by MHC II molecules if those proteins are degraded via the autophagy pathway.[16] The two pathways of antigen presentation are contrasted in figure 1.1.

MHC-II molecules are heterodimers of two variable α and β chains, each with two domains and a transmembrane helix. The peptide binding groove, formed by the two N-terminal domains of each molecule, is open at the ends allowing the binding of peptides 13-25 residues in length. Crystal structures of the extracellular domains of MHC-I and MHC-II molecules in complex with the extracellular domains of the TCR are shown in figure 1.2.

Genetics of the major histocompatibility complex

Both types of MHC molecules are encoded on a highly gene-dense 3Mbp stretch of chromosome six. Individual genes within this cluster are called human leukocyte antigen (HLA) genes. These include not only the classical MHC genes just discussed but also non-classical MHC molecules, genes for many cytokines such as TNFα, the complement proteins C2 and C4, genes involved in antigen presentation such as TAP and the genes for the stress-induced ligands MICA and MICB.

In humans, there are three classical class I α-chain genes called HLA-A, -B, and -C and three pairs of MHC class II α- and β-chain genes called HLA-DR, -DP, and -DQ. In many cases the HLA-DR cluster contains an extra β-chain gene whose product can pair with the DRα chain. These genes are highly polymorphic between individuals leading to huge population diversity. This diversity is thought to increase the range of peptides that can be presented by a heterozygous individual, maximising the chances of a cell being able to initiate an immune response from an infection.[19] Non-classical MHC products include HLA-DM involved in peptide exchange by MHC-II molecules and three non-classical type I MHC molecules that are thought to interact with the KIR receptors of NK cells.[20–22]

1.3.2 Dendritic cells as the interface between the innate and adaptive immune systems

Though certainly not the only cells capable of antigen presentation, dendritic cells play a critical role in initiating the adaptive immune response. Most importantly, they transfer antigens from the

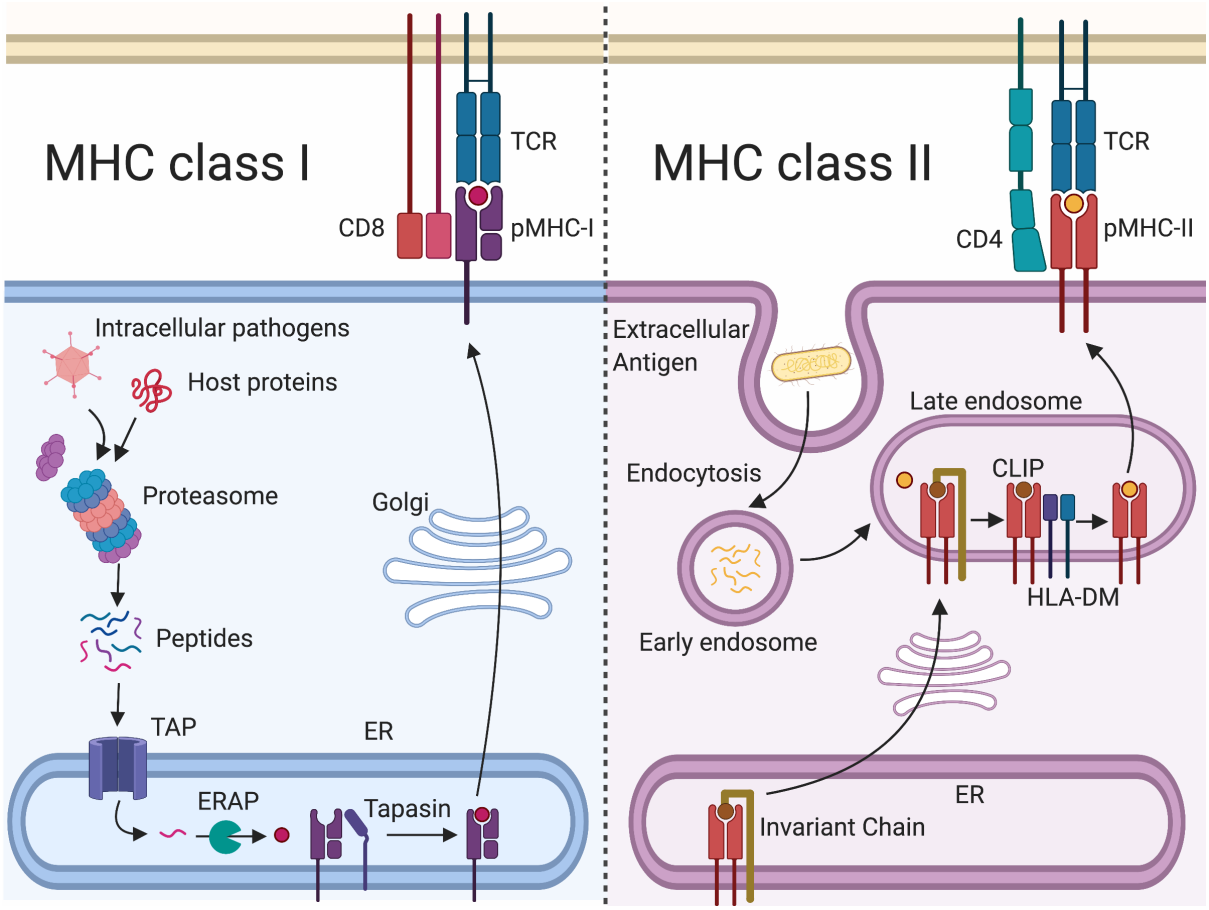

Figure 1.1: Pathways of antigen presentation for MHC-I (left) and MHC-II (right) molecules.
Cytoplasmic proteins are degraded via the proteasome and presented via MHC-I molecules (left). Digested peptides are imported into the ER by TAP1/2, processed by aminopeptidases such as ERAP1 and loaded onto MHC-I molecules by the peptide loading complex which includes the protein tapasin. The pMHC-I complex is then transported to the cell surface via the Golgi body and can interact with CD8+ T lymphocytes.
Proteins that have been ingested by phagocytosis can be digested in late endosomes and presented by MHC-II molecules (right). MHC-II are synthesised in the ER in complex with an invariant Ii chain which is trimmed away and the remaining CLIP peptide removed using HLA-DM. pMHC-II complexes at the cell surface can interact with CD4+ T lymphocytes.
Created using BioRender.com

.

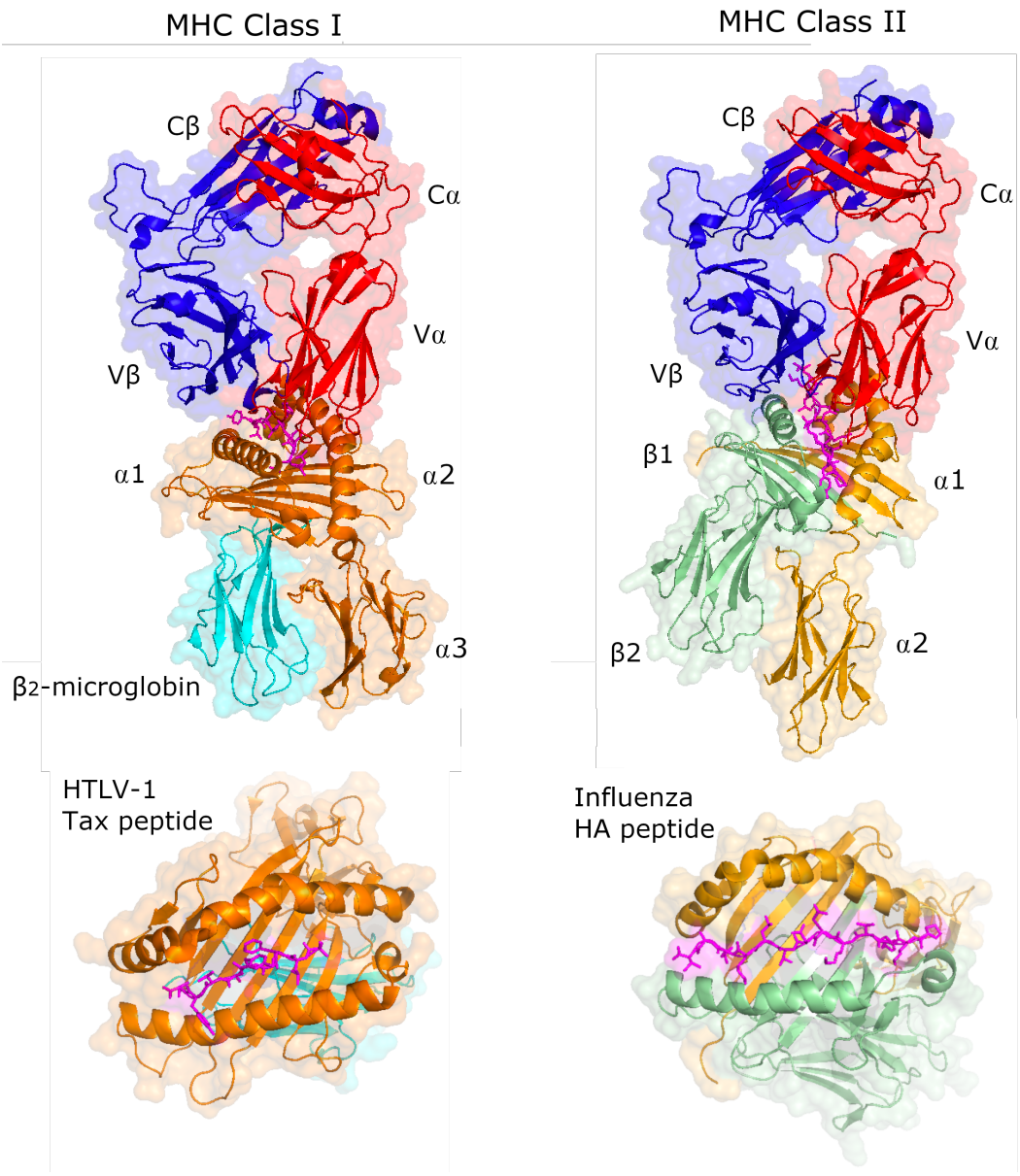

Figure 1.2: Structures of the extracellular domains of the $\alpha\beta$TCR with the extracellular domains of a MHC class I molecule (left) and class II molecule (right) with top views of viral peptides in the MHC peptide-binding groove of each complex.
The binding mode of the TCR is similar in both cases though differences are apparent especially in the hypervariable loops of the TCR variable domains. Binding to each MHC involves a coreceptor: CD8 in the case of MHC class I which interacts with the α3 domain and CD4 for the MHC class II which binds to the β2 domain.
Left: The $\alpha\beta$TCR (red and blue respectively) with a MHC class I molecule, here the HLA-A2 α chain (orange) with β_2-microglobin (cyan). The peptide-binding groove, made of the α1 and α2 domains is closed at both ends and here contains a 9 amino acid long Tax peptide from Human T-lymphotropic virus. Based of crystal structure 1QRN by Ding *et al.*[17]
Right: The $\alpha\beta$TCR with of a MHC class II molecule, here HLA-DR1 made of the HLA-DR α chain (yellow) and HLA-DR1 β chain (green). The peptide-binding groove, made of the α1 and β1 domains is open at both ends and here contains a 13 amino acid long Influenza HA peptide. Based on crystal structure 1FYT by Hennecke *et al.*[18]

peripheral tissues, where most infections arise, to the lymph nodes and the spleen where naive T cells are present.

Classical dendritic cells (cDC) follow the "Langerhans cell paradigm" of DC function[23,24] in which the cell starts out in the body peripheries with low expression levels of MHC molecules and T cell stimulatory molecules but high expression of innate immune receptors such as TLRs, C-type lectin receptors and cytokine receptors. When resident in tissues they behave similarly to tissue-resident macrophages including phagocytosis of opsonised material. In the absence of inflammation, cDC occasionally travel to the lymph nodes and present innocuous material or self-antigens to T cells via pMHC molecules to induce T cell tolerance.[25] However, cDC migration to the secondary lymphoid tissues is greatly increased upon tissue inflammation. During this migration the cells mature to increase their ability to activate T cells by increasing the surface expression of pMHC complexes and costimulatory molecules such as CD80 and CD86. Migration towards the lymph nodes is directed by chemokines CCL19 and CCL21 through receptor CCR7, the same system used for migrating T cells.

In lymph nodes, cDCs present pMHC molecules to naive T lymphocytes expressing clonal-variant TCRs. The consequences of these interactions are described in the next section. A key feature of DCs is their ability to also present peptides from endocytosed material using MHC-I molecules rather than only MHC-II molecules as with B cells and macrophages. This cross-presentation has important consequences for the activation of the adaptive immune system in response to viruses and intracellular bacteria.[26,27]

T lymphocytes are a hugely important type of immune cell that arise from the common lymphoid lineage in the bone marrow but continue their development in the thymus. During this process clonally-specific T cell receptors (TCRs), composed of α and β chains, are created by the recombination of short gene elements and tested. TCRs have unique ligand specificities for peptide-MHC complexes depending on their sequence. A more minor fraction of T cells express a TCR composed of $\gamma\delta$chains whose function and development will be described later in this chapter.

T lymphocytes circulate in the blood and reside in lymph nodes as mostly spherical cells without significant polarity. In routine circulation, T cells are relatively non-adhesive as a primary adhesive molecule called leukocyte function-associated antigen-1 (LFA-1) is kept in an inactive conformation. Once activated by chemokine signalling through other receptors, LFA-1 adopts an active conformation and interacts with the intercellular adhesion molecule (ICAM) family, primarily ICAM-1 which is widely expressed.[28] In this adherent state, protrusions from a T cell can rapidly sample the peptide MHC molecules expressed by nearby cells. Identification of a pMHC molecule initiates TCR triggering and the formation of a polarised cell-cell contact between the T cell and the antigen presenting cell called the immunological synapse. This process is discussed more in section 1.4. This contact appears as a series of concentric rings named supramolecular activation complexes (SMACs). The TCR and costimulatory receptors form the central supramolecular activation complex (cSMAC) together with smaller adhesion receptors such a CD2/LFA-2 which interacts with its ligand CD58/LFA-3 on the APC. Surrounding these molecules is a ring of larger adhesion receptors such as LFA-1 called the peripheral SMAC (pSMAC) and surrounding this a ring of molecules with larger extracellular domains such as anti-adhesion molecule CD43 and receptor phosphatase CD45 called the distal SMAC (dSMAC).[29,30] The immunological synapse is important in both T cell activation and T cell effector functions.

A great many T cell subsets can be described on the basis of phenotype and their expression of different transcription factors and surface receptors. The most significant distinction in terms of effector function is between CD4+ and CD8+ T cells, based on the coreceptors that accompany the $\alpha\beta$TCR. The CD4

coreceptor is a surface marker with four immunoglobulin domains (D1-4) and a short cytoplasmic tail. The N-terminal of these domains interacts with the β2 region of MHC class II molecules. The CD8 coreceptor meanwhile is a heterodimer of CD8α and CD8β chains, each with one extracellular domain. CD8 mainly interacts with the α3 domain of MHC-I molecules.[31]

1.3.3 CD4+ T helper cells

Naive CD4+ T cells, that are yet to encounter an antigen they respond to, can be activated by two simultaneous signals from antigen presenting cells, primarily cDCs. The first signal is engagement of the TCR and CD4 coreceptor with a pMHC-II complex presented by the APC, suggesting the presence of a non-self peptide. The second, called a co-stimulatory signal, is the expression of the ligands CD80 and CD86 by the APC which bind to the receptor CD28 of the T cell. These ligands are upregulated due to signalling from the antigen presenting cell's PRRs. In the absence of a simultaneous co-stimulatory signal, TCR engagement induces a state of low responsiveness called anergy. This ensures tolerance of peptides in the absence of tissue danger.

Once activated CD4+ T cells proliferate and differentiate into different effector cells depending on the cytokine signals from the antigen presenting cell.[32–34] Whilst some differences in the roles of these Th subsets is uncontroversial, particularly between Th1 and Th2 subsets first defined, it is debated to what extent many Th cell subsets coined more recently reflect actually distinct cell types rather than merely examples of a continuous spectrum of phenotypes.[35]

Th1 cells

The Th1 phenotype is induced by the cytokines IL-12, produced by DCs and macrophages after PRR activation and IFNγ produced by NK cells. These cytokines activate the transcription factor T-bet which enhances the Th1 fate and inhibits the Th2 and Th17 cell fates. Th1 cells are involved in the elimination of intracellular pathogens including both bacteria and viruses via cell-mediated immune responses. This is done through the secretion of IFNγ and TNF-β which activate macrophages and cytotoxic T lymphocytes and by inducing B cells to make opsonising IgG2a antibodies.[36]

Th2 cells

The Th2 phenotype is induced by the cytokines IL-2 and IL-4. These operate though the master regulator GATA3 which inhibits the Th1 fate whilst enhancing the Th2 fate. Th2 cells have an important role in the protection from extracellular parasites such as helminths through inducing a humoral immune response. This response involves the expression a wide variety of cytokines including IL-4, IL-5 and amphiregulin.

Th17 cells

The subset of CD4+ T cells with the Th17 phenotype was discovered after the Th1/Th2 division. This subset is specialised in eliminating extracellular fungi and bacterial pathogens. Development of this subset is enhanced by IL-6 and TGF-β which activate the retinoic acid receptor-related orphan receptor gamma-T (RORγt) master regulator. Th17 cells secrete the cytokines IL-21 which induces B

cell differentiation into plasmocytes, IL-22 which induces the expression of antimicrobial compounds and IL-17 which induces other cells to produce chemokines that recruit neutrophils to the site of infection.

iTreg cells

Induced T regulatory cells are a CD4+ T cell subset that negatively regulates the immune response similar to natural regulatory T (nTreg) cells that are produced in the thymus. Both cell types are characterised by the expression of the transcription factor FOXP3. Commitment to the iTregs cell type is increased by TGF-β and IL-2. Once this phenotype is induced, iTregs reduce inflammation through production of anti-inflammatory cytokine IL-10 and suppress the activity of DCs through engagement of CD80 with the cell surface receptor CTLA-4.[37]

Tfh cells

Follicular helper T cells are induced by IL-6 and IL-21 acting through the transcription factor Bcl6. These have a critical role of mediating the selection and survival of B cells in germinal centres, sites within the lymph nodes where B cells proliferate and mutate their antibody genes in a process called somatic hypermutation. These sites are important in generating plasma cells expressing higher affinity antibodies and memory B cells that can quickly re-activate the pathogen is encountered again.[38]

1.3.4 CD8+ Cytotoxic T lymphocytes

Cytotoxic T lymphocytes (CTLs) express a clonally-specific TCR and the CD8 coreceptor enabling them to recognise peptide antigens presented by MHC-I molecules. In the primary response to many pathogens, naive CTLs can be activated by DCs expressing a foreign pMHC-I molecule and costimulatory molecules, similarly to CD4+ T cells.[39–41] However, for a more robust response and for CD8+ T cell memory formation, the dendritic cell must be "licensed" by simultaneous interaction with a CD4+ T cell[36,42] (figure 1.3). Once activated CTLs undergo clonal expansion and leave the lymph nodes to search for infected cells.

The extra regulatory step of DC licensing is required for critical control over the effector functions of activated CTLs. Infected cells expressing a pMHC-I molecule can be recognised by the CTL using the TCR and an immunological synapse formed. CTLs can secrete death-inducing effector molecules into this synapse particularly perforin, granzymes and Fas-L which cause lysis of the target cell and controlled cell death via apoptosis. The rates of CTL-mediated cell killing have been quantified *in vivo* arriving at an estimate of 2-12 cells killed per CTL per day with sizeable heterogeneity within the CTL population.[43]

1.4 Signalling mechanism of the TCR

1.4.1 Structure of the TCR-CD3 complex

The TCR is made of α and β chains both consisting of two extracellular Ig-like domains, a transmembrane helix and a three residue cytoplasmic region. The membrane-proximal extracellular domain has a

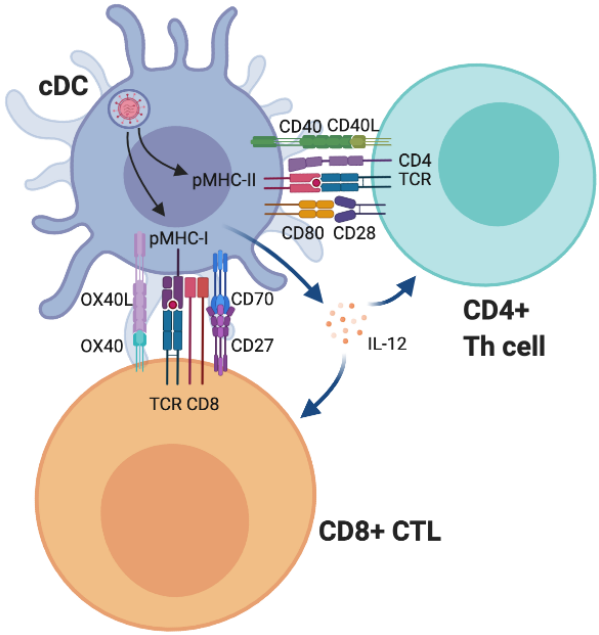

Figure 1.3: Robust CTL activation requires DC licensing through simultaneous interaction with a CD4+ Th cell.
Dendritic cells, as professional antigen presenting cells, can cross-present antigens on both MHC-I and MHC-II molecules. If a peptide is recognised by the TCR of a CD4+ Th cell in the context of costimulatory CD28 signals then the dendritic cell can be licensed through the ligand-receptor pair CD40 and CD40L. This DC can then activate CD8+ CTL through the TCR and TNF-receptor superfamily members including CD70 and OX40.

constant sequence whilst each of the variable domains, more distant from the membrane, contain three complementarity determining loops (CDRs) that vary in sequence due to genetic recombination. These loops contact the peptide-MHC complex.

Unlike other receptors families such as receptor tyrosine kinases, the TCR does not signal directly through its cytoplasmic sequences. Instead signalling occurs through a set of CD3 chains that are noncovalently-associated with the TCR complex. The full complex consists of TCR$\alpha\beta$, CD3$\gamma\epsilon$, CD3$\delta\epsilon$ and CD3$\zeta\zeta$(also called CD247). These dimers are all present in a 1:1:1:1 ratio. CD3γ, δ and ϵ subunits each have a single extracellular Ig-domain, a transmembrane helix and long cytoplasmic tail. CD3ζ has a very short extracellular sequence, a transmembrane helix and a much longer cytoplasmic sequence. A cryo-EM structure of the extracellular and transmembrane domains of the complex is displayed in figure 1.4.

The CD3γ, δ and ϵ cytoplasmic domains contain one copy of a sequence known as an immunoreceptor tyrosine-based activation motif (ITAM) with the consensus sequence: YxxL/I$x_{(6\text{-}8)}$YxxL/I. CD3ζ contains three such motifs. The phosphorylation of the tyrosine residues in these sequences is catalysed by the Src family kinases. In T cells, this is primarily lymphocyte-specific protein tyrosine kinase (Lck) and to an extent Fyn kinase.[45] Lck is tethered to the membrane due to N-terminal myristoylation and palmitoylation and is enriched by interactions with the cytoplasmic domains of the CD4 and CD8 coreceptors.[46] Once phosphorylated, ITAMs recruit Zeta-chain-associated protein kinase 70 (ZAP70) via SH2 domains. This association promotes ZAP70 kinase activity and triggers the activity of assembly protein Linker for activation of T cells (LAT). LAT acts as a scaffold for the assembly and activation of multiple intracellular signalling pathways that amplify the TCR signal and result in changes in gene expression. These changes ultimately result in T cell proliferation and differentiation. Changes in the actin cytoskeleton and in cell adhesion initiate the immunological synapse.[47,48] CD28 integrates co-stimulatory signals at the same interface as the TCR with overlapping but distinct downstream

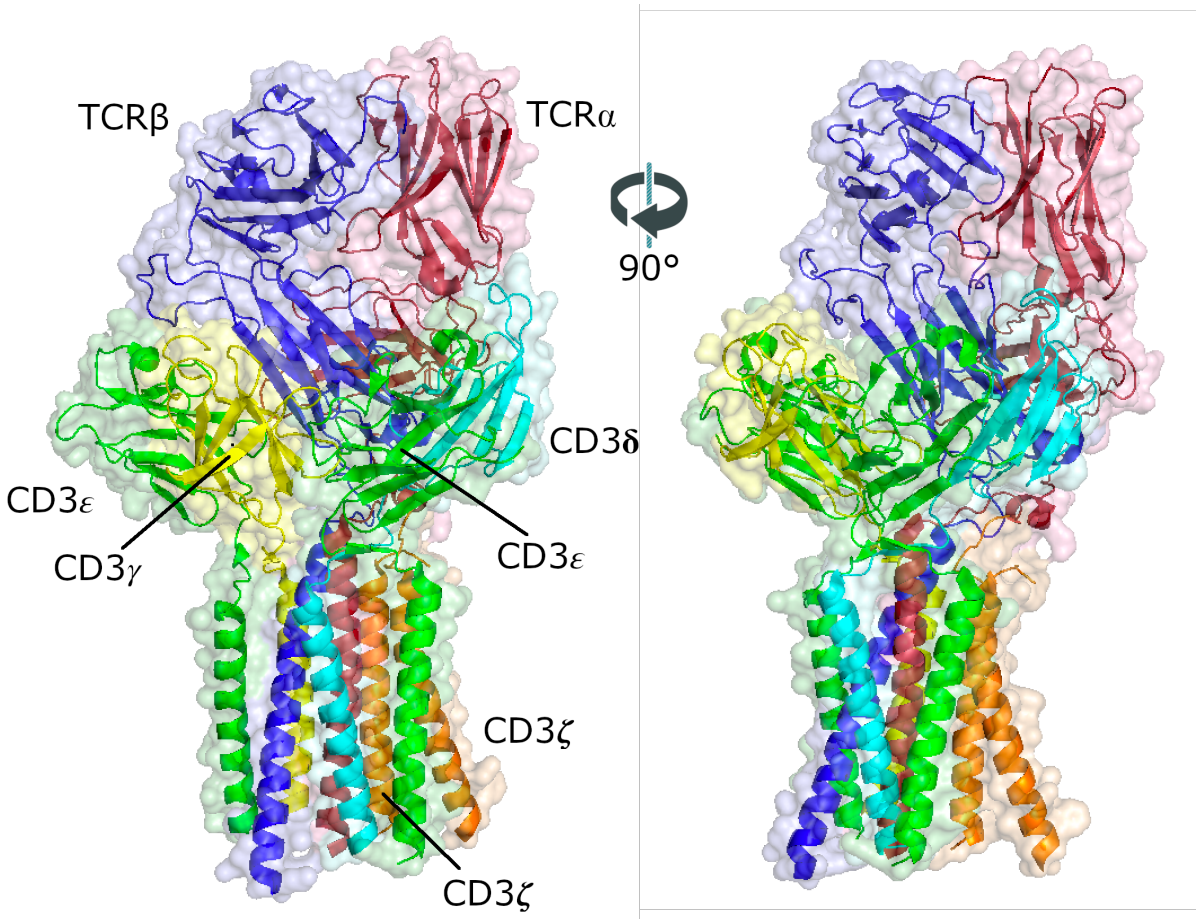

Figure 1.4: Cryo-EM structure of the extracellular and transmembrane domains of the octameric TCR-CD3 complex.
TCRα (red) and TCRβ (blue) each have two extracellular Ig domains, one constant domains and one variable domain with sequence generated by V(D)J recombination. These extracellular TCR domains interact with one dimer consisting of CD3γ (yellow) and CD3ϵ (green) and another dimer consisting of CD3δ(cyan) and CD3ϵ, all of which have a single extracellular Ig domain. The transmembrane part of the complex interacts with a dimer of CD3ζ subunits which are linked via a transmembrane disulphide bond. The short intracellular sequences of the TCR chains and the intracellular ITAM sequences of the CD3 chains are poorly defined suggesting they are disordered. No major conformation changes in the TCR extracellular domains are apparent compared with previous structures of isolated $\alpha\beta$-TCR extracellular domains bound to peptide-MHC molecules. Based on cryo-EM structure 6JXR by Dong *et al.*[44]

signals.[49]

The activity of Lck is controlled by phosphorylation at two tyrosine resides. Tyr394 is a site of autophosphorylation required for the activation of the kinase[50] whilst Tyr505 is phosphorylated by C-terminal Src kinase (Csk) that inhibits Lck. Both Lck and the ITAMs of the TCR complex are dephosphorylated by the receptor-like protein tyrosine phosphatases, primarily CD45[51] but also CD148.[52] CD45 is widely expressed across leukocytes[53] and accounts for up to 10% of the cell surface of a T cell.[51] The protein has two intracellular phosphatase domains, only one of which is functional, a transmembrane helix and a multi-domain extracellular sequence. Whilst this extracellular sequence varies between different isoforms it is heavily glycosylated and forms a rigid rod structure.[54]

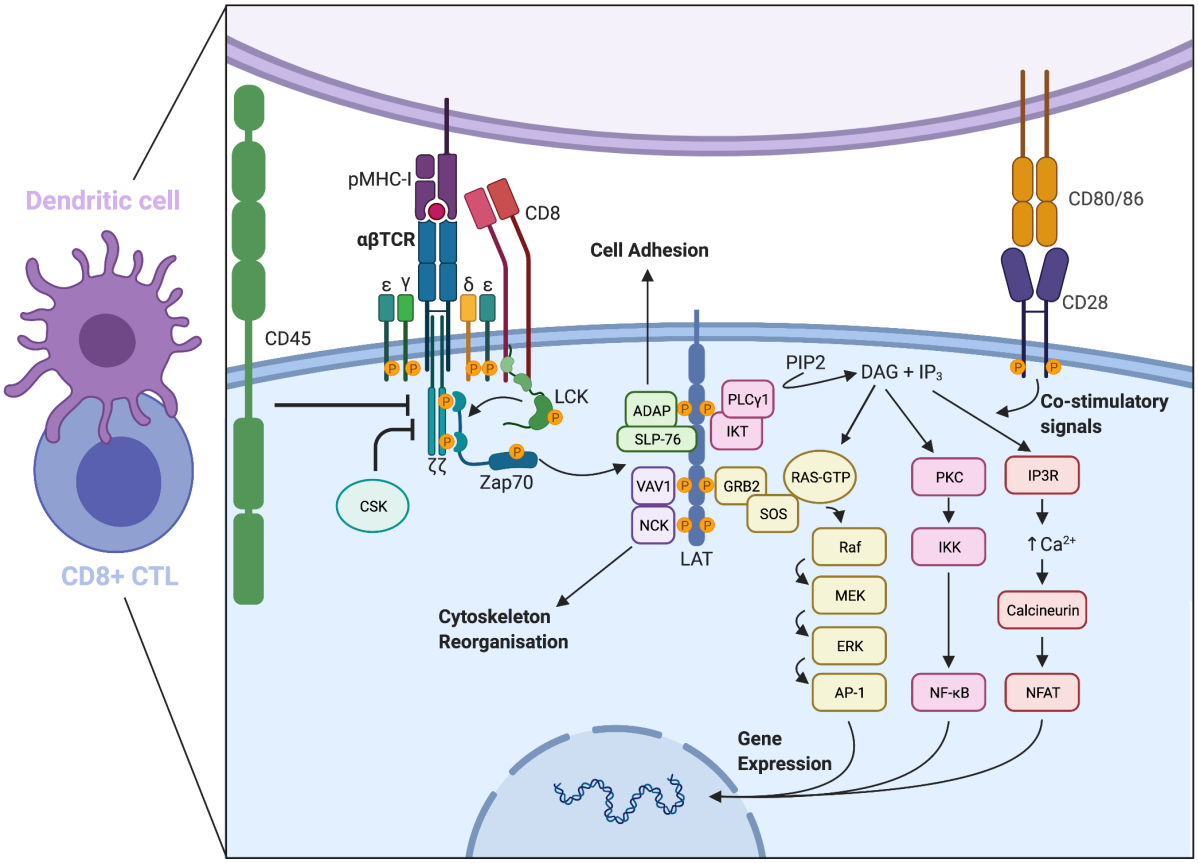

Figure 1.5: The TCR signalling pathway in a CD8+ Cytotoxic T lymphocyte.
Recognition of a pMHC complex by the TCR leads to phosphorylation of the ITAM domains of the CD3 chain by the kinase Lck. Lck is localised to the membrane by N-terminal palmitoylation and interacts with the cytoplasmic sequences of the CD4 and CD8 coreceptors. Lck activity is controlled via phosphorylation including by Csk. Phosphorylated ITAM sequences recruit the kinase ZAP-70 which is also phosphorylated by Lck. ZAP-70 activates the LAT signalosome which triggers cytoskeleton rearrangements and changes in cell adhesion to form an immunological synapse as well as changes in gene expression for T cell activation via the MAP kinase, PKC and calcium signalling pathways. CD28 communicates co-stimulatory signals via overlapping but distinct mechanisms. ITAM phosphorylation is reversed by the phosphatase CD45.

.

It is still unclear if the different CD3 subunit ITAM sequences provide different qualitative signals or whether the same TCR signalling signature is scaled in a quantitative fashion by each of the ten ITAMs. All the CD3 subunits are believed to interact with ZAP70 when doubly phosphorylated though CD3ϵ and CD3ζ are also reported to have other binding partners that could be important in transmitting different signals.[55–58] Earlier studies investigating mouse knockout lines reported a distinct role for CD3δ in transmission of TCR signalling in the DP stage of development but this could not be replicated more recently suggesting the CD3 ITAM domains are functionally redundant.[59] Reconstitution of

TCR signalling using an artificial receptor with a variable amount of ITAM sequences and measuring responses at a single cell level found that increasing the number of ITAMs enhanced the probability of signal transduction from each receptor, thereby increasing the fraction of activated T cells, rather that amplifying the downstream response from each receptor.[60]

1.4.2 Models for proximal TCR signalling

Though the downstream mechanism of TCR signalling are relatively established, there is still debate as to how pMHC binding by the TCR triggers cytoplasmic ITAM phosphorylation by Lck. The competing models for TCR signalling involve either receptor clustering, a conformational change in the receptor to transmit mechanical force, or the kinetic segregation of the receptor from phosphatases due to membrane dynamics.[48,61] Despite convincing evidence for the kinetic segregation model reflecting the dominant mechanism for the TCR, proponents of the competing models have published recently.[62,63]

Conformational change model

A number of different conformational changes in the TCR-CD3 complex have been suggested to be responsible for signal transduction. Some studies propose that pMHC ligand binding induces changes in the TCR constant domains causing relative movement compared to the extracellular domains of the CD3 dimers. This is supported by minor differences in TCR Xray crystal structures with and without ligands[64,65] or NMR data.[66] However, there is little agreement as to whether a pushing, pulling or twisting force on the TCR-CD3 complex is involved.[67] Recently a cryo-EM structure of the complete TCR complex attached to a pMHC ligand has been published showing little difference to the cryo-EM structure of the TCR complex alone.[68] This appears convincing evidence that the TCR operates without large structural rearrangements but is unlikely to end debate.

Clustering models

The clustering model of TCR triggering proposes that ligand binding causes TCR-CD3 complexes and coreceptors to cluster, increasing the local concentration of Lck and enabling trans-autophosphorylation of ZAP70 and ITAM domains. This model explains how TCR signalling is induced artificially by crosslinking the receptor with anti-TCR$\alpha\beta$ antibodies or oligomeric pMHC molecules but not by Fab fragments or monomeric pMHC molecules.[69]

Implicit assumptions in this model are the oligomeric states of the TCR before and after ligand binding which have been disputed by many authors. Whilst the formation of TCR dimers has been suggested[70,71] it seems likely that the glycosylation of the TCRα and β chains would impede lateral contacts between adjacent TCRs thereby preventing a stable dimer interface.[72–74]

Although larger clusters of TCR and pMHC in the micrometre scale are widely reported in and around the cSMAC of the immunological synapse, these are known to form after TCR triggering.[75] Smaller clusters of 5-20 TCR-CD3 complexes of around 35-70 nm have been reported in resting T cells and observed using both electron microscopy[76] and single molecule localisation microscopy super-resolution imaging techniques.[77] These nanoclusters were predicted to play a role in signal transduction via propagation of phosphorylation of pMHC-ligated TCRs to unligated TCRs within the cluster through a cooperative conformational change[76] or by increasing avidity for pMHC molecules.[78] Similar microscopy techniques

reported that the TCR and LAT are present in separate protein island clusters that concatenate upon TCR triggering, bringing the TCR together with its signalling components.[79] TCR nanoclusters are often discussed together with lipid rafts: membrane domains enriched in glycosphingolipids, cholesterol and lipid-modified proteins such as Src-family kinases. Whether TCRs belong within or outside these domains prior to or after pMHC ligation[80,81] and the role of these rafts in TCR signalling is still unclear.[82]

Other authors question the existence of TCR nanoclusters and their appearance in single molecule localisation microscopy attributed to over-counting artefacts.[83,84] Indeed studies using single molecule spectroscopy,[85] fluorescence recovery after photobleaching[86] and label-density-variation single molecule localisation microscopy as well as STED microscopy[87] have concluded that the majority of cell surface TCR is present as monomers.

A challenge for the clustering model of TCR triggering is that T cell activation is observed using very small or even single copies of cognate pMHC complexes.[88,89] This means few of the TCRs in a cluster could be bound by a cognate ligand and yet activation is still conferred throughout the cluster. An attempt to reconcile this issue with the clustering model proposed the formation of pseudo-dimers where the CD4 or CD8 coreceptors bridge pMHC-interacting and non-interacting TCRs.[88,90] This would seem incompatible however with the relatively low affinity of the CD4 or CD8 coreceptors to pMHC complexes[91] as well as evidence of TCR triggering in the absence of co-receptor binding such as with immobilised anti-CD3 antibodies.[92]

Kinetic segregation model

The kinetic segregation model proposes that at resting state, there is a balance between CD3 ITAM phosphorylation by Lck and dephosphorylation by CD45 and CD148 with the phosphatases dominating. Transient interaction between a T cell and an antigen presenting cell, initially via proteins such as LFA-1, form localised regions where the two cell membranes are close together for interactions. These close-contact regions would force the exclusion of proteins with large, highly glycosylated extracellular domains such as CD45 and CD148[93] but could be stabilised by smaller adhesion proteins such as CD2. TCRs and costimulatory receptors would be expected to diffuse into and out of these regions of CD45 segregation but remain confined to them if making a long-lived interaction with ligands within this zone. Once confined, ITAM phosphorylation by Lck cannot be reversed so signal transduction is initiated.[94] The CD4 and CD8 coreceptors are important though not essential in this model by recruiting Lck to the site of TCR-pMHC interaction.

This model is supported by several lines of evidence. Truncation or removal of the CD148 and CD45 ectodomains has been shown to reduce the segregation of these molecules from ligated TCR complexes and prevent TCR triggering.[95] Similarly increasing the dimensions of the TCR-pMHC complex by the addition of Ig domains to the extracellular region of the MHC molecule was found to reduce TCR triggering. Imaging showed an increase in the intermembrane distance between T cells and antigen presenting cells expressing the modified MHC molecule and more CD45 is present at the cell interface.[96] The kinetic segregation model is also consistent with the artificial triggering of TCRs by crosslinking reagents such as oligomeric soluble pMHC molecules, as clustering of TCR and pMHC at high density is sufficient to cause an increase in the kinase/phosphate ratio by crowding out the phosphatases. More recently a monovalent, single-domain antibody fragment that binds to a constant domain of TCRβ was converted from inert in solution to stimulatory by association with an artificial membrane, demonstrating T cell activation on a membrane without TCR clustering.[97]

A significant merit of the kinetic segregation model is providing a unifying mechanism for the TCR and similar ITAM-containing receptors such as the BCR or KIRs which have similar dimensions of ectodomains but lack sequence homology. These receptors could all be present at the same interface and integrate positive and negative signals through localising or not localising to sites of CD45 exclusion. The segregation of CD45 from the receptor by membrane forces has been attributed as a triggering mechanism for the $\gamma\delta$TCR[98] and Fcϵ receptor.[99,100]

The model also makes testable predictions about how an artificial receptor could be created i.e. by creating a receptor with small ectodomains that, through interaction with a ligand on another cell, can confine ITAM cytoplasmic signalling sequences to sites of CD45 exclusion. This has allowed TCR signalling to be reconstituted in a non-immune HEK cell expressing CD45, ZAP70, Csk and Lck through interactions with Raji B-cells.[101] The TCR in this system was mimicked by a construct expressed in the HEK cells that featured the FKBP domain fused to the extracellular domains of CD86 and the cytoplasmic domain of CD3ζ fused to the C-terminus. The pMHC molecule was mimicked by a protein tagged with the domain FRB with the transmembrane sequence of CLEC2. Dimerisation of FKBP and FRB induced by rapamycin was sufficient for the artificial ITAM-containing receptor and ZAP70 to be localised at small sites of contact between the two cells where CD45 was found to be excluded. Artificial receptors of similar architecture are used in the Chimeric antigen receptor (CAR) T therapy which currently sees extensive clinical use targeted against CD19 B cells in acute lymphoblastic leukaemia.[102] These receptors feature an extracellular ligand binding domain, typically a single-chain variable antibody fragment, a hinge and transmembrane region based upon transmembrane immune proteins such as CD28 and cytoplasmic signalling sequences from CD3ζ and costimulatory receptors such as CD28.[103]

1.4.3 TCR internalisation and the signalling role of the immunological synapse

In seconds to minutes after TCR triggering, T cells are observed to form a specialised cell-cell contact with the antigen-presenting cell called an immunological synapse.[104] The classical "bulls-eye" structure of the immunological synapse consists of the TCR and costimulatory receptors localised to the central cSMAC, adhesion molecules such as LFA-1 in the surrounding pSMAC and negative regulators in the distal dSMAC. This structure is observed in the interactions of activated CD4+ T cells with B cells and activated CD8+ CTL with their targets. It is not observed however in the interactions between naive T cells and DCs where the cell-cell contact appears to have multiple foci.[105] Formation of the immunological synapse involves changes in the T cell's cytoskeleton including actin polymerisation at the contact site and the repositioning of the centrosome (microtubule organising centre) to underneath the cSMAC.[106] The most widely accepted function of the immunological synapse is the directed secretion of vesicles, containing cytokines or death effectors for CTLs, towards the antigen presenting cells.

The immunological synapse is also believed to have functional signalling consequences. Whilst the TCR and components of the TCR signalling pathway such as ZAP70, Lck[107] and PKCθ[108] are observed to localise together to the cSMAC, the synapse is only fully formed after the initial TCR triggering has abated and TCRs at the cSMAC show decreased tyrosine phosphorylation. This implies the role of the immunological synapse in the termination of TCR signalling primarily via TCR internalisation.[109] In resting T cells, the complex of TCR$\alpha\beta$CD3$\gamma\delta\epsilon_2$ is continuously internalised and recycled via clathrin-dependent endocytosis through an interaction between a diLeucine motif in CD3γ and clathrin adapter AP-2.[110] This motif is normally blocked by the ζ chain.[111,112] TCRζ internalisation in resting cells appears to occur through clathrin-independent endocytosis. Both these processes are thought to contribute

to polarised endocytic recycling of the TCR complex back to the cell surface.

When engaged by pMHC, internalisation of the TCR via the clathrin-dependent pathway is increased and the TCR can also be internalised via a clathrin-independent mechanism involving the GTPases TC21/Rras2 and RhoG.[113] This pathway is not associated with recycling but instead TCR degradation in lysosomes.[114] The cSMAC appears devoid of clathrin when observed by electron microscopy so this pathway could be the dominant mechanism of TCR signal termination at the immunological synapse.

1.5 B cells and Innate-like lymphocytes

Whilst CD4 and CD8+ $\alpha\beta$T cells are central to the human adaptive immune response, other types of lymphocyte also play an important role in infection. Some of these have a level of innate-like behaviour characterised by low receptor diversity suggesting they respond to conserved antigens.

1.5.1 B lymphocytes

B cell development via recombination and testing of the BCR

The activation of the antibody-producing B lymphocytes has important parallels to that of T cells. B cells are produced from the same lymphoid lineage as T cells but complete their development in the bone marrow. Each B cell expresses a receptor called the B cell receptor (BCR) which is composed of two identical copies of a heavy chain (IgH) and two copies of a light chain (Igκ or Igλ). The complex is linked by disulphide bonds between the two heavy chains and between the heavy and light chains. The regions of these proteins where ligand binding occurs are variable due to VDJ gene recombination. This process is discussed in relation to the TCRα and TCRβ components of the TCR in section 1.6.2. The BCR is attached to the membrane via a transmembrane segment encoded by the final exon of the gene for IgH but alternative splicing produces an otherwise identical soluble version called an antibody. Similarly to the TCR CD3 chains with their cytoplasmic ITAM sequences, BCR signalling occurs through the cytoplasmic ITAM sequences of a heterodimer of Igα and Igβ chains known as CD79.[115]

In the course of B cell development, the IgH gene is recombined and transcribed using the exon for the μ heavy chain constant domain. If a functional heavy chain results from this rearrangement then the heavy chain is expressed on the cell surface in a complex called the pre-BCR with two surrogate light chains: a protein resembling the variable domain of the Ig light chain called VpreB and a protein resembling the constant domain of the Ig light chain called λ5. The pre-BCR is transiently expressed on the cell surface before being rapidly degraded. Signalling through the pre-BCR is thought to occur in a ligand-independent manner and results in proliferation, the recombination of the other IgH allele being stopped and the recombination of the light chains initiated.[116] If a functional light chain is produced then recombination of the other light chain gene is ceased and the full BCR is expressed at the cell surface (figure 1.6). Crosslinking of the BCR with a ligand at this stage indicates autoreactivity so to maintain self-tolerance the cell either dies by apoptosis, rearranges its light chain genes again or becomes anergic.

Surviving immature B cells leave the bone marrow and migrate to the spleen. These are called transitional B cells and are still susceptible to apoptosis or anergy upon antigen binding (negative selection) until they mature in the red pulp of the spleen.[117,118] These naive mature B cells have access to the heavy

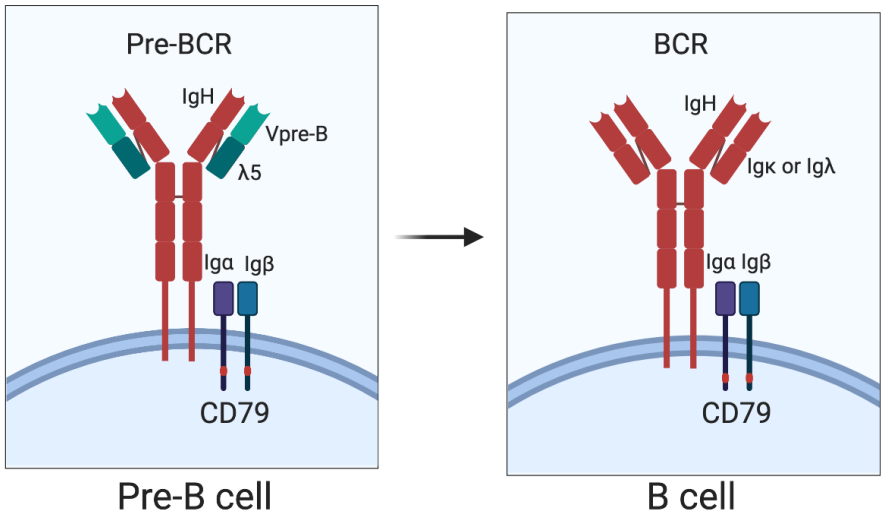

Figure 1.6: The pre-BCR enables B cell development.
During B cell development in the bone marrow the heavy chain of the BCR is recombined and expressed at the cell surface with a surrogate light chain consisting of Vpre-B and $\lambda 5$. Ligand-independent signalling through this pre-BCR triggers recombination of the Igκ and Igλ light chains. If they can be produced then the BCR is expressed at the cell surface which signals through CD79, a dimer of Igα and Igβ.

chain exon segments needed to produce IgM or IgD through alternative splicing. A fraction of these mature B cells, particularly those with BCRs with semi-invariant or poorly diversified VDJ segments, occupy the marginal zone (MZ) of the spleen whilst other migrate to the follicles of secondary lymphoid organs.

T cell-independent activation of B cells

The marginal zone of the spleen continuously samples the blood through a mesh network containing macrophages, DCs, granulocytes and MZ B cells.[119] MZ B cells with poorly diversified BCR genes typically recognise long, repeated antigens common to many pathogens such as bacterial LPS. Clustering of the BCR via these antigens is one of the two triggers for activation, the other being costimulatory signals via PRRs. This T-cell independent activation produces an immediate response of low affinity IgM antibodies but responses are typically short lived.

T cell-dependent activation of B cells

B cells in lymphatic follicles have a continuous influx of potential antigens via lymph. If a BCR binds to an antigen then the antigen can be internalised, digested and presented to CD4+ T cells on MHC-II molecules. If the T cell has been activated before and recognises the peptide via its TCR then the T cell can activate the B cell. Activation induces B cell differentiation down two different pathways: firstly the production of short-lived permablast cells that produce low affinity IgM antibodies giving an initial response to the pathogen and secondly the production of specialised structure called a germinal centre. Within this germinal centre B cells undergo multiple rounds of proliferation tied to a process called somatic hypermutation that randomises the BCR genes. Guided by follicular helper T cells, B cells with higher affinity BCRs persist throughout this process whilst cells with lower affinity BCRs are lost. This process produces long-lived plasma cells that are able to produce large amounts of high-affinity antibodies and also memory B cells that circulate and can rapidly reactivate if the same pathogen is encountered

again.[120] The germinal centres are also commonly the site of class switching whereby cytokine signals from T cells direct the swapping of the IgH chain constant exon with alternative downstream exons resulting in antibodies with different properties and adapted for different tissues.

1.5.2 $\gamma\delta$T Cells

Though much more significant in some animals such as cattle and pigs, in humans $\gamma\delta$T Cells are an unconventional and rare subset of lymphocytes in circulation with important roles in the defence of epithelial and mucosal tissues. Unlike all T cells mentioned before which have TCRs composed of α and β chains, these cells have a different TCR composed of γ and δ chains. Though the extracellular domains of the $\gamma\delta$TCR are similar in structure to the $\alpha\beta$TCR and signalling occurs through the same CD3 chains, $\gamma\delta$T cells do not typically express the CD4 or CD8 coreceptors and do not appear to recognise classical pMHC molecules. Similar to $\alpha\beta$TCRs, $\gamma\delta$TCRs are generated through recombination of genetic elements during T cell development but whilst $\alpha\beta$TCRs are very diverse in sequence across both chains, for $\gamma\delta$TCRs diversity is concentrated into a small loop section of TCRδ. In practice, many observed $\gamma\delta$TCRs are invariant or near invariant. This invariance implies that these $\gamma\delta$T cells recognise conserved ligands more typical of the innate immune system. Another unusual feature of $\gamma\delta$T cells is distinct localisation depending on the sequence components of the $\gamma\delta$TCR chains.[121]

In humans, the two more apparent and most studied $\gamma\delta$T cell subsets are identified by the variable sequences of the TCRδ chain.[122] Vδ1 T cells use a δ-chain with a Vδ1 variable segment paired with a range of TCRγ chains and are predominant in the thymus, intestinal epithelium and spleen. This subset has been reported to recognise non-classical MHC-I molecules including MICA and MICB, which are upregulated upon cellular stress such as viral infection[123] and CD1d, a molecule able to present a variety of bacterial lipid antigens from APCs.[124] The most common human $\gamma\delta$T cells in blood circulation all have a near invariant Vγ9Vδ2 TCR. These have been reported to recognise phosphorylated antigens such as HMB-PP, an intermediate in isoprenoid synthesis in many pathogenic bacteria, and intermediates produced by stressed mammalian cancer cells[125] presented by a non-classical MHC molecule Butyrophilin 3A1.[126] These two subsets do not cover the full range of $\gamma\delta$T cells and a wide variety of ligands have been suggested. The functionality of these ligands *in vivo* is disputed.[125,127]

$\gamma\delta$T cells express a number of another receptors in addition to their TCRs including CD27, TLRs and cytokine receptors. When activated, $\gamma\delta$T cell can kill infected or cancerous cells through perforin and granzymes similar to CTLs. Under different conditions they have been linked to anti-inflammatory and wound healing responses.[128]

1.5.3 NK cells

Natural Killer cells arise from common lymphoid progenitor cells like T cells and B cells but do not express TCR or BCR recombined immune receptors. Like CTLs, NK cells are able to detect and kill cancerous and virus-infected cells through the secretion of perforin that creates holes in cell membranes and granzymes that induce apoptosis.[129] NK cell activation is controlled by the balance of both activating and inhibitory signals the cell receives through surface receptors, particularly C-type lectin-like family receptors (CTLRs) and Killer-cell immunoglobulin-like receptor (KIRs).[130] Cell killing through a rise in activatory signalling represents an "infectious non-self" model. Meanwhile cell killing through a loss of inhibitory signals represents of a "missing self" model whereby infection or carcinogenesis removes markers of healthy cells such as MHC-I molecules.

1.5.4 Innate-like $\alpha\beta$T cells

Whilst most $\alpha\beta$T cells have extremely diverse TCR sequences, some populations are observed with much more restricted sequences and distinct functionality, mostly notably responding to nonpeptide antigens.[131] Like conventional $\alpha\beta$T cells, these subsets develop in the thymus but they undergo clonal expansion and acquire effector functions before thymic egress rather than in the periphery.[132,133] These include intrinsic Natural Killer T cells (iNKTs), germline-encoded mycolyl lipid–reactive (GEM) cells and mucosal associated invariant T cells (MAIT cells).

1.6 Mechanisms of T cell development and cell fate decisions

1.6.1 Commitment to T cell lineage

T cell development is intrinsically tied to the recombination of T cell receptor genes and the subsequent selection of the receptor.[134] The development process begins with cells from the common lymphoid lineage in the bone marrow but continues in the thymus, a thoracic organ located behind the sternum and in front of the heart. The thymus produces most of its T cell repertoire in early life, reaches its maximum size around puberty and subsequently declines in size and output throughout adulthood. Within the thymus, early T cell development begins in the dense outer thymus cortex and later development takes place in the thymic medulla. These regions are separated by the corticomedullary junction that is supplied with blood from the thyroid arteries.[135]

Cells of the thymus can be divided into cells of the haematopoietic lineage, primarily developing T cells called thymocytes and resident stromal cells, most crucially cortical and medullary thymic epithelial cells (cTECs/mTECs). Surface protein ligands and chemokines expressed by cTECs and mTECs guide thymocytes throughout the developmental process.[136] Cells in the thymus can be distinguished and subdivided in flow cytometry using antibodies to cell surface markers.[137] For thymocyte development the most major division is between the earliest thymocytes expressing neither CD4 or CD8 coreceptors called double negative (DN) cells, cells expressing both markers i.e. double positive (DP) cells and cells with only one of these markers called CD4 or CD8 single positive (SP) cells. DN cells, after careful removal of B lymphocytes and NK cells, can be subdivided into four groups (DN1-4) by surface expression of the Interleukin-2 receptor alpha chain CD25, adhesive glycoprotein CD44 and the receptor tyrosine kinase for stem cell factor CD117. DN2 cells can be further divided into DN2a and DN2b cells before and after final commitment to the T cell lineage. The DN3 subset can be divided into cells before and after successful rearrangement of TCRβ or TCRγ genes (table 1.1).

Surface Marker	DN1	DN2a	DN2b	DN3a	DN3b	DN4	DP	CD4+ SP	CD8+ SP
CD4	-	-	-	-	-	-	+	+	-
CD8	-	-	-	-	-	-	+	-	+
CD117	++	+	-	-	-	-	-	-	-
CD25	-	+	+	+	+	-	-	-	-
CD44	+	+	+	+	-	-	-	-	-
FSC	Low	Low	Low	Low	High	High	High	High	High
i.c. TCRβ	-	-	-	-	+	+	+	+	+
surface TCRβ	-	-	-	-	low	low	+	+	+
TCRα	-	-	-	-	-	-	+	+	+

Table 1.1: Flow cytometry markers for distinguishing thymocyte populations.
Forward scatter (FSC) is a measure of cell size which increases in the DN3a to DN3b transition during β-selection. (i.c. intracellular)

Thymic progenitor cells arrive at the thymus from the bone marrow through blood vessels in the corticomedullary junction. After entering the thymus, early thymic progenitors (ETPs) undergo cycles of proliferation but still retain the potential to differentiate into B cells, DCs, NK cells and macrophages.[138] Commitment to the T cell lineage and the exclusion of the other cell fates depends upon signalling through the Notch-family signalling pathway. Notch initially activates transcription factors T cell factor 1 (TCF1) and GATA-binding protein 3 (GATA3).[139] Once TCF1 is expressed it becomes self-sustaining and activates expression of T cell genes encoding LAT, Lck and CD3γ. GATA3 is involved in regulating cell proliferation and prevents DN2 cells adopting the B cell fate.[140]

TCF1, GATA3 and Notch all contribute to induce expression of B cell lymphoma–leukaemia 11B (Bcl11b), a key transcription factor for T cell commitment at the DN2b stage.[141,142] Bcl11b is a Zn-finger transcription factor able to operate both as an activator via association with chromatin remodelling, HDAC-containing complex NuRD and a repressor through interaction with histone acetyltransferase p300. Once it is expressed in the DN2b stage, expression is maintained through the T cell lineage and it has important roles at many stages of development. Its activity is controlled through post-transcriptional regulation particularly via phosphorylation and sumolation states.[143] In the DN2b stage its primary role is to prevent thymocytes adopting the DC or NK cell fates.[144]

Another key group of transcription factors involved in T cell identity is E-proteins including E2A and HEB.[145] If E2A is knocked out then T lymphocytes are severely reduced in number, though some persist likely due to a level of redundancy with HEB.[146]

1.6.2 Generation of novel TCR sequences by VDJ recombination

Commitment to the T cell lineage in the DN2b stage of development begins the process of VDJ recombination of TCR genes. The genomic locus encoding TCRα and TCRγ, as well as the loci for the Ig light chains, are split into numerous variable (V) and join (J) gene segments and a larger constant segment. Each V and J segment is flanked by recombination signal sequences (RSSs). The genes for TCRβ, TCRδ and the Ig heavy chain have additional diversity (D) gene segments with similar flanking sequences and duplication of the constant region. These gene segments were created by gene duplication and divergence creating a high level of redundancy. The number of each of these gene segments in the human genome is given in table 1.2. Of note, the gene segments encoding TCRδ are located entirely within the gene segments for TCRα and a few variable segments usually assigned to TCRα can also be

used in a TCRδ chain.

Gene	Locus	Variable	Diversity	Join	Constant
TRA	14q11.2	54 total, 43-45 functional	0	61 total, 50 functional	1
TRB	7q34	64-67 total, 40-48 functional	2	14 total, 12-13 functional	2
TRG	7p14	12-15 total, 4-6 functional	0	5	2
TRD	14q11.2	3 plus 5 shared with TCRα	3	4	1

Table 1.2: Number of V, D, J and C gene segments for TCR genes per haploid genome.
Numbers from ImMunoGeneTics (IMGT) information system.[147]

VDJ recombination is a site-specific recombination process that selects one of each of the types of gene sections making up a gene and joins them together in an error prone repair process to create genes with novel sequences. Mechanistically this involves creating a double strand break at the RSSs of two chosen segments using a lymphocyte-specific recombinase enzyme called recombination-activating genes (RAG). RAG is composed of RAG1 and RAG2 proteins, both upregulated by E2A proteins.[148] RAG is thought to cleave both DNA sequences simultaneously at a site called a recombination centre leaving the two coding ends terminated by a hairpin loop. Intervening DNA is ligated into a circle that is lost from the genome. The hairpin loops are resolved by cutting both hairpins and ligating the ends together using ubiquitously expressed DNA repair enzymes of the non-homologous end joining pathway. The random loss or gain of nucleotides at these junctions creates huge diversity in addition to the significant combinatorial diversity of selecting one V, D and J segment from the many encoded in the germline.

Due to the constant segment of each gene, diversity is not spread through each TCR chain but concentrated in the variable domain particularly in three complementarity determining regions (CDRs) that contact the MHC molecules (figure 1.7). Varied sequences in the loops confer unique peptide specificities onto TCRs from a limited number of genetic components. A downside to this largely random joining process is that two thirds of joins are expected to result in the gene segments being out of frame. Cells in which these rearrangements occur have to undergo new rearrangements with remaining gene segments or do not develop further. Amongst the in-frame sequences, many would be expected to not fold correctly or bind to the individual's selection of MHC molecules and thereby lack functionality. Worse many are likely to bind to MHC molecules presenting self-peptides. Mechanisms to eliminate cells expressing these TCR sequences are described in the subsequent sections. The theoretic diversity from VDJ recombination is thought to be as high as 10^{18} different sequences. In practice, due to gene segment preferences the actual repertoire of the thymocytes in a thymus is estimated at 40-70x10^6 unique TCRβ sequences and 60-100x10^6 TCRα sequences.[149] Sequence diversity is lost due to selection processes in the thymus that cause the death of over 95% of starting thymocytes. The means the potential TCR diversity is further limited to roughly 2x10^7 due to the limited number of T cells present in any one individual at a time.[150] Many of the T cells present will be of the same TCR clone.

1.6.3 The β-selection checkpoint and mechanisms of $\alpha\beta$ or $\gamma\delta$-lineage choice

The recombination of TCRβ, TCRγ and TCRδ genes is initiated in the DN2b developmental stage within the thymic cortex and completed during DN3. This point marks a critical stage for lineage commitment as the $\alpha\beta$T cell and $\gamma\delta$T cell lineages diverge. Whilst E2A and HEB are necessary for T lineage commitment in the DN2 stage, in the DN3 stage they enforce a developmental checkpoint and cell

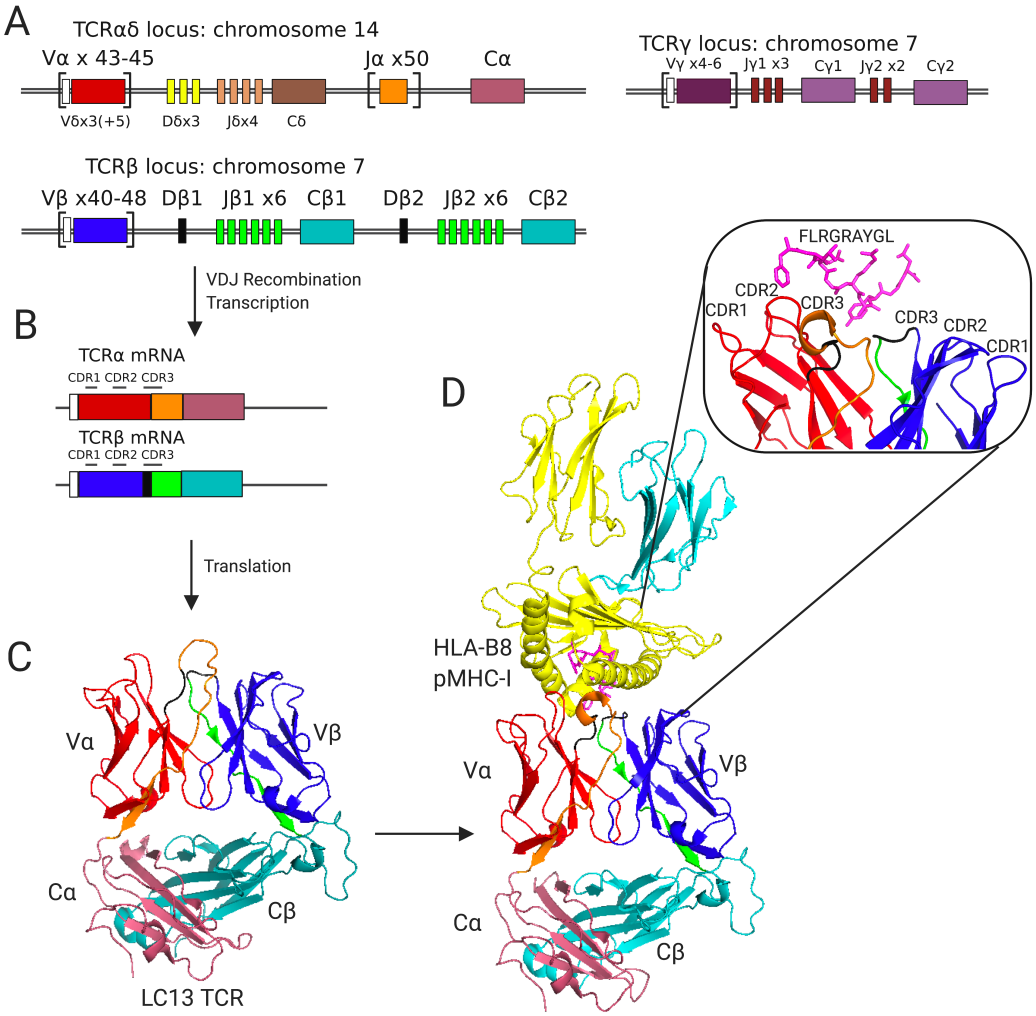

Figure 1.7: Diversity in TCR sequence is created by VDJ-recombination which is then reduced by positive and negative selection in the thymus.
A: Recombination is initiated by the RAG recombinase which creates double strand breaks around variable (V), diversity (D) and join (J) gene segments of TCR genes. Resolution of these breaks creates diversity in TCR sequence due to the combinatorial selection of one segment from the many genetically encoded and the random loss or gain of nucleotides in the DNA repair process. Though the genes are not shown to scale, the organisation of the gene segments is accurate with the genes for TCRδ located within the same locus as TCRα.
B: Transcription of the genes for TCRα and TCRβ involves the splicing in of a constant exon (C) for each chain. For TCRβ both chains are functionally identical. Sequence diversity is greatest in three Complementarity-Determining Regions (CDRs) in the variable domain. The leader peptide (white) is removed in protein synthesis.
C: Translation of TCRα and TCRβ mRNA produces a TCR receptor. The LC13 TCR is a public TCR common to most individuals with the HLA-B*08 allele and has been crystallised with and without ligand. As this TCR always uses the same gene segments, amino acids can easily be coloured based upon their origin. Within TCRα sequences arising from TRAV26-2*01 are coloured red, amino acids from N nucleotides in black, amino acids from TRAJ52*01 segment in orange and the sequence encoded by the constant domain in pink. In TCRβ protein sequence from the TRBV7-8*03 variable segment is coloured blue, from TRBD1*01 in black, TRBJ2-7*01 in green and Cβ2 domain in cyan. From structure 1KGC.[151]
D: The LC13 TCR recognises the peptide FLRGRAYGL (magenta) from the latent antigen EBNA 3A presented by the MHC-I molecule HLA-B*08 (yellow) with β_2-microglobin in cyan. The insert shows rotated cutaway of the binding interface with the MHC removed. A common diagonal docking mode of TCR binding is seen with TCRα near the N-terminus of the peptide and TCRβ near the C-terminus. The CDR1 and CDR2 loops of each chain contact the helices of the MHC groove whilst the longer and more diverse CDR3 loops contacts the presented peptide. From structure 1MI5.[64]

cycle arrest.[152,153] The exact mechanism of this arrest is still unclear but E protein mediated expression of cyclin dependent kinase inhibitors[154] and inhibition of IL-7 signalling are important.[146]

If recombination produces a productive, in-frame and well-folded TCRβ chain then it can be transiently expressed at the cell surface as a receptor called the pre-TCR, in which it is paired with an invariant surrogate TCRα chain called pre-Tα (pTα). Expression of the *ptrca* gene encoding pTα, is upregulated by E2A proteins in conjunction with Notch signalling.[148] Pre-TCR signalling induces cell survival, a burst of proliferation and upregulation of CD4 and CD8 to progress to the DP stage. This process of selecting a functional TCRβ chain, called β-selection, commits the thymocyte to the $\alpha\beta$T cell lineage. The structure and signalling of the pre-TCR will be discussed in greater depth in section 3.1.1.

Pre-TCR signalling also triggers the allelic exclusion of the TCRβ locus so that only one TCRβ sequence can be expressed. Allelic exclusion is enabled by the two step recombination of the TRB locus. First there is the joining of a D_β segment to a J_β with the loss of intervening DNA. This takes place quickly on both TRB alleles. A second, more difficult recombination event then joins a V_β segment onto $D_\beta J_\beta$. The TCRβ variable segments are located several hundred kb away from the cluster of D, J and constant segments and are in accessible chromatin in the DN stage. The successful recombination of TCRβ on one allele produces the pre-TCR which downregulates RAG1 and RAG2 preventing the V_β to $D_\beta J_\beta$ recombination on the second allele.[155] The TCRβ variable segments are compacted into less accessible chromatin before RAG expression is increased for TCRα recombination in the DP stage. The V_β-$D_\beta J_\beta$ recombination, but not D_β-J_β recombination, is blocked in Bcl11b -/- thymocytes which do not progress to the DP stage suggesting that Bcl11b is involved in V_β segment accessibility.[156]

If pre-TCR expression is preceded by successful recombination of TCRγ and TCRδ then both chains can be expressed on the cell surface as the $\gamma\delta$TCR. Signalling from this receptor also saves the thymocyte from programmed cell death but prevents further recombination of TCRα and TCRβ genes and guides the thymocyte down the $\gamma\delta$ T cell lineage. Notably there is no evidence for the involvement of a preantigen receptor in $\gamma\delta$ T cell development or functional pairing of pTα with TCRγ.[157] Like β-selection, $\gamma\delta$-selection causes an increase in proliferation but cells do not upregulate CD4 or CD8 to progress to the DP stage.[158]

How expression of the pre-TCR or $\gamma\delta$TCR relates to commitment to the $\alpha\beta$ or $\gamma\delta$ lineage is not immediately obvious.[159] Both receptors signal through phosphorylation of the ITAM sequences on the invariant CD3 chains and both seem to use the same cytoplasmic signalling components. The pre-TCR, being only half of a functional receptor, is not commonly thought to have a ligand so presumably signals in a ligand-independent manner. Meanwhile probable ligands for the $\gamma\delta$TCR (section 1.5.2) are not consistently present in the thymus though some may play a role in selection.[160]

Initially two models sought to explain how the two different receptors established different cell fates.[161] The "stochastic" or "pre-commitment model" predicted that the lineage fate was specified prior to TCR gene rearrangement and expression of the pre-TCR or $\gamma\delta$TCR only served to rescue cells from cell death if the recombined receptor matched the correct lineage. The "instructional model" instead proposed that different signals from the pre-TCR or $\gamma\delta$TCR direct uncommitted precursors to adopt either the $\alpha\beta$ or $\gamma\delta$ fate.

Evidence in support of the stochastic model focuses on the heterogeneity of different markers in subsets of DN1 thymocytes, prior to recombination, that seem to bias development towards different fates.[162] One study found there was strong lineage bias in DN2 cells based upon expression of Interleukin-7 receptor, an important Notch target, with IL-7R+ cells five times more likely to adopt the $\gamma\delta$T cell fate than IL-7R- or IL-7R low cells.[163] IL-7 signalling was linked to accessibility of the TRG locus for RAG recombinase.[164]

Similarly DN2 thymocytes expressing transcription factor Sox19 were found to be predisposed to the $\gamma\delta$ T cell fate and that Sox19 was required for development of $\gamma\delta$T cells but not $\alpha\beta$T cells.[165] More recently the same group linked Sox19 in a subset of DN1 cells to the adoption of a transcriptome consistent with a $\gamma\delta$T cell subset that produces IL-17.[166] These changes were adopted prior to recombination of the $\gamma\delta$TCR though TCR expression was necessary for the progenitors to mature. This implies a level of pre-commitment may be important in the development of $\gamma\delta$T cell subsets independent of TCR gene arrangement.[167,168]

Whilst a degree of lineage bias seems to be established by heterogenous Sox19 expression in early thymic T cell precursors,[169] the central claim of the stochastic model that lineage is established before TCR signalling is inconsistent with experiments using the OP-9 cell system to study T cell development *in vitro*. This model mimics T cell development in the thymus by the co-culture of mouse haematopoietic progenitor cells with cells from the OP9 stromal cell line expressing the Notch ligand Delta-like-1.[170] Using this system it was shown that expression of TCRβ to form the pre-TCR or expression of TCRγ and TCRδ to form the $\gamma\delta$TCR in RAG2 -/- DN3 cells efficiently generated cells of the correct lineage without the widespread cell death expected if lineage was committed prior to this stage.[171] Similarly tracking the progeny of individual pTα -/- DN3 cells expressing the $\gamma\delta$TCR revealed that cell fate was not established prior to TCR expression.[172] Whilst this late lineage commitment is not consistent with the stochastic model, neither is it consistent with a simple "instructional model" as expression of both the $\gamma\delta$TCR and pre-TCR can produce cells of the incorrect lineage.[173] This is further complicated by the observation that premature expression of an $\alpha\beta$TCR in transgenic mice suppressed recombination of TCRγ and TCRδ genes but supported development of CD4- CD8- T cells consistent with the $\gamma\delta$ lineage.[174]

Experiments using $\gamma\delta$TCR transgenic mice found that both $\gamma\delta$ lineage DN cells and small amounts of $\alpha\beta$ lineage DP thymocytes could be generated. As DP cells had consistently lower surface expression of the $\gamma\delta$TCR, mice with a CD3ζ +/- genotype, known to lower TCR surface expression were studied. These produced more DP cells and fewer cells of the $\gamma\delta$ lineage. Expression of a $\gamma\delta$TCR in mice with a full length CD3ζ transgene and a tailless CD3ζ transgene found that the proportion of T cells in the $\gamma\delta$ lineage was reduced in the absence of the CD3ζ ITAM domains whilst surface expression was similar.[175] These results supported a refined "TCR signal strength" model which proposes that stronger TCR signalling from the $\gamma\delta$TCR promotes uncommitted DN2 cells to adopt the $\gamma\delta$ T cell fate whilst weaker signalling from the pre-TCR promotes the $\alpha\beta$ T cell fate. This model is also supported by experiments in transgenic mice that found fewer $\gamma\delta$ lineage cells if signalling from the $\gamma\delta$TCR was reduced via Lck deficiency. Accordingly there was an increase in $\gamma\delta$ lineage if apparent signal strength was increased through ectopic expression of downstream transcription factor Egr.[176] Supporting this model whereby it is the accumulated TCR signalling at the DN stage that is important for lineage choice, the co-expression of the pre-TCR was found to rescue commitment to the $\gamma\delta$ lineage in $\gamma\delta$TCR-expressing DN cells in which TCR signalling strength had been artificially weakened.[177]

More recently, single-cell RNA sequencing of thymic progenitors has enabled β-selection to be studied in greater detail. A study by Oh *et al.*[178] found a high level transcriptional diversity in DN1 cells that is obscured by the lack of obvious lineage markers. However, the authors poorly justify their conclusion that this demonstrates a level of predetermination to the $\alpha\beta$ or $\gamma\delta$-lineage fates. More convincingly Scaramuzzino *et al.*[179] used single cell RNA sequencing to define the transcriptional profile of developing $\alpha\beta$ or $\gamma\delta$-lineage thymocytes and observe a relatively homogenous DN3 population. They repeated this analysis in mice deficient in LAT, a required adaptor for strong TCR signalling. These mice have a developmental block at the DN3 stage with a small proportion of trapped DN3 cells expressing the

$\gamma\delta$TCR. Single cell transcriptomics of this $\gamma\delta$-TCR^{+ve} population found no other signs of commitment to the $\gamma\delta$-lineage. This implies TCR signalling through the LAT-complex is required for the specification of uncommitted progenitors and that receptor expression alone is insufficient.

Whilst the signal strength model is now widely regarded as the best model for $\alpha\beta$ vs $\gamma\delta$ lineage commitment, how TCR signal strength should be defined is unclear.[180] An important sensor of TCR signal strength is thought to be the activation of the MAPK/ERK pathway which activates Early growth response protein transcription factors (EGR1-3). These transcription factors induce expression of Id3 which is critical in inhibiting the E2A-mediated block of proliferation.[148,181] Expression of Id3 was found to be necessary and sufficient for cells to adopt the $\gamma\delta$ fate independent of Notch ligands using the OP9 system.[182]

Although the $\gamma\delta$TCR appears capable of inducing sufficient Id3 to overcome the block on proliferation by itself to allow adoption of the $\gamma\delta$T cell fate, the pre-TCR seems to require input from other signalling pathways for the adoption of the $\alpha\beta$T cell fate. These pathways, including Notch signalling, Wnt signalling,[183] Hedgehog signalling[184,185] and signalling through the CXCR4 chemokine receptor,[186,187] likely support thymocyte survival and the proliferative burst accompanying β-selection. Recent imaging supports the idea that the signals are coordinated by an interface between thymocytes and thymic epithelial cells analogous to an immunological synapse of mature T cells.[188]

The role of Notch signalling in the $\alpha\beta$-$\gamma\delta$ lineage choice is difficult to determine due to the prior requirement for Notch signalling in the establishment of T cell identity and TCR gene recombination. In mice, Notch signalling is believed to be critical for the development of cells of the $\alpha\beta$ lineage whilst $\gamma\delta$-T cells can develop in the absence of Notch ligands.[171] In humans however, development of $\alpha\beta$ T cells seems to be favoured by low Notch signalling whilst $\gamma\delta$T cell development was found to be entirely dependent on Notch signals.[189] These differences might be related to the different Notch ligands expressed by human and mice thymic epithelial cells.[190] Whilst the relationship between Notch signalling and lineage choice is unclear, both Notch and pre-TCR/$\gamma\delta$TCR signals are required for the burst of thymocyte proliferation that accompanies β-selection and differentiation.[191,192]

1.6.4 $\alpha\beta$-selection

Thymocytes that pass the β-selection checkpoint proceed to upregulate the CD4 and CD8 coreceptors, progressing to the DP stage, and downregulate receptors characteristic of haemopoetic stem cells such as cytokine receptor IL-7Rα.[181] The burst of proliferation accompanying the selection of a TCRβ chain expands the pool of thymocytes with a productive TRB gene rearrangements. This also helps cells commit to the $\alpha\beta$-lineage by diluting out earlier regulatory molecules and resetting epigenetic marks. This burst of proliferation is however temporary and DP cells do not go through any more stages of intrathymic division.

Transition to the DP stage causes RAG1/2 to be re-expressed and recombination initiated again at the TRA/TRD locus. During this stage, gene segment accessibility is controlled by a more distant Eα enhancer element and typically each allele shows multiple rounds of Vα to Jα recombination until a productive TCRα chain is produced. All these recombinations result in the deletion of the entire TCRδ locus located in the intervening DNA. Expression of a functional TCRα chain seems to outcompete pTα for TCRβ leading to mature $\alpha\beta$TCR surface expression and declining pre-TCR expression.[193] Transcription of pTα is later stopped in the SP stage.

Surface TCR, CD4 and CD8 expression allows DP thymocytes to be selected through three developmental

mechanisms: positive selection that rescues from programmed cell death the thymocytes that can interact with self-peptide MHC complexes expressed by the thymic epithelium, the lineage choice between continued CD4 or CD8 expression matching the MHC specificity of the TCR and finally negative selection that eliminates thymocytes that interact too strongly with self-peptide MHC complexes such that they would become self-reactive in the periphery. From the tens to hundreds of haematopoietic precursors entering the thymus every day in neonatal life, the approximately 20 divisions in the DN stage produce of the order of $5x10^7$ DP thymocytes. As only $1\text{-}2x10^6$ mature T cells leave the thymus each day it is implied that over 95% of DP thymocytes fail these selection steps.[194]

Positive selection and coreceptor choice

On a species level TCR variable gene segments have evolved to recognise MHC molecules. However due to sequence diversity introduced by recombination and polymorphic diversity in the HLA repertoire of individuals, the TCR expressed by a DP cell may not interact successfully with any host MHC molecules. DP thymocytes with TCRs that fail to interact with self-peptide MHC molecules expressed by cortical thymic epithelial cells, die through apoptosis within three days of their generation. This is estimated to result in the death of roughly 90% of thymocytes cells reaching the DP stage. Survival is dependent on successful TCR-pMHC interactions which increase expression of anti-apoptotic molecules such as Bcl2 and Mcl1.

Positive selection, like β-selection, occurs in the thymic cortex and involves interactions with cTECs presenting self-peptide MHC molecules.[195] These cTECs express both type I and type II MHC molecules and appear to have a distinct peptide repertoire due to unique properties in their antigen presentation pathways. In the cytosolic/class I antigen presentation pathway cTECs express a catalytic subunit of the proteasome, β5t, forming the thymoproteasome that appears to produce lower affinity peptide ligands than the standard proteasomes or immunoproteasomes expressed by other cells.[196] cTECs also appear to have a high rate of constitutive macroautophagy, enhancing the loading of peptides from self proteins onto pMHC-II molecules via the endogenous route.[197] The unique MHC-I and MHC-II peptide repertoires of cTECs may allow positive selection to occur to host MHC molecules without enhancing affinity to the public self-peptides generated elsewhere in the body.[198,199]

During positive selection, DP thymocytes express both CD4 and CD8 co-receptors though one of these is redundant as TCRs are specific for either class I or class II molecules. During this stage in the cortex, thymocytes with MHC-I specificity silence expression of CD4 but continue CD8 expression whilst thymocytes with MHC-II specificity silence expression of CD8 and maintain CD4 expression. This transition from bipotential DP thymocytes to unipotential SP thymocytes by extension determines the effector function of the cell after leaving the thymus. Early models of the mechanism of this lineage fate decision were similar to those of the $\alpha\beta$ Vs $\gamma\delta$ lineage choice. The stochastic model proposed that co-receptor downregulation was made randomly and only thymocytes making the correct decision survived; instruction models proposed that interaction to MHC-I or MHC-II molecules created different signals which dictated whichever co-receptor was maintained.[200] Strong evidence against the stochastic model and in favour of an instructional model was measurements of the high efficiency of correct coreceptor choice, incompatible with random selection.[201] The most common instructional model of lineage choice proposed that lineage choice was determined by quantitatively higher TCR signalling through interaction with pMHC-II molecules than pMHC-I molecules as the CD4 cytoplasmic tail co-receptor binds more intracellular Lck than the tail of CD8.[202] However, later experiments attenuating TCR signalling via altering the number of ITAM domains were incompatible with this "strength of signal

instructional" model as decreasing TCR signalling strength was found to affect the total number of thymocytes passing positive selection but not the efficiency of lineage choice.[203, 204]

Evidence now favours a "kinetic signalling" model based on the observation that positively selected thymocytes of both lineages go through an intermediate stage of $CD4^{+}CD8^{low}$ in the DP to SP transition. If TCR signalling during this intermediate stage persists then the thymocyte will commit to the CD4+ lineage; if TCR signalling ceases then CD4 expression is terminated and CD8 expression re-established.[205] A set of transcription factors has been shown to be important in mediating and enforcing this lineage choice. The transcription factor ThPOK (T-helper-inducing POZ/Krueppel-like factor) is required to enforce the CD4+ T helper cell fate. In seems to function in part by repressing the transcription factor RunX3 which is required to enforce the CD8 CTL cell fate.[205] Expression of both RunX3 and ThPOK seems partially dependent on Bcl11b which binds to promotor regions of both genes. If Bcl11b is deleted after the β-selection checkpoint then CD4-CD8 lineage choice appears to occur randomly and haphazardly.[206]

Once thymocytes pass the positive selection checkpoint and commit to either of the CD4 or CD8 lineages, cells upregulate CCR7: the chemokine receptor for the ligands CCL19 and CCL21 that are expressed in the thymic medulla.[207] Conversely expression is reduced of CXCR4, the receptor for the ligand CXCL19 expressed in the thymic cortex.[208] These changes trigger the migration of cells from the cortex to the medulla[209] (figure 1.8).

Negative selection

Whilst the process of positive selection ensures that SP thymocytes entering the medulla have functional TCRs capable of recognising host MHC molecules, many of these TCRs will strongly recognise MHC molecules presenting peptides from widely-expressed host proteins leading to autoimmune recognition if these cells left the thymus. Negative selection ensures that if the affinity of TCRs to self-peptide MHC complexes is too high then these thymocytes are eliminated via apoptosis, a process known as clonal deletion. Alternatively, cells can be induced to differentiate into nTreg cells which express the transcription factor Foxp3. This selection begins in the cortex through interaction with cortical DCs but continues in the medulla through interactions with medullary epithelial cells.

In order to select against T cells with reactivity to the multitude of self proteins in the periphery, mTECs have promiscuous gene expression including expression of tissue-restricted antigens (TRAs) that otherwise found only in particular cells.[195] This promiscuous expression is attributed to the transcription factor AIRE thought to stochastically activate silenced genomic loci.[210] Peptides from the TRAs are presented to DP thymocytes in the medulla directly from mTECs and via surrounding residential dendritic cells.[198] Migratory cDC's additionally appear to present self-peptides, gathered from the periphery, to thymocytes in the medulla.

If SP thymocytes strongly recognise TRA peptides via their TCR then these cells upregulate pro-apoptotic molecules and are eliminated via clonal deletion. Strong but more transient TCR signalling or intermediate level signalling seems to instead divert thymocytes to the nTreg lineage that dampens effector T cell responses in the periphery.[211] It is somewhat unclear how TCR signalling enables thymocytes to be dependent on interactions with self-peptide MHC molecules during positive selection in the cortex but susceptible to clonal deletion in the presence of self-peptide MHC molecules during negative selection in the medulla. The difference in signals is thought to be related to the affinity of the TCR for the peptide MHC (with low affinity TCR clones failing positive selection and clones with too high

affinity suffering clonal deletion or diversion) rather than the distinct locations providing qualitatively different signals.[212]

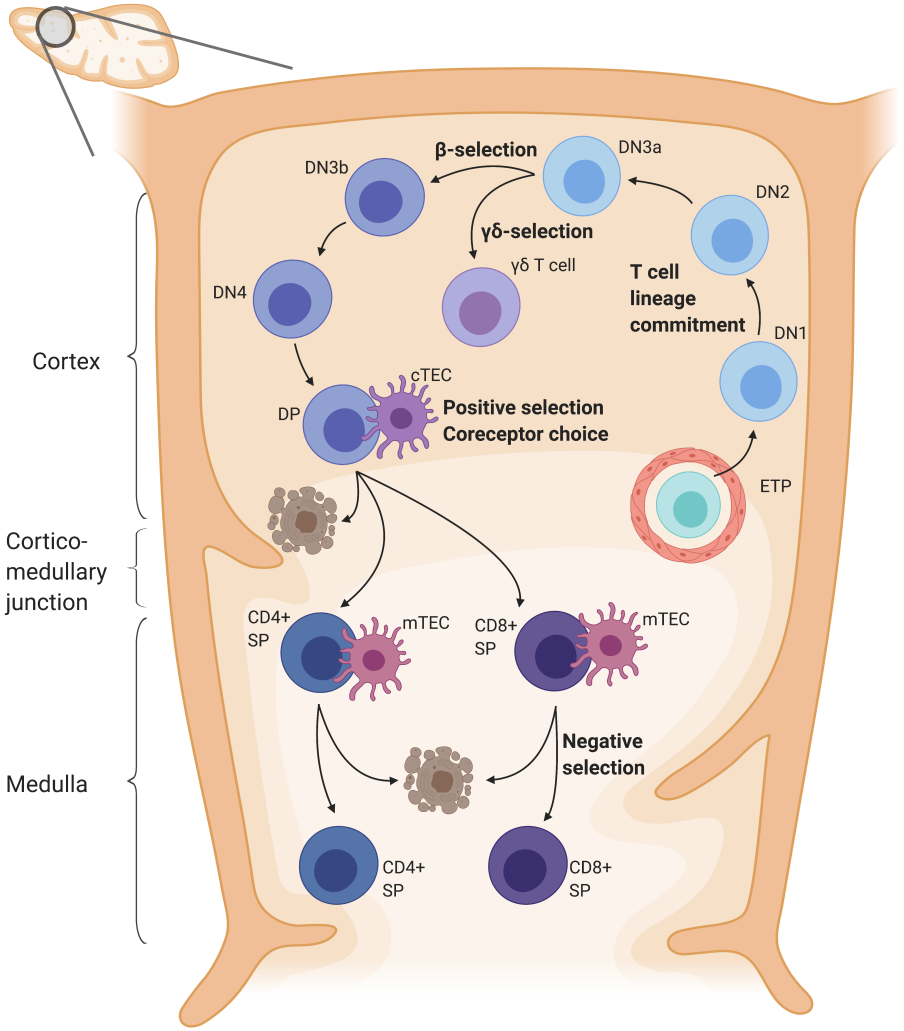

Figure 1.8: Thymocyte development, primarily recombination and selection of the TCR, accompanies cytokine-driven movement through the thymic cortex to the thymic medulla.
Early thymic progenitors arrive via blood vessels in the corticomedullary junction. In the thymic cortex a portion of these cells commit to the T cell lineage and recombine the genes for TCRβ, TCRγ and TCRδ. Surface expression of the $\gamma\delta$TCR causes thymocytes to continue development in the $\gamma\delta$T cell lineage whilst prior surface expression of TCRβ, as the pre-TCR, allows thymocytes to undergo a burst of proliferation and progress to the DP stage. In the DP stage, TCRα is recombined. If an $\alpha\beta$TCR is expressed at the cell surface, positive selection and co-receptor choice occur through pMHC interactions with cortical thymic epithelial cells. Chemokine signals then cause thymocytes to migrate to the medulla where negative selection occurs through interactions with medullary epithelial cells.

Developing SP thymocytes spent about 4-5 days in the medulla during which negative selection occurs.[213] Thymic egress is triggered by upregulation of S1PR1, a receptor for the lipid sphingosine-1-phosphate (S1P). A gradient of S1P that guides thymocytes to the blood vessels of the corticomedullary junction is established by its production by pericytes surrounding the blood vessels and degradation by thymic dendritic cells expressing S1P lyase.[214]

1.7 Aims of this thesis

Previous work within the James lab investigated the pre-TCR complex with the goal of explaining how signalling from the pre-TCR and $\gamma\delta$-TCR are distinguished at the β-selection checkpoint. This involved reconstituting receptor expression in non-immune HEK293T cells, allowing the localisation and trafficking of the receptor to be studied in greater detail than could be achieved in primary cells. The pre-TCR was found to be expressed at the cell surface but rapidly internalised unlike the $\alpha\beta$TCR which was surface stable. This difference in surface expression could not be explained by the cytoplasmic sequences of the pre-TCR complex. Instead surface expression was found to be dependent on the extracellular structure of the receptor through an unknown mechanism. This thesis aims to continues this work by investigating if an unknown binding partner, contacting the extracellular domains, is responsible for pre-TCR internalisation or if pre-TCR dimerisation might play a role. These experiments are described in chapter 3.

Previous work also established an interaction between the pre-TCR and a poorly characterised membrane protein called Transmembrane protein 131 (TMEM131). Although the molecular function of this protein is unknown, there is circumstantial evidence for a role in T cell development. Particular features about this protein, combined with its high level of conservation, justified further study. This project aims to characterise the cell biology of TMEM131 with the goal of understanding its localisation, degradation and interaction partners. These experiments are described in chapter 4.

A mouse knockout of *Tmem131* was commissioned with the expectation of an immune dysfunction phenotype that could be studied. Instead the knockout was found to be homozygous lethal, much more severe than anticipated. This strongly implies the protein has a vital role in more fundamental processes than T cell biology. This work aims to create a knockout of TMEM131 in Zebrafish so the phenotype can be studied throughout embryogenesis. This work is detailed in chapter 5.

Chapter 2

General Methods

2.1 Molecular cloning

2.1.1 Plasmids and DNA sequences

Cloning strategies were designed and optimised *in silico* using SnapGene software. For most constructs in this thesis, sequences were amplified by polymerase chain reaction to add flanking sequences containing restriction sites. This enabled insertion into the pHR or pHCM plasmid backbones. The pHR lentiviral plasmid backbone expresses proteins via the spleen focus forming virus (SFFV) promotor. It can be used for transient transfections of target eukaryotic cells or together with two plasmids encoding HIV genes produce non-replicating viral particles. These viral particles can be used in the generation of stable eukaryotic cell lines. The pHCM plasmid is a variant of pHR that can also be used for transient and stable transfections though expression of the protein of interest is controlled via the cytomegalovirus promotor. Both plasmids encode ampicillin resistance for cloning in *E. coli*.

2.1.2 Polymerase chain reaction (PCR), Restriction endonuclease digestion and ligation

Primer stocks (purchased as lyophilised powder from IDT or Sigma) were kept at 100µM then diluted to 10µM before use. DNA sequences of interest were amplified using KOD polymerase (Millipore) using a KOD master mix containing the KOD polymerase buffer, $MgSO_4$ and deoxyribonucleotides made according to the manufacturer's instructions and stored at -20°C.

Reaction Component	**Volume (µl)**
DNA Template	0.5
KOD polymerase	0.5
10µM Forward primer	2
10µM Reverse primer	2
KOD master mix	13
Water	32

Table 2.1: Standard reaction components for KOD PCR

For the generation of fusion constructs an overlap extension PCR method was used. This involved designing two complementary primers covering approximately 20bp either side of the fusion site. Using the standard KOD PCR protocol above, the 5' sequence is amplified with a forward primer and the overlapping primer with the antisense-coding sequence. The 3' sequence is amplified using a reverse primer and the overlapping primer with the sense coding sequence. This results in two products with a 40bp overlap. These products are then used as templates in a second round of PCR with two flanking primers, one in each sequence. This overlap extension PCR has the following reaction mixture:

Reaction Component	**Volume (µl)**
5' sequence	5
3' sequence	5
KOD polymerase	0.5
10µM Forward primer	2
10µM Reverse primer	2
KOD master mix	13
Water	22.5

Table 2.2: Standard reaction components for overlap extension PCR

Both standard KOD and overlap extension PCR mixtures were run in a thermocycler under the following reaction conditions:

Step		**Temperature (°C)**	**Time (minutes : seconds)**
KOD activation		95	2:00
32 cycles	Denature	95	0:10
	Annealing	60	0:10
	Extension	70	0:30 per kb
Final extension		70	1:00
Storage		10	Indefinite

Table 2.3: Reaction steps for KOD PCR

Amplified DNA was run on 1.5% agarose gels made by dissolving 1.125g of agarose powder in 75ml of TBE buffer, boiling and after cooling adding 3.8µl of Midori Green dye. Each 50µl PCR product was loaded together with 12.5µl of Orange G loading dye (1.7 mM Tris-HCl pH 7.6, 0.025% Orange G, 10% glycerol, 10 mM EDTA) and gels run at 130V for 25 minutes in TBE buffer. 10µl of DNA 1kb Hyperladder (Bioline) was used to estimate the size of the PCR products. Bands of the correct size were cut out and purified from the gel with a gel extraction kit (Qiagen) according to manufacturer's instructions.

I had more success with difficult PCR products using Q5 polymerase and achieved larger yields. Q5 mixes were prepared with the following components:

Reaction Component	Volume (µl)
DNA Template	0.5
Q5 polymerase	0.5
10µM Forward primer	2.5
10µM Reverse primer	2.5
Q5 buffer	10
10mM dNTPs	1
Water	33

Table 2.4: Standard reaction components for Q5 PCR

Overlap extension reactions with Q5 polymerase were set up using 5µl of the 5' and 3' sequence as above, adjusting the volume of water to maintain a 50µl reaction volume. Q5 PCRs were set up on ice and transferred to a PCR machine pre-warmed to 98°C. Reactions used the following conditions:

Step		Temperature (°C)	Time (minutes : seconds)
Initial Denaturation		98	2:00
32 cycles	Denature	98	0:10
	Annealing	65-72	0:10
	Extension	72	0:30 per kb
Final extension		72	2:00
Storage		10	Indefinite

Table 2.5: Reaction steps for Q5 PCR

Purified DNA fragments and desired vectors were digested with restriction enzymes (NEB) before being resolved on agarose gel electrophoresis and purified as above. Digestion with two restriction enzymes used the following reaction mixture incubated at 37°C for 90-120 minutes.

Reaction Component	Volume (µl)
Target	15
10X Buffer	5
Restriction Enzyme 1	1.5
Restriction Enzyme 2	1.5
Water	27

Table 2.6: Standard reaction components for restriction enzyme digestion

Cut and purified inserts and vector were ligated together using T4 DNA ligase (NEB) in volume ratios of insert to vector of 2:1, 4:5 and 0:1 where the latter controls for self-ligation events. These mixtures were incubated at room temperature from 30 minutes to overnight:

	Volume (µl)		
Reaction Component	High	Low	No
Insert	10	4	0
Vector	5	5	5
T4 ligase	1	1	1
Ligase Buffer	2	2	2
Water	2	8	12

Table 2.7: Standard reaction components for plasmid ligation

2.1.3 Heat-shock transformation of *E. coli*

Frozen aliquots of chemically competent DH5α bacteria were thawed on ice then 5µl of ligation mixtures added to the bacterial cell suspension and the mixture kept on ice for 30 min. These mixtures were then heated to 37°C for 5 min in a water bath then returned to ice for 2 min. 400µl of SOC without antibiotic was added and the cells incubated at 37°C. Cells were then plated onto agar plates containing ampicillin and incubated overnight at 37°C.

Bacterial colonies on plates were screened by colony PCR using the MyTaq Red polymerase (Bioline) and opposing primers within the insert and the vector. From the size and presence of the PCR product, colonies can be selected that contain the correct vector and correct insert in the correct orientation. This was done by picking several colonies from a plate using a pipette tip, swirling the tip in the reaction mixture then placing the tip into a labelled 50ml falcon tube containing 5ml of ampicillin-containing lysogeny broth.

Reaction Component	**Volume (µl)**
MyTaq red	7.5
Primer in insert	0.6
Primer in vector	0.6
Water	6.3

Table 2.8: Standard reaction components for colony PCR

These reaction mixtures were run on the following thermocycler program then run on an agarose gel to test for the presence of bands of the expected size. Successful colonies were grown up overnight at 37°C with 200rpm shaking and plasmid DNA was extracted using the Plasmid Miniprep Kit (Qiagen), according to the manufacturer's instructions. Plasmid were adjusted to 200ng/µl and stored at -20°C.

Step		**Temperature (°C)**	**Time (minutes : seconds)**
Cell lysis		95	5:00
35 cycles	Denature	95	0:05
	Annealing	60	0:05
	Extension	72	0:30 per kb
Final extension		72	2:30
Storage		10	Indefinite

Table 2.9: Reaction steps for MyTaq Red colony PCR

2.1.4 Generation of competent *E. coli* cells

50µl aliquots of DH5α *E. coli* were thawed and added to 5ml of sterile LB media. This starter culture was grown overnight shaking at 37°C. The next day, 1ml of this culture was used to inoculate 30ml of fresh LB, which was grown shaking at 37°C until the optical density reached 0.4-0.6. The culture was then centrifuged at 3000rpm for 10 minutes and the pellet gently resuspended in 2.5ml of TSS buffer (10% PEG-3350, 5% DMSO, 50 mM MgCl2 in LB). 50µl aliquots of these cells were then frozen down at -80°C.

2.2 Maintenance and transfection of HEK293T cell lines

Human embryonic kidney (HEK) 293T cells were cultured in T75 flasks at 37°C and 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM; Life Technologies) supplemented with 10% heat-inactivated foetal bovine serum (HI-FBS; Gibco) and a cocktail of 2mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin (Gibco).

Cultures were passaged when they reached 50-90% confluency by discarding the medium, washing with sterile PBS and detaching the cells from the surface by incubating with 3ml of 0.25% trypsin (Gibco) for up to 5 min at 37°C. 7ml of fresh supplemented DMEM was then added to quench the trypsin. A variable volume of this cell suspension was added with fresh medium to a new T75 flask to create a new stock of ~10-20% confluency.

2.2.1 Transient transfection of HEK cells and lentiviral transductions

For transient and stable HEK293T transfections, a T75 HEK stock was trypsonised as described above then a 20µl sample of this cell suspension mixed with 20µl of Trypan Blue (Gibco). This was then loaded onto a haemocytometer and cells counted by eye using a light microscope.

For transient transfections, cells were seeded into six well plates at $0.5x10^6$ cells per well. After ~24h, for each transfection a solution of 100µl of serum free medium and 3µl GeneJuice reagent (Millipore) was made up and incubated for 5 minutes at room temperature. 1µg total of plasmid DNA was then added to the mixture, incubated for 20 minutes at room temperature then added drip-wise onto the well. These cells were then left 24-48 hours before analysis.

For lentivirus transductions, cells were seeded into six well plates at $1x10^6$ cells per well. After 24 hours these cells were transfected a mixture of 100µl of serum free medium, 4.5µl GeneJuice reagent and a plasmid mixture containing 0.6µg of the plasmid of interest, 0.3µg of pMD2.G plasmid expressing the VSV-G envelope and 0.6µg of p8.91 plasmid expressing the HIV-1 packaging genes gag, pol, rev and tat. After 48-72h, the medium was removed from each well and centrifuged at 8000g for 3 min to sediment cellular components. This virus-containing supernatant was added to wells containing untransfected cells in various amounts (e.g. 1ml of each if two plasmids are required) and the lowest volume of plasmid resulting in a homogeneous population selected.

2.2.2 Freezing and thawing of cells for long-term storage

To store HEK cells long term, confluent T75 flasks were resuspended with trypsin as described previously. The cell suspension was then centrifuged at 1300rpm for 3 minutes and the cell pellet resuspended in 1ml of freezing media (DMEM with FBS and 10% DMSO (Sigma)). The cells were then added to 1.8ml cryotube vials (ThermoFisher Scientific) and stored in a freezing container containing isopropyl alcohol. When placed in a -80°C freezer, the chamber enables a gradual temperature decrease over 1-2 days after which cells were stored at -80°C or in liquid nitrogen.

To revive stored cells, vials were rapidly thawed in a 37°C water bath then slowly added to 9ml of warmed media. Cells were then centrifuged at 1300rpm for 3 minutes and the cell pellet resuspended in fresh media. Cells were then seeded in a T75 or T25 flask.

2.3 Flow cytometry

Samples were regularly prepared for flow cytometry to check the transfection efficiency of cell lines used for proximity labelling or pulldown experiments. Typically a proportion of cells was set aside on ice during the procedure then washed twice in 1ml of cold PBS by centrifugation at 800g for 3 minutes. This cell pellet was flick-resuspended and fixed in FACS-FIX buffer (1.6% formaldehyde, 2% glucose, 0.1% NaN_3 in PBS, pH 7.4). These samples were run on a LSRII flow cytometer (BD Biosciences) using BD FACSDiva software. Cells were gated and samples run until 10,000 events had occurred. Data was exported as .FCS files and analysed using FlowJo software.

2.4 Confocal Microscopy

For imaging transfected cells, I used a Nikon Ti inverted microscope with a CSU-X1 spinning-disk confocal head (Yokogawa) and an iXon Ultra EM-CCD camera (Andor). The imaging chamber (Okolabs) was maintained at 37°C and 5% CO_2. Imaging at 100X used a Plan Apo VC 100×/NA1.4 oil-immersed objective (Nikon). The microscope was controlled by µManager2 software, allowing the capture of multiple channels, timepoints and positions when required. A built-in perfect-focus unit corrected for drift in the z-plane over time. Images were analysed in FIJI or MATLAB. Montages were made using the LabCode ImageJ plugin from *quantixed*.

2.5 Polyacrylamide gels electrophoresis

Samples were prepared for western blotting according to the labelling procedure for each experiment. After the final wash and centrifugation step all the liquid was removed by pipetting and cells lysed on ice for one hour using 100µl of lysis buffer (150 mM NaCl, 1% NP-40, 100µM NaVi, 1x protease inhibitor cocktail (Life technologies), 50 mM Tris-Cl, pH 7.4). After lysis the samples were centrifuged at max speed for 10 minutes at 4°C. 90µl of each supernatant was mixed with 30µl of 4xLDS sample buffer (Thermo Fisher Scientific) and 12µl of 1M dithiothreitol (DTT) then heated to 70°C for 10 minutes.

10-15µl of each sample was loaded onto pre-cast NuPAGE 4-12% Bis-Tris gels (Thermo Fisher Scientific) and run at 200V for 55 minutes using a 1x MOPS running buffer with 500µl of NuPAGE antioxidant (Thermo Fisher Scientific) in the internal reservoir. Samples were compared to 1µl of BlueElf Prestained Protein Marker (Jena Bioscience). Samples were blotted onto nitrocellulose membrane using the iBlot2 apparatus (Thermo Fisher Scientific).

If gels were to be probed by antibodies the membrane was blocked in 2.5% skim milk powder in 1xTBS buffer (2.4 g Tris base, 8.8 g NaCl, 90 ml deionised H_2O, pH7.6) for one hour at room temperature with rocking or 4°C overnight. Primary antibodies were added in 2.5% skim milk powder in 1xTBS/T (1xTBS with 0.1% Tween-20) at a standard dilution of 1:5000 for two hours at room temperature with gentle rocking or 4°C overnight. After three washes in 1xTBS/T, secondary antibodies were added in 1xTBS/T with 2.5% skim milk powder at a standard dilution of 1:5000 for 2 hours at room temperature.

Blots were washed four times in 1xTBS/T and twice in 1xTBS before images were obtained with an Azure C600 gel imaging system (azure biosystems) with appropriate settings. Images were analysed and quantified using Fiji software.

When the blot was being probed using AlexaFlour-647 conjugated streptavidin to investigate biotinylation, skim milk powder could not be used as it contains too much endogenous biotin, resulting in high background. Instead membranes were blocked for 1 hour at room temperature with 1% BSA in 1xTBS. After blocking, primary antibody followed by secondary antibody and streptavidin staining took place as above except solutions of 1% BSA were used instead of 2.5% milk.

2.6 Mass spectrometry

2.6.1 In solution/on bead tryptic digest

Beads with enriched proteins were washed thoroughly to remove contaminants and detergent before being transferred to a new tube. 45µl of 50mM ammonium bicarbonate was then added to each sample. Cystine residues within the proteins were reduced and alkylated by incubation at 70°C for 5 minutes with 10mM tris(2-carboxyethyl)phosphine (TCEP) and 40mM 2-chloroacetamide (CAA). Proteins were digested overnight at 37°C with 100ng of proteomics-grade trypsin. The solution was then acidified with 1µl of 1% Trifluoroacetic acid for every 10µl of the reaction.

The liquid above the beads was then diluted to 100µl with 2% acetonitrile and the C18 stage tip protocol was used to remove remaining MS-incompatible salts and detergents.

2.6.2 C18 stage tip clean-up

Two layers of a C18 membrane were punched out then placed into a 200µl pipette tip held above a 1.6ml sample tube. The membrane was conditioned via centrifugation at 2,000rpm for 2 minutes with 50µl of 100% Methanol then 50µl of 100% Acetonitrile then 50µl of 2% Acetonitrile, 0.1% Trifluoroacetic acid. The tip was then moved into a new tube.

The peptides from overnight digest were diluted to 300µl with a solution of 2% Acetonitrile, 0.1% Trifluoroacetic acid then 150µl added onto the membrane. The tips were centrifuged at 2,000rpm for 10 minutes then this was repeated for the other 150µl. The tips were transferred into a new tube then washed via centrifugation at 2,000rpm for 4 minutes with 50µl of Ethyl acetate then 50µl of a solution of 2% Acetonitrile, 0.1% Trifluoroacetic acid.

Tips were then transferred to a new tube and peptides eluted in 50µl of 80% Acetonitrile by centrifugation at 2000rpm for 2 minutes. Acetonitrile was removed in a speed-vacuum centrifuge until the samples were nearly dry. Peptides were resuspended in 2% Acetonitrile, 0.1% Trifluoroacetic acid then transferred to mass spectrometry vials before storage at -20°C.

2.6.3 In gel tryptic digest

For in-gel digests, samples were run on NuPAGE gels and washed three times for 5 minutes with ultrapure water. The gel was stained with GelCode Blue Safe Protein Stain (Thermo Scientific, Coomassie G-250) for 1h at RT then background staining reduced by leaving the gel rocking in ultrapure water overnight. Bands of interested were cut out, diced into 2-4mm pieces and transferred to a 1.5ml sample tube.

Diced gel pieces were destained by four 20 minute incubations at RT with 650rpm shaking in a 50% ethanol, 50mM ammonium bicarbonate solution. Pieces were then dehydrated with 100% ethanol. Cystine residues within the proteins were reduced and alkylated by incubation at 70°C for 5 minutes with 10mM tris(2-carboxyethyl)phosphine (TCEP) and 40mM 2-chloroacetamide (CAA). The gel pieces were dehydrated again with 100% ethanol, rehydrated with 2.5ng/µl trypsin in 50mM ammonium bicarbonate and incubated overnight at 37°C for protein digestion.

The next day peptides were extracted in 25% acetonitrile, 5% formic acid by three 10 minute sonication steps with the liquid from each step combined into a new tube. Peptides were then concentrated in a speed vacuum to bring the volume down to 20µl then made up to 50µl with 2% acetonitrile, 0.1% Trifluoroacetic acid.

2.6.4 Sample analysis

Mass spectrometry was performed by the Warwick Proteomics facility. First tryptic peptides were separated using a reversed phase chromatography Ultimate 3000 RSLCnano system (Dionex). Separated peptides were ionised by electrospray ionisation and analysed on a Thermo Orbitrap Fusion mass spectrometer (Q-OT-qIT, Thermo Scientific).

Peptide fragment data was searched using MaxQuant version 1.6.2.6 software within a *Homo sapiens* protein database supplemented with the sequences of any expressed constructs or proteins added in the assay. A maximum of two missed trypsin cleavage sites specified. The carbamidomethylation of cysteine residues was treated as a fixed modifications whilst the oxidation of methionine residues and acetylation of the protein N-terminus were treated as variable modifications. Certain experiments included additional biotinylation, phosphorylation and ubiquitination modifications.

Scaffold version 5.1.1 was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least 1 identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm.[215] Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

Chapter 3

Trafficking of the pre-TCR complex

3.1 Introduction

3.1.1 Structural characteristics and signalling of the pre-TCR

The importance of pTα during β-selection

The pre-TCR is enigmatic as it contains essentially the same protein subunits as the mature $\alpha\beta$TCR, with only pTα instead of a TCRα chain, but has very different signalling properties.[216] The pTα protein was discovered as a 33kD glycoprotein interacting with TCRβ using surface biotin labelling on pre-T cells.[217] Sequencing of the human gene, located on chromosome 6 close to the MHC, revealed the protein was a type-I transmembrane protein with a four exon structure. The first exon encodes the signal peptide and first three amino acids of the protein whilst exon 2 encodes the extracellular domain with an Ig-like fold. This domain lacks homology with TCRα. Exon 3 encodes a short connecting peptide including a cysteine residue that forms a disulphide bridge with TCRβ, similar to TCRα in the mature TCR. Exon 4 encodes the transmembrane region, including basic arginine and lysine residues important to the interaction with the CD3 chains, as well as a unique cytoplasmic tail.[218]

In both humans and mice, alternative splicing creates a second mRNA transcript called pTα^{b} missing exon 2.[219,220] This isoform appears to be expressed by a minority of thymocytes, especially DP thymocytes, as well as in some peripheral T cells such as nTreg cells.[221] Despite the loss the Ig-like domain making up most of the extracellular part of the protein, pTα^{b} seems to enhance rather than reduce TCRβ surface expression compared to the full length construct.[222] A precise role for pTα^{b} is yet to be established.

Early genetic experiments found that pTα -/- knockout mice have less than 10% the normal number of thymocytes as WT mice, with an arrest at the DN3 stage.[223] Though the reduction in thymocytes is significant, it is not as severe as the reduction caused by knockout of TCRβ[224] or CD3ϵ[225] nor the complete block in a RAG1 -/- mutant which lack mature T and B cells entirely.[226] In a pTα -/- background remaining thymocytes include a three-fold larger proportion of $\gamma\delta$ T cells and a small number of DP thymocytes. These DP cells might be present due to premature expression of the $\alpha\beta$TCR which allows T cells to inefficiently pass the β-selection checkpoint in the absence of the pre-TCR.[227] These DP cells frequently have two rearranged TCRβ alleles suggesting a fault in allelic exclusion.[228]

Restoration of DP thymocyte production in a pTα -/- or RAG1 -/- knockout background through expression of transgenes can lead to interesting results about the importance of pre-TCR domains. However, these need to be interpreted with caution due to the level of overexpression compared to endogenous proteins. For instance, progression to the DP stage was partially restored in a pTα -/- background by transgenic expression of a tailless pTα mutant[229] and in a RAG -/- background by a pre-TCR lacking all external sequences in pTα and TCRβ beyond the disulphide bond. Similarly, progression in a pTα -/- mouse was enabled by a transgene of the natural pTα^{b} isoform, lacking the extracellular domain and a similar construct missing almost all of both extracellular and cytoplasmic domains.[230]

A common view that pTα acts as nothing more than a surrogate TCRα chain during β-selection is seemingly invalidated by experiments by Borowski *et al* where transgenes were compared with an element of competition.[231] This was done by injecting equal numbers of thymocytes expressing two different transgenes into the thymus of an irradiated pTα -/- mouse. Crucially the transgenes used the same p56^{Lck} proximal promotor to give roughly equal expression levels and the competition for space and resources in the thymus was expected to reveal differences between transgenes not apparent if they were expressed alone. This experiment found that neither TCRα nor a hybrid molecule, which had the extracellular domains of TCRα but the transmembrane and cytoplasmic domains of pTα, could completely recreate the functionality of wild-type pTα. Thymocytes derived from pTα-transduced precursors dominated over TCRα derived precursors by over 60 times and the hybrid by a factor of 10. Most surprisingly, TCRα-transduced thymocytes showed an increase in apoptosis, decreased proliferation and were biased towards the $\gamma\delta$-lineage rather than the $\alpha\beta$-lineage. These results strongly imply that unique factors about the pTα extracellular domain and cytoplasmic tail are crucial for its functionality *in vivo* that cannot be substituted by TCRα.[232]

The role of the pTα cytoplasmic tail

Although the pTα extracellular Ig-like domain has little homology to either extracellular domain of TCRα, this domain has approximately 80% amino acid similarity between the mouse and human proteins. The pTα cytoplasmic sequence in contrast is poorly conserved between human and murine homologs (figure 3.1).[233,234] The only features apparently conserved between the two proteins are a cysteine residue positioned near to the membrane and a proline-rich sequence that could act as a SH3-domain binding sequence. Even these features are questionable as *ptcra* homologs have been identified in non-mammalian animals with similar exon structures but lacking proline-rich sequences in cytoplasmic tail.[235]

```
Mouse: TCSHL----------------------------RLHVLAGQHLQPPPSRK---SLPPTHRIWT-
Human: TCSCLCDPAGPLPSPATTTRLRALGSHRLHPATETGGREATSSPRPQPRDRRWGDTPPGRKPGSP

Mouse: ------------------------------------------------
Human: VWGEGSYLSSYPTCPAQAWCSRSALRAPSSSLGAFFAGDLPPPLQAGAA
```

Figure 3.1: Sequence alignment between the murine and human pTα cytoplasmic tails.
Identical residues in both versions are highlighted in red while dashes indicate gaps in the sequence alignment.

Whilst mouse studies performed with pTα transgenes without competition found that the pTα tail was not required to pass β-selection,[229,230] in competitive conditions a more mixed picture is revealed. Aifantis *et al*[236] investigated versions of pTα lacking the cytoplasmic domain, lacking the cytoplasmic

proline sequences and in which the juxtamembrane cysteine residue was mutated to alanine. In a pTα -/- background, transgenic expression of the wild-type pTα or pTα with a juxtamembrane Cys to Ala mutation was found to restore function and thymocyte progression to the DP stage. Meanwhile expression of pTα variants lacking the tail or the proline-rich sequence only partially restored function and defects were observed in apoptosis, proliferation and $\alpha\beta$-$\gamma\delta$ lineage choice. This strongly implied a function for the proline-rich sequences but that the juxtamembrane cysteine was dispensable.

Carrosco *et al* attributed the function of the cytoplasmic tail to ER retention, finding that replacing the short cytoplasmic tail of TCRα with the human pTα tail sequence was sufficient to reduce the surface expression of the chimeric TCRα construct to levels closer to that of the wild type pTα.[237] Similarly fusing the pTα tail to TCRα seemed sufficient to direct the TCR complex towards internalisation and degradation, rather than internalisation and recycling, when measured by the decay of surface biotinylated receptor.[238] The SH3-containing adaptor protein CMS/CD2AP, a member of the wider family of CIN85 adaptor proteins,[239,240] was identified as a mediator of these behaviours through a GST-pulldown assay.[241] The adaptor likely binds to the proline-rich region of the pTα tail. However inability of a chimeric TCRα construct with the pTα cytoplasmic tail to entirely replicate the functionality of pTα during β-selection, implies that the cytoplasmic tail is not solely responsible for the different signalling of the pre-TCR and TCR.[231]

The consequences of pre-TCR extracellular domain asymmetry

As pTα has only a single extracellular domain and TCRβ has both C and V domains, the pre-TCR structure is inherently asymmetric. Though asymmetry in protein receptors is not usually problematic, here the lack of a partner for the TCRβ variable domain exposes a conserved hydrophobic patch, consisting of the residues Tyr35, Leu/Phe91, Pro43 and Pro108 (depicted in figure 3.2). This patch is hidden in the mature TCR as part of a $>200\text{Å}^2$ hydrophobic interface to the TCRα variable domain and is likely to significantly reduce the stability of the protein if exposed to solvent. Emphasising this, a recent effort to produce a single-variable domain chimeric antigen receptor from TCRβ found the hydrophobic patch severely restrained possible TCRβ sequences.[242] The only TCRβ sequences identified that could function without the TCRα domain used the Vβ5-8 variable segment which has a long CDR3 loop predicted to partially cover the hydrophobic surface. This sequence restraint is evidently not present in the pre-TCR where a wider range of TCRβ sequences are permissible.

This stability issue led to speculation that there was another extracellular subunit of the pre-TCR to partner the Vβ domain,[244–247] sometimes nicknamed VpreT in reference to the VpreB surrogate light chain that together with λ5 and the IgH chain makes the pre-BCR of developing B cells. Berger *et al.*[248] published evidence in support of this proposition with the identification of a pre-TCR isoform distinct from the conventional pre-TCR in that it was inaccessible to an antibody specific for the Vβ or Cβ domains. The level of surface expression of this "masked" pre-TCR isoform appeared to vary between different thymic lymphoma lines and was completely removed by appending a TCRα variable domain to pTα. The authors concluded that the antibody accessibility was control by the presence of an additional subunit pairing with the Vβ domain. However, no protein has been since identified and the pre-TCR is still most commonly depicted as an asymmetric dimer.

More recent structural studies have argued for the existence of a pre-TCR dimeric structure in which Vβ hydrophobic patch is shielded from solvent by another domain. Pang *et al.*,[243] based upon a crystal structure of the pre-TCR extracellular domains in solution, proposed the pre-TCR formed a dimer parallel to the plane of the membrane with both pTα domains sandwiched between the Cβ domain and

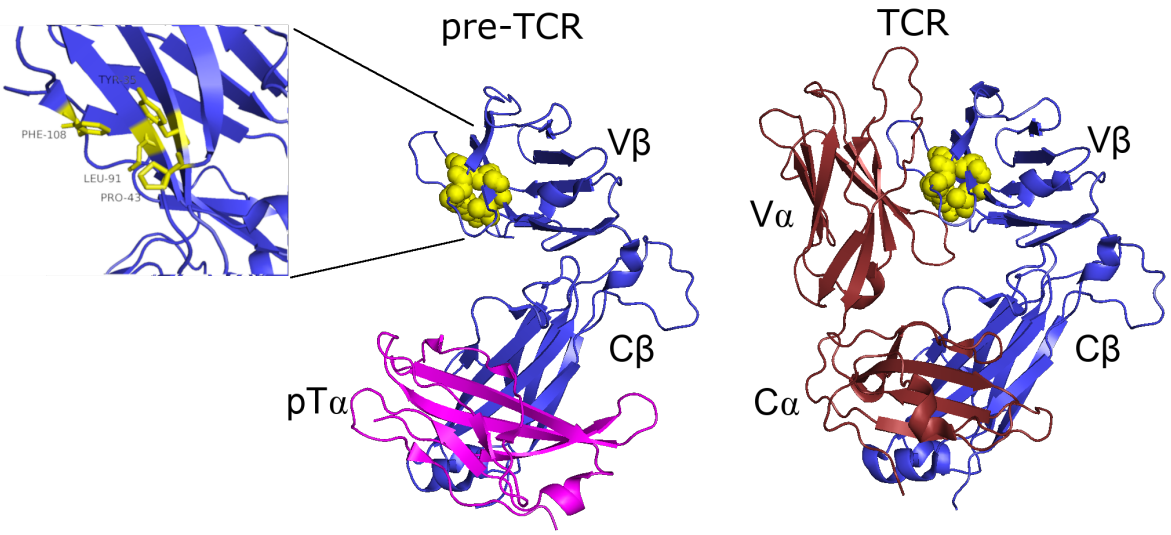

Figure 3.2: Structures of the pre-TCR (left) and TCR (right) extracellular domains highlighting the hydrophobic patch of TCRβ (yellow).
On the face of the Vβ Ig-like domain there is a hydrophobic patch consisting of highly conserved residues Tyr35, Leu91 or Phe91, Pro43 and Pro108. The patch is part of the interface to TCRα (red) in the mature TCR but exposed in the pre-TCR due to pTα (magenta) having only a single extracellular domain. (From protein databank structures 3OF6[243] and 1KGC[151] respectively.)

Vβ domain of the other TCRβ chain. This would hide the hydrophobic patch of each TCRβ V domain in the interface to C'CGF β-sheet face of the pTα domain. However this "head-to-tail" dimer model has a steep angle to the membrane which would seem to prevent association of the CD3 subunits. Although absent in the human protein, the murine pTα contains a glycosylation site at position N101 which would collide with the Vβ domain of the dimeric partner if the structure proposed by Pang *et al.* were adopted. Recent unpublished mutagenesis experiments would also imply that the only surface of pTα that is covered in the pre-TCR at the cell surface is the surface direct to TCRβ.[249] On the basis of such criticisms, as well as structures of TCRβ alone in solution, Zhou *et al.*[250] instead proposed the pre-TCR formed an upright dimer where the two Vβ domains are paired to shield their hydrophobic patches. This upright structure more closely resembles the mature $\alpha\beta$TCR, though with both Vβ domains rotated compared to the Cβ domains. Both models of dimerisation and the crystal structures supporting them are depicted in figure 3.3. Neither structure is widely accepted.

Although the pre-TCR is widely considered to signal in a ligand-independent manner,[251] it has been claimed that the pre-TCR interacts with pMHC ligands via the CDR loops and the hydrophobic patch. This was supported by solution NMR experiments[252,253] and optical tweezer experiments.[254] Li *et al*[255] claim to have captured the required "horizontal binding mode" this interaction would require in a crystal structure of the N15 TCRβ chain covalently linked to a MHC-I fragment (K^b-t2) presenting the VSV8 peptide (figure 3.4).

This interaction was said to act as a mechanism to not only select for productive TCRβ recombinations during β-selection but also select for recombinations with a level of pMHC affinity before the process continues together with TCRα chain during positive selection. Though theoretically compelling and would reconcile pre-TCR signalling with the steric exclusion model of TCR triggering, this hypothesis appears inconsistent with the low surface expression of the pre-TCR during β-selection. The genetic evidence that the β-selection checkpoint can be passed without the pre-TCR extracellular domains and in MHC-I and MHC-II deficient mice, would also indicate that if any interaction between the pre-TCR and pMHC occurs *in vivo* then it is not required for development.[257]

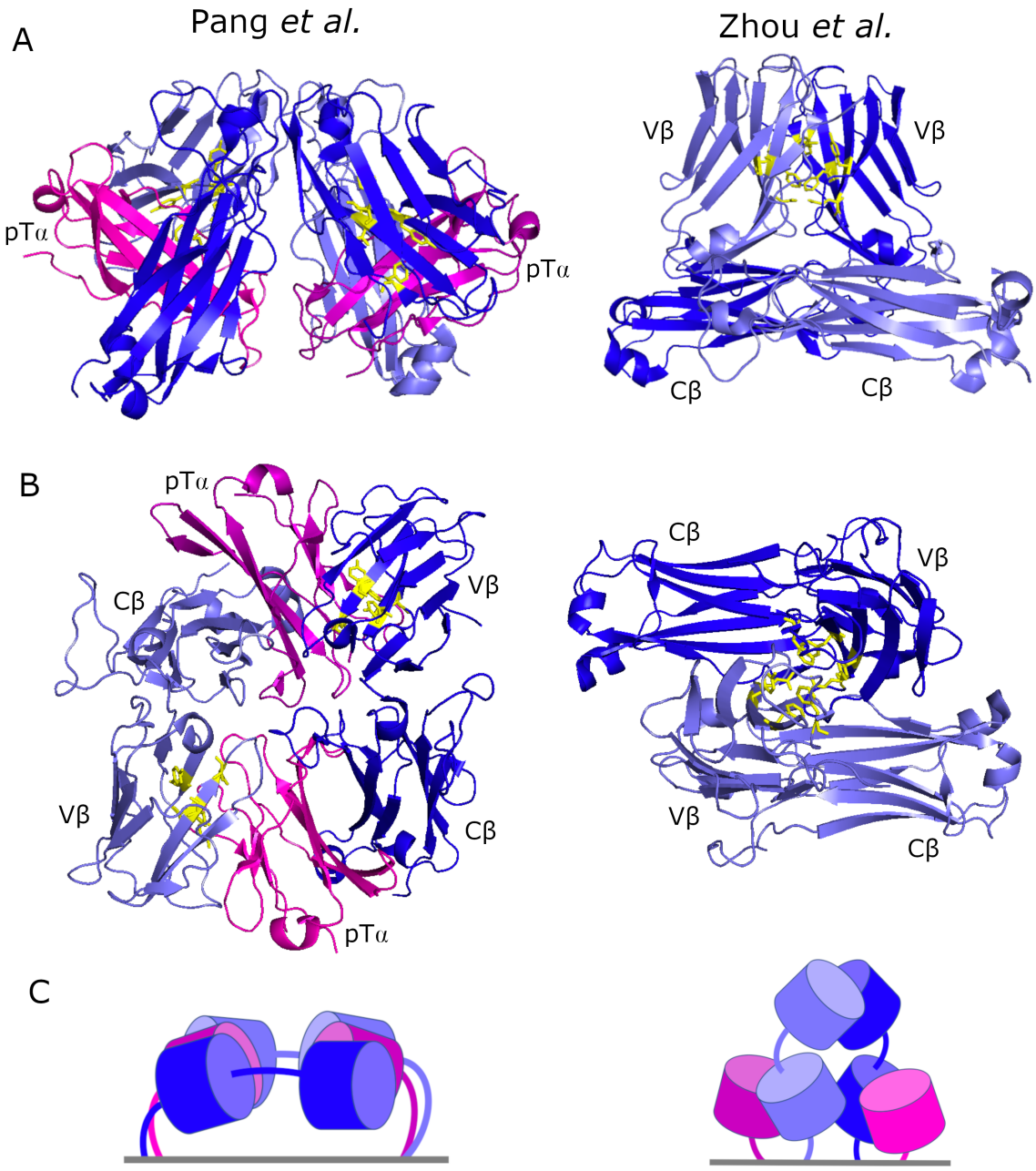

Figure 3.3: Structures of a pre-TCR dimer proposed by Pang *et al*[243] (left) and Zhou *et al.*[250] (right) based upon crystal structures of the pre-TCR extracellular domains (From protein databank structures 3OF6 and 3Q5Y).
A: Side views of the two structures with TCRβ chains shown in blue and pTα chains in magenta. Residues of the TCRβ hydrophobic patch are shown in yellow. Variable (Vβ) and Constant (Cβ) domains of TCRβ are labelled.
B: Top views of the same structures
C: Diagrammatic representation of the pre-TCR dimer models inferred by the respective authors.

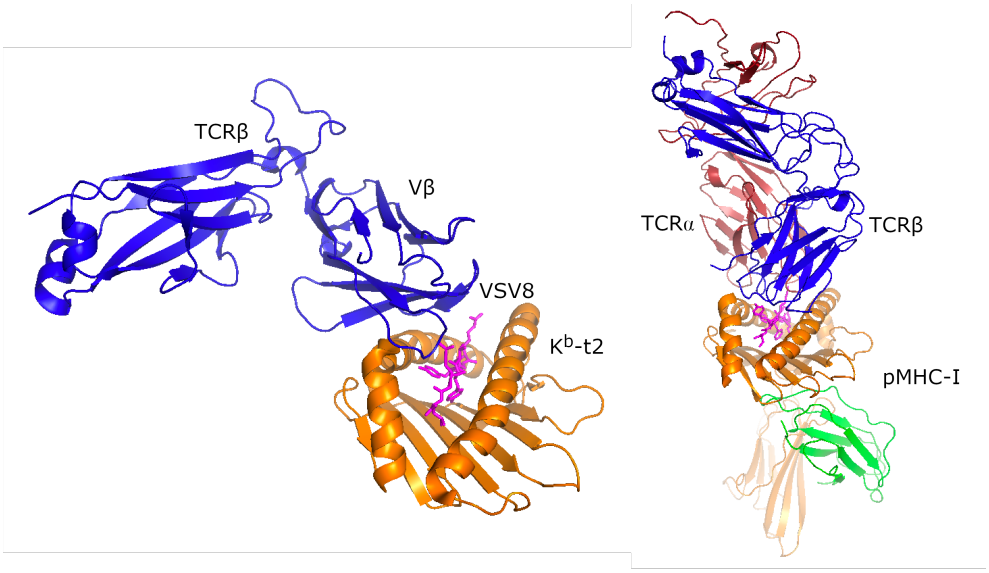

Figure 3.4: Structure of a proposed horizontal interaction between the pre-TCR and pMHC from Li *et al.*[255] (left) compared to the vertical interaction of the $\alpha\beta$TCR (right).[256]
The left structure shows the N15 TCRβ chain (blue) interacting with a fragment of MHC-I fragment (K^b-t2, orange) and the vericular stomatitis virus octapeptide (VSV8, magenta) via the C"C'CFG face of the Vβ domain and CDR3. The right structure shows the N15 TCR (TCRα red, TCRβ blue) interacting with the full pMHC structure (heavy chain orange, β2m green) presenting the same VSV8 peptide (magenta). Adapted from protein databank structures 6WL4 and 2CKB.

Pre-TCR localisation and signalling

The competing models for the mechanism of TCR signalling discussed previously are only further complicated when applied to the pre-TCR which has much lower surface expression, typically ~50-100x lower than the mature TCR on T cells.[258] Though there was some debate as to whether all four CD3 chains are strictly required, particularly CD3δ,[225,259–261] the pre-TCR appears to signal through identical signalling components as the $\alpha\beta$TCR and $\gamma\delta$TCR.

Whilst the TCR is known to internalise and be degraded upon ligand stimulation, the pre-TCR is observed to undergo constitutive internalisation and degradation in the lysosomes in the absence of ligand.[262] This continuous degradation, combined with retention of pre-TCR complexes in the ER, likely results in the lower expression at the cell surface. This in turn would seem be the source of the reduced level of TCR signalling initiated from the pre-TCR at the DN3 developmental stage compared to the $\gamma\delta$TCR with implications for the $\alpha\beta$-$\gamma\delta$ lineage choice.[263]

Consistent with clustering models of TCR triggering, an initial study observing asymmetric pre-TCR surface distribution proposed that the pre-TCR could signal without a ligand through constitutive movement in the plasma membrane to lipid rafts enriched in signalling molecules such as Lck.[264] This was thought to be controlled by the palmitoylation of a juxtamembrane cysteine in the pTα cytoplasmic tail however this residue was later shown to not be required for function[236] and enrichment of the pre-TCR is lipid rafts could not be replicated.[265] Other authors have proposed that the pre-TCR signals without a ligand through formation of oligomers, for instance through charged residues on the pTα extracellular domains.[266–268] However experiments demonstrating pre-TCR signalling with pTα and TCRβ chains missing their extracellular domains[257] and the existence the pTα^b isoform that lacks most of the residues thought necessary for dimerisation, would imply that pre-TCR oligomerisation via the pTα extracellular domains is not the primary mechanism of pre-TCR triggering.

An emerging but not consensus view, is that the ability of the pre-TCR to signal without a ligand has

less to do with any intrinsic property of pTα but more to do with the signalling state of DN cells when the pre-TCR is expressed. In mature T cells, the CD4 or CD8 coreceptors coreceptors recruit Lck to sites of TCR activity through binding to pMHC molecules. In DN cells their absence increases the pool of free Lck at the cell membrane.[257] The larger pool of free Lck means that more can associate with surface CD3 chains biasing the phosphorylation state of ITAM sequences towards the phosphorylation. Under these conditions DN cells would have a lower activation threshold for TCR signalling that mature $\alpha\beta$T cells where dephosphorylation from CD45 is thought to dominate.[265,269] The expression of spleen tyrosine kinase Syk, in addition to ZAP70 at the DN3 stage[270] and the inputs of other signalling pathways may also help to render DN thymocytes more sensitive to low TCR signals. In this primed state, TCR signalling strength might be dependent solely on the surface abundance of CD3 chains which varies between the pre-TCR and the $\alpha\beta$TCR due to structural differences.

3.1.2 Previous experiments in the James Lab

The pre-TCR is difficult to study *in vivo* due to its low expression at a brief point in thymocyte development. Previous work within the James lab therefore reconstituted pre-TCR signalling in a non-immune system as had been done previously for the $\alpha\beta$TCR.[101] The overall aim was to establish how signalling from the pre-TCR and TCR are distinguished at the β-selection checkpoint. The pre-TCR seems to result in quantitively lower signalling that the $\alpha\beta$TCR or $\gamma\delta$TCR and has much lower surface expression. However it was unclear whether its transient localisation at the cell surface is intrinsic to the receptor itself or a consequence of receptor signalling.[271] These two effects were distinguished by expressing the pre-TCR in non-immune HEK293T cells with the CD3 chains required for the complex to reach the cell surface but without other proteins needed for signal transduction such as Lck and ZAP70. Reconstituted expression in HEK also enables the receptor trafficking with a wider range techniques such as microscopy which is more difficult in small, non-adherent T cells.

When the pre-TCR and mTCR were expressed in HEK with a C-terminal GFP tag on TCRβ, both receptors were expressed at comparable amounts. However when measured with an anti-TCRβ antibody on ice, the pre-TCR was near undetectable at the cell surface and only slightly at high expression levels (figure 3.5A). Microscopy showed the majority of the pre-TCR was present in the ER whilst the mTCR was present in the ER and at the cell surface (figure 3.5B).

Once explanation might be that HEK cells are incapable of forming and secreting the pre-TCR complex. However, if cells expressing the pre-TCR were incubated in 37°C media with anti-TCRβ antibodies, conditions under which trafficking can still occur, then there was increasing fluorescence over time (figure 3.5C). This shows that pre-TCR is able to transit the cell surface when expressed in HEK cells and internalise bound antibody. Similar effects were seen in HeLa and Jurkat T cells. This seemed to mimic the expression seen in thymocytes where surface expression of the pre-TCR is much lower that the TCR. This implies that the distinct localisation of the pre-TCR is signalling-independent and intrinsic to the protein itself.

If the pre-TCR complex is able to transit the cell surface then why is its steady state expression so low? An internalisation assay was performed where cells were incubated with brefeldin A (BfA) which inhibits all protein export from the ER (figure 3.5C). The steady state expression of the $\alpha\beta$TCR at the surface fell gradually over 4h. Surface expression of the pre-TCR dropped much more quickly suggesting it was more rapidly internalised. Co-expression of clathrin-inhibitor AP180C partially stabilised the pre-TCR suggesting that internalisation was at least partially clathrin-mediated (data not shown).

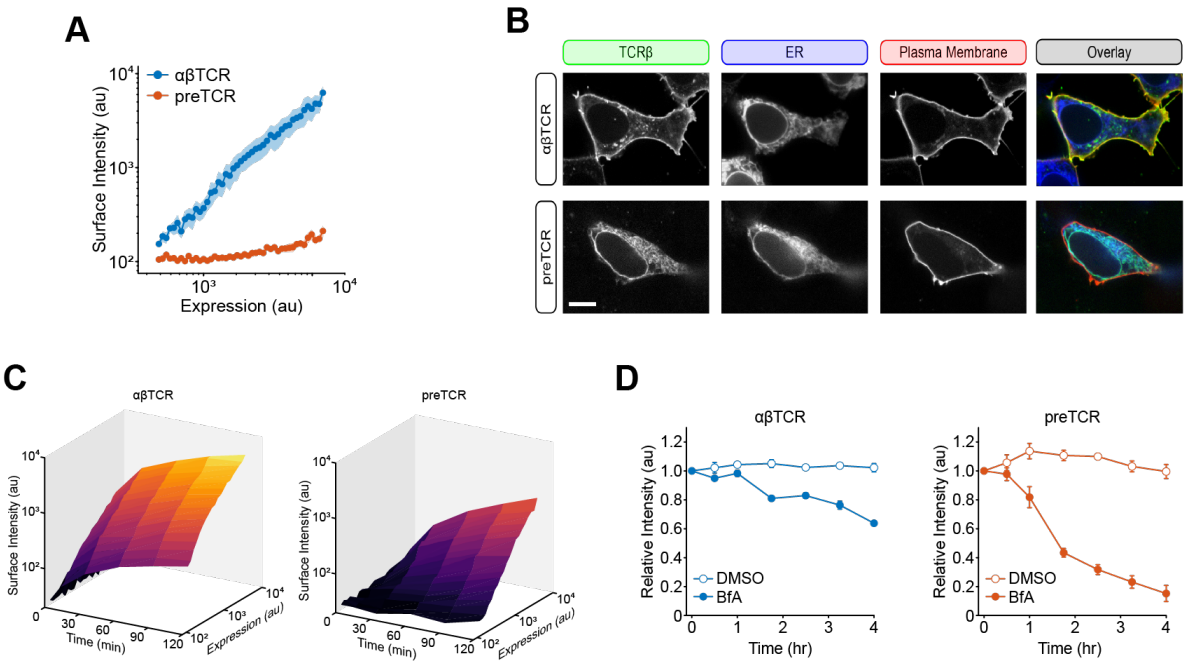

Figure 3.5: The pre-TCR is rapidly internalised from the cell surface.
A: HEK cells transiently transfected with GFP-tagged pre-TCR and $\alpha\beta$TCR with CD3$\gamma\delta\epsilon\zeta$ chains. The lines show the correlation between expression of TCRβ and surface staining on ice with an anti-TCRβ-af647 antibody. The shaded area shows mean ± SEM (n=3).
B: Microscopy images of HEK cells expressing the $\alpha\beta$TCR and pre-TCR with an BFPkdel ER marker and mCherryCaaX membrane marker.
C: HEK cells expressing the pre-TCR and mTCR were incubated with anti-TCRβ antibody in media at 37°C. There is an expression dependent and time-dependent accumulation of antibody over time for both receptors.
D: HEK cells expressing the $\alpha\beta$TCR or preTCR were incubated with either BfA or DMSO control before receptors still present at plasma membrane were detected with an anti-TCRβ antibody. The decrease in intensity from the initial timepoint infers the internalisation of the receptors. Datapoints show mean ± SEM (n=3).
Experiments performed by Andrei Smid with analysis by John James.

What features of the pre-TCR explains its divergent trafficking from the $\alpha\beta$TCR? As the pre-TCR cytoplasmic tail has been previous reported to have a role in ER retention, a new construct was made with the tail removed from pTα (preTCRΔTail) and a fusion construct made with the pTα tail on the C-terminus of TCRα ($\alpha\beta$TCR+Tail). These are depicted in figure 3.6A. The surface expression of these constructs relative to the original receptors is shown in figure 3.6B. The tail did not seem to greatly change surface expression in either case with preTCRΔTail (purple line) difficult to distinguish from the pre-TCR (red) and $\alpha\beta$TCR+Tail (yellow) very similar in surface expression to the $\alpha\beta$TCR (blue).

Differences in surface expression were instead attributed to the extracellular structures of the receptors. New constructs were made with the variable domain of TCRα attached to the N-terminus of TCRα (preTCR+Vα), the pre-TCR missing the Vβ domain (preTCRΔVβ) and the $\alpha\beta$TCR missing the Vα domain ($\alpha\beta$TCRΔVα). The surface expression of these constructs is shown in figure 3.6C. Complexes with the TCRβ variable domain removed or paired with TCRα appeared stable at the cell surface much like the αTCR. In contrast, the construct in which the variable domain of TCRβ was unpaired, was found to be poorly expressed at the cell surface like the pre-TCR.

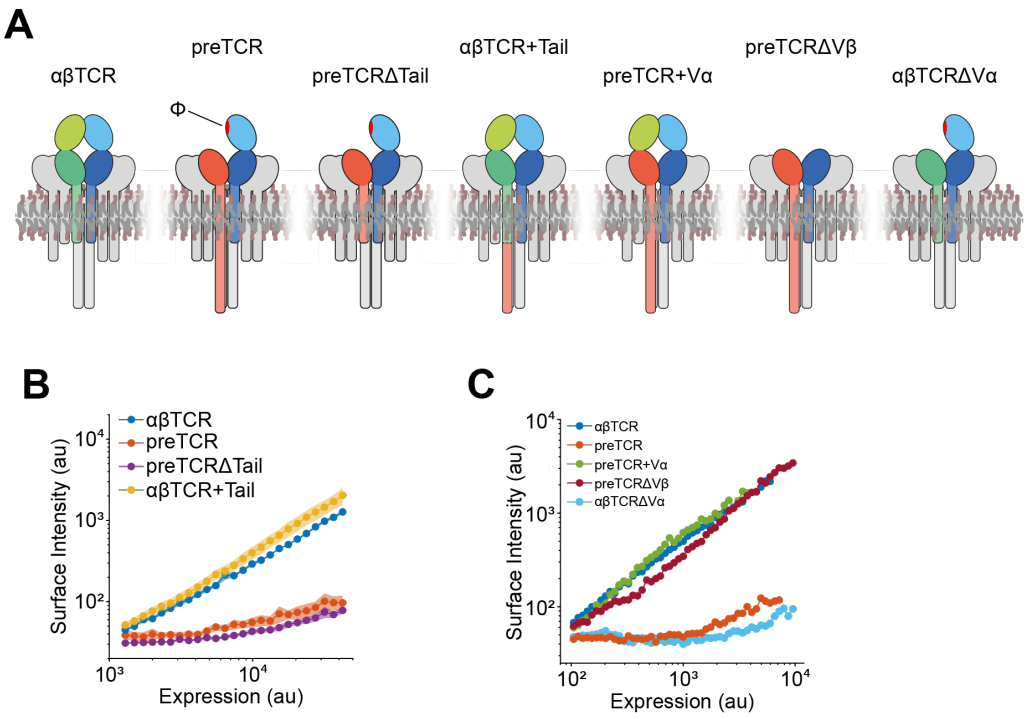

Figure 3.6: Pre-TCR internalisation is not affected by removal of the cytoplasmic tail but is dependent on the extracellular structure of the receptor.
A: Schematics showing the $\alpha\beta$TCR and pre-TCR complexes and receptor variants. The TCRβ chain is shown in blue, TCRα in green, pTα in orange and the CD3$\gamma\delta\epsilon\zeta$ chains in grey. The hydrophobic patch (Φ) on the TCRβ variable domain is shown in red.
B: Surface expression of $\alpha\beta$TCR and pre-TCR complexes with and without the pTα cytoplasmic tail. Receptor expression was measured by fluorescence of a GFP-tag on the C-terminus TCRβ whilst surface expression was measured by the binding of anti-HA antibody.
C: Surface expression of $\alpha\beta$TCR and pre-TCR and receptors with modified extracellular domains.
Experiments were performed by Andrei Smid with analysis by John James.

These result strongly implied that the extracellular domain structure of the pre-TCR was responsible for its localisation in HEK cells rather than properties of the cytoplasmic tail. Two hypotheses sought to explain how the extracellular domain structure was communicated across the plasma membrane to cytoplasmic internalisation machinery:

Hypothesis 1: The asymmetric domain structure enables receptor oligomerisation which leads to internalisation, for instance via inducing membrane curvature.

This hypothesis is consistent with the pre-TCR dimer models suggested by Pang *et al*[243] or Zhou *et al.*[250] though these are yet to be demonstrated to form *in vivo*. Work within the James lab to test this hypothesis used the Nanobit protein-protein interaction assay, a variant of the split luciferase assay development by Promega[272] with components on the cytoplasmic side. The assay suggested that the pre-TCR primarily existed as a monomer *in vivo* though this measured total receptor, the majority being in the ER, rather than the small portion of the receptor at the cell surface.

Hypothesis 2: HEK293T cells, and presumably other cell types, express a transmembrane interaction partner for the pre-TCR that is able to interact with the extracellular regions of the pre-TCR and communicate with the cytoplasm.

This hypothesis is consistent with the work of Berger *et al.*[248] that predicted a widely-expressed pre-TCR interaction partner and could also reconcile the stability of the pre-TCR and the exposed hydrophobic patch on the TCRβ variable domain. This interaction partner or chaperone would presumably transit with the pre-TCR from the ER. The relocation of ER chaperones, including BiP, calreticulin and protein disulphide isomerase, to the cell surface has previously been reported when cells are under ER stress.[273] A number of ER chaperones, including calnexin and BiP, appear to be expressed on the cell surface of immature thymocytes suggesting they have incomplete or insufficient ER-retention mechanisms.[274]

3.1.3 Proximity Labelling Assays

From their synthesis to their degradation proteins are accompanied by a changing set of interaction partners. Some of these interaction partners are near universal to all proteins such as the protein components of the ribosome; some like the membrane protein export machinery are specific to a class of proteins and some are highly specific to a particular protein. Investigating the identity and behaviour of these protein-protein interactions (PPIs) enables a deeper understanding of protein function and mechanism of action.

Challenges in studying protein-protein interactions include separating specific interaction partners of a protein, for instance the pre-TCR and a membrane chaperone, from non-specific interaction partners throughout the rest of the protein's lifetime. Another challenge is that many functional PPIs are relatively low affinity, meaning the interaction falls apart upon the reduction in protein concentration that accompanies cell lysis with detergents.[275] This prevents these interactions being studied by co-immunoprecipitation or tandem affinity purification methods. Many PPI's are also very transient meaning only a small proportion of the target protein is bound to its partner at any given time. Proximity labelling assays overcome these issues by attaching an enzyme to the protein of interest that generates reactive chemical tags that covalently label proteins nearby. Proteins can then be enriched via this label and identified using mass spectrometry techniques.

The commonly used BioID assay utilises a promiscuous biotin-ligase $BirA^{R118G}$ domain, tagged to the protein of interest, to generate a short-lived biotin adenylate ester. This ester (biotin-AMP) reacts with the lysine residues on nearby proteins with a range of ~10nm. Biotinylated proteins can be pulled down by affinity of biotin to streptavidin-beads and identified using mass spectrometry.[276] Biotin is membrane-permeable so can be added externally and the strong interaction to streptavidin enables high

stringency washes and protein extraction.[277] The modified BirA domain was later refined to make the domain smaller with better labelling (BioID2).[278]

The main restriction with the BioID assay is that labelling is quite slow so needs to accumulate over 18-24h. This prevents the study of short-lived processes such as the acute response of a cell to a drug or stimulus. The accumulation of biotinylated proteins could also lead to cellular stress potentially changing interactions.

An alternative to BioID is proximity labelling assays that rely on the action of peroxidase enzymes to generate short-lived radicals of biotin-tyramide/ biotin-phenol in the presence of H_2O_2.[279,280] These radicals react with tyrosine residues on nearby proteins with a range of 10-100nm. Labelling can be targeted by expression of a fusion construct of the protein of interest with ascorbate peroxidase (APEX).[281] The ascorbate peroxidase domain has also been modified for better activity (APEX2).[282]

Radical-based labelling methods are significantly quicker than BioID with labelling, taking place in seconds rather than minutes. This enables study of faster processes such as the trafficking of GPCR proteins.[283] One major disadvantage is that labelling primarily occurs on tyrosine residues which are less frequent and usually less accessible in proteins than lysine residues. The exposure to H_2O_2, whilst brief, also raises concerns about activating the cellular stress response.

Previous work in the James lab to identify pre-TCR interaction partners used the BioID proximity labelling assay. However, due to the small amount of time the pre-TCR spends at the cell surface, the majority of hits were ER-resident proteins. No protein consistent with being the internalisation factor was conclusively identified though the assay did suggest an interaction between the pre-TCR and a transmembrane protein of unknown function called transmembrane protein 131 (TMEM131) that will be discussed more in chapter 4 of this thesis.

It was thought that a membrane-impermeable assay would decrease the number of non-specific interactions by restricting labelling to the interaction partners of pre-TCR complexes that had been secreted from the ER and arrived at the cell surface. Two such assays are described below:

Selective proteomic proximity labelling assay using tyramide (SPPLAT)

Selective proteomic proximity labelling using tyramide (SPPLAT) is a peroxidase-based labelling method developed by Rees *et al.*[284] Instead of fusing APEX onto the protein of interest, radical labelling is targeted though the external addition of an antibody conjugated to horseradish peroxidase (HRP). Antibody binding, followed by washes to remove the excess, restricts peroxidase labelling to around the target protein. As labelling is targeted through an antibody, the target protein does not need to be modified. This avoids artefacts relating to the incorrect localisation of fusion constructs which are possible in BioID and APEX assays. Although biotin-tyramide is membrane permeable, the antibody is not so labelling is restricted to proteins on the cell surface.

HRP catalyses the same labelling chemistry as APEX with faster kinetics and greater resistance to H_2O_2-induced inactivation. It is however, much larger (44kD with glycosylations vs 27kDa for APEX) and has four structurally essential disulphide bonds. This means it cannot operate in the reducing environment of the cytoplasm.

Surface proximity labelling with Nanobody-APEX

One disadvantage of SPPLAT is that it is limited to proteins with available, high-affinity antibodies. Antibodies are also large, potentially increasing the labelling radius, and can crosslink their targets causing artefacts. A convenient way around these issues was recently published by Zavodszky and Hegde[285] while investigating the trafficking of misfolded prion protein PrP*.

PrP* is a GPI-anchored protein yet still transits the cell surface and is degraded via the lysosomes. The trafficking of the protein was studied using a GFP-tag which was used in imaging to track the protein through microscopy. Surface and internalised GFP-PrP* was distinguished using a fluorescently-labelled anti-GFP nanobody. This nanobody (Nb) is small (~18kd), monovalent and has sub-nM affinity. The GFP-tag was also used for pulldown experiment using anti-GFP beads (the same Nb attached to sepharose). For PrP, identified proteins were mostly ER residents, consistent with the bulk of GFP-PrP being in the ER.

To identify interactions of the membrane-localised PrP, a proximity labelling assay was performed that used the anti-GFP Nanobody fused to APEX2. This protein was added to the external media to target peroxidase-labelling to around GFP-tagged proteins at the cell surface. This assay found that misfolded PrP* was present at the cell membrane in complex with a set of ER-derived chaperones including TMED10, BiP and calnexin. This behaviour seemed very similar to that of the pre-TCR, which we believed was also unstable at the cell surface and suspected to be trafficked with an ER chaperone.

The two assays are depicted in figure 3.7 applied to the pre-TCR.

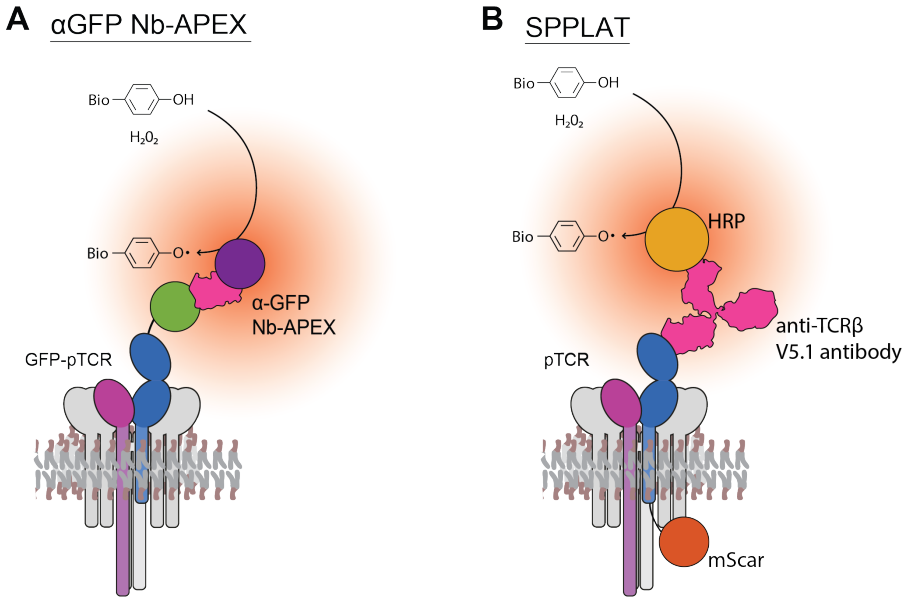

Figure 3.7: The Nanobody-APEX and SPPLAT proximity labelling assays.
Both assays use a peroxidase enzyme to catalyse the formation of biotin phenol radicals with hydrogen peroxide. These radicals react with tyrosine residues on nearby proteins. The approximate sphere of biotin labelling represented by circles with an orange-white gradient.
A: Labelling is targeted to the pre-TCR by an anti-GFP nanobody (pink) fused to APEX2 (purple). The pre-TCR is tagged with GFP (green) on the N-terminus of TCRβ.
B: Labelling is targeted to the pre-TCR by an anti-TCRβ antibody (pink) fused to HRP (orange). The pre-TCR can be unmodified though here a C-terminal mScar tag (red) was used to monitor expression.

3.2 Aims

When the pre-TCR is reconstituted in HEK cells, its localisation and trafficking are distinct to that of the structurally similar $\alpha\beta$TCR. This appears to copy the localisation of the receptor observed in thymocytes but without signalling kinases. Whilst the $\alpha\beta$TCR appears stable at the cell surface, the pre-TCR appears to be rapidly internalised resulting in low surface expression. This localisation was not explained by the intracellular sequences of the receptor but was instead dependent on the structure of the extracellular domains. The mechanism by which the pre-TCR is distinguished from the $\alpha\beta$TCR, in regards to localisation and trafficking, is currently unknown.

This project aims to:

- Apply proximity labelling assays to the pre-TCR to identify potential interaction partners at the cell surface. The ideal candidate to mediate pre-TCR internalisation would be a membrane protein that appears in the list of pre-TCR interactions partners but not in the list of TCR interaction partners.
- Investigate pre-TCR dimerisation as a mechanism of internalisation.
- Investigate the secretion of the pre-TCR from the ER.

3.3 Methods

3.3.1 Expression of TCR constructs

The pre-TCR and $\alpha\beta$TCR were expressed in HEK293T cells from pHR plasmids using the SFFV (spleen focus forming virus) promotor. For most experiments pTα/TCRα were expressed in the same bicistronic construct as TCRβ with the two proteins, complete with signal peptides, separated by the P2A sequence (YQGPGATNFSLLKQAGDVEENPGP). 2A sequences cause the ribosome to skip when forming the peptide bond between the final Gly and Pro residues.[286] This produces two separate peptide chains in a 1:1 ratio.

In many experiments, the TCRβ chain was modified with epitope tags. When a HA-tag was used on the N-terminus of TCRβ, the endogenous signal peptide was replaced with the *Gaussia* luciferase secretion signal (MGVKVLFALICIAVAEAK). N-terminal modifications were made using overlap extension PCR reactions. Tags on the C-terminus of TCRβ, typically GFP and mScar fluorescent proteins, were introduced using BamHI/NotI restriction endonuclease sites.

For most experiments, CD3$\delta\gamma\epsilon$ and ζ-chain proteins were co-expressed from a separate pHR plasmid as one quad-cistronic construct in the order CD3δ-E2A-CD3γ-P2A-CD3ϵ-T2A-CD3ζ. Frequently this was followed by the IRES sequence of EMCV before the fluorescent protein TagBFP. This allowed the expression of CD3 chains to be measured separately to the TCR chains. Other lab members reported slight improvements in expression by increasing the amount of CD3 plasmid relative to pTCR/TCR plasmid. In my testing I did not see an increase in expression or surface staining (supplementary figure S4) and achieved consistent results using equal amounts of the plasmids in each transfection.

Unless otherwise stated, experiments used the G10 TCR clone comprised of Vα28.1 and Vβ5.1 TCR chains. This TCR recognises the SLFNTVATL peptide derived from the HIV gag p17 presented by

HLA-A2.[287] The variable domain of the G10β chain can be detected as part of the pre-TCR or $\alpha\beta$TCR, using a mouse anti-TCRβ V5.1 antibody (clone LC4).

Some experiments also used the 1G4 TCR clone that was made by directed evolution to the tumour-associated peptide SLLMWITQC presented by HLA-A*0201.[288] The G115 $\gamma\delta$TCR clone was briefly investigated that uses Vγ9JPC1 and Vδ2D3J1C gene segments.[123]

3.3.2 Peroxidase-based proximity labelling methods

The protocol for peroxidase-based labelling methods was based on the detailed protocol of Hung *et al.*[280]

Before the experiment, Biotin phenol (BP, Sigma Aldrich) was dissolved in DMSO to form a 500mM stock. This was stored as 10µl aliquots at -80°C. 30% wt/wt Hydrogen peroxide (~10M H_2O_2, Sigma Aldrich) was stored at 4°C. This was diluted in PBS to form a 100mM working solution immediately before use. A fresh Quencher solution was prepared (10mM sodium ascorbate, 5mM Trolox, 10mM sodium azide in PBS).

For small-scale experiments, HEK cells in six well plates were transiently transfected with pre-TCR and mTCR constructs tagged with the APEX2-domain on the N-terminus of TCRβ. Cells were suspended using 0.25% trypsin then made up to 1ml of fresh media. Biotin phenol was added to a final concentration of 500µM and cells incubated for 20minutes. Labelling was initiated with hydrogen peroxide to a final concentration of 1mM. After 1 minute the reaction was quenched with three washes of cold Quench solution. Cells were lysed for 30 minutes on ice with a NP-40 lysis buffer.

For labelling with Nanobody-APEX the following modifications to the protocol were made. Cells expressing the pre-TCR or mTCR constructs with N-terminal GFP tags were resuspended in trypsin then incubated in 1ml of fresh medium containing 1:200 anti-GFP Nanobody-APEX at 37°C with occasional flicking. Excess Nanobody-APEX was removed by three washes 1ml in cold PBS. The cells were then resuspended in 1ml of PBS with 500µM biotin phenol for 5 minutes. Labelling was initiated with hydrogen peroxide and the reaction quenched as before.

For the SPPLAT experiments, 70 µg of anti-TCRβ V5.1 antibody (clone: LC4) in 100 µl was conjugated to HRP using a 100µg HRP conjugated kit (Abcam) following the manufacturer's instructions. Labelled antibody was stored at 4°C. Labelling used the same protocol as described above except the anti-GFP Nanobody-APEX was replaced with the same volume of HRP-conjugated antibody.

To scale up the SPPLAT assay for mass spectrometry, T75 flasks were transfected with proportional increases in the volumes of plasmids, GeneJuice and serum-free media. Labelling was performed as before scaling up the reagents to account for the larger input volume. After cells were lysed the lysate was clarified by centrifugation at max speed for 10 minutes at 4°C. The clarified lysate was then incubated with gentle rocking overnight at 4°C with 40µl of streptavidin-sepharose beads (Thermo Scientific).

These beads were then washed once with lysis buffer, once with 1M KCl, once with 0.1M Na_2CO_3, once with 2M urea in 10mM Tris-HCl pH 8.0 then twice with lysis buffer without detergent. Each wash involved addition of 1ml of ice-cold solution and a short centrifugation to pellet the beads before the solution was removed. The washed beads were transferred to a fresh tube with a final wash in detergent-free lysis buffer then processed with the on-bead tryptic digest protocol described in section 2.6.1.

3.3.3 Chemical crosslinking of surface proteins

Cells were suspended then washed three times in cold PBS to remove amide-containing media. After the final wash, cells were resuspended in 1ml of cold PBS pH 8.0 with 10µl of 100mM bis(sulfosuccinimidyl)suberate (BS3, Thermo Scientific) (~1mM final concentration). Cells were incubated on ice for 1h with gentle flicking. The reaction was quenched with 1µl of 1M Tris-HCl pH 8.0 before cells were lysed as normal.

3.3.4 RUSH assay

HEK cells were seeded into an 8-well microscopy dishes (Ibidi) at low density. After 24h, four wells were transfected with GeneJuice reagent with a mixture containing 2µl of pHR_pTα_SBP-GFP-G10β, 2µl of pEPT_CD3$\delta\gamma\epsilon\zeta$, 2µl of pHR_Streptavidin-BFPkdel and 1µl of pHR_mCherryCaaX to mark the plasma membrane. Four wells were transfected with the corresponding pHR_G10α_SBP-GFP-G10β construct.

48h post-transfection, cells were imaged using a spinning disk confocal microscope, maintained at 37°C and 5% CO2, using a 40X/0.75 NA air objective. Nine positions were selected in each well allowing four wells, two expressing the pre-TCR and two expressing the TCR, to be imaged simultaneously every 2 minutes. After the first frame, imaging was paused and 50µl of a 1:1000 dilution of GFP-Booster Alexa Flour 647 (Chromotek) in media was gently added to each well[†]. After allowing 10 minutes (5 frames) for the nanobody to diffuse evenly, 15µl of 5mM biotin was added to two of the four wells. Imaging proceeded for the next 90 minutes (45 frames). This procedure was then repeated for the other 4 wells. Imaging cells transfected with identical plasmids apart from receptor constructs without the SBP-tag indicated that maximum nanobody binding occurred in 10 minutes in the absence of Streptavidin-mediated ER-retention.

A small number of images were manually removed if cells left the frame. Image stacks were subsequently analysed in MATLAB. For each image, the number of foci in the nanobody channel exceeding a defined threshold were measured and results were averaged for each group. The Pearson correlation coefficient between the pixel intensities of the GFP channel and the BFP and antibody channels was also recorded for each frame.

3.4 Results

3.4.1 Proximity labelling of surface pre-TCR using anti-GFP Nanobody-APEX

HA-EGFP was cloned onto the mature N-terminus of TCRβ after the gaussia luciferase signal peptide. This chain was expressed as part of both the pre-TCR and $\alpha\beta$TCR. A plasmid encoding His-tagged anti-GFP Nanobody-APEX2 as well as some purified Nano-APEX protein were obtained as a gift from Dr Eszter Zavodszky.[285] The plasmid sequence was used as a template to clone the APEX2 domain onto TCRβ in the same position as the GFP. The two domains are comparable in size and the direct fusion should allow the comparison between labelling throughout the lifetime of the receptor to surface-restricted labelling. A mScar tag was added to the C-terminus of APEX2-TCRβ so that the expression of the

[†]Cells seemed happiest if this media was taken from cells grown at the same time rather than fresh but this was not tested extensively.

construct could be monitored via flow cytometry. A dual-tagged GFP-TCRβ-mScar was also cloned which should be labelled identically to the GFP-TCRβ version but have similar mass to the APEX fusion. An internalisation assay with anti-TCRβ V5.1 antibody, suggested that fusion to GFP did not impeded receptor trafficking to the cell surface (figure 3.8). There was lower surface expression of APEX2-tagged pre-TCR but this was not seen for the mature TCR.

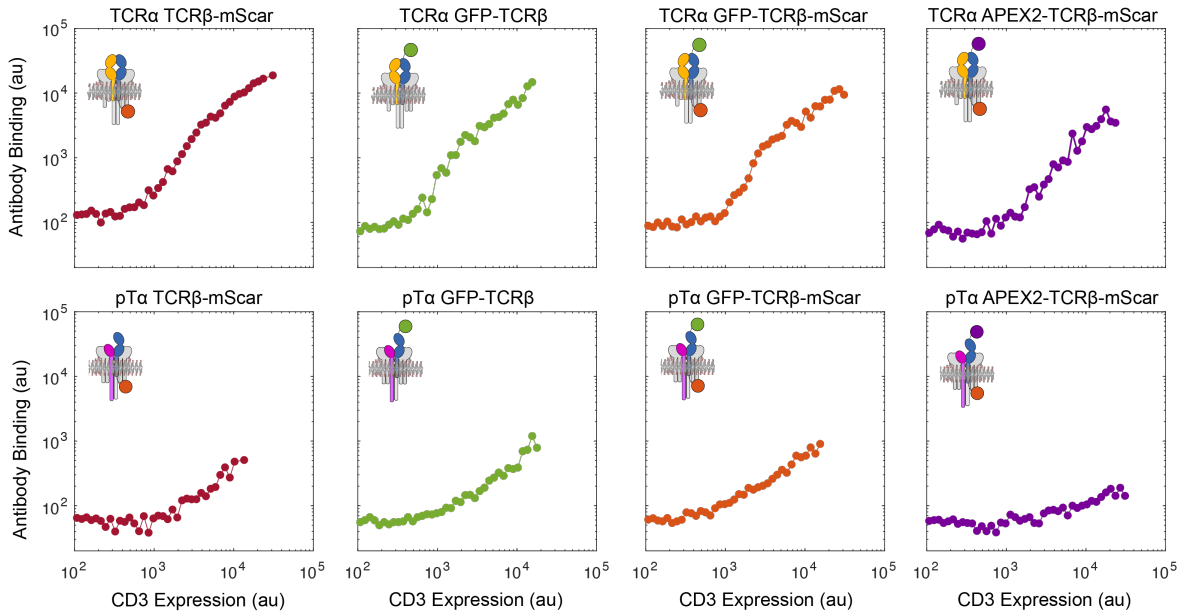

Figure 3.8: Fusing GFP to the N-terminus of TCRβ as part of the pTCR or mTCR does not affect trafficking to the cell surface.
HEK cells expressing the indicated constructs with CD3$\gamma\delta\epsilon\zeta$-BFP were suspended with trypsin then incubated in 37°C media for 30 minutes with 1:200 anti-TCRβ APC-conjugated antibody. Cells were washed, fixed then analysed by flow cytometry. The data was sorted into 50 log-distributed bins between 10^2 and 10^5 along the "B 530/30" axis. The mean value in the "R 685/35" channel for each bin was plotted on the y-axis in the middle value of each bin. Bins with less than 10 events were excluded. Inserts show the structures of these constructs with the TCRβ chain shown in blue with GFP (green), mScar (red) or APEX2 (purple) tags. TCRα and pTα are shown in yellow and pink respectively. All complexes include the CD3 chains (grey).

To conserve Nanobody-APEX protein, the assay was first tested on small volumes of cells. HEK cells in a six well plate were transfected with the GFP-tagged constructs, suspended with trypsin then incubated with 1:200 µl anti-GFP Nanobody-APEX for 30 minutes at 37°C. After time for the nanobody to bind, the excess was removed with three washes in PBS. All cells were then incubated in PBS with 500µM biotin phenol before labelling was initiated with hydrogen peroxide. The reaction was quenched by four washes in quenching buffer before lysis. The APEX-TCRβ-mScar constructs were used as positive controls for the APEX reaction whilst a sample expressing the BioID2-tagged mTCR was used a control for streptavidin staining. These samples were run on a 4-12% BisTris gel and analysed by western blotting (figure 3.9).

Lysate from cells expressing GFP-tagged receptor constructs treated with the nanobody-APEX protocol produced a series of streptavidin-labelled bands after blotting. The pattern of bands was similar for the pre-TCR (lanes 1 and 2) and mature TCR (3 and 4) implying that similar proteins had been labelled in each case. Additional streptavidin-labelled bands, for instance at 41kD, are seen in lanes containing lysate from cells expressing the same receptors fused to APEX2 (lanes 5 and 6). This is consistent with these constructs having labelling throughout the receptor lifetime rather than just at the cell surface. There are only three bands visible in lanes containing untransfected HEK that were consistent in mass with three endogenously-biotinylated proteins (Pyruvate carboxylase, MCCA and PCCA).

In the cf790 channel, there are three HA-positive bands for GFP-TCRβ-mScar and GFP-TCRβ. In the two receptors with GFP-TCRβ-mScar these bands are larger, consistent with the increase in mass, and quite far apart. The heaviest band is most intense. In both receptors with GFP-TCRβ the three bands are closer together and the middle is most intense. In all four lanes, the heaviest band appears to be biotinylated whilst the others do not. These bands might represent different glycosylation states of TCRβ with the heaviest being present at the cell surface. There is one band in each of lane for the APEX2-constructs. It is worth noting that the sequence for the HA-tag (YPYDVPDYA) contains three tyrosine residues that can be biotinylated in the labelling reaction. This means that intense labelling can interfere with recognition of the tag by anti-HA antibodies.

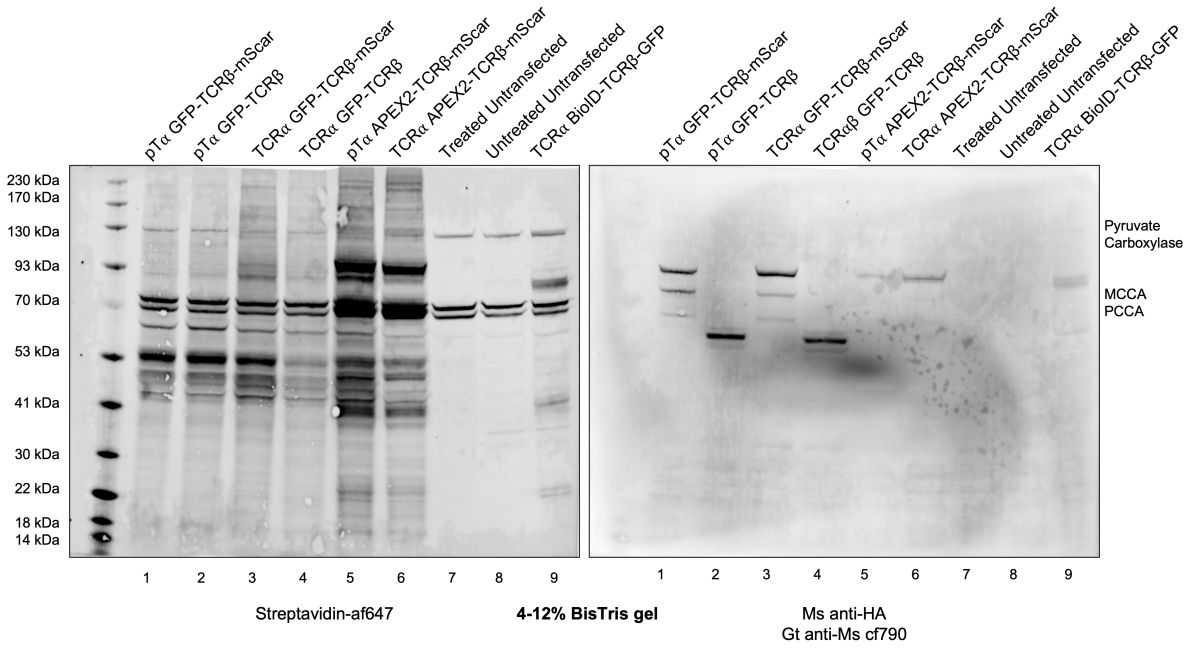

Figure 3.9: Biotinylation from the Nanobody-APEX assay analysed by western blotting.
HEK cells expressing GFP-tagged receptor constructs were incubated with Nanobody-APEX, washed then labelled using the protocol described in section 3.3.2. Cells expressing APEX2-tagged receptors were labelled without added Nb-APEX. Cells expressing the mTCR with a BioID2 domain were treated with 50µM biotin overnight. Lysate was analysed by western blotting. Bands consistent with three endogenously-biotinylated proteins are shown.

Whilst the assay seemed to be working as designed and producing more labelling that the BioID2-tagged mTCR, I was still keen to understand the dynamics of the assay to see if it could be optimised further. Biotin phenol is reported to be poorly membrane-permeable so I wondered if reducing the incubation time with biotin phenol might also help restrict labelling to the cell surface. To test this without using up stocks of Nanobody-APEX, I set up the assay with just the APEX2-fusion constructs (figure 3.10). Cells were incubated with biotin-phenol for 30, 15 or 5mins before the reaction was initiated with hydrogen peroxide or the chemical was added with the peroxide. Reducing the incubation times did not noticeably change the pattern of biotinylated bands but did decrease its intensity. The intensity decrease was more noticeable for the pTCR than the mTCR consistent with the APEX2-domain in the pTCR being largely localised the ER lumen and the mTCR having a larger proportion of the receptor at the cell surface. Five minutes of biotin phenol incubation seemed to produce good labelling. This experiment also confirmed that there was no background due to endogenous hydrogen peroxide production (lanes 1 and 7) or without biotin phenol addition (lanes 6 and 12).

I scaled up the assay to mass spectrometry, hoping to find a membrane protein labelled by proximity to the pre-TCR but not the mature TCR. A pre-TCR construct with a GFP-tag on the cytoplasmic side was used as a background control. The assay started with a T75 flask of HEK cells for each construct

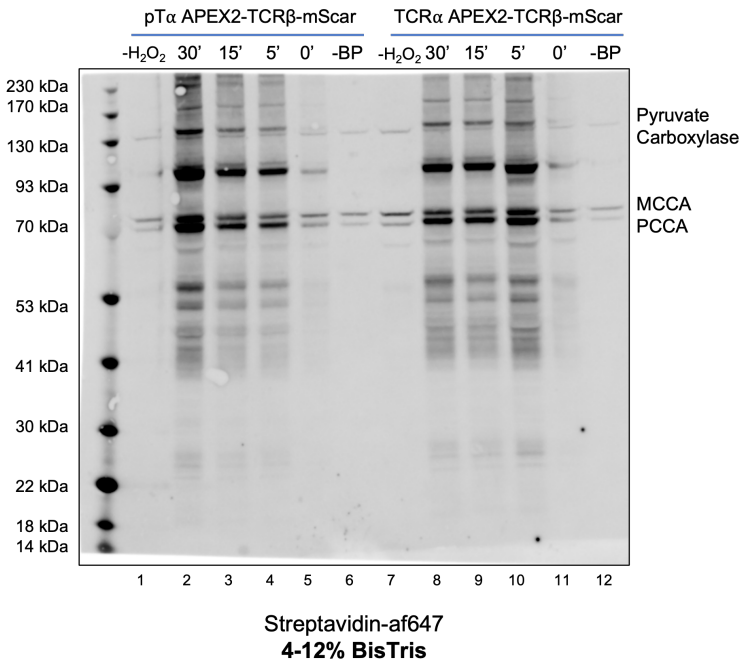

Figure 3.10: Biotin is poorly membrane-permeable. Little biotinylation is seen without addition of Biotin phenol or hydrogen peroxide.
HEK cells expressing APEX2-tagged pre-TCR and mTCR were suspended and incubated in media with biotin for the number of minutes indicated. Labelling was then initiated with hydrogen peroxide for 1 minute. The cells were quenched with three washes of Quench buffer then lysed. Lanes 1 and 7 contain samples incubated for 30 minutes with biotin phenol but no hydrogen peroxide was added. Lanes 6 and 12 were treated with hydrogen peroxide but no biotin phenol was added.

with proportional increase volumes of the other reagents. This used up the last of the Nanobody-APEX protein I had been sent. Biotinylated proteins were enriched from cell lysate with streptavidin beads then the beads washed five times with lysis buffer. Not knowing the size of the protein we were seeking, an on-bead tryptic digest was performed before mass spectrometry analysis using a protocol detailed in section 2.6.1.

The list of peptides identified in this experiment is given in supplementary section A.4.1. More proteins were identified in the GFP-mTCR sample, consistent with more intense labelling from the receptor with greater surface expression. Four peptides of HA-TCRβ-GFP were found in this sample along with two peptide of TCRα. Ten peptides from the Nanobody-APEX were also found. Two peptides from the Nanobody-APEX were also found in the pTCR-GFP control. No peptides from pTα or TCRβ were found in the GFP-pTCR sample and the CD3 chains were not found in any samples.

Unfortunately the results appeared to have been dominated by contaminants including keratins and histones. There were also peptides from the endogenously biotinylated proteins and from proteins that appear to be preferentially biotinylated in BioID control experiments such as Parp1, EEF1A1, TOP1, PRKDC, PKM, FLNA.[289] The high background suggested the beads needed more stringent washes.

I made plans to purify more Nanobody-APEX protein. This would involve transforming the pRSET_ nanobody-APEX-3xFLAG-6xHis plasmid into an *E. coli* strain positive for T7 RNA polymerase then purification via the His-tag. Plans to do this were interrupted by the COVID-19 lockdown and never resumed.

3.4.2 Proximity labelling of surface pre-TCR using SPPLAT

Before trying the Nanobody-APEX labelling assay again, I wanted to try the SPPLAT assay to see if labelling with a HRP-conjugated anti-TCRβ antibody gave better results. Whilst the antibody is larger than the nanobody, the faster kinetics of HRP compared to APEX might allow a higher proportion of nearby proteins to be labelled.

An unlabelled anti-TCRβ V5.1 antibody was conjugated to HRP then the assay tested on a small scale. Variants of the pre-TCR and mTCR with C-terminal mScar-tags were the main test samples and cells expressing these constructs without the antibody were used to show the background. The APEX2-fusion receptor constructs were labelled as positive controls. When the lysate was analysed via western blotting (figure 3.11), there appeared to be excellent biotin-labelling for both the pre-TCR and mTCR and these looked distinct from the APEX2-fusion constructs.

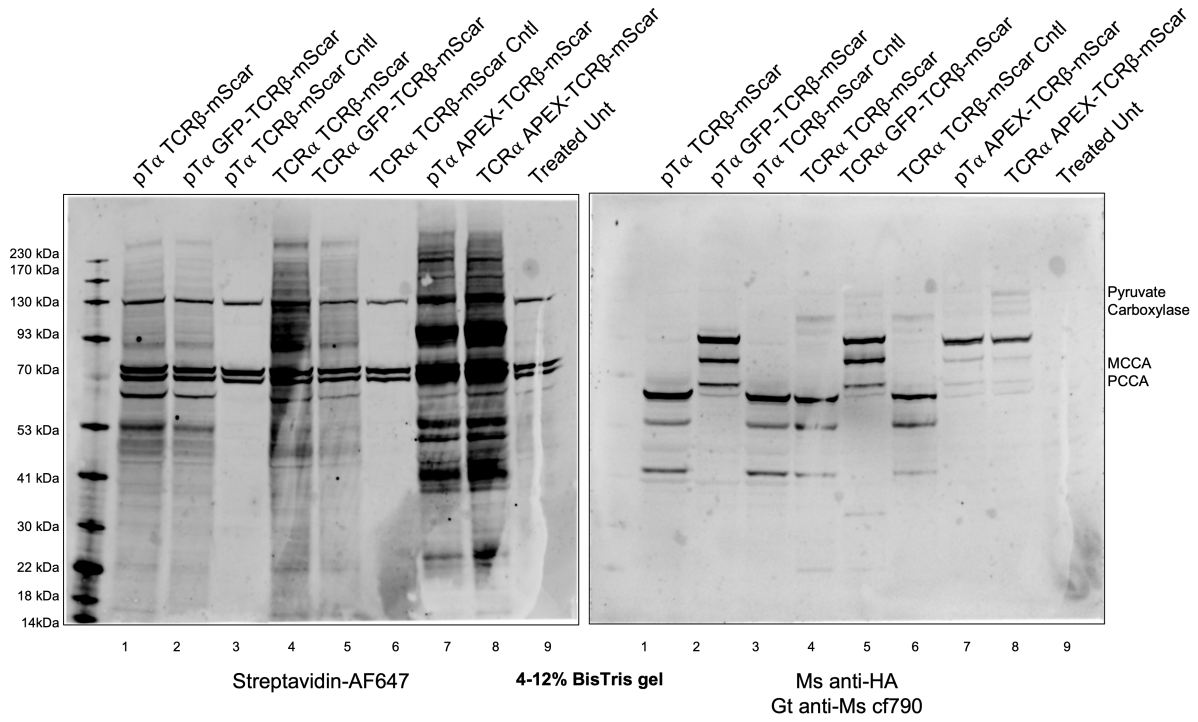

Figure 3.11: Biotinylation from the SPPLAT assay analysed by western blotting.
HEK cells expressing receptor constructs were incubated with HRP-conjugated anti-TCRβ antibody, washed then labelled using the protocol described in section 3.3.2. Cells expressing APEX2-tagged receptors were labelled without added HRP-conjugated antibody. No conjugated antibody was added to the Cntl samples. Lysate was analysed by western blotting.

The assay was scaled up for mass spectrometry starting from confluent T75 flasks of transfected cells. For a negative control I used a pre-TCR construct expressing the 1G4 TCRβ chain. This clone should not be recognised by the HRP-conjugated anti-TCRβ V5.1 antibody I was using. Labelling samples of the same cells with an APC-conjugated version of the same antibody showed this was the case (figure 3.12A). The cell lysate from SPPLAT-labelled samples was analysed by western blotting before and after incubation with Streptavidin-sepharose beads (figure 3.12B). This suggested labelling had occurred correctly though background biotinylation in the 1G4 pTCR lane was higher than desired and a few biotinylated proteins seemed to remain in the mTCR unbound fraction.

The streptavidin-sepharose beads were washed extensively using the protocol described in section 3.3.2 then processed with the on-bead tryptic digest protocol described in section 2.6.1. The list of peptides

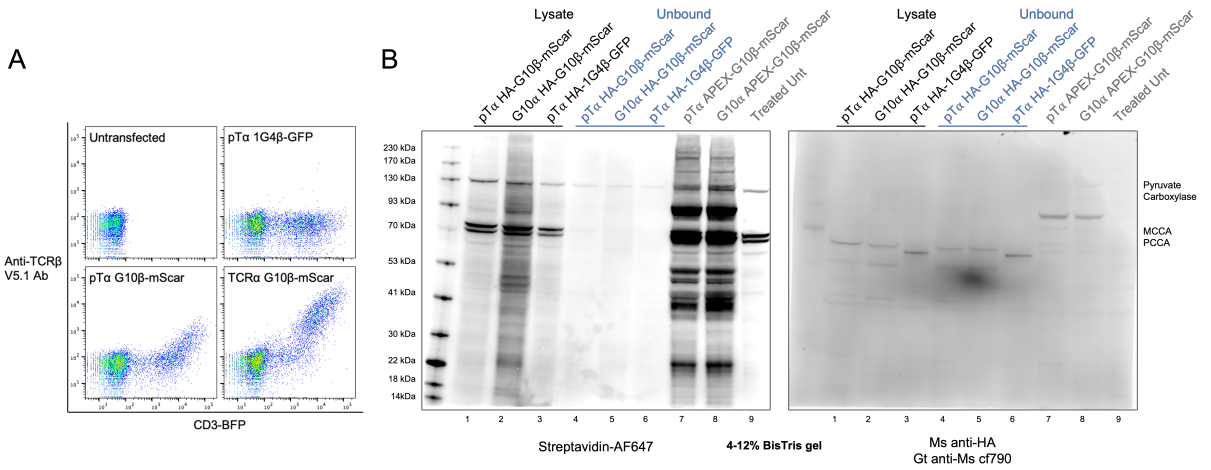

Figure 3.12: Samples from the SPPLAT assay taken to mass spectrometry analysis.
A: HEK cells expressing the G10 TCR, G10 pre-TCR and 1G4 pre-TCR were incubated with APC-conjugated anti-TCRβ V5.1 antibody (LC4 clone) in media for 30 minutes at 37°C then washed and analysed via flow cytometry.
B: The same cells labelled with HRP-conjugated anti-TCRβ V5.1 antibody (LC4 clone) in media for 30 minutes at 37°C then washed and labelled with biotin phenol and hydrogen peroxide. The lysate fraction and the unbound fraction (lysate after overnight incubation with beads) were analysed by western blotting. The last three lanes contain samples from previous experiments for comparison.

identified is presented in supplementary section A.4.2.

This experiment gave more encouraging results with the TCRβ chain as one of the top hits for both the G10 pTCR and mTCR. Peptides from CD3γ and CD3ϵ were also found in both samples whilst CD3γ was found in the mTCR sample only. Two peptides from TCRα were found in mTCR sample along with two peptides from NOTCH2, one from HLA-A and one from B_2-microglobin. This sample also identified one peptide of mouse Ig gamma chain which is likely from the anti-TCRβ antibody. One peptide from pTα was found in the pTCR sample. The most interesting proteins found in the pTCR sample but not in the mTCR sample or 1G4 pTCR control, were two 14-3-3 proteins (YWHAZ and YWHAB). Comparing the list of hits to the Contaminant Repository for Affinity Purifications database[290] suggests these are not uncommon contaminants.

Whilst these results were encouraging, the inability to detect all components of the pre-TCR bait in the pTCR sample suggested I needed to increase the input volume to be confident of capturing all nearby proteins. I scaled up the input by using three T75s for the G10 pTCR and 1G4 pTCR and used my remaining HRP-conjugated antibody. This attempt however did not result in an increase in proteins identified (data not shown).

3.4.3 Chemical crosslinking at the cell surface using BS3

Before I planned to revisit the proximity labelling assays, I wanted to try an alternative approach. This was to crosslink the pre-TCR to nearby proteins and see which proteins are pulled down with the GFP-tagged TCRβ chain. Crosslinking would enable more stringent washes so that contaminants could be removed without also loosing proteins from weak or transient interactions. Proteins can be crosslinked at the cell surface using bis(sulfosuccinimidyl)suberate (BS3) which has two amine-reactive, sulpho-NHS ester group separated by an 8-carbon linker. Unlike the similar disuccinimidyl suberate reagent, BS3 is water soluble but membrane-impermeable. This means that only receptors at the cell surface should be crosslinked.

Whilst crosslinking can be targeted to the cell surface, the GFP-pulldown would capture TCRβ chains throughout the lifetime of the receptor. For the pre-TCR, the majority of the receptor lifetime appears to be in the ER so these chains would take up the majority of sites on the beads leaving little room for crosslinked surface receptors. To see if I could enrich surface receptors in the pulldown, I would need a way to label the membrane fraction with GFP before crosslinking. I had made FKBP-TCRβ for a previous experiment and had a supply of purified FRB-GFP protein. FKBP and FRB domains can be dimerised with the addition of the drug rapamycin. I demonstrated rapamycin-dependent binding of FRB-GFP domain with kinetics similar to an anti-TCRβ antibody (supplementary figure S5).

HEK cells in six well plates were transfected with HA-GFP-tagged and FKBP-tagged receptor constructs. After 48h, cells were suspended and the cells expressing FKBP-tagged constructs were incubated with FRB-GFP with and without rapamycin. The tubes of cells were then washed to remove amide containing media, divided evenly then half crosslinked using the protocol described in section 3.3.3. Cell lysates were analysed by western blotting (figure 3.13).

As might be expected from the small fraction present at the cell surface, the lanes from cells expressing the pre-TCR (lanes 1-2 and 5-6) are not noticeably different with or without crosslinking. Lanes containing the mTCR meanwhile (lanes 3-4 and 7-8), have an obvious difference with BS3 treatment producing a smear of crosslinked protein at high molecular weight. This is visible in both the anti-GFP and anti-HA channels.

In the lanes containing the FKBP-tagged constructs incubated with FRB-GFP (lanes 9-14), there is a visible smear of crosslinked protein only for the mTCR with both rapamycin and BS3 treatment. In the two lanes where the receptor has been labelled with FRB-GFP in the presence of rapamycin but without crosslinking, there are two bands at ~35kDa. These are consistent with the mass of FRB-GFP, implying that the protein was bound and persisted through the wash steps but without crosslinking dissociated when the lysate was denatured with sample buffer. It is reasonable to assume that the FRB-GFP protein is bound to the FKBP-tagged pre-TCR in lane 9 but the smearing due to crosslinking has make the signal too spread out to be visible. The FKBP-tagged receptors lack the HA tag so are not visible in that channel.

Whilst BS3 crosslinking with this protocol seemed a viable strategy for finding membrane interaction of the TCR, it did not seem viable for the pre-TCR even if GFP-tagged proteins were enriched with GFP-trap beads. The limiting factor is that the crosslinking needs to occur in PBS rather than 37°C media, so secretory trafficking is paused. Under such conditions, the amount of crosslinked proteins is limited by the low surface expression of the pre-TCR. The two proximity labelling assays tested have the antibody binding occur under conditions where secretory trafficking can occur and the nanobody/antibody can accumulate. The biotin phenol labelling also acts as a form of signal amplification as one bound peroxidase can biotinylate multiple nearby proteins.

Given difficulties of identifying a candidate pre-TCR membrane interaction partner through proximity labelling, we reconsidered pre-TCR dimerisation as a hypothesis for determining pre-TCR internalisation. Dimerisation between pre-TCR molecules would not have been identified by the proximity labelling assays I had performed so far as the TCRβ chain was used as the bait so peptides from this chain would be expected in the list of hits. Other members of the James lab were working on an assay that would address this issue: crosslinking with non-canonical amino acids (ncAAs). This involved replacing a residue in the TCRβ sequence with an amber TAG stop codon then transfecting with an artificial aminoacyl-tRNA synthetase. These enzyme can combine a tRNA with a photocrosslinker amino acid analog that is supplemented into the medium. This amino acid then incorporates into the growing TCRβ chain at the

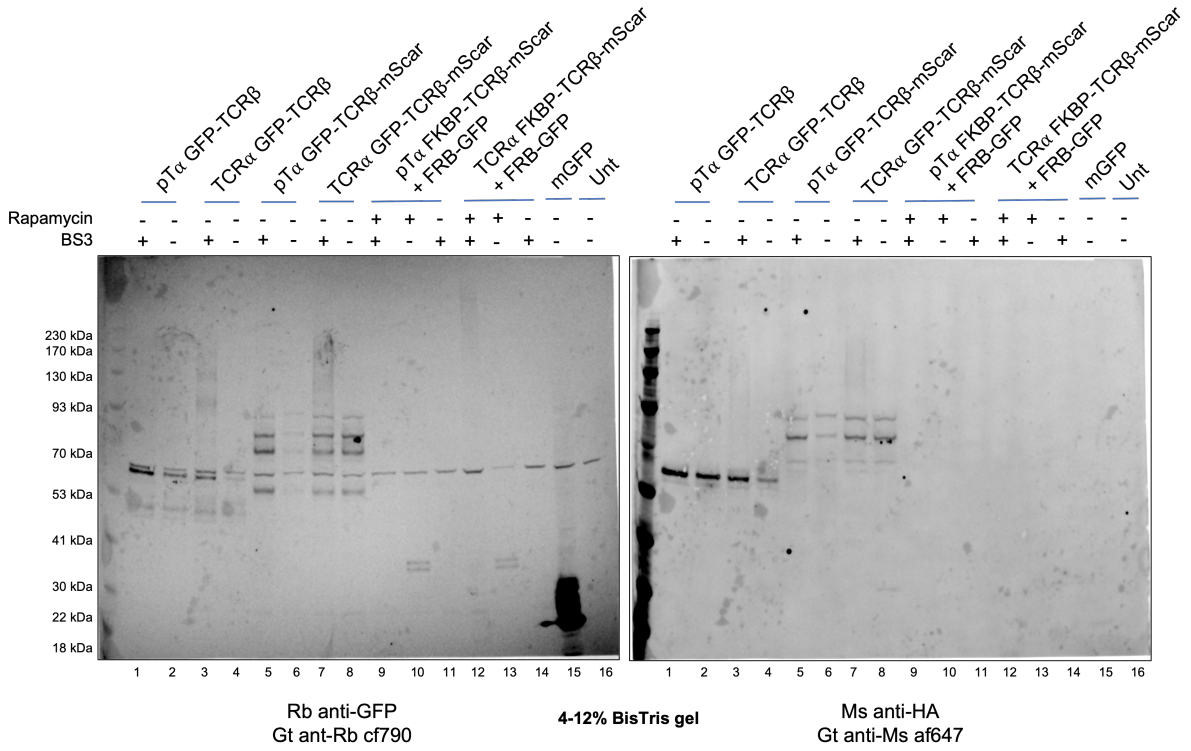

Figure 3.13: Crosslinked TCR can be detected using membrane-impermeable BS3 but crosslinked pre-TCR is not detectable.
Two wells of 6 well plates were transfected for each of the indicated GFP-tagged receptors. These were suspended, washed four times in PBS then split evenly. These were incubated for 1h on ice in pH8 PBS ±1mM BS3. The reactions were quenched with Tris-HCl then lysed as normal.
Cells expressing the FKBP-tagged receptors were incubated for 30mins in media with 1:100 FRB-GFP ±100nM rapamycin. Crosslinking was performed as before.

site of the stop codon. Once incorporated UV light can be used to trigger crosslinking of the residue to neighbouring proteins. This technique was previously used to map the interaction sites between TCRα and TCRβ[291] and we wondered whether it would work for the pre-TCR. If two different TCRβ chains were co-expressed, one with the ncAA photocrosslinker and one without, then crosslinking between them would allow the two chains to be co-immunoprecipitated.

Multiple different sites in TCRβ were tested along with different photocrosslinking residues and crosslinking conditions. Despite a lot of effort we were unable to replicate the photocrosslinking of TCRβ to TCRα in the $\alpha\beta$TCR complex making use with the pre-TCR unfeasible. The main issues were low expression, readthrough of the stop codon, low crosslinking efficiency and poor solubility of the crosslinkers.

I thought these issues might be circumvented by using BS3 to chemically crosslink TCRβ chains. Previously it appeared that insufficient pre-TCR could crosslinked with this method to cause detectable smearing in cell lysate. However if two TCRβ chains were co-expressed, one with a GFP tag and the other with a HA-tag, then crosslinking with BS3 would allow both to be pulled down with GFP-trap beads. I would then just have to detect the HA-tag in the bead elution. I thought an increase in the input volume, combined with bead enrichment in concentration, might allow the crosslinking to be detected.

T75 flasks of HEK were transfected with CD3$\gamma\delta\epsilon\zeta$ and two different plasmids encoding the pre-TCR or $\alpha\beta$TCR, one with a GFP-tagged TCRβ chain and one with a TCRβ chain with HA and mScar tags. A construct with a HA-tagged pTα chain and GFP-tagged TCRβ chain was used a positive control. The cells were suspended using trypsin then half of each sample crosslinked with BS3. GFP-tagged proteins were enriched with GFP-trap agarose beads, the beads washed extensively then proteins eluted from the beads by heat in sample buffer. Eluted proteins were analysed by western blotting (figure 3.14).

The blot was probed with mouse anti-HA and rabbit anti-GFP antibodies, washed then probed with goat anti-mouse AF647 and goat anti-rabbit CF790 antibodies. In lane 1 containing lysate from cells expressing HA-pTα TCRβ-GFP that was chemically crosslinked there was faint smearing in the CF770 channel at high molecular weight smearing consistent with non-specific crosslinking of GFP-tagged TCRβ chains. This smearing was also observed in lane 5 containing cells expressing the mTCR with BS3 crosslinking but lane 3 containing cells expressing the pre-TCR. Surprisingly however HA-tagged TCRβ chains appeared in the GFP-pulldown with or without crosslinking for both the pre-TCR and TCR. Comparing the band intensities suggested the interaction between TCRβ chains was comparable to the co-immunoprecipitation of HA-pTα and TCRβ.

3.4.4 Co-immunoprecipitation of TCRβ chains

Normally the co-immunoprecipitation of two proteins would be convincing evidence for an interaction between them but here the amount of TCRβ chains that could be pulled down seemed discordant with the amount of pre-TCR complex that is detectable at the surface. I suspected this instead represented dimerisation of TCRβ chains in the ER. I replicated this experiment without the crosslinking steps and a smaller input volume (figure 3.15).

This experiment used a different set of constructs: firstly to confirm that HA-TCRβ-mScar chains were not pulled down in the absence of TCRβ-GFP (lane 1) and no HA-positive bands are visible without co-expression of a HA-tagged TCRβ chain (lane 2). In this run HA-TCRα appeared to efficient precipitate with TCRβ-GFP (lane 3) whilst a band for HA-pTα was only faintly visible at around 30kDa when the gel was overcontrasted. HA-TCRβ was readily pulled down when co-expressed with

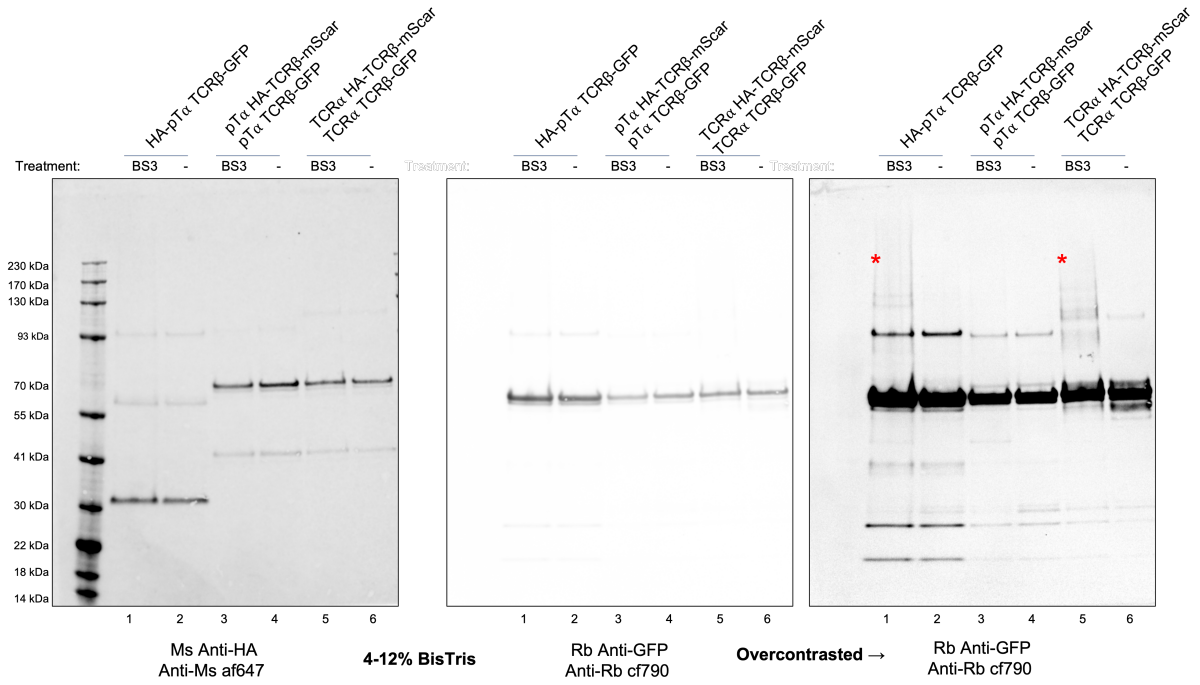

Figure 3.14: TCRβ chains co-immunoprecipitate without chemical crosslinking.
T75 flasks of HEK were transfected with the indicated constructs then half were crosslinked with BS3 using the protocol described in methods section 3.3.3. After lysis, GFP-tagged proteins were enriched by incubation with GFP-trap agarose beads for 1h at 4°C with rocking. Beads were washed five times with 1ml of GTA wash buffer then proteins were eluted with heat in sample buffer. High molecular weight smearing of GFP-tagged TCRβ chains in two lanes in highlighted by red asterisks.

TCRβ-GFP and pTα but not as efficiently when expressed with TCRβ-GFP and TCRα. Although this gel was run under reducing conditions with DTT in the sample buffer, there is high molecular weight bands around 100kDa in lanes 2 and 5, likely disulphide-linked protein dimers.

This data suggests that TCRβ chains have some level of affinity for each other and that TCRα and pTα chains can disrupt this structure to form the proper TCR or pre-TCR complex. TCR complex assembly is not a trivial process, requiring eight different subunits to come together to neutralise the charged residues present in the transmembrane domains.[292] The TCRα transmembrane helix is poorly hydrophobic and can be transported entirely into the ER lumen during translation if not paired with TCRβ. The constant domain of TCRα seems to require pairing with constant domain TCRβ in order to fold and in a process monitored by the chaperones calnexin and BiP.[293] Chains that fail to fold and assemble are quickly degraded by the ER-associated protein degradation (ERAD) pathway. It seems reasonable that this process is imperfect when the proteins are overexpressed in HEK. Perhaps TCRβ chains are looking to pair with any similar Ig-like domains and pair with other TCRβ chains when TCRα or pTα are not available.

I wanted to investigate the co-immunoprecipitation of TCRβ chains further to verify it was due to defects in complex assembly rather than dimerisation of functional pre-TCR complexes. I set up the experiment again including mixes with and without the CD3 chains of the complex. I also wanted to demonstrate that any interaction between the GFP and mScar fluorescence protein tags was not causing the effect so included samples expressing HA-TCRβ chains without a fluorophore attached.

To better match expression levels, best practice would be to replace the CD3 chains with another similar plasmid. I tried this in some attempts at this experiment (using pHR_DRB7+CLIP, another polycistronic vector), however after struggling with transfection efficiency, I decided to simplify the mixes

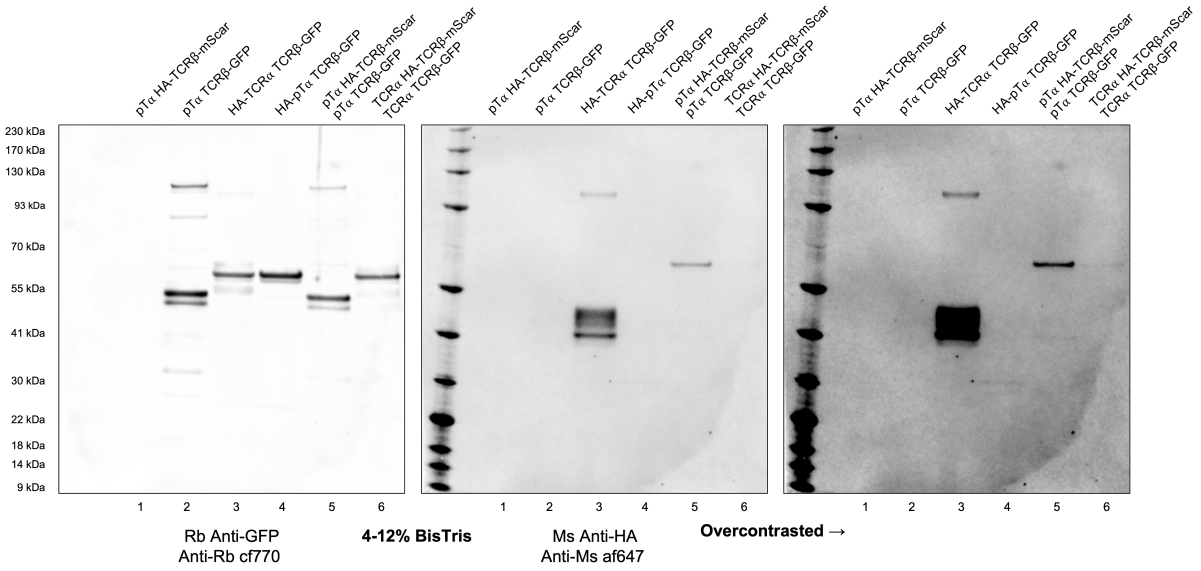

Figure 3.15: Co-immunoprecipitation of TCRβ chains.
T25 flasks of HEK were transfected with the indicated constructs with CD3$\gamma\delta\epsilon\zeta$-BFP making up half of the transfection mix. After 48h, cells were lysed and GFP-tagged proteins were enriched by incubation with GFP-trap agarose beads for 1h at 4°C with rocking. Beads were washed five times with 1ml of GTA wash buffer then proteins were eluted with heat in sample buffer.

as much as possible and just adjusted the volumes with water. I also wanted to be able normalise the expression levels of the constructs somehow. The simplest way to do this seemed to be to run the cell lysate on a western blot with the bead elutions. I could then compare the enrichment of HA-tagged proteins between the lysate to bead elution relative to the enrichment of GFP. I tried running the lysates and bead elutions together on the same gel but had issues with the gel distorting, likely due to difference in the amount of salts and sample buffer between the two types of sample. These are instead run as two separate gels in parallel.

The co-immunoprecipitation experiment with these modifications is shown in figure 3.16. Lanes 1 and 2 of the blots contain samples where TCRβ chains were expressed without any other components of the pTCR or TCR complex. Whilst the TCRβ chain is not especially intense in the cell lysate (top) it appears to have been enriched in the bead elution. HA-TCRβ-mScar and HA-TCRβ are also present in the bead elution showing that the co-immunoprecipitation of TCRβ chains does not require the full receptor complexes. Apart from the HA-TCRα TCRβ-GFP control, the other lanes appear mostly empty in the anti-HA channel though there is a faint band in the co-expressed pTCR complexes. The signal appeared too low to use band intensity to reliable quantify enrichment from lysate to bead elutions.

Having established that the co-immunoprecipitation of TCRβ chains did not require the full complex, I was curious as to which domains of the receptor were important. I set the experiment up again using HA and GFP-tagged versions of the receptor with modified extracellular domains (depicted in figure 3.6). I also swapped the wavelengths of the secondary antibodies so I could use of the better signal:background of the cf770 channel for the more interesting HA-staining. This experiment is shown in figure 3.17.

Comparing lanes 1 and 2, which should be identical except for the Vβ domains, suggests that TCRβ chains can be efficiently pulled down together with just the Cβ domains. This implies that these domains are most important for this interaction. Lanes 3-6 contain the components of the pre-TCR and mature TCR with and without CD3 chains. If the blot for the bead elutions is overcontrasted then faint band is visible at the same size as the HA-TCRβ-mScar band in lane 1. The band in lane 5 with TCRα but

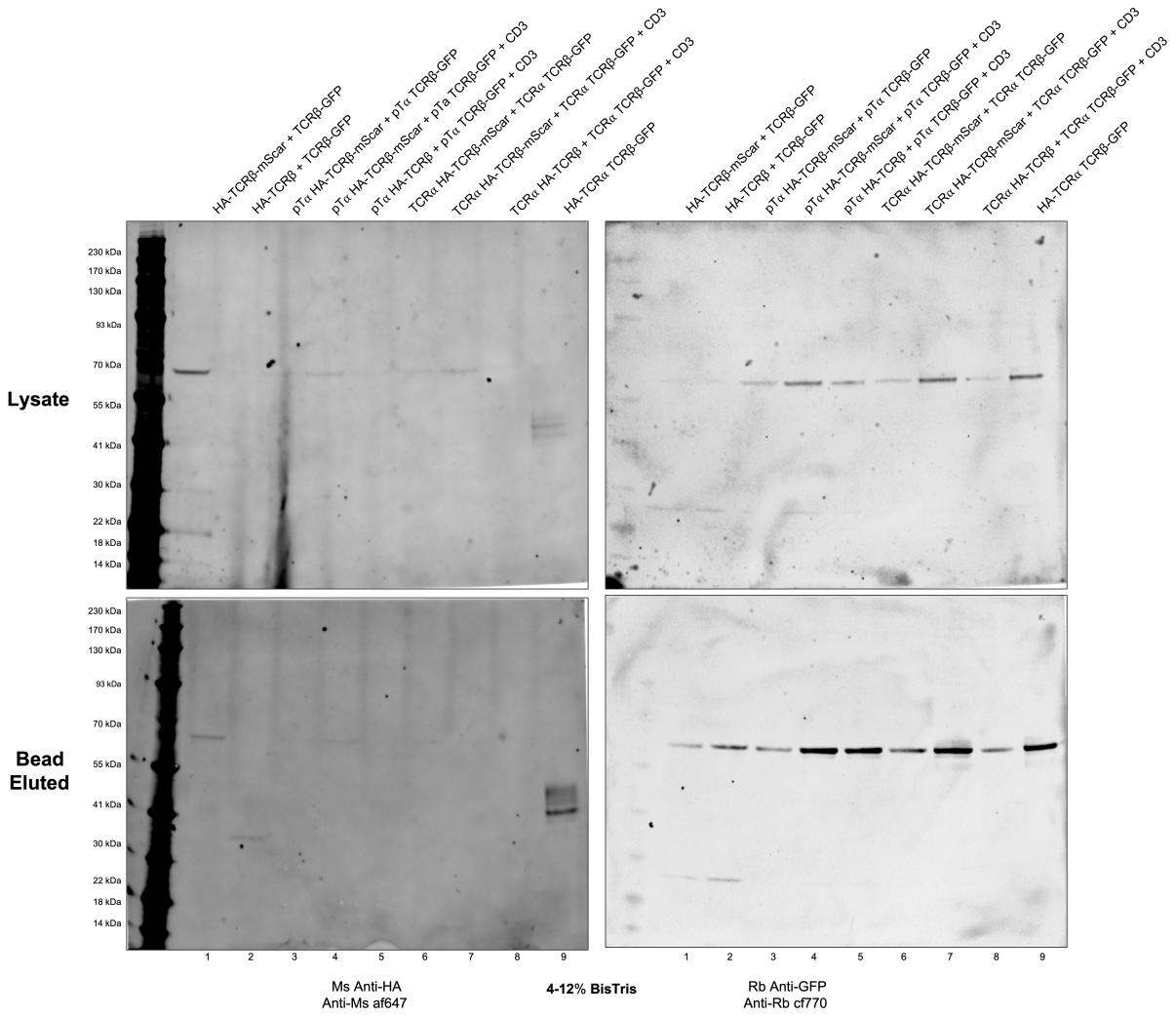

Figure 3.16: Co-immunoprecipitation of TCRβ chains does not require pTα, TCRα or CD3 chains.
T25 flasks of HEK were transfected with the indicated constructs. Samples marked with CD3 were transfected with pHR_CD3$\gamma\delta\epsilon\zeta$ making up half the transfection mix. In the rest this volume is replaced with extra water. After 48h, cells were lysed and GFP-tagged proteins were enriched by incubation with GFP-trap agarose beads for 1h at 4°C with rocking. Beads were washed five times with 1ml of GTA wash buffer then proteins were eluted with heat in sample buffer. Lysate (top) and bead elutions (bottom) were analysed by western blotting.

without CD3 is slightly brighter whilst the others are approximately equal. The lane expressing pTα paired with the constant domains of TCRβ (lane 7) has a weaker HA^{+ve} band in the bead elution than lane 2. Comparing the GFP bands for these lanes suggests the difference might be due to expression rather than pTα successfully separating Cβ domains.

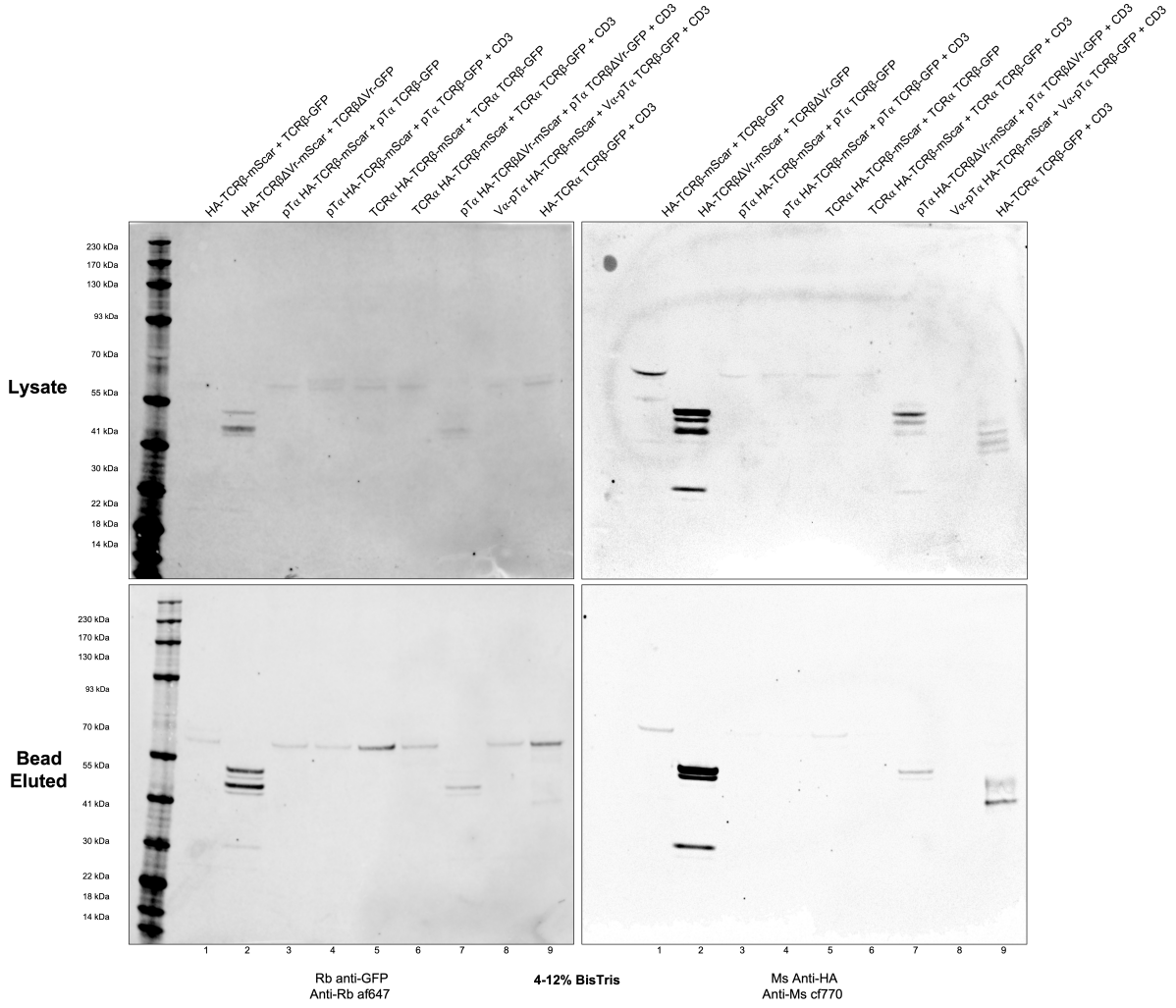

Figure 3.17: TCRβ dimerisation is mediated by the Cβ domains.
T25 flasks of HEK were transfected with the indicated constructs. After 48h, cells were lysed and GFP-tagged proteins were enriched by incubation with GFP-trap agarose beads for 1h at 4°C with rocking. Beads were washed five times with 1ml of GTA wash buffer then proteins were eluted with heat in sample buffer. Lysate (top) and bead elutions (bottom) were analysed by western blotting.

Taken together these blots suggest that HEK have difficulty assembling the pre-TCR and TCR complexes due to affinity between Cβ domains. The pTα and TCRα chains can disrupt this interaction especially with co-expression of the CD3 chains. It's possible this is an artefact of overexpression but should be considered when interpreting other experiments with HEK. Perhaps a cell line with the correct ER chaperones for TCR complex assembly such as Jurkat T cells might not display the same effect.

There did appear to be a lot of variation between experiments in the proportion of HA-tagged TCRβ chain pulled down with TCRβ-GFP. I expect that some of this is due to different expression levels of the two plasmids. Whilst two proteins on the same plasmids should be expressed in 1:1 stoichiometry due to the P2A sequence, the two plasmids may not be inherited evenly so there could be variation in the amounts of each TCRβ chain between cells. Experiments using transient transfections tend to be dominated by the highest expressing cells where this imbalance would be most significant. A contributing

factor, discovered after these experiments were concluded, is that the 2A sequences might be disrupting the endogenous signal peptide of TCRβ. This signal peptide was used in all constructs without the HA tag (which instead used the gaussia luciferase signal peptide). When the 2A sequence is cleaved it leaves a proline at the N-terminus of the new protein which seemed to disrupt recognition of the endogenous signal peptide of TCRβ but not affect the gaussia luciferase signal peptide. This effect might amplify the difference in expression levels though it did not seem to entirely prevent formation of pre-TCR or TCR complexes without the HA tag.

3.4.5 Dimerisation investigated with split fluorophore assays

Whilst dimerisation appears to occur between TCRβ chains during assembly in the ER, this does not necessarily mean dimerisation is uninvolved in pre-TCR internalisation at the cell surface. We therefore wanted to investigate pre-TCR dimerisation through other techniques.

I repurposed the split GFP assay, discussed in more detail in section 4.4.6 of the next chapter, to investigate receptor dimerisation. This assay has two components, GFP11 and GFP(1-10), that are non-fluorescence alone but produce GFP fluorescence when complemented. These parts were attached to the N-terminus of TCRβ then expressed with other components of the pre-TCR or TCR (figure 3.18). This assay found strong GFP signal when TCRβ chains with GFP11 and GFP(1-10) tags were co-expressed alone. There was less signal with co-expression of pTα or TCRα. Controls with the components expressed individually as well as a control with both the two tags on opposite sides of the ER membrane found no fluorescence. Microscopy images of the four samples expressed on the top row is shown in supplementary figure S6. Little GFP fluorescence was apparent for either the pre-TCR or mTCR receptors but fluorescence consistent with ER-dimerisation was seen for the TCRβ chains alone.

Whilst this experiment also implied a level of affinity between TCRβ chains, it is severely limited by the high affinity of the GFP11 and GFP(1-10) components. This means that, once complemented, the two chains are permanently stuck together. Whilst this might be advantageous in capturing transient interactions, it is also likely to interfere with normal complex assembly and receptor trafficking. A better assay would use split domains with lower affinity for each other but still produce a signal when fused to two proteins that oligomerise.

Two such assays were then performed by John James. The first assay used fused the two components (SmBiT and LgBiT) of a split luciferase enzyme (SmBiT and LgBiT) to the N-terminus of TCRβ chains expressed as part of the pre-TCR or mTCR. Complementation between the two components produces luminescence. Only a weak signal was observed for the pre-TCR and TCR compared to the positive control with the components fused to TCRα and TCRβ (figure 3.19A). A similar assay used two halves of the Venus fluorescent protein (VN and VC) also fused to the extracellular domains of TCRβ. These domains have weak affinity for each other but produce fluorescent when combined. There was only faint Venus fluorescence observed when the two TCRβ chains were expressed as part of the pre-TCR or mTCR. The combined Venus protein was not detectable at the cell surface with the anti-GFP nanobody (figure 3.19B).

This data supports the pre-TCR and TCR being monomeric when expressed in HEK cells. This is not necessarily inconsistent with the co-immunoprecipitation of different TCRβ chains as these experiments were performed with pTα, TCRα and CD3 components of the receptors which appear to efficiently separate TCRβ chains.

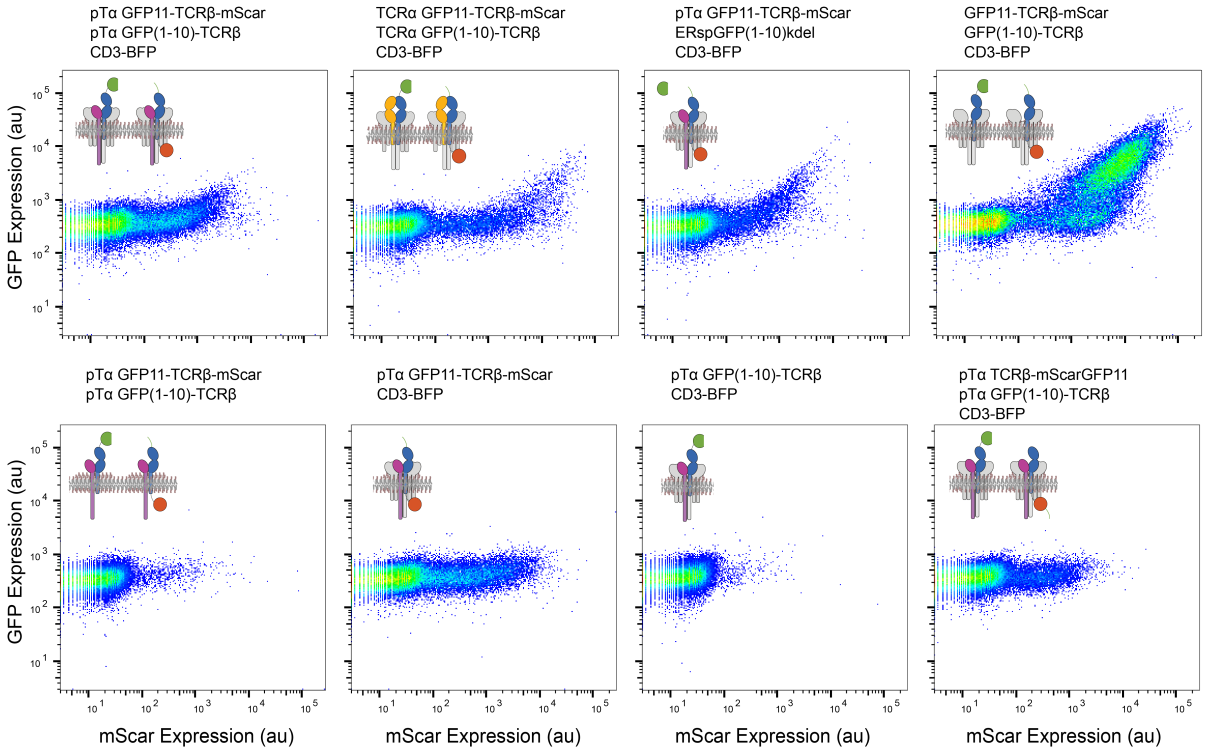

Figure 3.18: A split GFP assay implies affinity between TCRβ chains.
Components of the TCR and pre-TCR complex were expressed with GFP11 and GFP(1-10) domains on the N-terminus of TCRβ. Cells were analysed by flow cytometry. Inserts show the structures of these constructs with the TCRβ chain shown in blue, GFP(1-10) and GFP11 both shown in green and mScar in red. TCRα and pTα are shown in yellow and purple respectively. CD3 chains are show in grey.

I designed one further experiment to investigating pre-TCR dimerisation as a potential mechanism of pre-TCR internalisation. As the pre-TCR constructs missing the Vβ domains had high steady state surface expression, we wondered if two of these receptor could be artificially dimerised through FKBP/FRB domains to mimic the pre-TCR dimer structures shown in figure 3.3. I fused a HA-tagged FKBP domain and the FRB domains onto Cβ in place of the Vβ domains and expressed these constructs in HEK. I then used B/B homodimeriser and A/C heterodimeriser rapalogs to hopefully trigger dimerisation before measuring surface expression with an anti-HA antibody. The A/C homodimeriser seemed to have no effect on surface expression of the constructs whilst the B/B homodimeriser appear to increase the stability of the complex at the cell surface, opposite to expectations (supplementary figure S7). The main issue with this experiment is the length of the linkers between the Cβ domains and the FKBP/FRB domains. If these linkers were too long then the complex might be two flexible for dimerisation to mediate internalisation but if they were too short the dimeric structure might not form. Whilst I could experiment with linkers of different lengths, this was not considered a good use of time.

Experiments performed by John James and Maryam Obaid investigating the fate of the pre-TCR after internalisation concluded around this time in the project. This experiment tracked the pre-TCR through the endocytic pathway using an af647-conjugated anti-GFP nanobody added to the medium and internalised by affinity to GFP-tagged pre-TCR and TCR constructs I had cloned. By measuring the colocalisation of the nanobody foci with the Rab5, Rab7 and LAMP1 markers, vesicles containing the pre-TCR appeared to be directed to the lysosomes for degradation. Internalised vesicles of the $\alpha\beta$TCR appeared to be recycled back to the cell surface (figure 3.19C). This would seem to replicate the behaviour observed in thymocytes but in the absence of signalling kinases.

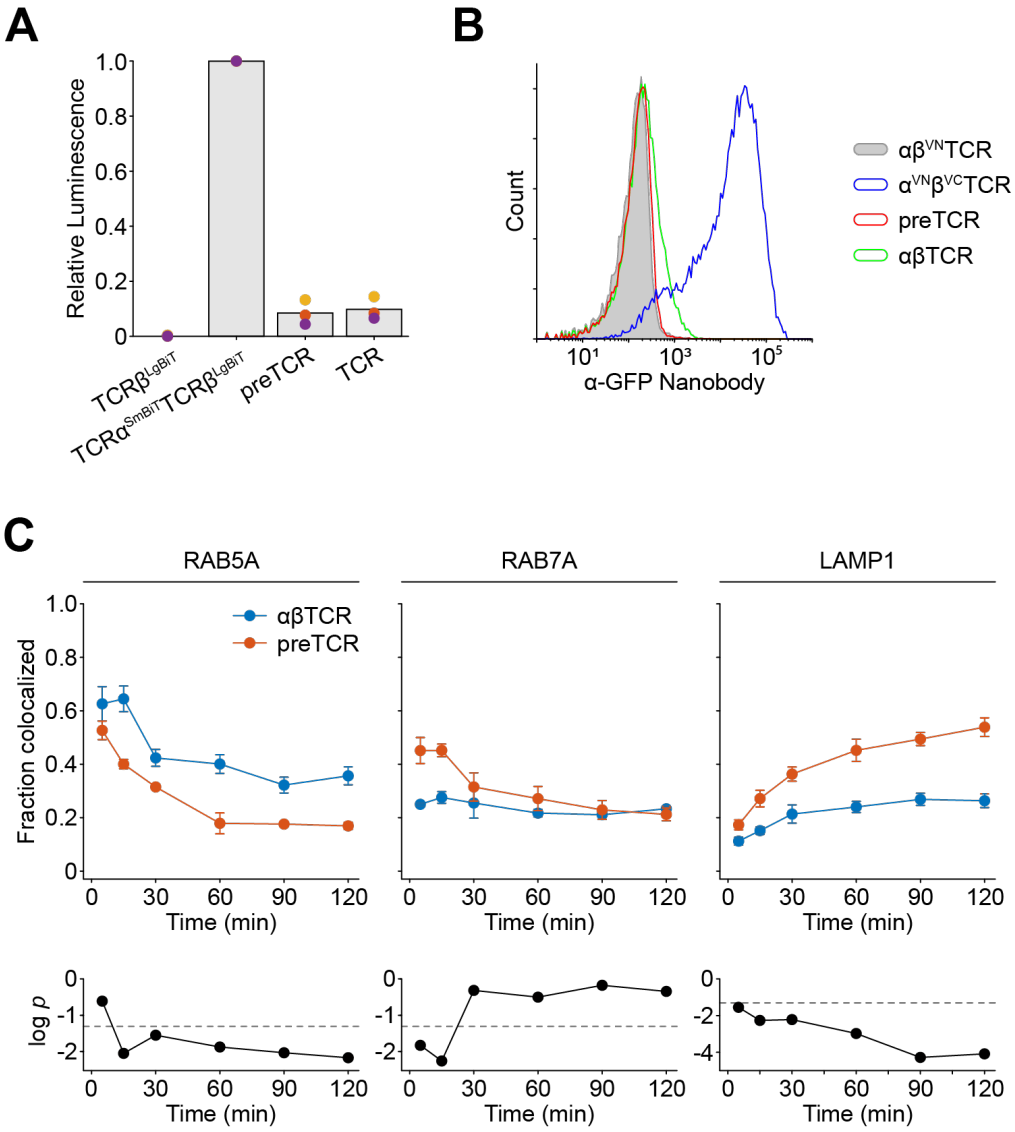

Figure 3.19: The pre-TCR is monomeric at the cell surface and is constitutively trafficked to lysosomes.
A: Luminescence from a NanoBiT split-luciferase assay with the component "BiTs" localised to the outside of the cell. The data is shown relative to the mTCR positive control with three biological replicates.
B: A split Venus assay does not show Venus fluorescence when VC/VN components are attached to the pre-TCR (data not shown). Reconstituted Venus is not detectable at the cell surface with an anti-GFP nanobody when both components are fused to the pre-TCR.
C: AlexaFluor647-conjugated anti-GFP nanobody was used to track the pTCR and mTCR through the endosomal pathway. Colocalisation of internalised nanobody with early endosomes (RAB5A), late endosomes (RAB7A) and lysosomes (LAMP1) was measured over time. Upper panels show mean ± SEM (n=3 or 6) and lower panels show a t test statistic comparing mTCR and preTCR datasets. Dashed lines indicate p=0.05.
Experiments performed by John James and Maryam Obaid.

3.4.6 The pre-TCR and mTCR are capable of being secreted at similar rates

The above experiments made us pause and reconsider how pre-TCR and $\alpha\beta$TCR surface expression could be distinguished. Our data thus far indicated the pre-TCR has a low steady state surface expression and is predominantly localised to the ER. The $\alpha\beta$TCR meanwhile appears to have a higher fraction present at the cell surface. The surface expression level of any receptor is determined by three factors: the rate of secretion towards the surface, the rate of internalisation away from the surface and the rate of receptor recycling. For these receptor complexes, the rate of secretion includes the rate of complex assembly as well as the rate of trafficking from the ER to the surface via the Golgi.

I wanted to design an experiment to investigate the rate of receptor secretion and see if this differed between the pre-TCR and $\alpha\beta$TCR. If the rates were different then it would implicate secretion as a contributing factor in determining surface expression. If the rate of secretion were identical then it would suggest that receptor internalisation and recycling are more important determinants for the surface expression.

Investigating protein secretion is technically challenging as it is a continuous process and imaging the cells at any point would show molecules at all stages in the secretory and endocytic pathways. To isolate the rate of a single process, trafficking can be synchronised so that all the molecules undergo secretion at the same time. Synchronisation can be performed through reducing the temperature or through fusing the protein to conditional aggregation domains but these techniques have wider impacts on cellular metabolism. A better assay, called the retention using selective hooks (RUSH) assay, uses the affinity between a streptavidin-binding peptide (SBP) tagged to the reporter protein of interest to a streptavidin-kdel "hook".[294] This hook is retained in the ER lumen by kdel receptors with the reporter attached. Biotin can be added to the medium which outcompetes the SBP and initiates the synchronised trafficking of the protein. A fluorescent protein tag on the reporter protein can then be used to track the protein as it is secreted. As only tagged proteins are retained, the assay has minimal effects on other secreted proteins.

I made several changes to the RUSH assay, as it was originally described, to be better adapt it for our purpose (shown in figure 3.20). The assay expressed both the streptavidin hook and SBP-tagged reporter from the same plasmid, separated by an IRES sequence. As the TCR plasmids were already quite large, I decided to express the hook from a separate plasmid and monitor its expression through a BFP protein tag. This BFP was used to mark the ER. The cell surface was marked with mCherry-CaaX. I designed my pre-TCR and $\alpha\beta$TCR reporters with SBP-GFP on the N-terminus of TCRβ. As the SBP-tag is separated from the rest of the complex by the GFP domain any steric effects should be identical between the two receptors. The extracellular GFP also allows the arrival of the receptor at the cell surface to be detected with high affinity af647-conjugated anti-GFP nanobodies.

I briefly tested this assay using flow cytometry but encountered a few issues. I was unable to suspended the cells before adding biotin and take out aliquots over time as suspension impeded receptor secretion. If biotin was added when the receptor was adherent, it would be too difficult to resuspend, wash and fix cells with the time-resolution I wanted. I therefore decided to approach the assay through imaging. Using 100X magnification prevented imaging many cells at the same time as the field of view was very narrow and I couldn't move between positions quickly due to lagging of the immersion oil. Cells also appeared to be bleached and photodamaged over the timecourse. Imaging at 20X allowed a lot more cells to be imaged but it proved difficult to distinguish intracellular and membrane receptors. Imaging at

40X was a compromise as the air objective allowed quick movements with reasonable resolution.

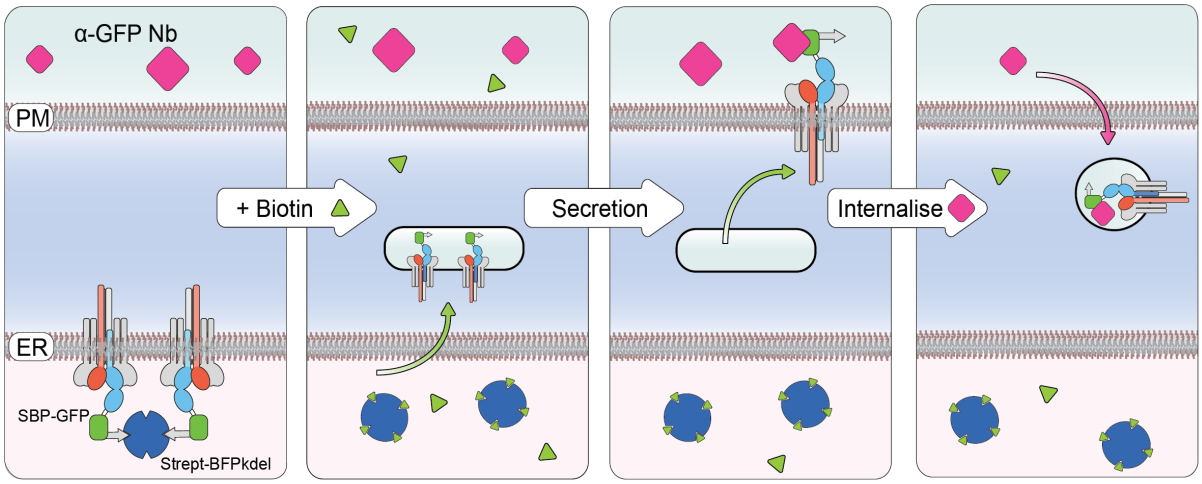

Figure 3.20: The components of the Retention Using Selective Hooks (RUSH) assay to investigate rate of pre-TCR secretory trafficking.
SBP-tagged proteins are retained in the ER lumen (light pink) due to affinity to a Streptavidin-kdel "hook" (dark blue). Addition of biotin (green triangles) initiates the synchronised secretion to the cell surface. This assay was adapted by using Streptavidin-BFPkdel to both retain the receptor and act as marker for the ER. SBP-GFP was fused to the N-terminus of TCRβ (cyan) and af647-conjugated anti-GFP nanobodies (pink squares) were added to the external medium. The rate of internalisation of the nanobody was recorded over time via microscopy. The pTα chain is shown in orange and CD3 chains in grey.
Adapted from a figure by John James.

The final protocol used for this experiment is described in section 3.3.4. This protocol enabled four wells to be imaged simultaneously, two expressing the pre-TCR and two expressing the $\alpha\beta$TCR. After finding good positions in each slide, the nanobody was added to the media and allowed ten minutes to diffuse evenly. Imaging constructs without the SBP-tag or without the streptavidin hook showed this time was sufficient for maximal binding in the absence of ER-retention. Synchronised secretion was then initiated in two wells through addition of biotin. Positions were imaged every 2 minutes for the next 90 minutes. The protocol was repeated for a total of six technical replicates of each condition with 9 images in each set.

Representative images from this experiment are shown in figure 3.21. At the start of the experiment the GFP-tagged receptors appear localised inside of the cell, overlapping with the Streptavidin-BFPkdel hook. The streptavidin appears to be distributed in sharp punctae. As streptavidin forms a tetramer, aggregation of this protein was not unexpected. I suspect the protein had a similar distribution in the original protocol which lacked the BFP-tag to visualise the hook. After addition of biotin, the GFP-tagged receptors and streptavidin separate. This is most obvious in the frame of the TCR at 40 minutes where the receptor clusters near the centre of multiple cells in the image. I suspect these clusters are the Golgi though I lack a marker for this organelle. The GFP-tagged receptor is less visible after 40 minutes. This is likely due to GFP-fluorescence being quenched by nanobody-binding rather than photobleaching. For both receptors, the system was slightly leaky with faint surface binding without added biotin. This might be due to trace biotin in the growth media.

Images like the ones shown in figure 3.21 were analysed using MATLAB. The first analysis I wanted was to monitor the uptake of the nanobody which should indicate the arrival the receptor at the cell surface and its internalisation. I tried detecting intracellular foci using the mCherryCaaX to indicate cell boundaries. However the marker proved too inconsistent for an algorithm to find neat cells outlines. Instead I recorded the number of distinct particles within the nanobody channel that exceeded a certain intensity threshold. Comparing computer-identified foci to the images through a visual overlay suggested

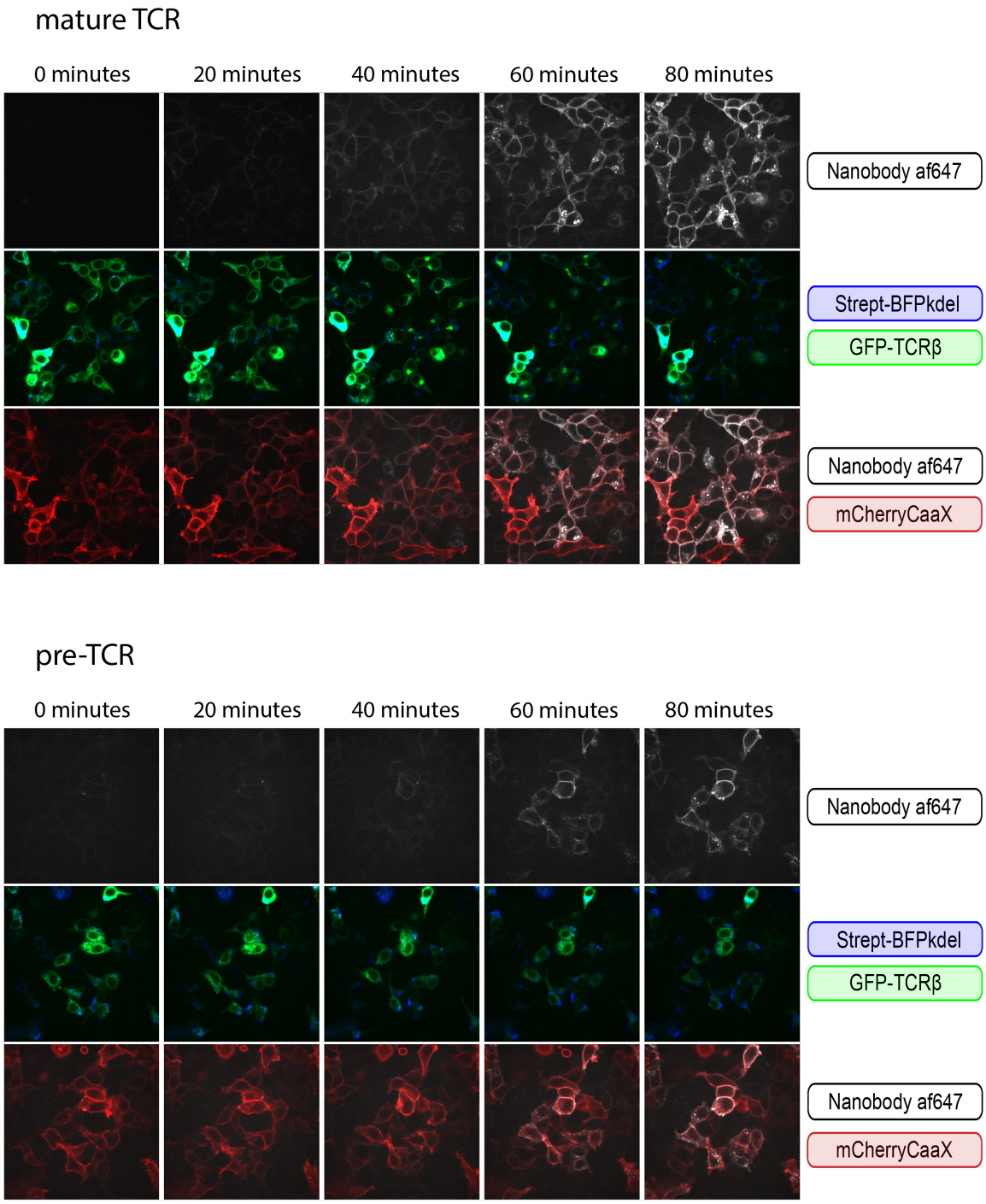

Figure 3.21: Secretory trafficking of the pre-TCR and mTCR can be synchronised using the RUSH assay. HEK cells were transfected with equal amounts of SBP-GFP-tagged pre-TCR or mTCR, CD3$\gamma\delta\epsilon\zeta$ and Streptavidin-BFPkdel. A smaller amount of mCherryCaaX was used to mark the cell boundaries. Anti-GFP nanobody af647 was added to the external medium and allowed to diffuse for 10minutes. Biotin was added at time 0 then cells imaged under 40X magnification every 2 minutes.

the chosen threshold value was appropriate. The number of nanobody particles found over time is plotted in figure 3.22.

For both the pre-TCR (blue) and mTCR (red) the number of detected particles increases 40 minutes after biotin was added. The two receptors appear to level out at different numbers of foci per image (~140 for the mTCR and ~30 for the pre-TCR after 90 minutes). Both are consistent with the final number of foci found after 10 minutes in the controls without ER-retainment. There is a slight, near linear increase in nanobody foci in images of the mTCR without biotin due to the leakage of the receptor from ER retainment.

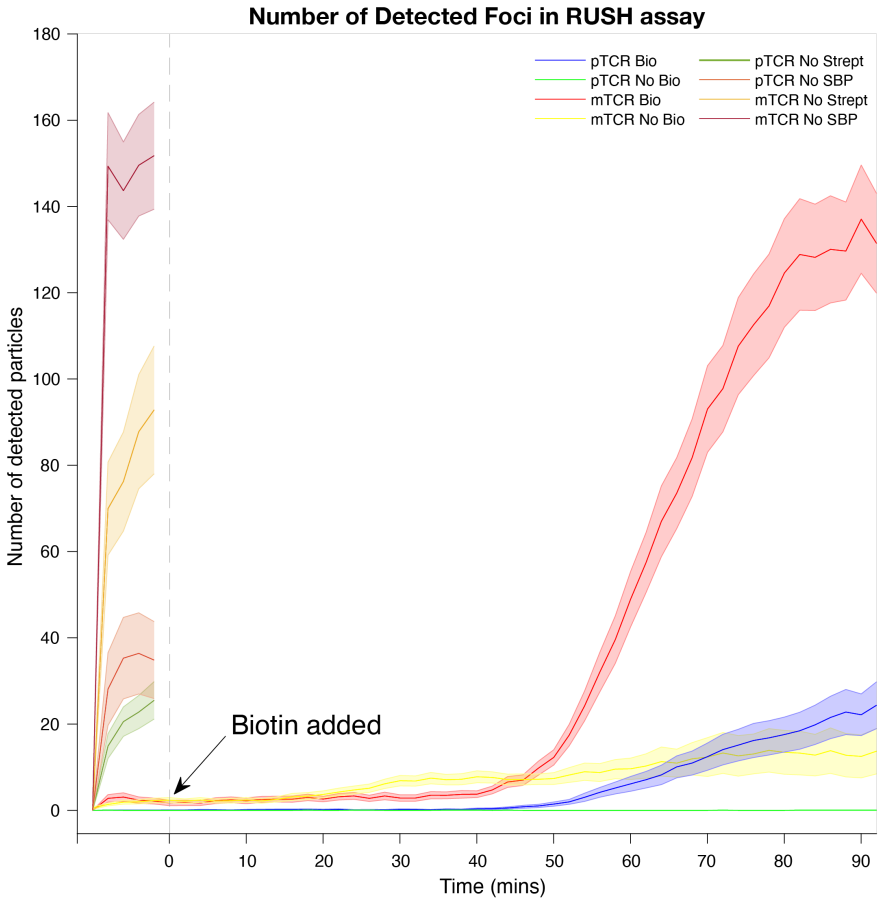

Figure 3.22: The pre-TCR and mature TCR are both detectable at the cell surface from 40 minutes after secretion is initiated.
Images like the ones presented above were analysed in MATLAB. The number of distinct foci in the af647 channel that were brighter than a pixel threshold value (10,000 for all groups), was recorded for each frame. The mean number of foci per image is shown here with shaded areas representing the standard error. The time axis has been shifted so that biotin was added at time 0 (grey dashed line). Six biological replicates were performed for the biotin/no biotin groups.
The blue and green lines show data for the SBP-tagged pTCR with and without biotin. The red and yellow lines show data for the SBP-tagged mTCR with and without biotin. These were imaged for all 52 frames.
Dark green and orange lines shown data for the pTCR without Streptavidin on the BFPkdel and without the SBP-tag. The gold and maroon lines show similar for the mTCR. All four lacked ER-retention and maximum binding was achieved within 10 minutes.

I managed to analyse the same dataset using the Pearsons correlation coefficient as an independent output. Normally the correlation coefficient is a poor metric of co-localisation as the value is dominated by the amount of empty space in the image where the fluorescence is low in both channels. Here though the empty space in the images is approximately constant throughout the timecourse. This means changes in the correlation coefficient over time, averaged over many images, become meaningful though the absolute values may not be.

For both the pre-TCR and mTCR, the pixel intensity for the GFP-tagged receptors is reasonably well correlated with the BFP-tagged Streptavidin hook at the start of the experiment (figure 3.23 left). Biotin causes both receptors to separate from the hook with the correlation beginning to fall after around 30 minutes, approximately 10 minutes before the receptors start being detected at the cell surface. Without biotin the correlation coefficient stays constant. The correlation coefficients between the af647 nanobody channel and the GFP-tagged receptor start low for all samples then seem to rise quickly to an early plateau between 10 and 30 minutes (figure 3.23 right). I suspect this is due to the amount of receptor that has leaked past the ER-retainment to the cell surface as it appears to be similar for the two receptors with and without biotin. Biotin addition causes another increase beginning around 40 minutes corresponding to the arrival of synchronised trafficking at the cell surface. The correlation might become less accurate after this point as GFP-fluorescence is quenched by nanobody binding. These correlations end at similar values to the controls without ER-retainment but do not seem to have reached a plateau after 90 minutes.

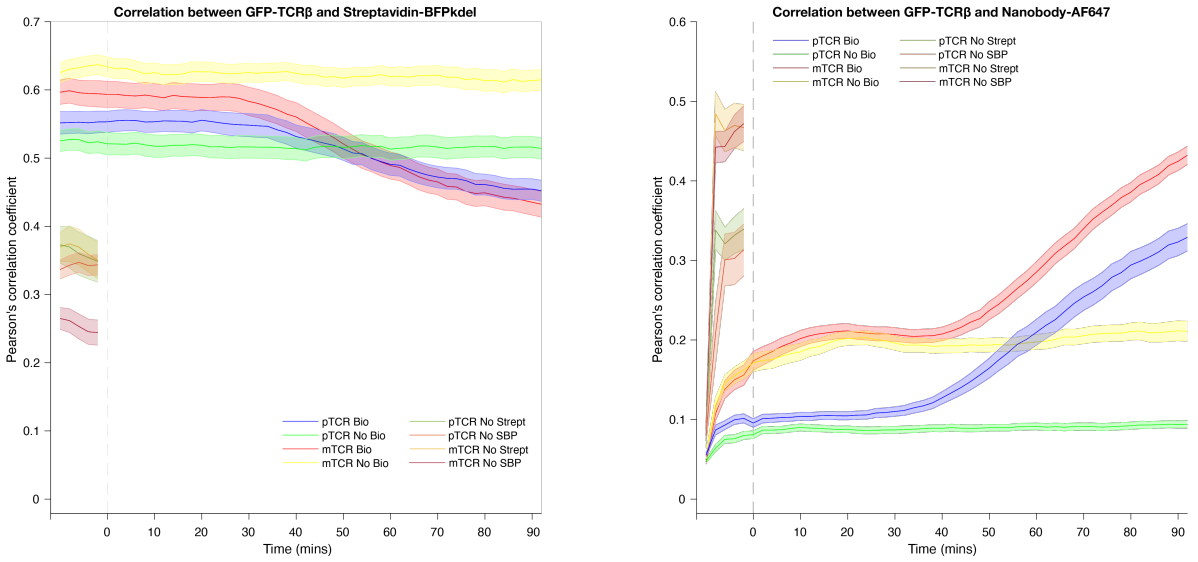

Figure 3.23: Both the pre-TCR and mature TCR leave the ER ~10 minutes before they are detected at the cell surface.
The Pearson's correlation coefficient between the pixel intensity of the GFP-channel and the BFP channel as well as the correlation between the GFP-channel and the af647-channel was recorded for every frame of the set of images discussed above. The correlations are presented here with shaded areas representing the standard error. The time axis has been shifted so that biotin was added at time 0 (grey dashed line).

This experiment suggests that secretory trafficking from the ER to the cell surface occurs at a similar rate for the pre-TCR and mTCR. Both receptors begin export from the ER ~30 minutes after biotin addition and are first detectable at the cell surface 10 minutes later. This is most obvious when the data is normalised by the number of foci present near the end of the experiment and the results for both receptors are overlaid (figure 3.24). This would imply that the rate of trafficking through the secretory pathway is not responsible for the different surface expression levels at steady state.

Why then does the $\alpha\beta$TCR have a greater number of internalised foci than the pre-TCR after synchronised secretion? I suspect that the RUSH system has trapped the receptors in the ER after complexes have been assembled. Whilst the two receptors are comparable in their ability to traffic through the secretory pathway, this does not mean that prior complex assembly occurs at the same rate or reliability. If a larger proportion of pre-TCR receptors failed this assembly stage and were degraded by ERAD, there would be fewer trapped complexes ready to be secreted as soon as the streptavidin-mediated retainment

is released.

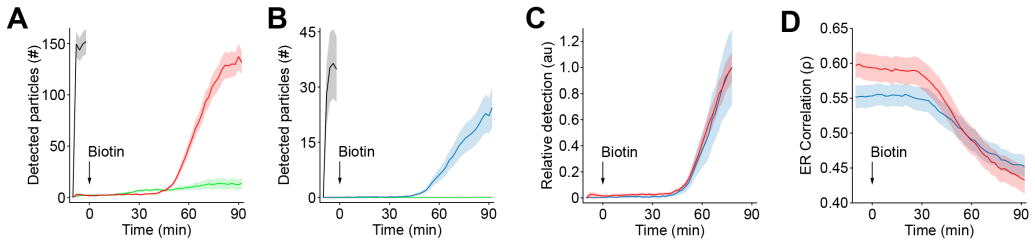

Figure 3.24: The pre-TCR and mTCR are capable of comparable rates of secretion.
A: The mean number of detected foci per image of the mTCR with biotin (red) and without (green) compared to the no SBP control (black).
B: Equivalent data for the pTCR with biotin (blue) and without (green) compared to the pTCR without SBP (black).
C: Data for the mTCR and pTCR, normalised to the amount of foci present at 75 minutes.
D: The correlation between the receptor and Streptavidin-BFPkdel hook for the mTCR (red) and pTCR(blue).
Bounding area around datapoints shows mean ± SEM of multiple fields of view from 6 biological replicates.
Addition analysis performed by John James.

3.4.7 Signalling consequences of pre-TCR and $\alpha\beta$TCR divergent trafficking

The combined data from the James lab suggested that the reduced surface expression of the pre-TCR, compared to the $\alpha\beta$TCR, was an intrinsic property of the receptor rather than a consequence of autonomous receptor signalling. We wondered how these distributions affected receptor signalling. Differences, either qualitative or quantitative, could explain how the pre-TCR is distinguished from the $\alpha\beta/\gamma\delta$TCR at the β-selection checkpoint. Our ideal system would produce three states: an "off" state in the absence of TCR signalling and two distinct "on" states induced by signalling by the pre-TCR and $\alpha\beta$TCR. As HEK cells lack the required kinases for TCR signalling, these experiments were performed in Jurkat T cells.

Jurkat cells were transduced with plasmids encoding cas9 with guides targeted to the *TCRA* gene. This produced a cell line lacking surface $\alpha\beta$TCR expression. A fraction of these cells were then transduced with pTα-mScar to reconstitute pre-TCR expression (figure 3.25A). As this pre-TCR uses the endogenously expressed TCRβ and CD3 chains, its expression should more closely match the level *in vivo* than if all chains of the pre-TCR were expressed exogenously. Surface staining of CD3 confirmed there was intermediate levels of receptor expression compared to wildtype Jurkats or the TCRα knockout (figure 3.25B).

Our original intention was to tightly couple these cells to a target cell to recreate the exclusion of CD45 phosphatases observed in T cell contacts to antigen presenting cells. If the small amount of pre-TCR at the surface in these contacts was capable of signalling this would align the receptor with the steric exclusion model. I investigated the formation of these contacts using the affinity between an anti-GFP Nanobody-CD86 transmembrane construct expressed on Jurkat cells and GFP-tagged CD86 or CD28 expressed on target Raji cells. A version of the Nanobody-TM with the intracellular sequence of the ζ-chain, similar in structure to a CAR-T receptor, was to be used as a positive control. These components were found to create tight cell conjugates. To our surprise however, conjugate formation was not required for signalling for the TCR and pre-TCR to be distinguished.

The $\alpha\beta$TCR is known to produce low-level, tonic signalling in the absence of a ligand which maintains background expression of activation markers such as CD69.[295] Wildtype Jurkat cells, the TCRα knockout

line and pTα-transduced cells were mixed together then labelled with markers of tonic signalling (figure 3.25C+D). TCR^{+ve} cells were seen to have consistently higher surface expression of CD69 than TCR^{+ve} cells with pTα-transduced cells expressing an intermediate value. The same pattern was observed with the markers CD137 and PD-1. This trend was not observed however for CD5, a marker of tonic signalling in thymocyte development. Cells transduced with pTα had lower surface CD5 than TCR^{+ve} or TCR^{-ve} cells suggesting that signalling from the pre-TCR was inhibiting CD5 expression (figure 3.25E).

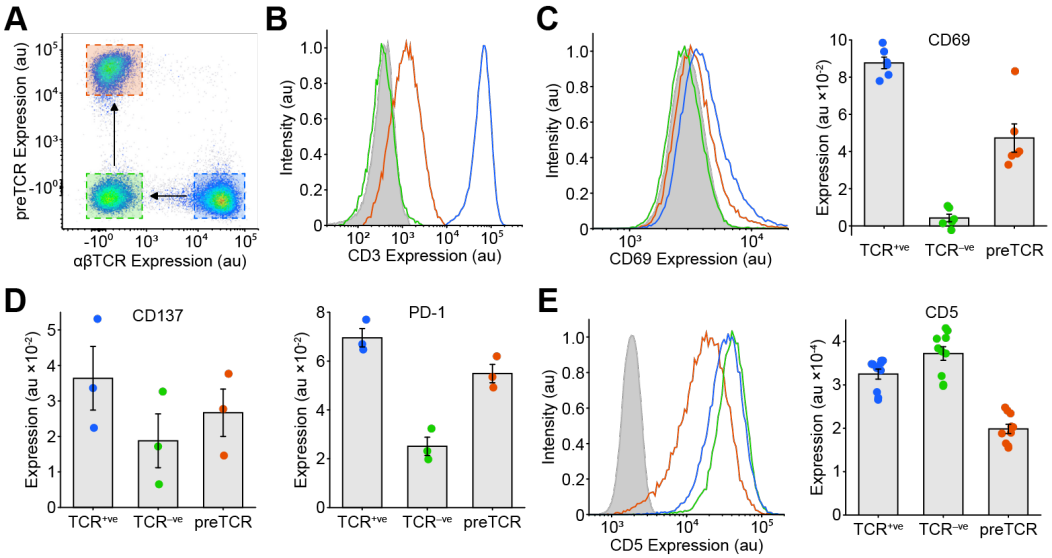

Figure 3.25: The pre-TCR generates intermediate levels of tonic signalling compared to wildtype Jurkat T cells or a *TCRA* knockout line.
A: Jurkat T cells (blue) were transduced with CRISPR/Cas9 plasmids targeted toward the *TCRA* gene. This produced a TCRα negative population (green) that was transduced with pTα-mScar to induce pre-TCR expression (red). These populations were mixed then labelled with the antibodies indicated below.
B: Comparing surface CD3ϵ expression in these three populations compared to an isotype control (grey).
C: Comparing surface expression of activation marker CD69 in these three lines with the mean of 6 repeats shown to the right. Error bars show mean ± SEM.
D: Similar data for CD137 and PD-1 (n=3).
E: Similar data for CD5 (n=6)
Experiment and analysis performed by John James.

Differences in tonic signalling could also be seen in the phosphorylation state of proteins in the proximal TCR signalling pathway. Cell lysate from the TCR^{+ve}, TCR^{-ve} and pTα-transduced populations was analysed by western blotting with phosphospecific antibodies (figure 3.26). Total protein normalisation was used to correct the band intensity to the amount of cell loaded. Cells transduced with pTα were found to have intermediate levels of phosphorylated CD3ζ, Lck, ZAP70 and LAT compared to wildtype and TCR^{-ve} cells. Expression of pTα did seem to reduce total CD3ζ suggesting active degradation of the pre-TCR. Taken together this shows that the pre-TCR is capable of generating low, but detectable, tonic signalling. The low intensity and transience of this signal might allow the receptor to be distinguished from the TCR during β-selection.

3.5 Discussion

This project started with the observation that when reconstituted in HEK, the asymmetric extracellular structure of the pre-TCR seemed to determine its steady state surface expression. The cytoplasmic tail of pTα was not found to be important contrary to previous work.[237] As HEK cells lack signalling kinases this behaviour is signalling independent and intrinsic to the receptor itself. I verified that the

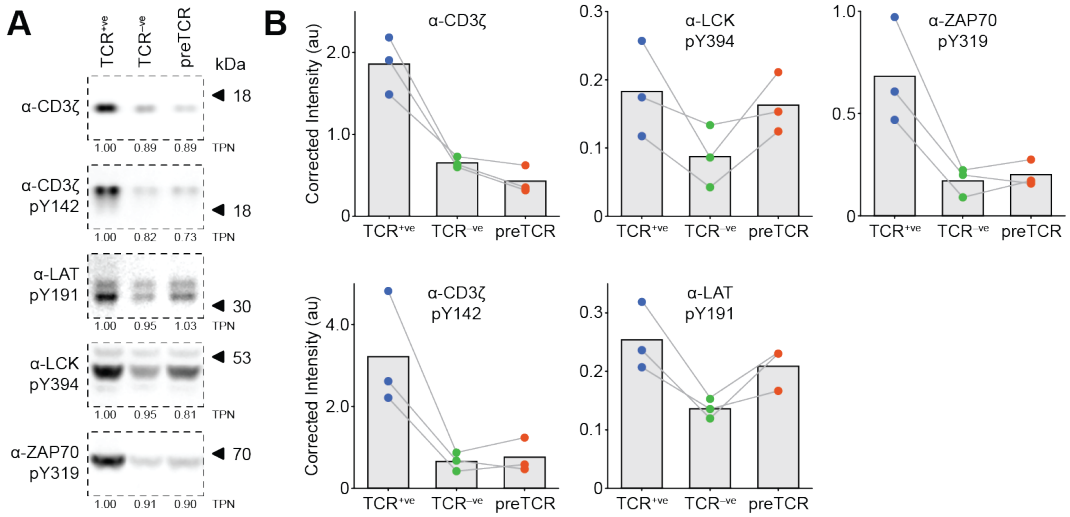

Figure 3.26: Pre-TCR tonic signalling can be observed in the phosphorylation status of proteins in the proximal TCR signalling pathway.
A: Cell lysate from the Jurkat populations discussed above was analysed by western blotting using phosphospecific antibodies to components of the proximal signalling pathway. Total protein normalisation (TPN) was used to account for difference in the amount of cells loaded. These values are presented beneath each blot. One repeat of three shown.
B: Band intensity in western blots as presented in A was quantified and normalised using the total protein measurements. Each replicate is shown with connecting lines. Bars represent the mean for each measurement.
Experiment and quantification performed by the author. Data presentation by John James.

pre-TCR is not detectably phosphorylated in HEK cells (supplementary figure S9A). Two hypotheses were investigated to explain how information about the structure of the extracellular domains of the receptor could be communicated to cytoplasmic internalisation machinery.

The first hypothesis was that an unknown membrane protein was present with the pre-TCR at the cell surface. This protein could contact the TCRβ variable domains and signal for the receptor to be internalised. I used two membrane-targeted proximity labelling assays to biotinylate proteins in the vicinity of the pre-TCR. Biotinylated proteins were enriched with streptavidin-sepharose then identified by mass spectrometry. No protein consistent with this model of internalisation was identified. However the signal obtained in these assay was unfortunately poor and peptides from the pre-TCR itself were not consistently identified. There are multiple reasons why the number of peptides identified might not correspond to the abundance of nearby proteins in a proximity labelling experiment. These include the accessibility of target residues, in this case tyrosines, the distribution of trypsin sites and how well different peptides are ionised and detected. Improvements to these assays could be made to remove background whilst keeping genuine interactions.

The other hypothesis was that the asymmetric structure of the pre-TCR allowed the receptor to dimerise which triggers internalisation. Two models of pre-TCR dimers have been proposed based on crystal structures. Whilst I detected some affinity of TCRβ chains for each other through co-immunoprecipitation assays, this interaction is disrupted by pTα or TCRα chains. NanoBit and split Venus complementation assays suggest the pre-TCR is monomeric both in bulk and at the cell surface.

How then are pre-TCR and TCR complexes trafficked differently? I suspect the biggest contribution is that the assembly of the pre-TCR is slower than assembly of the $\alpha\beta$TCR or more prone to ER-degradation. This limits the amount of pre-TCR ready to be secreted from the ER to the cell surface. Once the complexes are assembled, the RUSH assay suggests the pre-TCR and $\alpha\beta$TCR can be trafficked at comparable rates.

I suspect the pre-TCR and $\alpha\beta$TCR are actually internalised from the cell surface at the same rate

through normal membrane turnover. I designed one experiment to test this hypothesis which compared the internalisation of fluorescent anti-GFP nanobodies between GFP-tagged pre-TCR, $\alpha\beta$TCR, CD86 and CD28 (supplementary figure S8). This experiment was quite challenging technically due to large difference in expression levels. However for all four proteins, foci of the nanobody were detected in the cytoplasm very quickly. A better experiment here would be to label surface receptors (for instance through biotinylation or antibodies), allow internalisation to occur then strip the label from the surface. Measuring the drop in the amount of labelled receptor gives the amount internalised in this time.[296]

If the pre-TCR and $\alpha\beta$TCR are internalised at the same rate then why does the surface expression of the TCR fall slowly when trafficking from the ER is blocked whilst the surface expression of the pre-TCR rapidly falls (noted in figure 3.5D). An overlooked feature might be that if internalisation is limited by the endocytic machinery then a receptor expressed at low levels would appear to be internalised more rapidly than one that is more abundant. For instance, octameric TCR complexes have been reported to be constitutively internalised at ~1% per minute.[111] If the same absolute number of pre-TCR receptors were internalised per minute then this would be a much larger proportion of the total when pre-TCR surface expression is ~50-100x lower than the mature TCR.[258] Whilst it seems unlikely for the endocytic machinary to be limiting in HEK, this might not be the case for all cell types.

The difference in steady state surface expression is also due to the effects of receptor recycling. When the pre-TCR is internalised it appears to be quickly degraded in lysosomes whilst a large proportion of internalised $\alpha\beta$TCR is recycled back to the cell surface. This suggests that the pre-TCR and $\alpha\beta$TCR are distinguished not at the cell surface but in the endosomal pathway. Assuming the peroxidase-conjugated antibodies of the SPPLAT is internalised at the same rate as fluorophore-conjugated nanobody, it should have been present in intracellular vesicles before labelling was initiated with hydrogen peroxide. Proteins usually associated with endosomal trafficking were not enriched in this dataset though most are on the cytoplasmic side.

If the pre-TCR is primarily monomeric then how is solubility maintained given the hydrophobic patch on the TCRβ variable domain? I still don't think it unreasonable that the domain could be shielded by a chaperone protein such as BiP that is trafficked with the pre-TCR. BiP and other HSP70 chaperones were detected with my Nanobody-APEX and SPPLAT assays. These are however very abundant and frequently appear in control experiments. The hydrophobic patch on TCRβ has been reported to mediate an interaction with pMHC complexes.[252,253,255] I designed guides to exon 1 of *B2M* and these were expressed in HEK cells with plasmids encoding cas9-BFP. Individual clones were acquired via single-cell sorting. I found a line with no detectable surface MHC expression through screening with an anti-MHC antibody. Sequencing this line found two *B2M* alleles with disruptive indels. There appeared to be no effect on pre-TCR surface expression or trafficking in this line which implies the pre-TCR does not require MHC molecules for surface expression in either *cis* or *trans* (supplementary figure S9B-D).

Our data suggests that the pre-TCR and $\alpha\beta/\gamma\delta$TCR might be distinguished by tonic signalling. Due to the low residence time of the pre-TCR at the cell surface signalling from the pre-TCR could be weaker than the TCR, more transient or a combination of both. Whilst the pre-TCR and $\alpha\beta$TCR have been contrasted in these experiments due to their structural similarity, I note the more relevant comparison *in vivo* is between the pre-TCR and the $\gamma\delta$TCR during β-selection. I cloned a construct of the G115 $\gamma\delta$TCR which is typical of human Vγ9/Vδ2 subsets.[297] Flow cytometry and microscopy (supplementary figure S10 and S11) implied that when expressed in HEK, the receptor was similar in surface expression to the $\alpha\beta$TCR. The pre-TCR with the pTα^{b} isoform, another biologically relevant comparison, was similar to the pre-TCR.

Chapter 4

Cell biology characterisation of Transmembrane Protein 131

4.1 Introduction

Previous experiments within the James lab searching for pre-TCR interaction partners, linked the receptor to a poorly characterised membrane protein called Transmembrane Protein 131 (TMEM131). These experiments are discussed later in this introduction. Interesting observations about TMEM131, as well as tentative links to T cell development, justified further study.

In humans, the gene encoding TMEM131 is located on chromosome 2, immediately adjacent to the gene for ZAP70. The longest protein isoform has 41 exons encoding 1883 amino acids with a theoretical molecular weight of 205kD. Its Uniprot entry (Q92545) describes it as having two TM helices just over halfway through the protein in positions 1091-1111 and 1118-1138. The C-terminal tail is described as disordered.

Homology mapping suggests that the protein has two or three Ig-like domains in its N-terminal section that are described as TMEM131-like domains or domains of unknown function (DUF) 3651. These domains have sequence similarity to the domains of a bacterial periplasmic chaperone called PapD, that uses these domains in the assembly of pilus subunits using a "donor-strand-exchange" mechanism.[298–300]

TMEM131 appears to be very highly conserved throughout the animal kingdom with recognisable orthologs even in non-vertebrates such as *C. elegans* and *Drosophila melanogaster* (figure 4.1). In humans there is one recognisable paralog, *TMEM131-like* (TMEM131L)/KIAA0922 that is discussed later.

4.1.1 Evidence suggesting a role for TMEM131 in T cell development

The Human Protein Atlas database states that TMEM131 mRNA is widely expressed at low levels but is enriched in the brain, endocrine and lymphoid tissues. Circumstantial evidence however, supports a role for TMEM131 in T cell development. The publicly available, mouse RNA-seq ImmGEN database[303,304] finds that TMEM131 has higher expression in DN3 cells, at the point of β-selection (figure 4.2). The same dataset also suggests that TMEM131 is more highly expressed in B cells located in the germinal

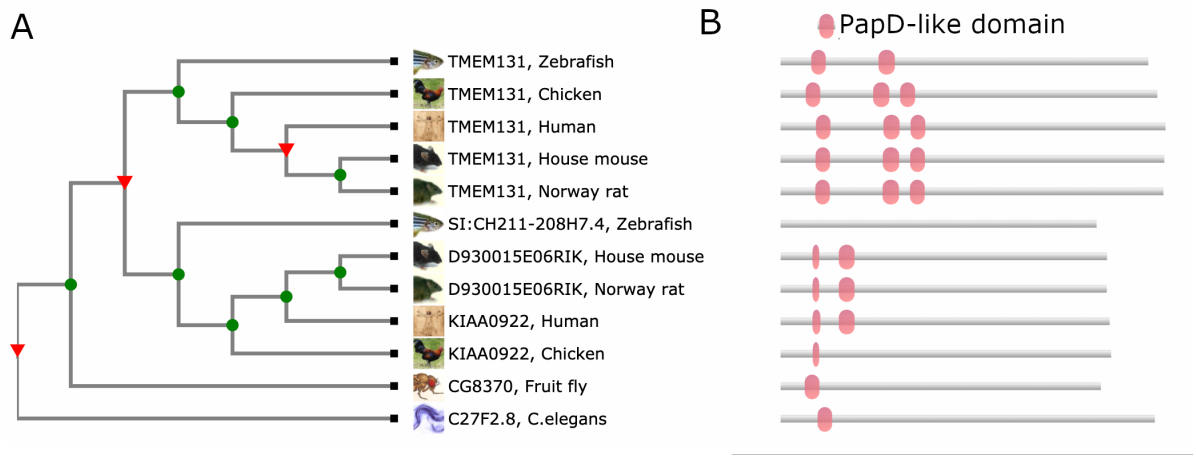

Figure 4.1: Cladogram and domain architecture for the TMEM131 family in model species.
A: Cladogram showing the relationship between TMEM131 and KIAA0922/TMEM131L genes in model species. A full tree with more species is available at the source.
B: The predicted protein domain architecture for these genes with identified PapD-like domains shown in pink. Different models identify different numbers of these domains. Figure adapted from TreeFam which featured 143 sequences from 91 species.[301,302] Sequences were on average 34% conserved.

centres where somatic hypermutation takes place. *TMEM131* was identified as having a large change in expression throughout thymocyte development[305] and in a genome wide transcription model was clustered with Notch1, Erg1 and RunX1.[306] A study of the transcription factor Bcl11b, that is highly important in establishing T cell lineage commitment, identified *TMEM131* as a gene that was significantly downregulated in a Bcl11b-KO line compared to WT cells.[144] Finally, a screen for genes with altered methylation patterns in lymphocytes from adults with Down syndrome identified *TMEM131* as one of 8 candidate genes that were significantly hypomethylated compared to controls.[307] This list also included *CD3ζ* and *Tcf7*, important factors in thymocyte development. Downs syndrome features a number of immunological abnormalities.

Though protein homology is no guarantee of similar function, the TMEM131 paralog TMEM131L has been attributed functions in thymocyte development. TMEM131L is encoded on chromosome 4 and shares ~36% amino acid identity to TMEM131. Maharzi *et al.* found that the expression of TMEM131L increases during the development of DN thymocytes and peaks during DN3 at the point of β-selection.[308] Antibody staining showed that TMEM131L is present at the cell surface in DN3 cells and localises to cell-cell contacts together with Axin, a component of the β-catenin degradation complex. The paper concluded that TMEM131L was a negative regulator of the Wnt signalling pathway through triggering lysosome-dependent degradation of p-LRP6. A subsequent paper investigating *Drosophila* mutants, also suggested TMEM131L affected signalling in the Notch pathway.[309]

4.1.2 Previous work in the James lab

In the James lab, a connection between the pre-TCR and TMEM131 was first made via a BioID assay. The pre-TCR and TCR were expressed with and without the BioID2 domain on the N-terminus of TCRβ and biotinylated proteins precipitated with streptavidin beads. Proteins were eluted from the beads then identified via mass spectrometry using an in-gel digest approach. Two peptides of TMEM131 were found in the lane corresponding to pre-TCR+BioID but not in the controls without the BioID2 domain or the mTCR+BioID sample. Due to the literature evidence for a role of TMEM131 in T cell development, this protein was investigated further.[271]

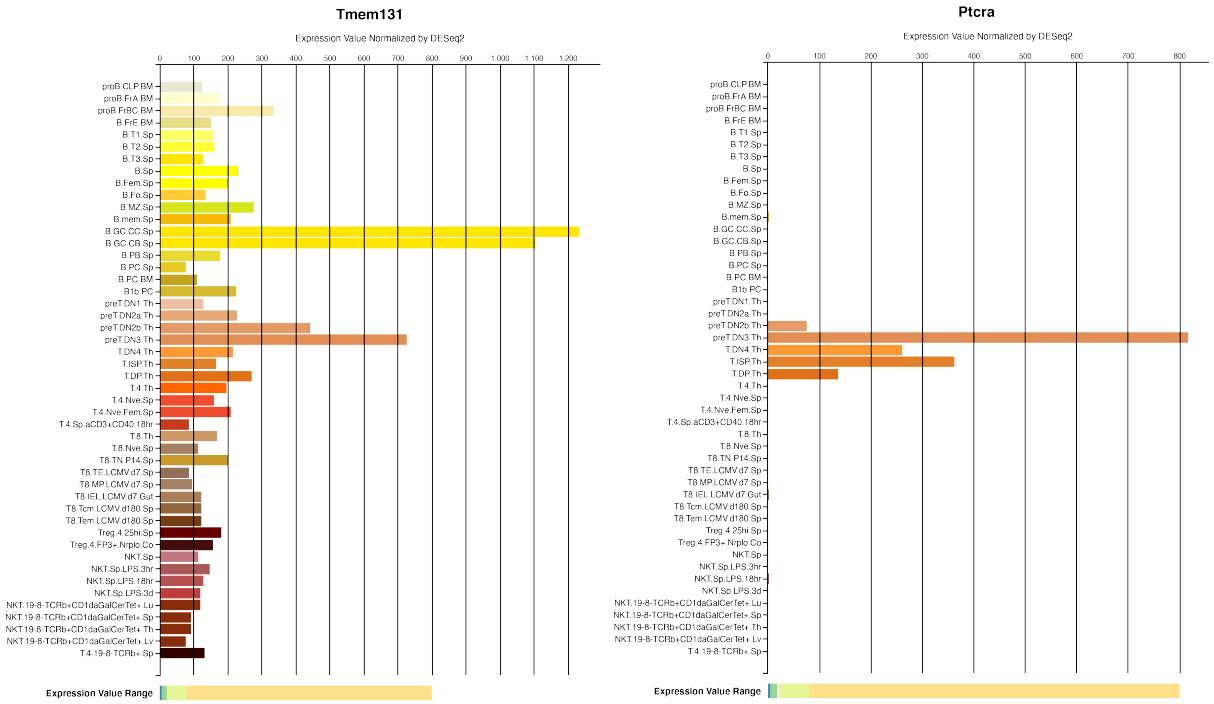

Figure 4.2: ImmGEN mouse RNAseq data for TMEM131 (left) and PTCRA (right).
Only cells in the $\alpha\beta$ and B cell lineages are shown. Normalised expression values are characterised as 0-5 (Trace), 5-20 (Very Low), 20-80 (Low) and 90-800 (Medium) and 800+ (High).

Using the one commercially available anti-TMEM131 antibody (A302-129A), that recognises a segment of the TMEM131 C-terminus, TMEM131 was immunoprecipitated from untransfected HEK cells and from biotin-treated cells expressing BioID2-tagged pre-TCR and TCR receptors. Biotinylated bands at 240kDa were observed in the pulldowns from cells expressing both receptors but not untransfected cells. This band is larger than TMEM131's predicted molecular weight of 205kD suggesting the protein is glycosylated. This experiment implied that TMEM131 interacted with both the pre-TCR and TCR and was not endogenously biotinylated.

The pre-TCR and TCR receptors with TCRβ-GFP chains were expressed in HEK cells, as well as pTα-GFP, TCRβ-GFP, TCR$\beta\Delta$Vr-GFP and ICAM1-GFP proteins individually. Proteins were immuno -precipitated via GFP-trap agarose beads then eluted from the beads via heat in sample buffer. Bead elutions were blotted and probed for TMEM131. A band around the molecular weight of 240kD was observed in the lanes from all samples but not in the bead elution from untransfected cells. The band in the sample from ICAM1-GFP was noticeably weaker than the bands from the other proteins. The investigator concluded that TMEM131 interacts with the chains of the pre-TCR and TCR in particular and not with proteins with Ig-domains more generally. I do not think the small range of constructs tested provide sufficient evidence for this conclusion.

The OP9 co-culture system was used to differentiate mouse haematopoietic stem cells through thymocyte development. After 20 days cells were harvested and labelled with CD4, CD8, CD44 and CD25 antibodies. This allowed the cells to be sorted into DN/DP populations then DN cells further segregated into DN1-4. These fractions were then lysed and diluted lysate run on Tris-acetate gels. These were blotted and probed for TMEM131. The amount of TMEM131 protein in each fraction was quantified via band intensity with an attempt to normalise for the different amount of cells in each group by the intensity of an anti-actin band. This protocol suggested that TMEM131 was expressed in both DN and DP cells. Within DN cells there was slightly higher expression in DN2 and DN3 cells than DN1 cells but no

detectable expression in DN4 cells. This protein distribution is consistent with the previously mentioned RNAseq data which found an increase in *TMEM131* expression in the DN2/DN3 stages. The experiment was limited however by poor segregation of DN1-4 cells, simplistic normalisation and apparent protein degradation in the samples.

An attempt was made to knockout TMEM131 in HeLa cells and measure changes in the expression of pre-TCR and TCR constructs. Constructs were made with Cas9-BFP and CRISPR guides targeting TMEM131 exon 4 and 5. These constructs were expressed in HeLa cells by lentiviral transduction then clonal BFP-positive lines isolated. A couple of cell lines were produced that appeared to have reduced TMEM131 expression. However, the surface expression or internalisation of the pre-TCR and mTCR did not appear to be affected. These lines were never genotyped and due to the hypotriploid haplotype of HeLa cells, it is possible that TMEM131 expression was not completely eliminated in these lines. I do not consider this convincing evidence that TMEM131 is dispensable for the trafficking of pre-TCR or TCR.

The human TMEM131 cDNA sequence was synthesised in four sections and assembled. This sequence was used to express the protein in HEK cells with a C-terminal GFP tag. Cell lysate was analysed by western blotting with the antibody to the TMEM131 tail and an anti-GFP antibody. The anti-TMEM131 antibody proved to be non-specific with a lot of background bands visible in the lysate from untransfected HEK. Two additional bands were present in the lysate from cells expressing TMEM131-GFP. These were at ~240kD, consistent with the molecular weight of the protein, and ~95kD. More surprisingly the GFP antibody also labelled a distinct ladder of bands, from ~30kD to 95kD with the highest band overlapping with the band labelled by the anti-TMEM131 antibody. No GFP^{+ve} band was present that could correspond to the full-length protein with the tag. This implies the C-terminal tail of the protein is cleaved producing a series of GFP-positive peptides with variable amounts of the TMEM131 tail still attached.

TMEM131 was expressed again with an N-terminal HA tag and a C-terminal GFP tag. The endogenous signal peptide was also exchanged with the signal peptide for *Gaussia* luciferase to enhance expression. A similar construct was cloned and expressed where the tail was truncated at a convenient SpeI site at position 1242 (called hereafter TMEM131dC). When the lysate was analysed via western blotting with anti-HA and anti-GFP antibodies, the lane with HA-TMEM131-GFP showed only a GFP-positive band around 90kD and no HA-positive bands. The lane with HA-TMEM131-dC had a GFP and HA-positive band at ~170kD which is correct for its now reduced molecular weight. This suggested that the protein was cleaved in the last 641 amino acids of the tail and the N-terminal part of the cleavage somehow degraded. A panel of TMEM131 constructs with truncated tails of different lengths was made and expressed with C-terminal GFP tags. Each of these truncations produced a ladder of distinct GFP-positive bands. As these bands were distinct they were not thought to be due to non-specific aggregation or degradation. A pattern in the bands was apparent but difficult to interpret. As the tail was shortened the intensity of the GFP and HA-positive band consistent with the intact protein seemed to increase. I replicated this experiment using different antibodies and constructs with results presented in section 4.4.1.

The TMEM131 tail with C-terminal GFP tag was fused to CD86 which also produced a ladder of GFP-positive bands when blotted implying that the cleavage was an intrinsic property of the tail. Ultimately no specific cleavage site in the TMEM131 tail was identified that could explain these results and we do not know the protease responsible. Protein cleavage after synthesis is not without precedent (for example zymogens or pro-caspases) but is unexpected in this context.

4.1.3 Recent evidence suggesting a role of TMEM131 in ER exit

In early 2020, after the work that would form this thesis was initiated, a paper was published by Zhang *et al* that related TMEM131 to the intracellular recruitment and secretion of collagen.[310] These authors first identified TMEM131 from an RNAi screen in *C. elegans* for genes involved in ER stress. In *C. elegans*, TMEM131 was found to be widely expressed but enriched in the intestine and hypoderm. TMEM131-GFP was found to localise to the ER. In a *C. elegans* TMEM131 knockout, expression of GFP-tagged collagen protein COL-19 was absent and two other collagen markers, COL-101 and LON-3, were reduced. Knockout of the TMEM131-homolog in *Drosophila* also caused a defect in collagen assembly. Using lentiviral expression of shRNAs to knockdown TMEM131 in the human U2OS bone osteosarcoma cell line found a decreased secretion of collagen fibres. The authors performed yeast two hybrid screens which found that the TMEM131 PapD-like domains interacted with C-terminal propeptide domains of human collagen families. The C-terminal tail was found to interact with the protein TRAPPC8 (trafficking protein particle complex 8). This interaction was abolished if the terminal residues 1741-1883 of TMEM131 were removed. TRAPPC8 is part of a complex acting as a guanine nucleotide exchange factor for the Rab GTPases promoting ER to Golgi trafficking of COPII-vesicles. The authors concluded that collagen secretion was an evolutionary conserved function of TMEM131 proteins.

A follow up paper described a similar role for the multipass membrane protein TMEM-39. This protein was also identified by a RNAi screen for ER stress and knockout was shown to affect collagen secretion. A yeast-two hybrid assay found a cytoplasmic loop of human TMEM-39 bound to the COPII component Sec23A. TMEM131 and TMEM39 were proposed to operate together in collagen secretion.[311]

Regarding a role in the ER exit of collagens, TMEM131 has strong functional similarities with a separate family of proteins: the Transport and Golgi organization (TANGO) family. This family includes TANGO1 itself as well as TANGO1-like protein (TALI) and its shorter isoform cTAGE5.[312] These proteins are encoded by the genes *MIA3* and *MIA2* (Melanoma inhibitory activity protein) respectively. The TANGO family is involved in the transport of proteins from ER exit sites (ERESs), sub-domains of the rough ER lacking ribosomes, to the ER to Golgi Intermediate Compartment (ERGIC). The conventional model of ER transport between these two compartments is mediated by cytoplasmic coat protein complex II (COPII) vesicles that bud off the ERES. Vesicle formation is initiated by the GTPase Sar1 which is activated by membrane bound protein Sec12. This complex initiates membrane curvature and recruits Sec23–24 heterodimers which form the vesicle inner coat. This inner coat can recruit protein cargo via adaptor proteins and allows the formation of the outer coat. The outer coat consists of Sec13–31 heterotetramers which promote fission of the vesicle from the ER membrane. The vesicles are guided to the ERGIC by microtubules and are tethered by the TRAPP complex. Once there, proteins can proceed through the Golgi or return to the ER in COPI vesicles.[313]

Whilst COPII vesicles have been observed in lower eukaryotes and with synthetic membranes, the extent they are used by mammalian cells has been questioned.[314] A longstanding issue with COPII vesicles is that their 60-100 nm diameter is insufficient for common cargos such as pro-collagen fibres which are greater than 300nm in length. Instead an emerging view is that the ERESs and ERGIC are connected by a direct tubular connection or tunnel that is mediated by the TANGO family.[315,316]

TANGO1 has 1907 amino acids and consists of a N-terminal lumenal segment, a transmembrane domain and a 709-residue cytoplasmic tail. The lumenal sequence consists of a coiled-coil domain and a SH3-like domain. The TANGO1 membrane domain consist of one full transmembrane helix proceeded by a hydrophobic loop that is partially inserted into the inner leaflet of the ER membrane. The cytoplasmic tail contains two coiled-coil domains and a C-terminal proline rich domain. TALI has similar domain

architecture whilst cTAGE5 lacks the hydrophobic loop and SH3 domain. TANGO1 and cTAGE5 are believed to operate in a complex with 1:1 stoichiometry. TANGO1 is thought to form into rings through lateral associations that act as a neck at the ERES side of these transport tunnels. These tunnels have been observed via STED microscopy. The lumenal SH3 domains are believed to recruit collagens via the chaperone HSP47. Meanwhile the proline-rich domains in the C-terminal tail are thought to recruit COPII components such as Sec23 and Sec16 or recruit ERGIC membranes directly. The interesting membrane topology of the protein is thought to have two functions: lowering the surface tension at the tunnel entrance enabling the membrane curvature[317] and acting as a fence barrier to the motion of membrane lipids.[318] Unlike cargo receptors, TANGO1 proteins are not transported to the Golgi along with cargo proteins.

Whilst TANGO1 has been conclusively linked to the transport of collagens, it is less clear to what extent it is required for the transport of other cargos. CRISPR knockout of the long isoform of TANGO1 which has the lumenal SH3 domain caused relatively minor defects in collagen secretion in hTERT-RPE1 cells whilst the loss of both the long and short forms of TANGO1 results in major defects in cell organisation and secretion.[319] Knockout was also associated with widespread changes in the transcriptome and proteome of edited cells. It is unclear whether these defects are due to direct interactions of TANGO1 with other proteins or due to the build-up of collagen in the ER that impedes the transport of smaller proteins.[320] TANGO1 was specifically linked to the transport of bulky VLDL lipid droplets also likely too big for COPII vesicles.[321]

TMEM131 could fit into this model of ER exit in two main ways. Firstly whilst *C. elegans* expresses a variety of collagens and has homologues of COPII and TRAPP proteins, it has no TANGO1 homologues. TMEM131 could be acting as direct substitutes in the collagen secretory pathway in this organism and potentially others. Secondly TMEM131 could enable the secretion of a wider range of cargos either acting in conjunction with TANGO1 or independently. In humans TANGO1 is widely expressed in most tissues. However, it was found to be notably lacking in cells from the haematopoietic lineage including PBMCs and potentially T cells.[322] Whilst these cells are unlikely to be producing collagen, they are likely to be producing other secreted proteins and large receptors complexes that need chaperone or cargo receptor assistance to be secreted efficiently. This possibility is worthy of extra study. Whilst the Zhang *et al* paper makes a convincing case for TMEM131's role in collagen secretion, it does not mention any of the links to thymocyte biology or the pre-TCR described earlier in this chapter.

Other evidence supports the idea that TMEM131 interacts with a wider range of clients than pro-collagen fibres. TMEM131 was identified as a negative regulator of MR1 expression in a gene-trap experiment using the near-haploid HAP1 human cell line.[323] An unbiased CRISPR screen using the microglia-derived BV2 cell line identified TMEM131 as a negative regulator of surface expression of microglial phospholipid receptor TREM2.[324] TREM2 is an immunoreceptor expressed on a variety of immune cells including microglia and macrophages. It signals through the ITAM-motifs of its accessory protein DAP12 which has high homology to CD3ζ. Gene variants of TREM2 are linked to Alzheimer's disease. Other myeloid markers such as CD33, CD44/HCAM, and CD45 were not affected by the knockout implying TMEM131 was relatively specific in this context.

Other mentions of TMEM131 are less enlightening but detailed here for reference. It is listed as a gene under positive selection in cold adapted snow sheep[325] and Mesolithic Scandinavians humans.[326,327] One paper looking for phosphorylation events that are induced by EGFR stimulation identified TMEM131 as a candidate along with desmoplakin and LAD1.[328] A recent paper using machine learning to study the intersection of genes with missense single nucleotide variants in human genes and proteins with phosphosites effected by SARS-CoV-2 infection, identified TMEM131 as a hit along with other

proteins involved in ER export.[329] TMEM131 was identified as a differentially expressed gene in colorectal cancer.[330] A genome-wide association analysis suggest TMEM131 is a potential risk factor for hypothyroidism.[331,332] Finally in an APEX proteomic screen in HEK293T cells for proteins involved in ER-mitochondrial contact sites, peptides from TMEM131 were identified in both ER membrane and mitochondrial outer membrane datasets.[333] Although this might suggest dual-localisation of TMEM131, due to the length of the protein it is also possible that it reaches a significant way into the cytoplasm from the ER membrane, potentially interacting with other organelles.

4.1.4 TMEM131 structural prediction

At the start of this project some structural information could be predicted about TMEM131 from sequence analysis. Its N-terminus features an unusually long, 59-residue signal peptide. As mentioned previously, sequence homology predicted two or three PapD-like domains in its N-terminal region. Transmembrane helix prediction software predicted either one or two TM helices near the middle of the protein with the former less hydrophobic. The C-terminal tail was predicted to be entirely unstructured.

Keen to understand the protein better I ran the sequence through Consurf, a bioinformatics tool that uses the phylogenetic relations between homologous sequences to estimate the conservation of amino acid positions. This program outputs the protein sequence with a relative conservation score that is binned from 1 (most variable) to 9 (most conserved). In its graphical outputs these are conventionally coloured blue to maroon respectively. These scores are normalised for each input so the entire range is used. This means the results show which residues of an input protein are more conserved relative to the rest of that protein and not how conserved the residues are relative to other proteins in the genome. The results for human TMEM131, shown in figure 4.3, show a higher level of conservation in the N-terminal regions of the protein, presumably containing the PapD-like domains. The signal peptide is poorly conserved, though the predictions here were made with very few comparable sequences. The two putative TM helices in positions 1091-1111 and 1118-1138 appear quite highly conserved as is a PRP sequence between them that would act as a helix breaking motif. The C-terminal tail of the protein is much less conserved though there are notable short sequences of high conservation. Counting only the residues after the second putative TM helix, 132 of the 743 residues are serine with roughly equal number of positively charged (Arg + Lys) and negatively charged (Asp + Glu). Closer inspection of the sequence finds a proline rich sequence between 1300-1330 that is poorly conserved and a Lys-rich region at 1380-1480 that is more conserved.

In July 2021 a collaboration between the company Deepmind and EMBL's European Bioinformatics Institute (EMBL-EBI) released structural prediction for the human proteome made by AlphaFold, an advanced machine learning algorithm.[338,339] The predictions as first published had a number of limitations: most significantly the algorithm was only trained on monomeric proteins and is ignorant of non-protein components such as glycosylations or the lipid membrane of transmembrane proteins. Nonetheless, the predicted structures have been widely received as believable and useful by the scientific community. The AlphaFold prediction for the TMEM131 is shown in figure 4.4.

Not knowing how to model the disordered C-terminal tail, the AlphaFold program has randomly coiled it around the rest of the structure. The N-terminal sequence meanwhile is modelled as eight Ig-like domains each with a sandwich-like structure made two sheets of antiparallel beta strands. The 6th of these domains is unusual as it appears to have a long helix interrupting the beta strands. This helix loops down to contact the face of the 7th domain and interacts with a loop on domain 3 via salt bridges

Figure 4.3: Consurf predictions of TMEM131 residue conservation and function.
The human TMEM131 sequence was compared to 200 identified homologs with by the Consurf server[334–337] run with the sequence only mode. More conserved residues are shaded in pink whilst less conserved residues are shaded blue. Exposed and buried residues are indicated by e (orange) and b (green) respectively. Conserved and exposed residues are labelled as functional (f, red) whilst conserved and buried residues are described as structural (s, blue).

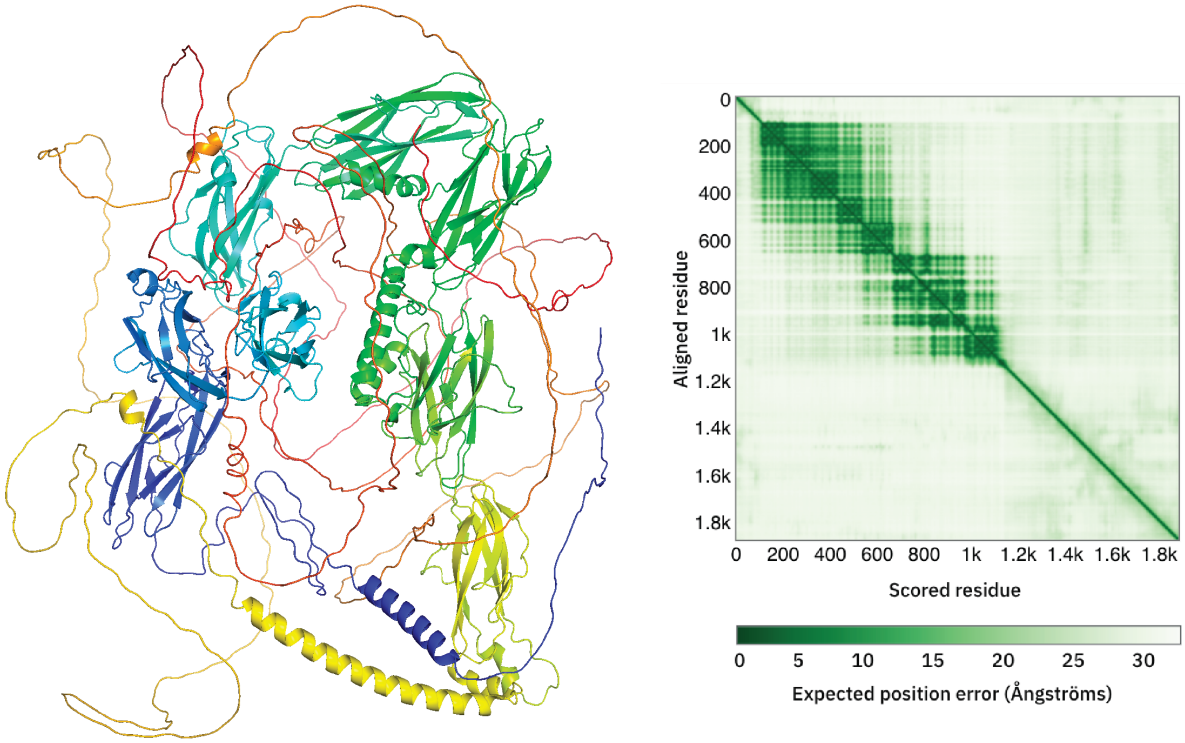

Figure 4.4: The AlphaFold model of human TMEM131.
The structure has been coloured from its N-terminus including the signal peptide (dark blue) to the C-terminus (dark red) via green, yellow and orange. The accompanying plot shows the predicted alignment error which is the expected position error at residue x, when the predicted and true structures are aligned on residue y.

between charged residues. These details are highlighted in figure 4.5C. This gives the protein the shape of a hook or a question mark with the helix helping to anchor the structure closed. The predicted alignment error plot is also produced by the AlphaFold program and provides a measure of the uncertainty in the position of residue X if the model is aligned on position Y. For human TMEM131, the program predicts a relatively low error between pairs of residues within domains 1-5 and similar for pairs of residues within domains 6,7 and 8. This would suggest that these two halves of the protein are somewhat able to move relative to each other with a hinge between domains 5 and 6. Comparing the human model to the models for other model organisms (shown in 4.5B) would seem consistent with flexibility at this point in the sequence. Of the two predicted TM helices annotated in the Uniprot entry (positions 1091-1111 and 1118-1138) only the second is modelled as a long helix and it extends to 1154. I applied the colour-graded conservation scores from the Consurf prediction onto the AlphaFold model to see if conserved residues neatly mapped onto the modelled domains (figure 4.5A). Inspection suggested that the first few domains were more conserved than the latter domains. The latter domains seem somewhat striped as residues with side chains facing into the domain seem more conserved than those facing out. The helix mentioned above is not notably conserved though the salt bridges appear to be.

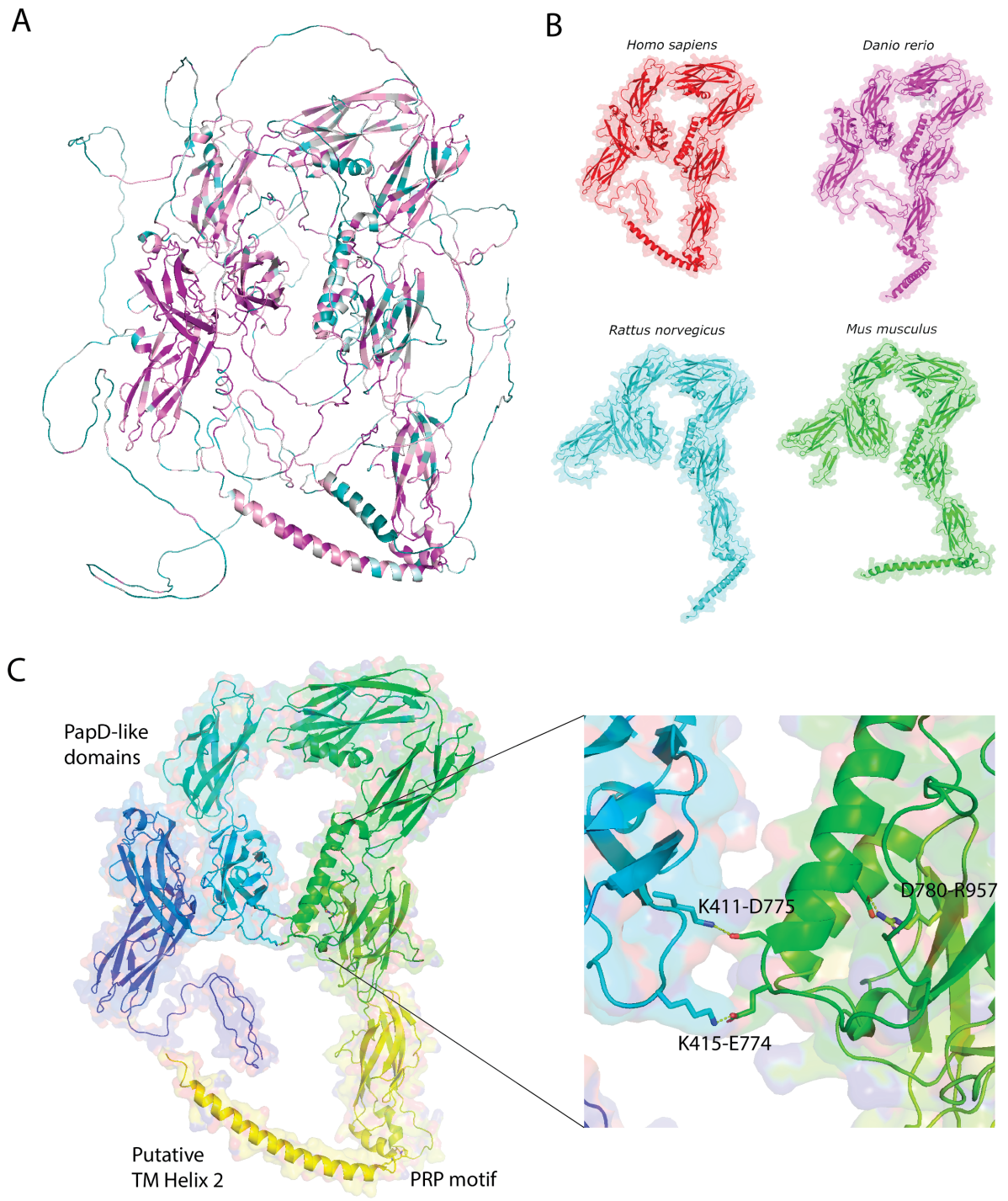

Figure 4.5: Comparisons between the AlphaFold model of human TMEM131 to that of other organisms.
A: The human TMEM131 model coloured with by the conservation factors of Consurf (cyan most variable and maroon most conserved).
B: The AlphaFold model of human TMEM131 (red) with the models for TMEM131 from *Danio rerio* (pink), *Rattus norvegicus* (cyan) and *Mus musculus* (green). For simplicity, the flexible signal peptides and C-terminal tails of the proteins have removed and the models aligned in pymol using the cealign command.
C: The AlphaFold model of human TMEM131 with the signal peptide and tail removed and side chains presented as a surface. The insert shows the residues forming the salt bridge between the loop of domain 3 and the helix coming out of domain 6.

4.2 Aims

Relatively little is known about TMEM131 cell biology in terms of localisation or function. Although there is published evidence for a role in collagen secretion, there is also evidence implying a role in T cell development particularly during β-selection. Computational predictions are inconsistent about the topology of TMEM131 and previous experiments in the James's lab have found unusual proteolysis in the C-terminal tail.

This project aims to:

- Investigate how the structure of TMEM131 determines its localisation
- Determine how and where protein cleavage occurs in the TMEM131 tail
- Use proteomic techniques to identify TMEM131 interaction partners.

Experiments to address these aims are described in separate results sections but will discussed together in section 4.7.

4.3 Methods

4.3.1 GFP pulldown and 3C protease digestion

A single PCR reaction was used to clone the HRV 3C protease site onto the N-terminus of GFP such that it could be inserted into vectors encoding TMEM131-GFP and TMEM131dC-GFP using BamHI/NotI restriction enzymes, replacing the existing fluorophore.

Four T25 flasks were transfected with full-length TMEM131-GFP, TMEM131-3C-GFP, TMEM131dC-GFP and TMEM131dC-3C-GFP whilst one flask was not transfected. After 48h, flasks were suspended in 1ml of trypsin then made up to 4ml with media. 0.5ml of suspension was washed three times in cold PBS then fixed for flow cytometry. The remaining 3.5ml was lysed for 30mins on ice in 200µl of lysis buffer (150 mM NaCl, 1% NP-40, 100µM NaVi, 1x protease inhibitor cocktail (Life technologies), 50 mM Tris-Cl, pH 7.4). This lysate was then diluted with 300µl of cold GTA dilution buffer (10mM Tris/Cl pH7.5, 150mM NaCl, 0.5mM ETDA).

Five tubes of 25µl of GFP-trap agarose beads (Chromotek) were equilibrated by three washes with 500µl of GTA-dilution buffer. Each of these washes consisted of centrifugation at 2,500g for 5 minutes at 4C before removing as much supernatant as possible without disturbing the pelleted beads. After saving 50µl of each samples as the diluted lysate fraction, each lysate was incubated with the beads for 1 hour at 4°C with gentle rocking. The beads were then pelleted and washed four times with GTA wash buffer (10mM Tris/Cl pH7.5, 150mM NaCl, 0.5mM ETDA, 0.05% NP40). A final wash in 500µl of GTA dilution buffer was then used to transfer the beads to a fresh tube. The beads were then incubated in approximately 20µl of dilution buffer with 1µl of PreScission protease overnight at 4°C with gentle rocking.

The next day the supernatant was collected from the beads (protease eluted fraction). Proportional volumes of LDS sample buffer with DTT were added to each fraction and samples heated to 70°C for 10 minutes. Fractions were run on polyacrylamide gels which were analysed by western blotting or labelled with SYPRO Ruby protein stain according to manufacturer's instructions. SYPRO Ruby gel staining can either be done using the rapid (90min) protocol or overnight. Overnight staining produces

the widest dynamic range between high and low abundance proteins whilst the rapid protocol leads to a lower background and fewer speckling artefacts.

4.3.2 Scale up of GFP-pulldown and 3C protease treatment for mass spectrometry

To scale the experiment up for mass spectrometry, I transfected two T175 flasks of HEK cells with HA-TMEM131-GFP and HA-TMEM131-3C-GFP, increasing the volumes of DNA, GeneJuice and serum-free media accordingly. After 48h, these were suspended with trypsin and lysed for 30mins on ice in 1ml of lysis buffer. After centrifugation at max speed to remove the nuclear fraction each supernatant was transferred into a 15ml falcon tube and diluted with 3ml of GTA dilution buffer. 20µl was saved as the lysate fraction then this liquid was incubated at 4°C rocking with 200µl of GFP-trap agarose beads conditioned in GTA dilution buffer. After 90 minutes the beads were pelleted and 20µl collected as the "unbound fraction". The beads were washed three times in GTA wash buffer before three more washes in GTA dilution buffer. After the final wash, the buffer was removed and 200µl of GTA dilution buffer added with 2µl of PreScission protease. The beads were left with protease overnight.

The next day the GFP-trap agarose beads were pelleted. 20µl was saved as the "Protease eluted" sample whilst the rest added to 100µl glutathione-magnetic beads that had been acclimatised with GTA dilution buffer. Proteins were eluted from the GFP-trap agarose beads using 70°C heat in sample buffer (the "Bead fraction"). Meanwhile the supernatant was incubated with magnetic glutathione beads for 90 minutes at 4°C. The liquid was separated from the magnetic beads using a strong magnet. 60µl was saved as the "Cleared sample" and the rest transferred into 10kD molecular-weight cutoff dialysis columns (Slide-A-Lyzer MINI Dialysis Devices 10K MWCO; Thermo Scientific). These columns were left in 1.6L of 40mM Ammonium bicarbonate buffer at 4°C. After 24h the buffer was replaced with another 1.6L of 40mM Ammonium bicarbonate. After another 8h of dialysis the liquid was transferred to mass spectrometry vials for Native Mass spectrometry analysis.

The lysate, unbound, protease eluted, beads and cleared samples were denatured using appropriate volumes of sample buffer then run on a 15 well 4-12% BisTris gel. This was blotted and labelled with anti-HA and anti-GFP antibodies. Two lanes of 15µl of each of the protease eluted and bead samples for each construct were run on a parallel 4-12% BisTris gel that was labelled with SYPRO Ruby using the longer overnight protocol. The gels are presented in figure 4.28.

4.3.3 EndoH and PNGaseF treatment

T150 flasks of HEK were transfected with HA-TMEM131-GFP, HA-TMEM131dC-GFP and pTα HA-TCRβ-GFP with CD3$\gamma\delta\epsilon\zeta$. These were suspended with trypsin then lysed in 1ml of lysis buffer for 30 minutes at 4°C. The lysate was centrifuged for 10 minutes at max speed then 900µl of supernatant collected and diluted with 2ml of GTA dilution buffer. Each diluted lysate was then incubated for 1h at 4°C with 60µl of conditioned GFP-trap agarose beads (Chromotek). Beads were washed four times with 1ml of cold GTA wash buffer. After the final wash the liquid was removed then proteins were eluted from the beads by heating to 100°C for 10 minutes in 4µl of 10X Denaturation buffer (NEB) and 36µl of water. After cooling on ice, 10µl of the liquid was added to three PCR tubes. The first tube was treated with the enzyme EndoH through the addition of 2µl l of 10X Glycobuffer 3 (NEB), 6µl of water and 2µl of EndoH (NEB). The second tube was treated with PNGaseF through the addition of 2µl of

10X Glycobuffer 2 (NEB), 2µl of 10%NP-40, 6µl of water and 1µl of PNGaseF (NEB). The final tube ("control") just had 10µl of water added. All three tubes for each construct were then incubated at 37°C for 1h. Finally, samples were heated to 70°C for 10 minutes with 8µl of LDS sample buffer with DTT then run on PAGE gels as described previously.

4.3.4 Wet blotting

1L of transfer buffer was prepared containing 80ml of 12.5X buffer (312.5mM Tris, 2.4M glycine, 1.25% SDS; pH8.3), 720ml of water and 200ml of 100% methanol. This buffer was cooled to -20°C. The membrane was briefly washed in 100% methanol. The polyacrylamide gel was run and trimmed as normal.

The gel, membrane, sponges and filter paper were then soaked in the cold transfer buffer for 20minutes on ice in the 4°C cold room. After equilibration the transfer was set up in the order: positive electrode, sponge, 2x filter paper, membrane, gel, 2x filter paper, sponge, negative electrode with effort taken to remove air bubbles. The transfer tank was then filled with transfer buffer then run for 60 minutes at 100V. The blot was then cut, blocked and labelled as normal.

4.3.5 BioID assay with TMEM131

T175 flasks were transfected with BioID2-TMEM131-GFP, BioID2-TMEM131ntd-GFP, TCRα BioID2-TCRβ-GFP and TMEM131ntd-GFP. After waiting 24h for the constructs to be expressed, biotin was added to the media to a final concentration of 50µM. Cells were treated for 16h overnight.

Each flask was washed with PBS then resuspended using 6ml of trypsin. The cells were harvested with 15ml of fresh media then pelleted and washed three times with 10ml of cold PBS by three centrifugation steps at 800g. A small volume of each sample was used to confirm expression of the GFP via flow cytometry. Cells were lysed in 2ml of lysis buffer for 30minutes at 4°C. 20µl samples of these lysates were run on a 4-12% BisTris gel (shown in figure S21). The rest was incubated overnight at 4°C with shaking with 50µl of streptavidin-agarose beads (Thermo Scientific) that had been pre-conditioned with lysis buffer.

Beads were pelleted then washed with twice with 1ml of Lysis buffer, once in 2ml of 1M KCl, once in 2ml of 0.1M $NaC0_3$, once in 2ml 2M urea in 10mM TrisHCl pH 8.0 and twice in 1ml of lysis buffer without detergent. The beads were then processed with the on-bead digest protocol detailed in methods section 2.6.1.

4.4 Results: TMEM131 structure and localisation

4.4.1 The TMEM131 C-terminal tail appears to be cleaved at multiple sites

As mentioned in section 4.1.2, previous work in the James lab identified TMEM131 as a possible pre-TCR interaction partner and found some unusual behaviour relating to cleavage of the C-terminal tail of the protein. I was keen to repeat these experiments before investigating potential mechanisms.

TMEM131 with a HA tag on the N-terminus and GFP on the C-terminus was expressed in HEK293T cells along with a range of similar constructs with truncated C-terminal tails. A well expressing pTα HA-TCRβ-GFP construct (without CD3 chains) was included as a control of known molecular weight. A third of these cells were incubated with anti-HA antibodies at 37°C to see if the constructs were able to reach the cell surface whilst the rest were lysed and analysed by western blotting. The blot for this experiment is shown in figure 4.6 and the flow cytometry in figure 4.7.

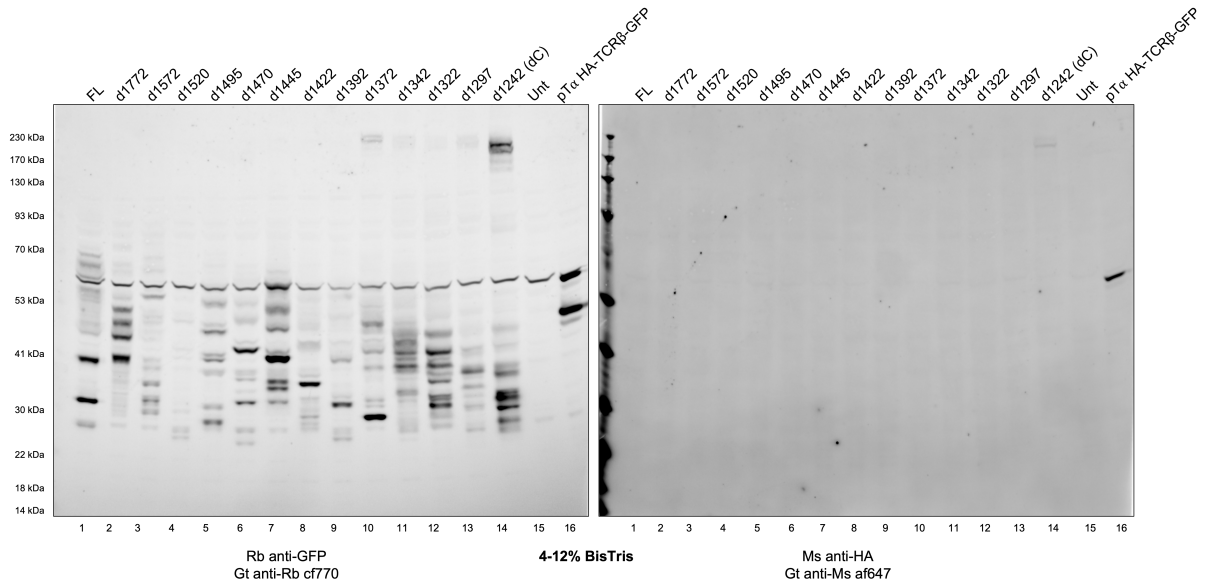

Figure 4.6: The TMEM131 tail is cleaved at multiple sites.
Cell lysate from cells transiently expressing TMEM131 constructs with N-terminal HA and C-terminal GFP tags was run on a 4-12% BisTris gel and blotted using the iBlot2 apparatus. The blot was stained with Mouse anti-HA and Rabbit anti-GFP primary antibodies then goat anti-Mouse af647 and goat anti-Rabbit cf770 secondary antibodies.

All TMEM131 truncations seem to produce a series of GFP-linked fragments ranging from about ~27kDa (the molecular weight of the GFP alone) to ~72kDa, presumably GFP attached to a considerable proportion of the TMEM131 tail. Close inspection of the bands revealed a few patterns. Most obvious was a series of bands between lanes 5-10 in which a prominent GFP-positive band seems to decrease in mass from around 45kD in lane 5 (d1495) to about 28kD in lane 10 (d1372). This is consistent with the tail being cleaved at a specific site in the tail that the GFP was moved closer towards as the tail was truncated from the C-terminal end. A similar series of bands is present between lanes 10 and 14. A few bands meanwhile do not appear to change size between lanes, for instance a band at around 32kDa between lanes 6 and 8. This would be consistent with the protein being cleaved at a set distance away from the GFP tag. Taken together this suggests a complex combination of distance and sequence effects.

There is a band at 60kD in all lanes including the untransfected HEK implying the anti-GFP polyclonal antibody recognises another protein expressed by HEK cells. Until we changed our supplier of the GFP-antibody, this band is present whenever cell lysate is blotted but not when GFP-tagged proteins are pulled down with anti-GFP beads.

At the top of the gel, the only higher molecular weight bands visible are in lanes 10-14. The shortest truncation d1242/TMEM131dC produced the strongest bands. These bands are likely the entire sequence of the proteins complete with HA and GFP tags. The double banding here might be the result from protein glycosylation consistent with an ER protein. Strangely the HA antibody only seems to label the lower of the two GFP-positive bands visible at the top of the dC lane. The same is true however in

the pTα TCRβ-GFP control lane where the anti-HA antibody appears to label the highest of the three bands present in the cf770-channel. This might imply some error in the staining with these antibodies. The IR700 channel used for af647 has a higher background that the IR800 channel used for the cf770 dye so it is possible more HA-staining would be apparent with better signal to noise.

In the TMEM131 constructs with tails beyond residue 1422, bands at the top of the gel consistent with the mass of the double-tagged construct are not observed and HA staining is absent. The simplest explanation would seem to be that the protein is cleaved somewhere in the C-terminal tail and the N-terminal fraction degraded leaving only the GFP-tagged fragments. Truncating the tail seems to stabilise the double-tagged construct though fragments are still present in the shortest truncation used here. The apparent degradation of the N-terminal part of the protein combined with a lack of understanding of the structure of these domains, made us reluctant at this stage to make modifications to this part of the molecule.

We were keen to learn exactly where in the tail the protein was cleaved as this might reveal the cleavage mechanism. One method might be to estimate the mass of the GFP^{+ve} bands and compare these to the protein sequence. The distance travelled by a peptide on a gel of consistent percentage is approximately proportional to a peptide's mass once the peptide has been denatured by the dodecyl sulphate detergent. However, there is variation in displacement based upon the number of charged residues and non-peptide additions such as glycosylation and phosphorylation modifications. This means getting precise mass measurement from western blotting is not trivial to do. An added complication here is the use of a variable percentage polyacrylamide gel. Although this allows the smallest and highest-GFP bands to be seen on the same blot, it makes mass estimation even more challenging.

Looking at the lane for the full-length construct, the highest molecular weight GFP^{+ve} band appears to be around 70kD. This would be consistent with the mass of the GFP-tag plus a segment of tail ranging as far as the end of the Lys-rich region at position 1475. The smallest band would seem consistent with the GFP tag alone without much of the tail attached. Meanwhile in the lane for TMEM131dC, the GFP-positive band at 45kD would seemingly be larger than GFP attached to the entire sequence after the putative TM helices (38.4kD total). At a minimum however, the pattern did not seem consistent with TMEM131 being cleaved at only a single or a couple of sites. Inexact mass measurements, combined with the number of bands and complex pattern, made it unfeasible to determine exact cleavage sites with this method.

4.4.2 The full-length TMEM131 protein does not leave the secretory pathway

The same cells that were lysed and run on the western blot above were labelled with anti-HA antibodies for 30 minutes in 37°C media. Internalisation of the antibody would indicate that the protein reaches the cell surface whilst a lack of staining would imply the protein does not progress through the secretory pathway. I used MATLAB to bin cells by their GFP-expression level and plotted the mean antibody binding for each bin. This allowed the flow cytometry results for each truncation to be overlaid (figure 4.7). Surprisingly all TMEM131 constructs shorter than the full-length protein (dark red) seemed to reach the cell surface quite readily in the highest expressing cells. This might imply that the protein is normally retained in the secretory pathway but this mechanism is exhausted at higher expression levels.

Differences in localisation of the full-length and the truncated TMEM131 constructs can also be seen via

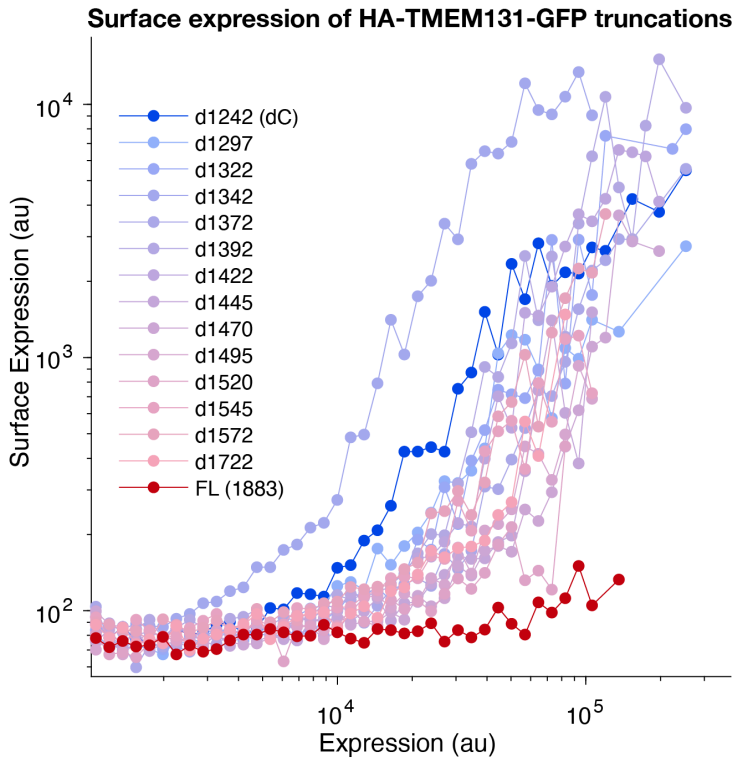

Figure 4.7: Surface expression of TMEM131 truncations.
HEK cells expressing HA and GFP-tagged TMEM131 constructs were suspended using trypsin then incubated with 1:200 anti-HA antibody for 30 minutes in 37°C media. Samples were run on the LSRII. Data was sorted into 50 log-distributed bins between 10^3 and $5x10^5$ along the "B 530/30" axis. The mean value in the "R 685/35" channel was plotted on the y-axis in the middle of each bin. Bins with less than 10 events were excluded. Lines were then coloured from the construct with the shortest C-terminal tail (blue) to full-length construct (red).

microscopy (figure 4.8). Cells expressing TMEM131 with a C-terminal GFP tag appear to have GFP delocalised in the cytoplasm. As the signal is diffuse it often appeared rather weak and out of focus. Cells also had GFP-foci that appeared co-localised with the ER marker. No GFP at the cell surface was apparent. Meanwhile cells expressing GFP-tagged TMEM131dC showed more numerous foci of GFP and the fluorescence appeared localised to the ER and the cell surface.

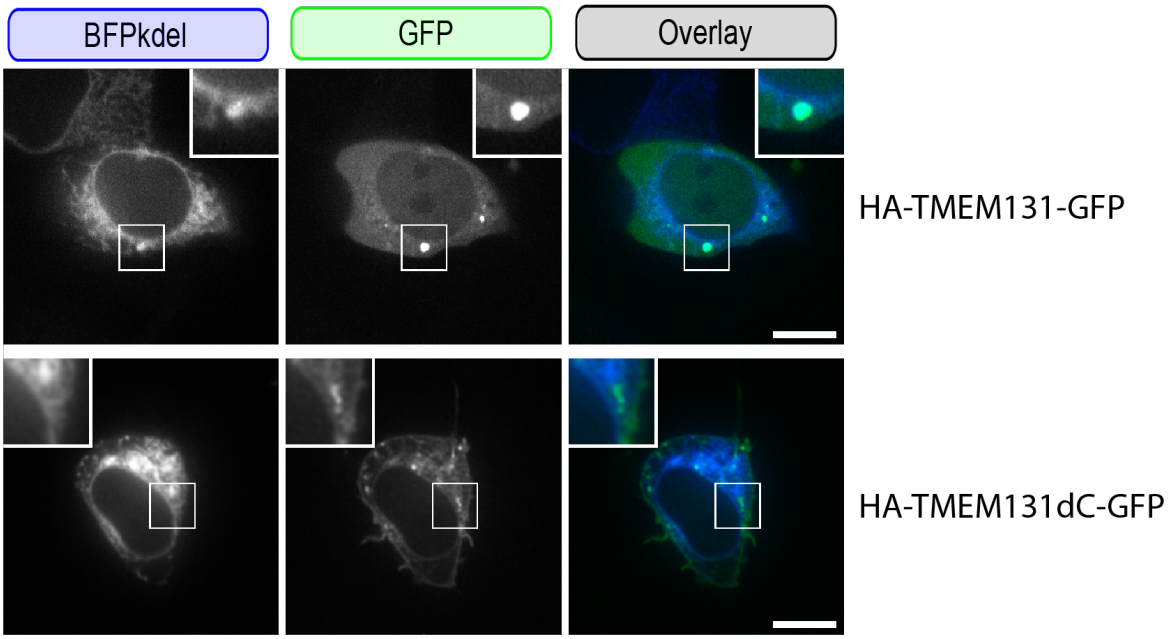

Figure 4.8: Microscopy images of TMEM131 and TMEM131dC.
Representative images of HEK cells expressing HA and GFP-tagged TMEM131 constructs and BFPkdel ER marker under 100X magnification. Inserts show a 2x zoom of selected area.

My hypothesis at this point in the investigation was that a sequence in the tail was acting as an ER retention motif and that cleavage of the tail was a mechanism for releasing the protein from this retention. I wanted to learn which exact sequence in the protein was causing the differences in localisation. The internalisation of anti-HA antibodies showed that TMEM131d1772 was present at the cell surface but full-length TMEM131 was not. This would seem to imply that a sequence involved in ER retention is located in the very C-terminus of TMEM131. I fused the final 162 amino acids of the TMEM131 tail onto the dC construct to see if this would restore retention in the secretory pathway (TMEM131dC+End).

I also searched the tail sequence for classical ER retention motifs. A typical ER retention motif is a KDEL or KKXX sequence at the very C-terminus. Although neither are present at the C-terminus, KKXX is present three times in the Lys-rich region in the TMEM131 tail between 1379 and 1475. These could feasibly become the new C-terminus when the protein is cleaved. To see if this region is involved in ER retention, I designed a construct missing just the Lys-rich region of the tail between 1375 and 1475 (TMEM131dLysRich). These two new constructs were compared to additional TMEM131 variants described below.

4.4.3 Protein cleavage is intrinsic to the TMEM131 C-terminal tail

The TMEM131dC construct was truncated at a convenient SpeI site in the gene rather than at a particular feature in the protein tail. Whilst this truncation seemed to increase stability of the HA^{+ve}/

GFP^{+ve} band consistent with the intact species, it still appeared to produce smaller GFP^{+ve} bands. In this construct there is still 100 amino acids between the GFP and the end of the putative transmembrane section of the protein. To see if this small fragment of the tail was important three new constructs were designed. These contained the N-terminal domains of TMEM131 connected to: neither of the putative TM domains (NTD), only the one at position 1091-1111 (NTD-TM1) or with both helixes (NTD-TM1-TM2). These are effectively even shorter truncations than TMEM131dC and are equivalent to d1082, d1111 and d1155 respectively.

To see if the cleavage of the TMEM131 tail was intrinsic to the protein itself we made a construct with the tail of TMEM131 fused to the end of CD86, a well-expressed membrane protein (CD86TM-131tail). Conversely a construct was made with the TMEM131 N-terminal domains with the transmembrane and cytoplasmic domain of CD86 (131ntd-CD86tm). We also designed a construct missing the first 382 amino acids of the N-terminal domains of the protein which had the highest homology to PapD (TMEM131dPapD). This was designed before the AlphaFold model was available so may not correspond to any meaningful domain boundary in the protein. Lysates of cells expressing these constructs were analysed via western blotting as before (figure 4.9).

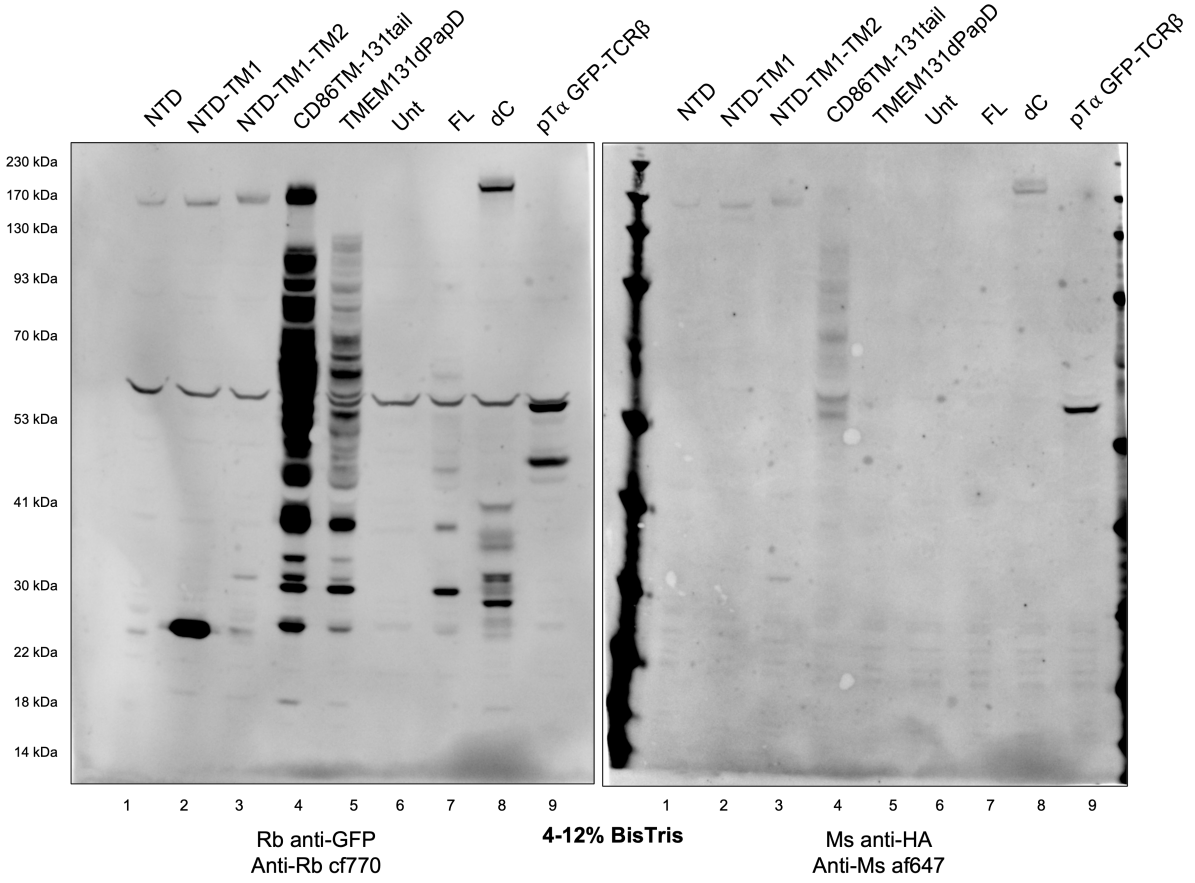

Figure 4.9: Banding is not observed in TMEM131 constructs lacking sequence beyond the putative TM domains. Banding is observed if the tail is attached to CD86.
Cell lysate from cells transiently expressing TMEM131 constructs was run on a 4-12% BisTris gel and blotted using the iBlot2 apparatus. The blot was labelled with the indicated antibodies.

All three tailless-constructs have a high molecular weight band that is labelled by both GFP and HA antibodies. Whilst some GFP-positive bands are visible at the bottom of the gel in these lanes, especially in the lane for the NTD-TM1 construct, there is not the same banding pattern observed with the full-length protein or TMEM131dC. These small molecular weight bands could be due to translation

initiation within the protein or products of misfolded protein (more likely in NTD-TM1 due to the presence of the hydrophobic helix). There is a strange band at about 30kD in the lane for NTD-TM1-TM2 that appears both GFP and HA-positive. This I have no explanation for.

The GFP-positive bands seen in the lanes for CD86-131tail and TMEM131dPapD have a similar pattern to the full-length construct though with different expression levels. This would suggest that the cleavage is an intrinsic property of the tail sequence and relatively unaffected by the protein it is attached to. For the CD86-131tail construct there is also a series of HA-positive bands presumably the N-terminal part of the protein with a variable amount of the tail still remaining. Again, this would seem to imply that the tail is cut at multiple sites as a single cut site would produce only a single HA-band.

4.4.4 Truncation of the TMEM131 C-terminal tail increases surface expression

The dLysRich and dC+End protein variants were expressed and blotted via the same procedure (figure 4.10). As HA-TMEM131dC-GFP was detectable at the cell surface via the HA tag, I was interested in whether this fraction of the protein was intact including the GFP tag or had already been cleaved. I decided to express the protein then crosslink surface proteins using the reagent BS3. If the GFP was attached when the protein was at the cell surface then the GFP-band would be smeared through the irreversible crosslinks to other proteins. This was also performed with the HA-CD86tm-131tail-GFP construct that had much stronger surface expression.

The GFP-positive banding pattern for the dLysRich construct was very similar to original sequence consistent with the cleavage sites of the protein being downstream of the Lys-rich region. The dC+End had a different pattern of bands than the dC construct but the differences could be explained by the increase in mass corresponding to the final amino acids of the protein.

Comparing TMEM131dC with and without surface crosslinking suggests a faint smearing of the anti-GFP signal. This was not detectable in the anti-HA channel. For the CD86 fusion protein smearing of the GFP^{+ve} band with the highest molecular weight was more noticeable. The HA^{+ve} bands appeared to disappear entirely perhaps indicating the HA-tagged species were largely crosslinked at the cell surface. Alternatively, this might reflect a big difference in the amount of protein loaded.

The same cells run on westerns above were labelled with anti-HA antibodies to check for surface expression (figure 4.11). Once again, the full TMEM131 protein does not seem to reach the cell surface whilst TMEM131dC does. The dC+End variant has about a log reduction in surface staining suggesting that the very terminal residues might be important for retainment in the secretory pathway. The dLysRich construct has slightly higher surface staining than the full-length construct but not as much as dC implying this region is of some importance.

Fusing the tail of TMEM131 to CD86 does not prevent the protein from reaching the cell surface but replacing the CD86 external domains with the TMEM131 N-terminal domains does not allow surface expression. Only the NTD variant with both putative TM helices was able to reach the cell surface. This suggests that either the second helix is transmembrane or that both helices are required together.

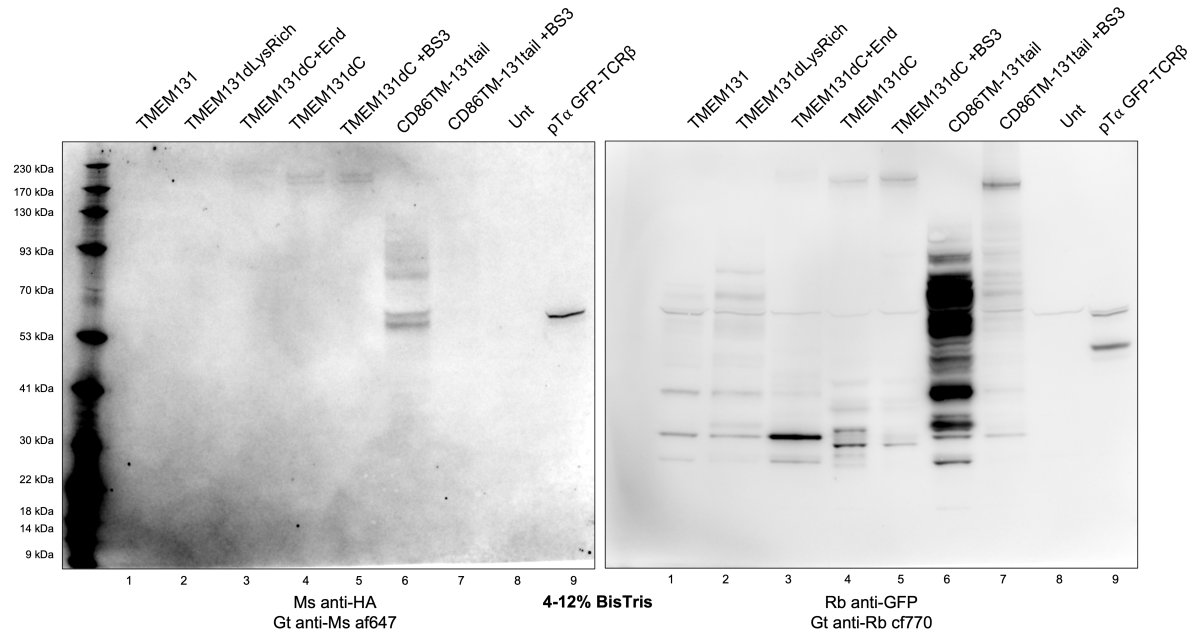

Figure 4.10: Banding is not greatly affected by removing Lys-rich region or fusing the final residues to TMEM131dC. Crosslinking suggests a detectable fraction of the surface TMEM131dC is uncleaved.
Cell lysate from cells transiently expressing TMEM131 constructs was run on a 4-12% BisTris gel and blotted using the iBlot2 apparatus. The blot was labelled with the indicated antibodies. Two wells of TMEM131dC and CD86-131cyto were transfected before being mixed and divided equally. One tube was then crosslinked with BS3 with the protocol in section 3.3.3.

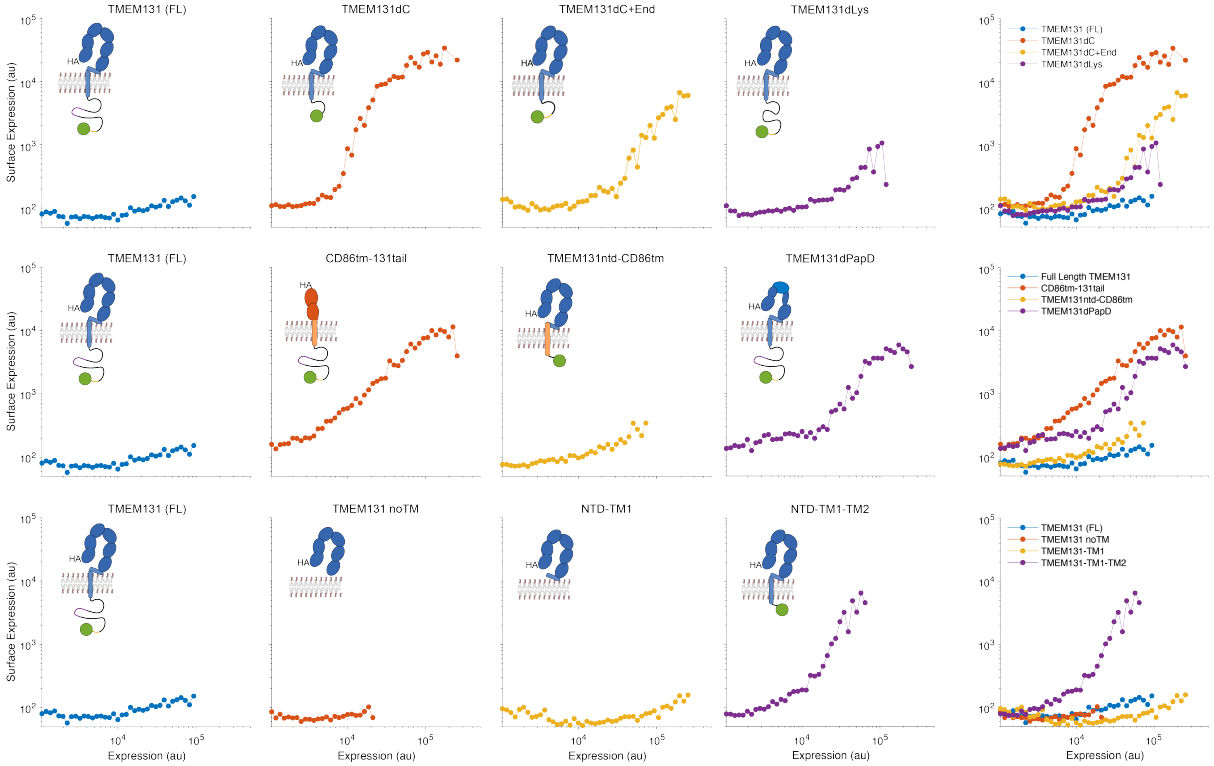

Figure 4.11: Surface expression of TMEM131 tail variants.
HEK cells transiently expressing HA and GFP-tagged TMEM131 constructs were suspended using trypsin then incubated with 1:200 anti-HA antibody for 30 minutes in 37°C media. These were analysed by flow cytometry then data sorted into 50 log-distributed bins between 10^3 and 5^5 along the 'B 530/30' axis. The mean value in the 'R 685/35' channel was plotted on the y-axis in the middle of each bin. Bins with less than 10 events were excluded. The diagrams within each plot show a representation of each construct based on the sequence and AlphaFold model. The TMEM131 N-terminal domains are shown in blue and the C-terminal GFP tag in green. The disordered tail of the protein is shown in black with Lys-rich and end regions in purple and yellow respectively. The exact structure of the putative TM helices relative to the membrane is unknown.

4.4.5 The apparent disappearance of the higher molecular weight bands of TMEM131 is an artefact of gel blotting

The number of GFP^{+ve} bands and the pattern resulting when the C-terminal tail was truncated were the most surprising finding from the western blots of TMEM131. Continued efforts to find the site or sites of cleavage and the protease responsible will be discussed later. The disappearance of the higher molecular weight bands for the longer TMEM131 truncations however is just as curious. If a protein with a HA-tag on the N-terminus and a GFP-tag on the C-terminus is cleaved then both HA and GFP-tagged fragments should be produced. This appeared to be the case with the CD86-131tail fusion construct. The lack of HA-tagged fragments for TMEM131 however suggests that the N-terminal product of cleavage was either degraded or undetected via western blotting.

How could this be reconciled with the internalisation of labelled anti-HA antibodies by TMEM131 constructs with truncated tails of length d1772 or shorter? This implies there was a significant fraction of the HA-tagged N-terminal domains present that did not appear on the western blots. Some of this might be attributable to the weaker signal in the af647 channel combined with sub-optimal staining from the anti-HA antibody. However, staining did not appear to correlate with the expression of the constructs measured by internalisation of the antibody or GFP fluorescence.

Whilst 4-12% BisTris gels are useful for fitting high and low molecular weight proteins on the same blot, the resolution is not ideal for higher molecular weight species, which the HA-linked N-terminal fragments would be. I re-used the lysate samples from previous experiments and ran two lower percentage pre-cast gels (8% BisTris and 3-8% TrisAcetate). These gels were blotted and the blot labelled with anti-HA and anti-GFP antibodies like the westerns above. Both are shown in figure 4.12.

In both gels there are still few bands visible in the anti-HA/af647 channel and these all corresponds to the constructs with the very shortest C-terminal tails (dC/d1242 and the NTD constructs). The image in the GFP/cf770 channel is more interesting. In both gels, lower molecular weight proteins have travelled further and some of the smallest GFP-tagged fragment have travelled off the end. The 8% Bis-Tris gel is similar to the 4-12% BisTris gel as high-molecular weight, GFP-tagged bands are observed for the constructs from d1242 until d1372 but not beyond. On the 3-8% TrisAcetate gel however, run with the same samples, these bands continue until the second longest construct (d1772 which is only 100 amino acids shorter than the full-length construct). No band for the full-length construct is apparent even if the gel is overcontrasted.

Of interest, in the TrisAcetate gel there appears to be a faint but distinct band above 240kD in the final lane containing TMEM131dC-GFP protein cross-linked at the cell surface with BS3. This suggest TMEM131dC is interacting with a specific protein at the surface as non-specific crosslinks would produce a smear.

As these gels were all run with the same samples and blots were stained with the same antibodies, we considered whether our western blotting procedure was leading to a loss in signal specifically for higher molecular weight species. The most likely issue is with the gel-transfer system as higher molecular weight proteins not only migrate more slowly as a gel is run but also migrate more slowly out of the gel during transfer. The blotting protocol for the semi-dry iBlot2 I had been using thus far transferred proteins at 20V for 1 min, 23V for 4 min then 25V for a further 2 min for 7 minutes of total transfer time. This protocol is recommended for the transfer of proteins from 30-150kD which encompasses most of the GFP bands observed. Increasing transfer time can increase the transfer of higher molecular weight proteins whilst loosing signal for smaller proteins. I compared different blotting conditions with the same set of

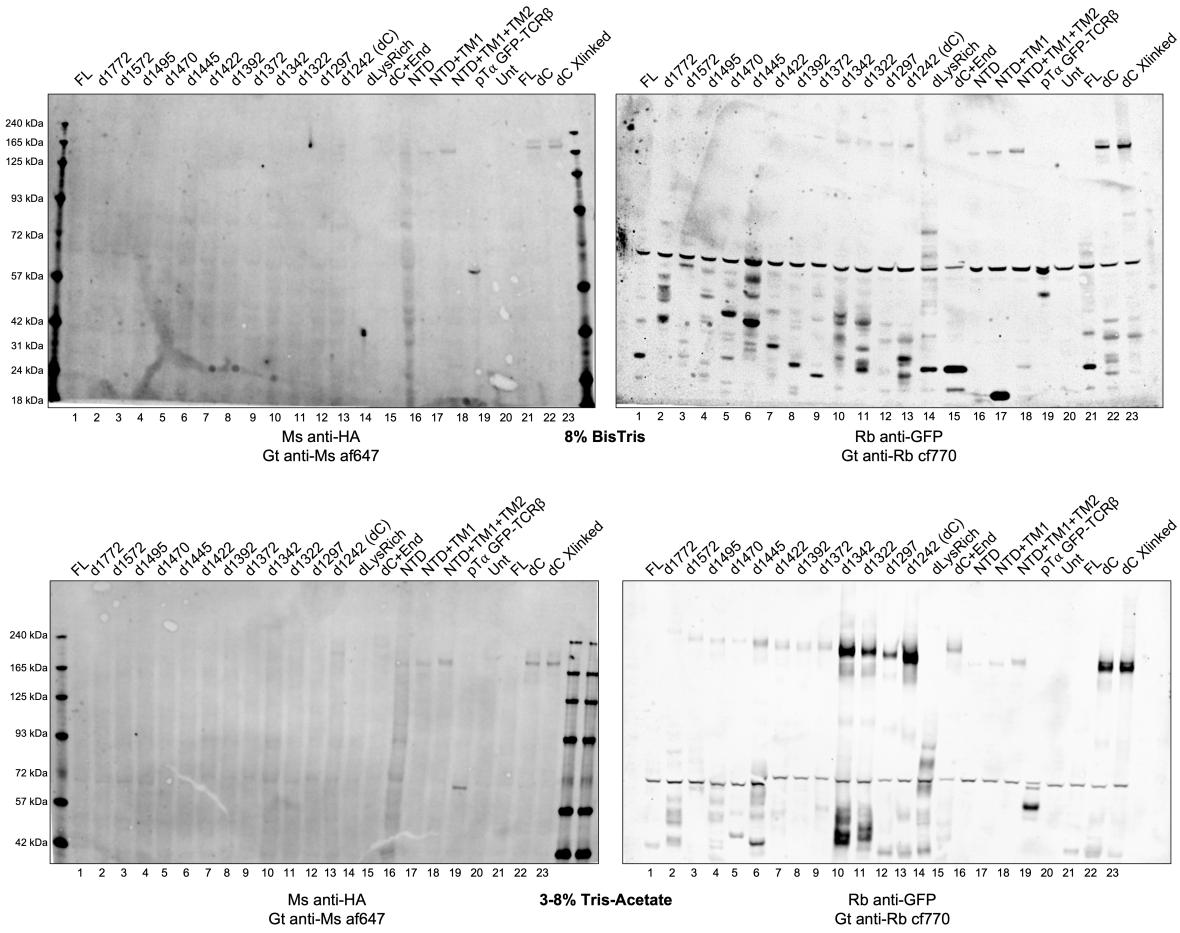

Figure 4.12: Higher molecular weight TMEM131 bands can be resolved with lower percentage gels. The same samples used in experiments above were run on an 8% BisTris gel and a 3-8% TrisAcetate gel. Gels were blotted using the iBlot2 apparatus. The blot was labelled with the indicated antibodies.

samples and found slight improvements upon increasing transfer time (supplementary figure S12).

I concluded that failure to optimise for higher molecular weight protein during gel transfer, combined with weaker signal from the Ms anti-HA/Gt anti-Ms af647 antibody pair, was contributing to the apparent disappearance of bands corresponding to the intact HA-TMEM131-GFP protein. Nonetheless it was still surprising that the HA-tagged N-terminal domains of full-length TMEM131 were not detected at the cell surface via flow cytometry unlike all of the truncated constructs and a band corresponding to the entire construct could not be observed either with a 3-8% TrisAcetate gel or with a 4-12% BisTris gel with a longer transfer time. Whilst higher molecular weight species are blotted more slowly it would seem strange if the cutoff was so sharp to preclude transfer of the full-length construct whilst allowing transfer of truncated constructs that are not that much smaller. I cannot eliminate the possibility that something unique in the very C-terminus of TMEM131 prevents surface expression and decreases the lifetime of the intact, HA-tagged molecule.

4.4.6 TMEM131 is a single transmembrane protein

Whilst TMEM131 appeared to be cleaved somewhere in its C-terminal tail, the actual topology of the protein was not conclusively established. As mentioned the protein was annotated with two TM helices but TM helix prediction algorithms were unclear whether both regions were genuine transmembrane helices or only the second region of the two. The former model would produce a protein with both termini on the same side of the ER membrane, either both in the lumen or both on the cytoplasmic side, whilst the latter model would produce a protein with termini on opposite sides of the ER membrane.

An easy way to determine protein topology is a split fluorescent protein assay.[340] This technique relies on a version of GFP that is split into two parts: a larger fragment consisting of the first 10 strands of the beta-barrel called GFP(1-10) and a smaller fragment consisting of the 11th strand called GFP11.[341] Neither of these two components is fluorescent alone but if both fragments are allowed to complement then the fluorophore can form and mature. This stably joins the two fragments together. To use as an assay for protein topology the 16 amino acid GFP11 tag can be attached to the protein of interest whilst GFP(1-10) can be expressed separately in the cytoplasm or targeted to the ER lumen. Whilst the modifications needed for ER-targeting could feasibly affect the maturation of the fluorescent protein after complementation with GFP11, in practice the termini and loops of the GFP barrel are flexible enough to accommodate minor modifications. Likewise, GFP11 can complement with GFP(1-10) if it is at the N-terminus, C-terminus or within a flexible region of a protein.

I used overlap extension PCR to tag TMEM131 with GFP11 on the N-terminus immediately after the signal peptide but before any PapD-like domains. I tagged the C-terminus with mScar to monitor the protein's expression. I also made a mScar-GFP11 part that could be easily cloned onto the C-terminus of other constructs using restriction enzymes. Separating the GFP11 tag from then rest of the protein via mScar domain adds greater confidence that all these constructs are directly comparable. These GFP11 and mScar-tagged proteins were co-expressed with GFP(1-10) or a version of GFP(1-10) with an ER signal peptide and C-terminal KDEL motif. Flow cytometry of these various constructs is depicted in figure 4.13.

The first row shows that GFP(1-10) or ER-targeted GFP(1-10) do not produce more GFP fluorescence than untransfected cells in the absence of GFP11. I designed three control constructs of known topology to establish that the assay worked: a cytoplasmic mScarGFP11 and pre-TCR variants with GFP11 on the N-terminus or C-terminus. The cytoplasmic mScar gave a very strong signal in both "YG 610/20" and

"B 530/30" channels necessitating compensation to adjust for the bleed-through. This compensation causes the plots to spread at the higher expression levels. Both pre-TCR variants were expressed without CD3 chains and were as expected with GFP11-tagging on the N-terminus producing greater GFP fluorescence with ER-targeted GFP(1-10) than cytoplasmic GFP(1-10) and C-terminal tagging producing the reverse. The "incorrect" pairings for both cases appeared to be subtly different from the samples with the GFP11-construct expressed alone perhaps indicating that a small portion of the GFP(1-10) is not targeted to the correct compartment. The expression of the TCRβ-mScar is not as high as expected, likely due to the absence of CD3 chains, but this was set up so the same amount of DNA was used per transfection as the other constructs.

When tagged on the N-terminus with GFP11 (5th row), the shortened TMEM131dC gives convincing signal when co-expressed with ER-targeted GFP(1-10) suggesting that the N-terminal domains are in the ER lumen. This is consistent with prior expectations. The C-terminus of appears to be localised to the cytoplasm as there is strong GFP fluorescence with cytoplasmic GFP(1-10) when the protein is tagged with mScarGFP11. This model is consistent with TMEM131 having a single transmembrane helix.

The results are less clear however when the full-length protein is tagged in the same manner. When tagged with mScarGFP11 at the end of the C-terminal tail there is greater GFP fluorescence with cytoplasmic GFP(1-10) but still considerable signal with ER-targeted GFP(1-10). When tagged on the N-terminus with GFP11 (5th row), TMEM131 produced more GFP signal with ER-targeted GFP(1-10) than cytoplasmic GFP(1-10) however the effect is less pronounced than expected. Compared to the other samples where GFP-fluorescence seem to increase at a steep gradient with mScar-fluorescence here the increase has a much lower gradient. This effect is more noticeable if the plots are overlaid (figure 4.14). Two explanations could produce this effect. Firstly, the N-terminal domains including the GFP11 tag are not present in the same ratio to mScar as the other constructs. Normally you would expect both termini of a protein to be equimolar but this might not be the case when the protein is cleaved as is the case with TMEM131. After cleavage the two parts could have different lifetimes and degradation pathways. A decrease in the amount of GFP11-tagged N-terminal TMEM131 compared with the mScar-tagged C-terminus is consistent with the apparent lack of HA-bands compared to the numerous GFP-positive bands present on prior western blots of HA-TMEM131-GFP. A second explanation for the lack of GFP fluorescence observed, which I consider less likely, is if GFP11-TMEM131 localised to a compartment less accessible to GFP(1-10)kdel such as later in the secretory pathway.

Whilst flow cytometry allowed the sampling of a larger number of cells to see the overall trend, the data lacked spatial information. I transfected cells with similar constructs and examined cells under microscopy (figure 4.15). Cells transfected with "incorrect pairings" (e.g. cytoplasmic GFP(1-10) with the pre-TCR containing TCRβ tagged on its N-terminus with GFP11) did not show discernible GFP-fluorescence above background.

When TMEM131 was expressed with GFP11 on the N-terminus and mScar on the C-terminus with ER-targeted GFP(1-10), mScar seemed to fill the cytoplasm whilst GFP fluorescence appeared restricted to the ER. Cells had foci present which appeared to be BFP, GFP, mScar-positive suggesting there was a concentration of double-labelled, uncleaved protein at these specific locations of the ER. When TMEM131 was tagged with mScarGFP11 on the C-terminus, there was considerable fluorescence for both mScar and GFP throughout the cytoplasm. There were obvious foci of mScar that co-localised with the BFP ER-marker but these were not noticeably enriched in GFP fluorescence. These images are consistent with the flow cytometry which implied TMEM131 had its N-terminal domains in the ER lumen and C-terminus in the cytoplasm.

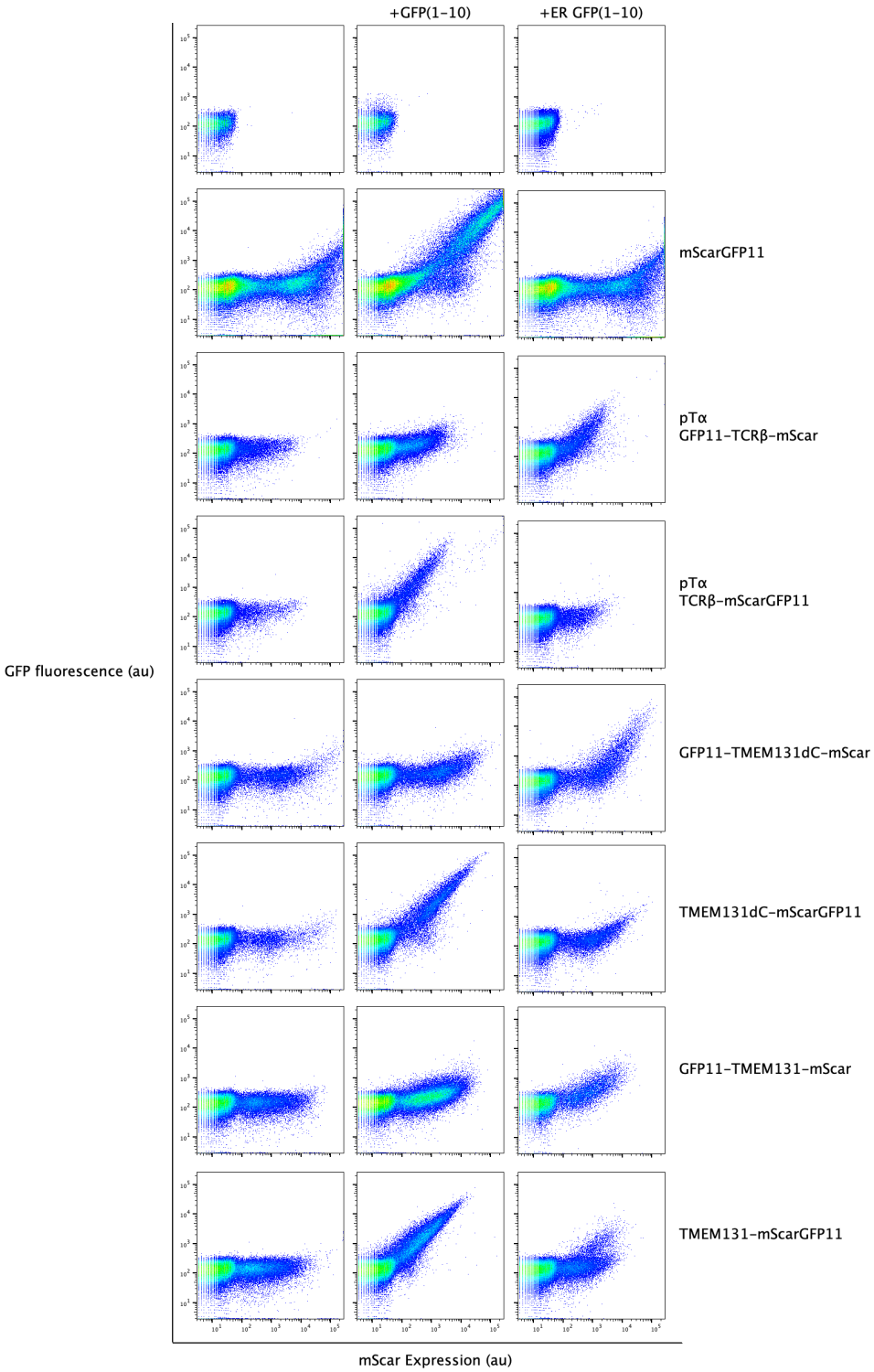

Figure 4.13: A split GFP assay to determine TMEM131 topology.
GFP11 and mScar-tagged constructs were transiently expressed in HEK cells alongside cytoplasmic GFP(1-10) or ER-targeted GFP(1-10). Samples were analysed by the LSRII flow cytometer then compensation was applied in FlowJo.

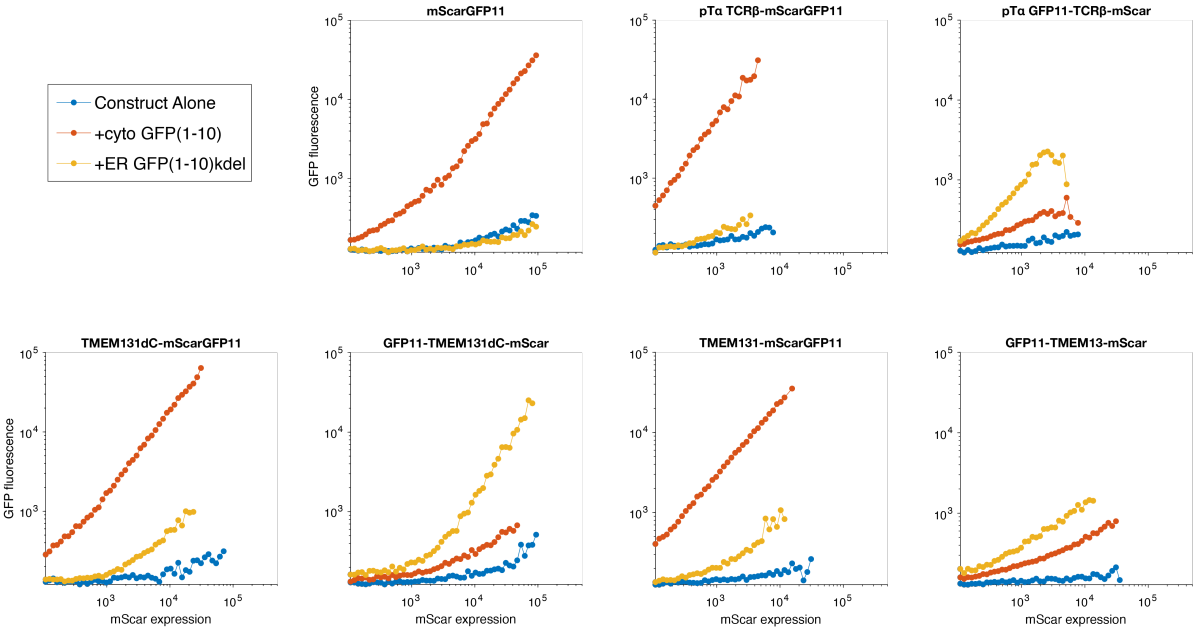

Figure 4.14: The Split GFP assay results for each GFP11-tagged construct averaged and overlaid.
The same compensated data plotted as scatter plots above was sorted into 50 log-distributed bins between 10^2 and 10^5 along the "YG 610/20" axis. The mean value in the "B 530/30" channel was plotted on the y-axis in the middle of each bin. Bins with less than 10 events were excluded. Lines for each construct were then overlaid with the construct expressed alone coloured blue, co-expression with cytoplasmic GFP(1-10) coloured red and co-expression with ER-targeted GFP(1-10) coloured yellow.

Imaging the tailless TMEM131ntd-TM construct, which was not obviously cleaved on westerns, with a C-terminus mScarGFP11 tag produces images surprisingly similar to full-length construct and once again GFP appears to fill the cytoplasm though it appears a lot dimmer and the ER is faintly visible. There was a large number of foci of mScar that did not colocalise with the ER-marker and moved rapidly through the cell. These vesicles were also noticeably GFP-positive though they moved so rapidly these appeared slightly streaked and the channels were slightly offset due to the frame rate. The concentrations of full-length TMEM131 in the ER might be clusters of aggregated protein or could be actual function sites such as sites of ER-export. Removal of the tail allows these clusters to move freely around the cell. Similar images for both TMEM131 and TMEM131ntd were obtained with GFP fused to the N-terminus instead of GFP11 (supplementary figure S13).

As full-length TMEM131 is cleaved, any tag on its C-terminus appears delocalised throughout the cell. To see if this background was obscuring a smaller fraction of uncleaved protein in the ER, I designed a form of GFP(1-10) with an N-terminal myristoylation-sequence which should localise the GFP(1-10) to the cytoplasmic face of cell surface. I wondered if complementation between this surface-localised GFP(1-10) and the free TMEM131 tails tagged with mScarGFP11 could deplete some of the cytoplasmic background, to reveal ER-enriched uncleaved protein. Whilst there did appear to be an enrichment of GFP and mScar fluorescence at the cell surface, the cytoplasmic mScar background did not appear to be greatly reduced and the Myr-GFP(1-10) appeared to be internalised (figure 4.15 row 4).

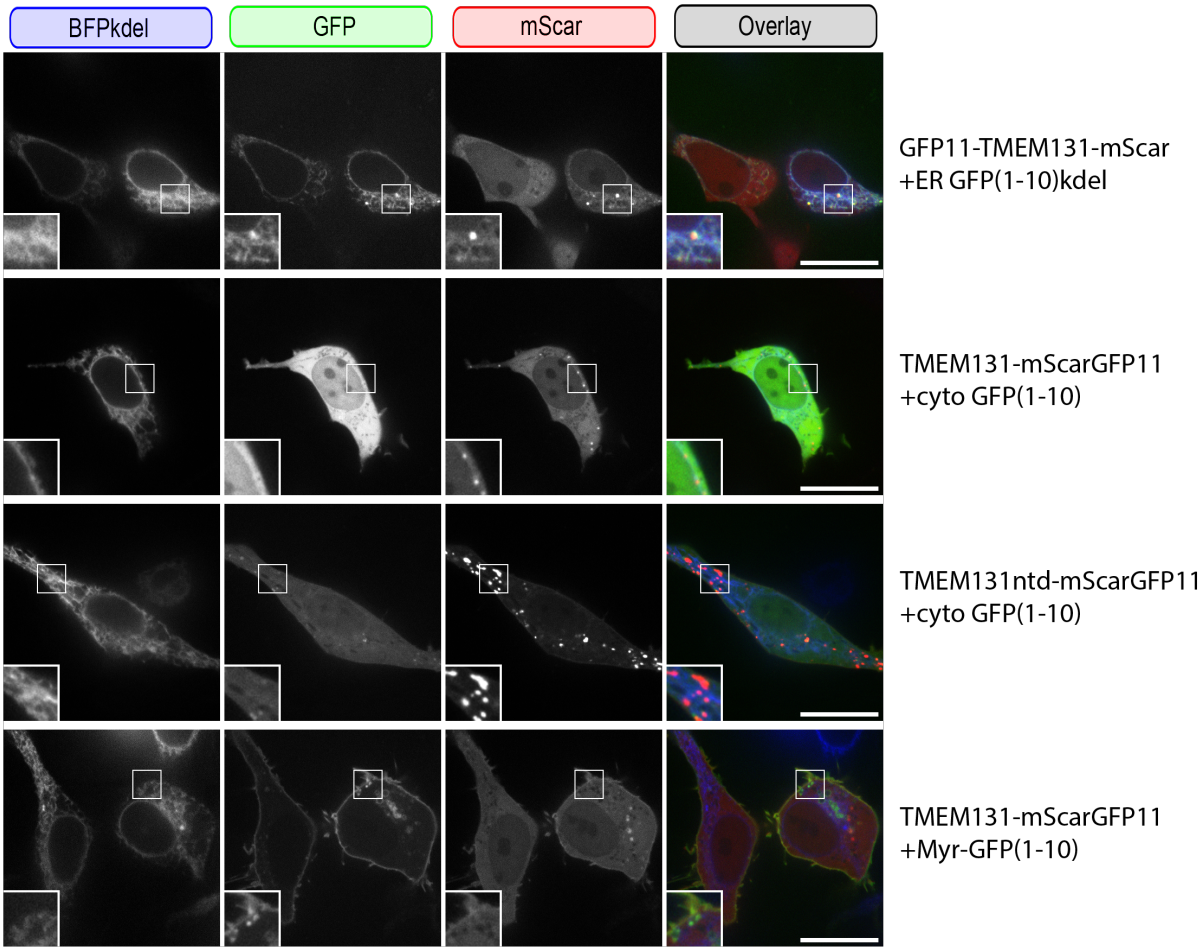

Figure 4.15: Representative images of selected GFP11-tagged constructs.
The indicated constructs were expressed with a BFPkdel ER marker and imaged under 100X magnification. The inserts show a 2X zoom of the selected area.

4.4.7 The first part of the endogenous signal peptide of TMEM131 is dispensable for ER targeting

Previous work in the James lab had replaced the endogenous signal peptide of TMEM131 with the signal peptide of gaussia luciferase (GLuc) to improve the protein's expression and all HA-tagged versions of TMEM131 mentioned above use this sequence. Signal peptides are normally considered to be transferable as they are cleaved off during translation by signal peptidase enzymes. However, some interesting features of the TMEM131 endogenous signal peptide justified further study.

Human TMEM131 appears to have a very long signal peptide (59 amino acids) with unusual architecture. Most signal peptides are 15-30 amino acids with a positive N-terminus, a hydrophobic region then a polar region with smaller side chains near the cleavage site, typically an AXA motif.[342] The endogenous signal peptide of human TMEM131, however seems to have a two-domain organisation that is seen in a few other ER proteins. These long signal peptides appear to consist of two components, an N-terminal part that appears to direct the protein to the mitochondria or elsewhere in the cell and a C-terminal part that is a functional ER-targeting signal peptide. These two components as separated by a short, turn-rich transition linker. Peptides with this "NtraC model" organisation were first described by Hiss *et al.*[343] through computational prediction. Due to their length and rarity, these long signal peptides may not be recognised by online prediction tools such as DeepLoc, TargetP and Signal-P (these all gave inconclusive results for human TMEM131). Why the signal peptides of certain proteins have this organisation is unclear but this could potentially reflect dual targeting.[344]

Through direct correspondence with Dr. Jan Hiss, we inferred that the human TMEM131 signal peptide fulfils the domain organisation of the "NtraC model". Software predicted a beta-turn in the peptide at position 31 with the former part being a functional mitochondrial targeting peptide and the latter part a regular ER-signal peptide. We were interested if any of these sequences affected the localisation of TMEM131. As the full-length protein is cleaved meaning C-terminal fluorescent tags just show non-specific cytoplasmic localisation, I decided to use the tailless TMEM131ntd construct with both putative TM helices as this construct did not appear to be cleaved on Western blots. I expressed this construct with four different N-terminal sequences: the endogenous signal peptide (SP12-ntd), the N-region of the endogenous signal peptide with the GGPR turn sequence (SP1), only the C-region of the endogenous signal peptide (SP2) and finally the GLuc signal peptide with HA tag (glSP-HA). The N-terminal sequences for these constructs are shown in figure 4.16. It is important to note that whilst this long signal peptide is present in humans and some close homologues such as mice, it is not as well conserved as the rest of the N-terminal part of the protein and many homologues such as the protein in *Danio rerio* have more conventional signal peptides.

When expressed with a BFPkdel ER marker, the construct with the *gaussia* luciferase signal peptide and HA-tag appeared to be mostly distributed in the ER with a fainter surface staining. Most cells had also had mobile GFP foci that did not colocalise with the ER marker. The construct with just the first part of the endogenous signal peptide (SP1) had much lower expression and GFP appeared to be distributed throughout the cytoplasm with slight enrichment in the ER. All GFP-positive cells observed had a cluster of immobile GFP-foci in one part of the cell perhaps representing aggregated protein. The protein distribution of the constructs with the second half of the endogenous signal peptide (SP2) and the full endogenous signal peptide (SP12) were very similar to that of the gaussia luciferase signal peptide with most GFP co-localised with the ER marker, some surface GFP and numerous mobile foci.

These images would suggest that the entire endogenous signal peptide or just the second half of the

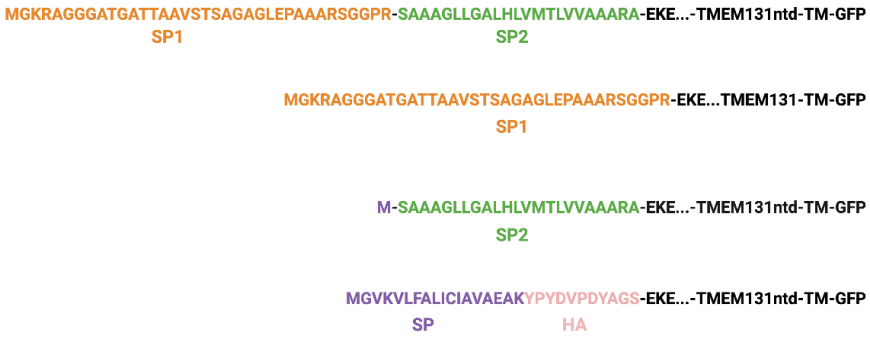

Figure 4.16: Sequences of the endogenous and modified TMEM131 signal peptides.
The endogenous signal peptide (top) consists of an N-terminal region and turn (SP1, orange) and a C-terminal region (SP2, green). These components were also used separately and in contrast to the GLuc signal peptide with HA tag.

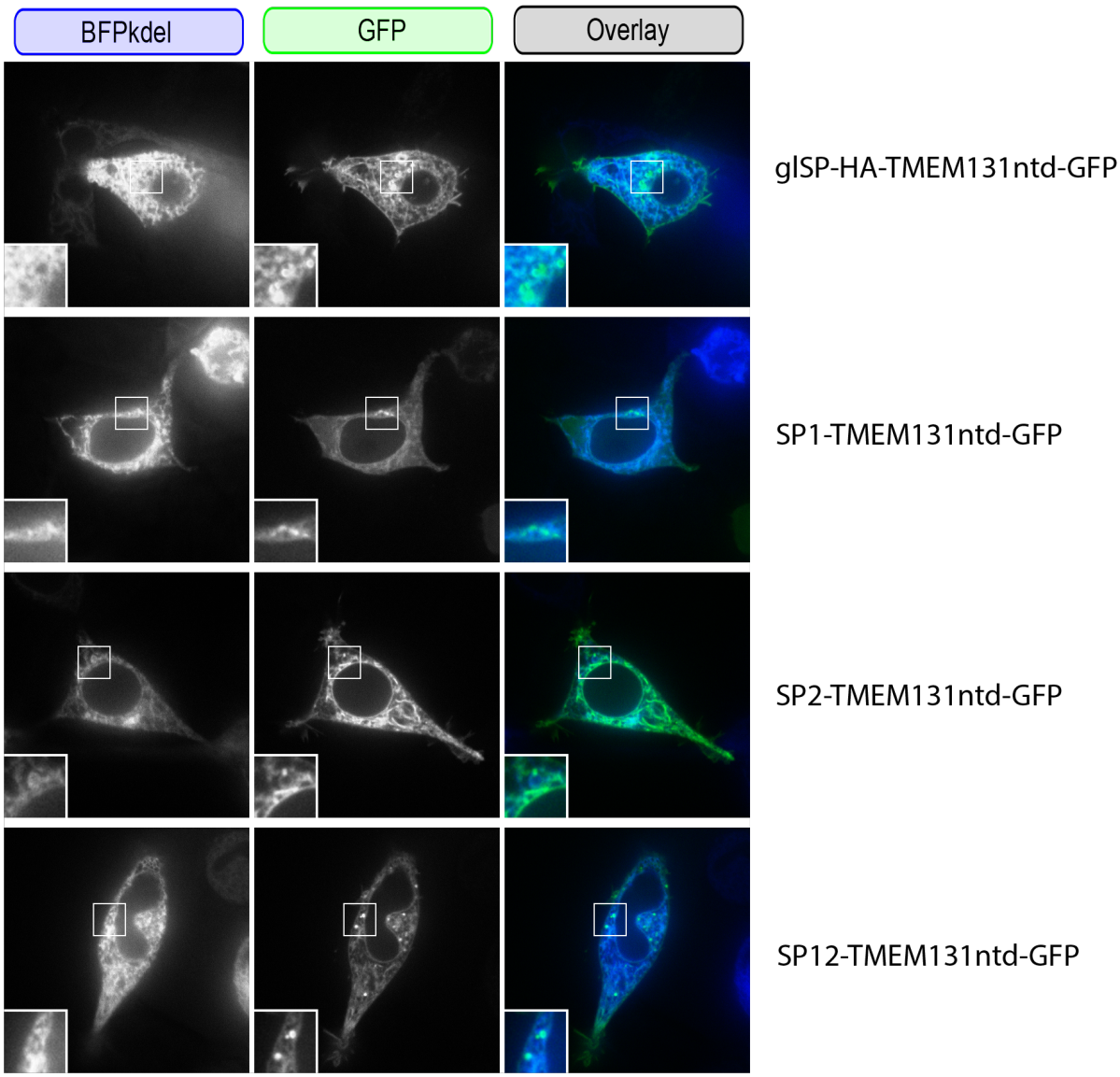

Figure 4.17: Representative images of TMEM131ntd with modified signal peptides.
HEK cells expressing tailless TMEM131 with four different signal peptides together with a BFPkdel ER marker imaged under 100X magnification. Inserts show a 2X zoom of the selected area.

endogenous signal peptide are sufficient for proper TMEM131 targeting. This is consistent with other studies of NtraC signal peptides that found the N-terminal part is dispensable.[345,346] Although it should not have been done without confirmation, this data supports the decision to substitute the signal peptide with the gaussia luciferase signal peptide to improve expression. Whilst all versions appear to be predominantly distributed in the ER, different cells have different numbers of mobile GFP-positive vesicles. These vesicles were most noticeable in cells with higher expression levels; it is unclear whether they are brighter and easier to detect with higher expression or whether higher expression leads to more vesicles after ER-retainment mechanisms are saturated.

4.4.8 Co-localisation of TMEM131 with other proteins

When investigating the localisation of TMEM131, I was curious as to how the foci of tailless TMEM131 compared to the pre-TCR given evidence for an interaction between the two proteins. I co-expressed TMEM131ntd-GFP in HEK with mScar-tagged pre-TCR, CD3 and an ER-marker then imaged cells under 100x magnification. It became apparent that there was a degree of overlap between the two proteins with two examples shown below in figure 4.18. The first row shows the most dramatic example I observed, a cell with GFP-fluorescence throughout the ER as well as large number of GFP foci together with an unusual number of foci of the pre-TCR. Some of these foci appear in both channels though some do not overlap completely due to the framerate of the imaging. The bottom row shows a more typical example with fewer foci for each protein.

Co-localisation between two proteins can be evidence for a functional interaction between them though needs to be investigated with caution. Co-localisation is resolution dependent as two points appearing co-localised can be resolved if the resolution is increased. Simple measurements of co-localisation such as the Pearson's correlation coefficient between the two channels are strongly affected by the amount of empty space surrounding cells where signal is low in both channels. Other measurements of correlation such as Mander's overlap coefficients are harder to implement and interpret. In this case I was keen to investigate the overlap between TMEM131 and the pre-TCR further due to the atypical number of pre-TCR foci outside the ER in TMEM131-expressing cells. My hypothesis was that TMEM131 expression was enabling more pre-TCR complexes to leave the ER.

The above examples were a biased selection of cells. I preferred imaging cells with higher expression levels, and adjusted to find the best focal-plane, as this produced images with better signal to noise. Transient transfections generally produce a range of expression levels across a few orders of magnitude and cells at the higher end may have different behaviour to the mean. When multiple plasmids are transfected together, the expressions levels of each usually appear correlated in flow cytometry when viewed on log scales. However when the same cells are examined via imaging, with a linear scale between expression level and intensity, variation in the expression levels of co-expressed plasmids becomes apparent. In this case, some cells appear enriched in TMEM131-GFP with many foci but TCRβ expression was a lot lower and the opposite arrangement. A complicating factor is the expression of CD3 chains from a separate plasmid which are unlabelled and could be limiting formation of the pre-TCR.

As these observations were being made the AlphaFold model of TMEM131 became available. Whilst we were previously reluctant to tag TMEM131 on the N-terminus, given the apparent disappearance of HA-tagged TMEM131 N-terminal fragments in western blotting, the model gave us more confidence in making edits to the N-terminal side of the protein. Around this time I cloned GFP, BioID and APEX domains onto the TMEM131 N-terminus (discussed in section 4.6.1). I used the version of TMEM131 tagged with GFP on the N-terminus for the next set of experiments as it seems to more consistently

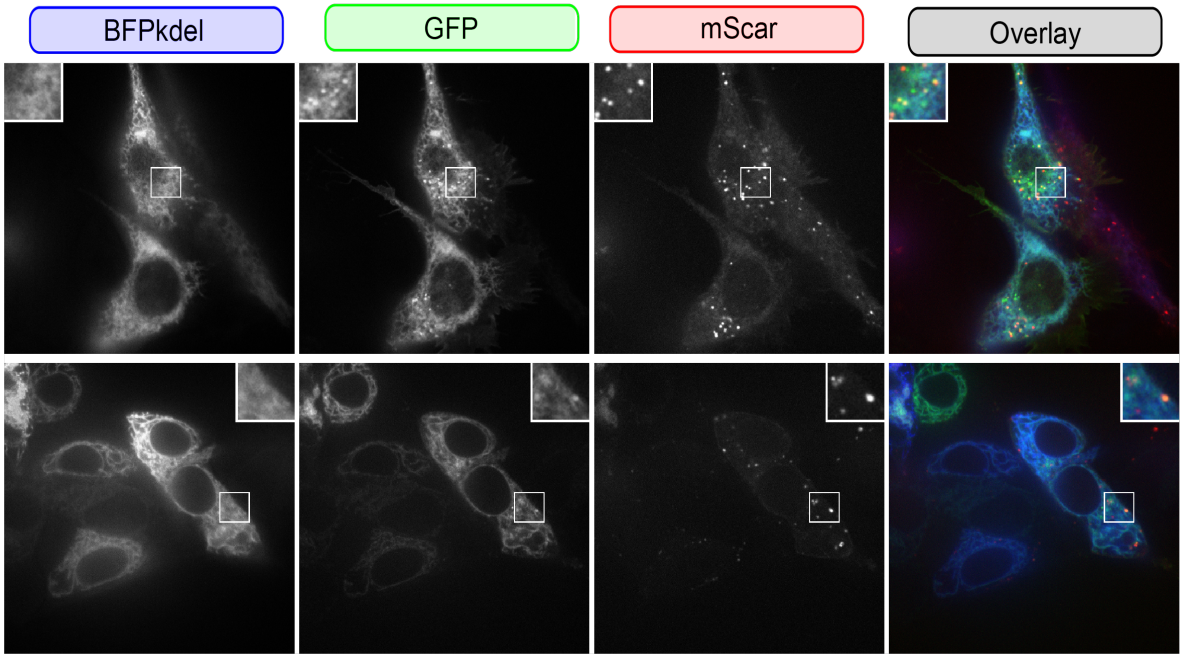

Figure 4.18: Co-localisation between TMEM131ntd-GFP and mScar-tagged pre-TCR.
TMEM131ntd-GFP, pTα TCRβ-mScar and CD3$\delta\gamma\epsilon\zeta$ were expressed in HEK cells with a BFPkdel ER-marker and cells imaged under 100X magnification. Inserts show a 2x zoom of the selected area.

produce GFP-foci outside of the ER, enabled the N-terminal domains of the full-length TMEM131 to be studied without the tag being cleaved off and to prevent the possibility of co-localisation due to multimerisation of the fluorescence protein tags. Whilst the mScar and GFP sequences used had monomerising mutations, the possibility of forming oligomers remains and putting GFP and mScar on opposite sides of the membrane should remove this possibility.

GFP-TMEM131ntd and mScar-tagged pre-TCR were co-expressed and there appeared to be reasonable overlap between GFP and mScar-positive foci. When anti-GFP nanobody was added to the external media after 30 minutes most GFP-positive foci as well as most GFP and mScar-positive foci were nanobody labelled (figure 4.19). This implies that the proteins within these foci had reached the cell surface and were therefore present in vesicles of the endocytic pathway. I expressed the same constructs with BFP-tagged Rab5, Rab7 and LAMP1 which should act as markers of early endosomes, later endosomes and endolysosomes respectively (figure 4.20). Overlap of GFP/mScar-labelled foci with Rab5 was not particularly notable though there were examples of overlap of these foci with Rab7-labelled endosomes. GFP and mScar did overlap convincingly with lysosomal-marker LAMP1 though high expression of LAMP1 did appear to greatly change the morphology of cells with GFP-TMEM131 and pTCR-mScar present in large, immobile clusters near the cell centre.

I attempted to investigate TMEM131 and pre-TCR co-localisation more rigorously by imaging a random selection of cells to remove bias in cell selection. This was done by transfecting cells, plating them on a slide aiming for a reasonable density and imaging all the cells in a grid at the same focal plane. The four samples I chose to image were cells transfected with pTα TCRβ-mScar with CD3, pTα TCRβ-mScar with CD3 and GFP-TMEM131ntd, pTα TCRβ-mScar with GFP-TMEM131ntd but without CD3 chains and pTα TCRβ-mScar with CD3 and full-length GFP-TMEM131.

Images obtained through this random sampling were generally low quality, often with cells partially out of frame or focus. Whilst a range of expression levels was captured, the need to keep the same laser

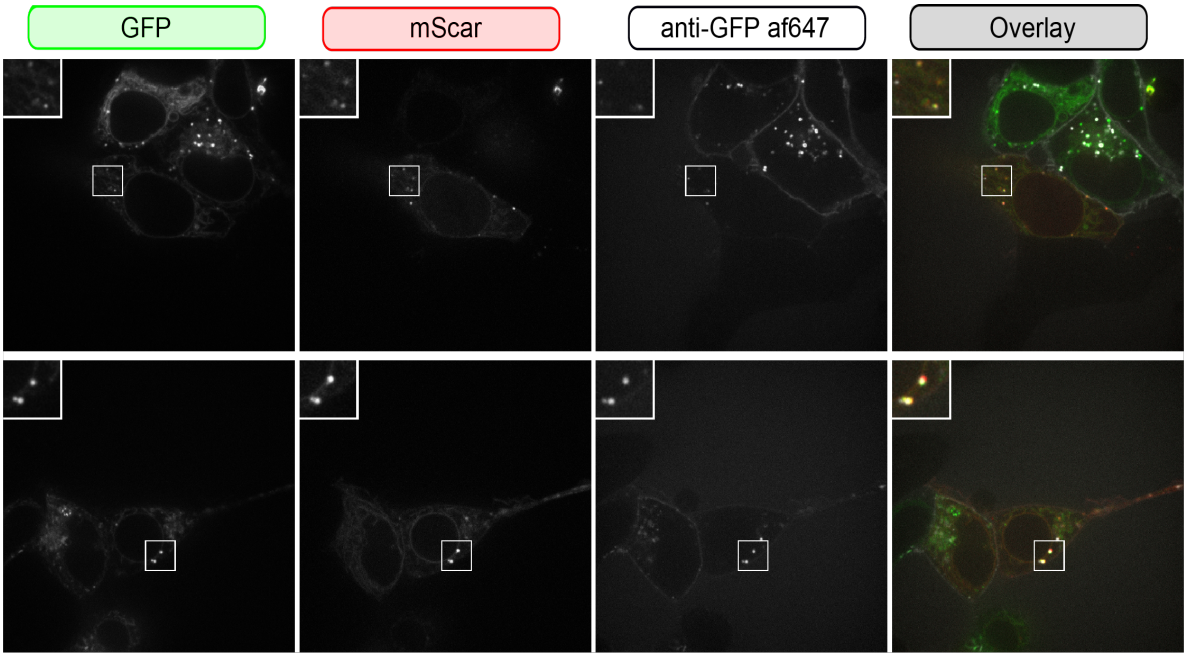

Figure 4.19: A fraction of the vesicles containing GFP-TMEM131ntd and mScar-tagged pre-TCR have been to the cell surface.
HEK cells expressing GFP-TMEM131ntd, pTα TCRβ-mScar and CD3$\delta\gamma\epsilon\zeta$ were imaged under 100X magnification after AlexaFlour647-conjugated anti-GFP nanobody was added to the medium. Inserts show a 2x zoom of the selected area.

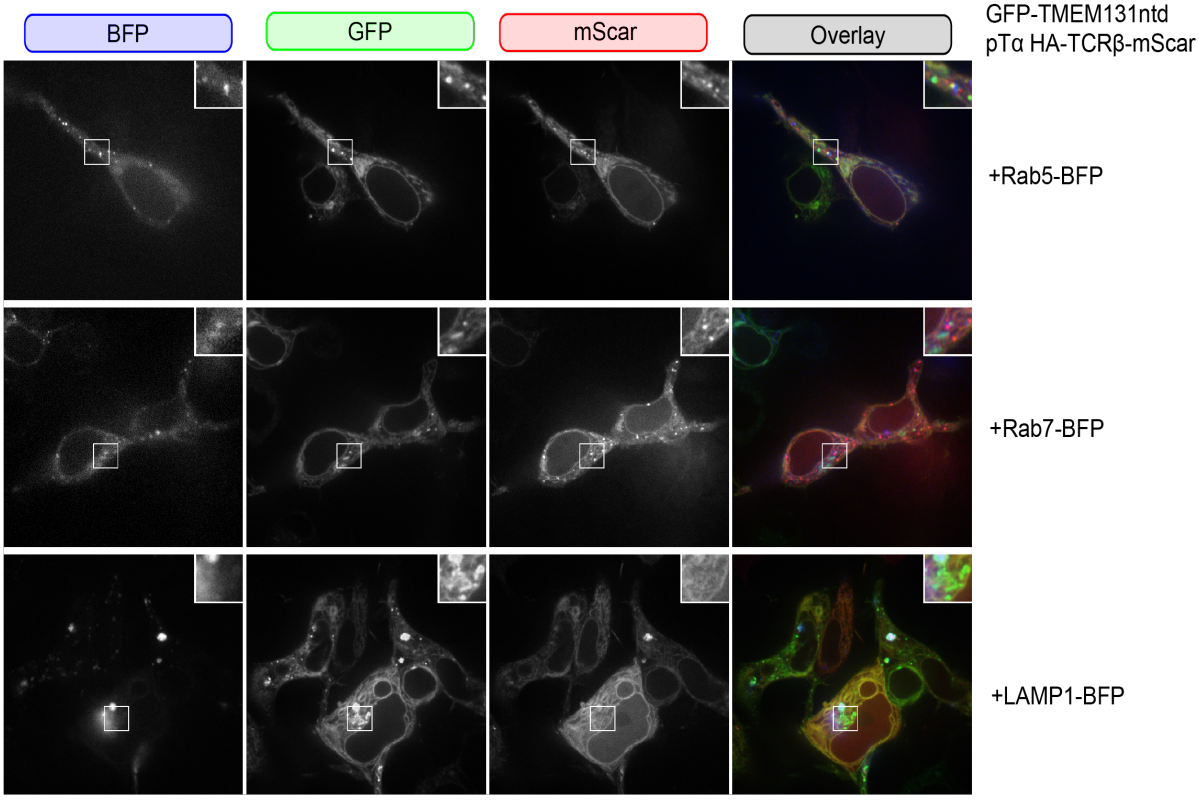

Figure 4.20: Co-localisation between TMEM131 and pTCR-containing vesicles with markers for the endocytic pathway.
GFP-TMEM131ntd, pTα TCRβ-mScar and CD3$\delta\gamma\epsilon\zeta$ were expressed in HEK cells with Rab5-BFP, Rab7-BFP and LAMP1-BFP markers and cells imaged under 100X magnification. Inserts show a 2x zoom of the selected area.

powers for the whole set of images meant that lower-expressing cells were dim with a small dynamic range in both channels.

I attempted an entirely automated image analysis pipeline to remove bad images then identify foci in each channel as well as foci that overlapped but this did not lead to reproducible results. A more manageable strategy was a semi-automated analysis on a subset of images that were in focus and had reasonable expression levels of both proteins. As this assessment was done manually it reintroduced bias into image selection. Although foci could be identified computationally the assessment of overlap between mScar and GFP foci was also performed manually as the algorithm was unable to distinguish spherical foci from brighter regions of ER or membrane. The counts for this analysis are presented in figure 4.21. Whilst there appears to be slightly more foci of the pre-TCR on average if it was expressed with the truncated GFP-TMEM131ntd or full-length GFP-TMEM131, the increase was not drastic and the variance was high. Some cells had 10-20 foci of the pre-TCR with and without TMEM131. For the full-length construct especially, the two overlap between red and green foci was relatively weak. Though TMEM131 might have a role in ER exit, I cannot conclude that co-expression with the pre-TCR greatly increases or decreases the number of foci of either protein given the caveats involved in the image analysis.

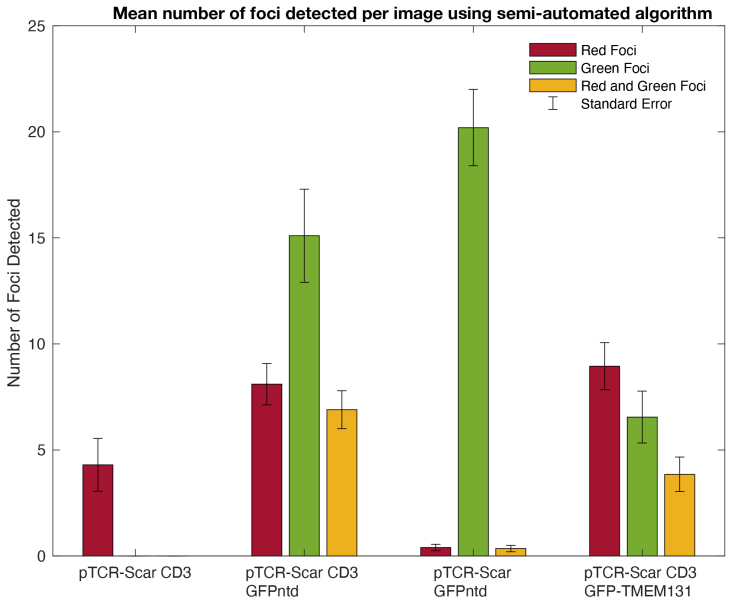

Figure 4.21: The mean number of foci of pTCR-mScar and GFP-TMEM131 detected per image using a semi-automated analysis algorithm of selected images. HEK cells were transfected with: the mScar-tagged pTCR, the mScar-tagged pTCR with GFP-tagged TMEM131ntd, the same without the CD3 chains or the mScar-tagged pTCR with GFP-tagged TMEM131 full-length protein. The number of foci counted in a selection of images in the mScar channel, GFP channel and in both channels is represented by the red, green and yellow bars respectively. The error bars show the standard error for this selection.

Having difficulty generating a reliable quantitative metric of co-localisation between TMEM131 and the pre-TCR, I decided to co-express TMEM131 with some other membrane proteins I had access to, to see if co-localisation was qualitatively similar or different (Figure 4.22). Superficially similar-looking co-localisation could be observed between GFP-TMEM131ntd and mTCR, CD86 and ICAM1 but not LAT-mRuby. The former are all transmembrane proteins with extracellular Ig-like domains whilst LAT lacks an external domain.

I concluded that whilst TMEM131 and the pre-TCR appear to partially co-localise, this co-localisation was not unique to the pre-TCR. As all membrane proteins are transported to and from the cell surface via

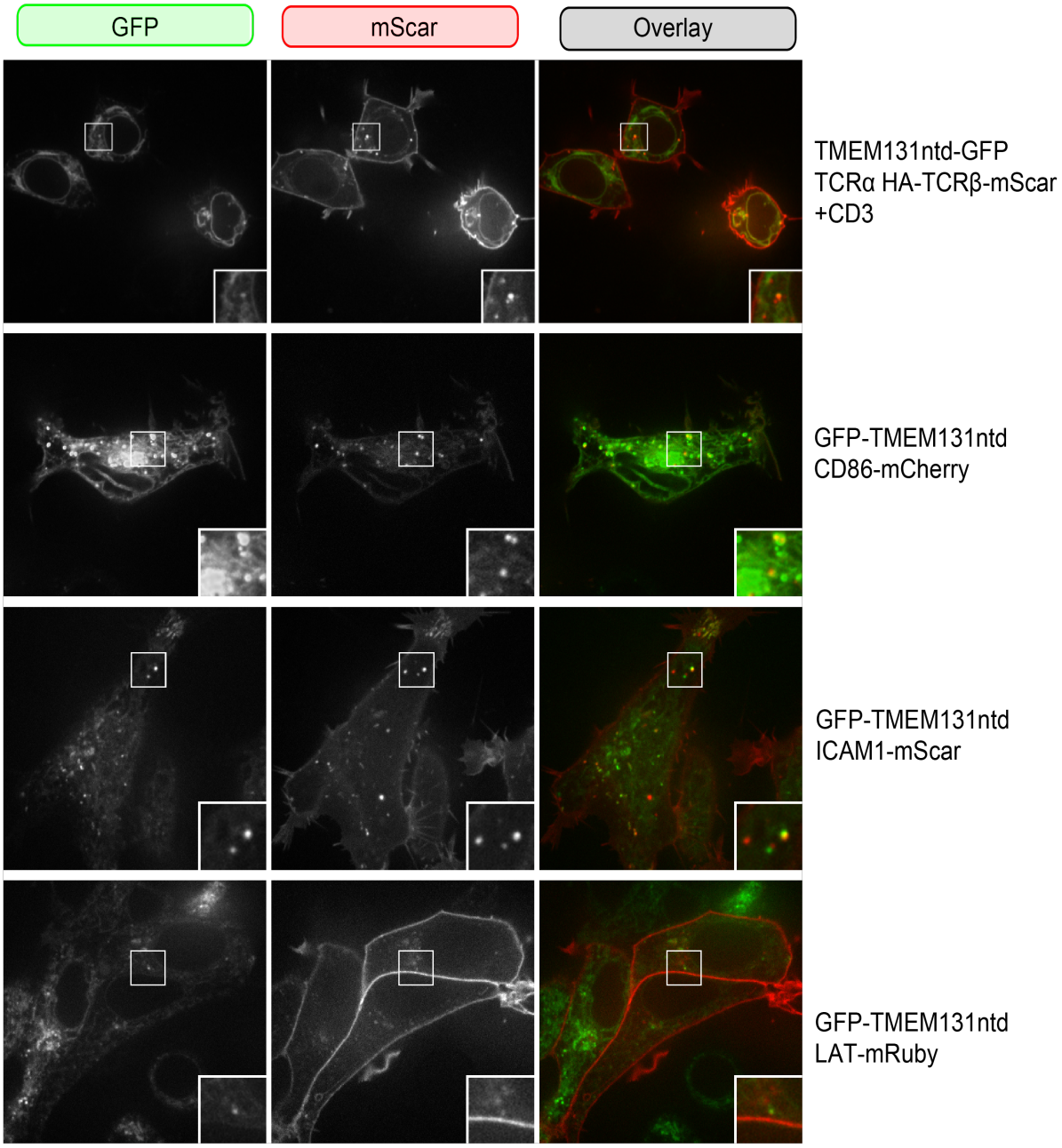

Figure 4.22: Co-expression of tailless TMEM131 with a selection of other membrane proteins. GFP-TMEM131ntd was expressed with mScar-tagged mTCR, CD86-mCherry, ICAM1-mScar and LAT-mRuby and cells imaged under 100X magnification. Inserts show a 2x zoom of the selected area.

the secretory/endocytic pathway, a mixing of cargos is expected even without protein-protein interactions.

4.4.9 TMEM131 is weakly oligomeric

Work in the James Lab investigating pre-TCR dimerisation (previous chapter) used the split Venus bimolecular fluorescence complementation assay.[347] This assay operates similarly to the split GFP assay, with a fluorescent protein is reassembled by complementation between two non-fluorescent fragments. The two components of split GFP have a high affinity to each other so irreversibly join if present in the same compartment without regard to the affinity of the sequences each component is fused to. This is undesirable in an assay for oligomerisation as it raises the background and decreases signal/noise. This assay instead uses two components of the yellow fluorescent protein Venus: VN encoding strands 1-7 and VC encoding stands 8-11. These fragments have a lower affinity to each other which reduces the fluorescent background and makes complementation dependent on the affinity between the two proteins the fragments are attached to. Once complemented however the two fragments are joined irreversibly allowing the method to detect transient or lower affinity interactions.

I fused VN (156 residues) and VC (81 residues) to the N-terminus of TMEM131ntd with 18 amino acid linkers then cloned the fluorophores BFP and mScar respectively onto the C-terminus to ensure both fragments were expressed. I used the tailless TMEM131ntd rather than the full-length TMEM131 for this experiment as the expression of the N-terminal domains of the molecule appeared higher in multiple other experiments when the C-terminal tail was removed. I transfected HEK cells with these constructs individually then together. For a positive control I used the mature TCR with VN and VC components on the two TCR chains.

Flow cytometry of the constructs is shown in figure 4.23 as dot plots of Venus (vertical axis) against BFP (top) and mScar (bottom). There is no Venus fluorescence above background for VN-TMEM131ntd-BFP or VC-TMEM131ntd-mScar but there is appreciable signal at higher expression levels when the two constructs are co-expressed. The trend is clearer when plotted against BFP rather than against mScar where the cells begin to diverge at higher mScar expression. Plotting BFP against mScar suggest that the VN-TMEM131ntd-BFP was limiting in these cells, consistent with the larger component of Venus being less well tolerated.

I binned these results by the expression levels of BFP/mScar then plotted the mean Venus fluorescence for each bin (figure 4.24). Plotted like this it becomes more apparent that the Venus fluorescence for VN-TCRα VC-TCRβ is strongly correlated with mScar expression as expected for a dimeric complex. The combination of VN and VC-tagged TMEM131 meanwhile stays near baseline at lower expression levels of BFP/mScar and only increases at higher expression levels. This is more consistent with TMEM131 being present as a weaker or more transient oligomer.

Imaging suggested that the oligomers were present only in the ER and the complemented Venus domain was not detected at the cell surface with labelled anti-GFP nanobodies (supplementary figures S15 and S16). VC and VN-tagged TMEM131ntd constructs were co-expressed with pre-TCR or mTCR constructs with compatible tags. Although an interaction between TMEM131 and the pre-TCR and mTCR was found previously interact, there was no increase in Venus fluorescence. It is possible the VN/VC tags are not positioned correctly for Venus to form despite the relatively long linkers (supplementary figure S14).

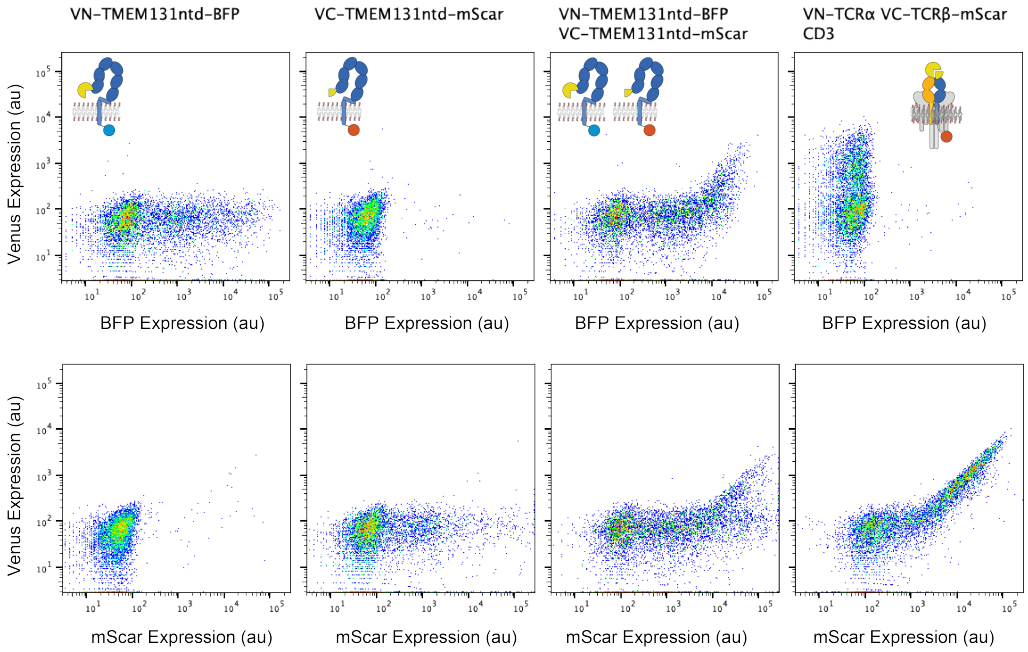

Figure 4.23: A split Venus bimolecular fluorescence complementation assay suggests TMEM131ntd forms a weak oligomer.
TMEM131ntd was expressed in HEK cells with the VN and VC components of Venus on the N-terminus and BFP/mScar fluorophores on the C-terminus. Presented as dot plots of Venus against BFP (top) and mScar (bottom).

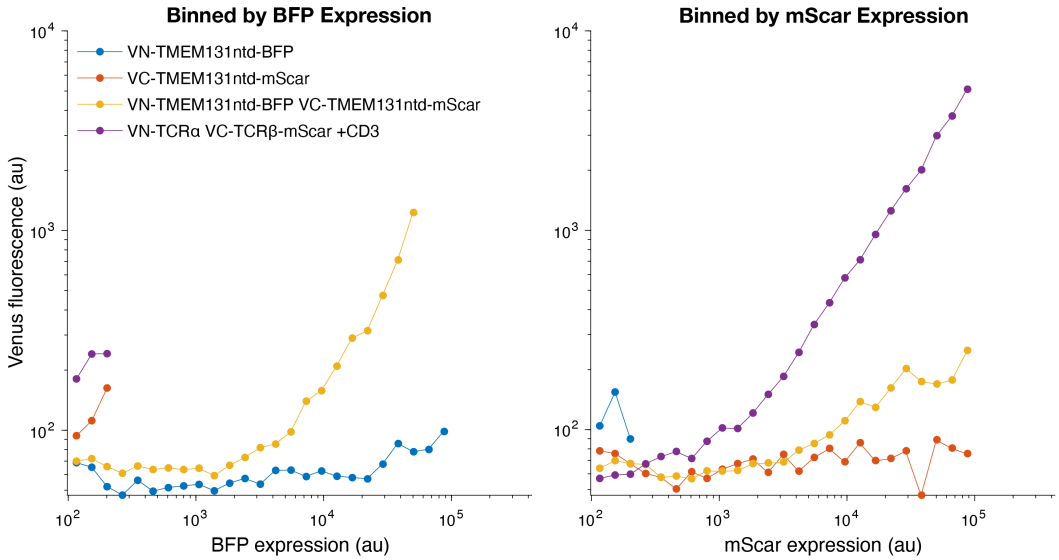

Figure 4.24: Previous flow cytometry results binned by BFP/mScar expression then overlaid.
Compensation was applied in FlowJo then data was sorted into 25 log-distributed bins between 10^2 and 10^5 along the 'B 530/30' and 'YG 610/20' axes. The mean value in the 'B 530/30' channel for each bin was plotted on the y-axis in the middle value of each bin. Bins with less than 10 events were excluded.

4.4.10 TMEM131 is not detectable at the cell surface if its secretory traffic is synchronised

The prior flow cytometry and microscopy experiments found that full-length TMEM131 was undetectable at the cell surface. Meanwhile constructs with truncated cytoplasmic tails were able to reach the cell surface, albeit to different extents. If TMEM131 is retained in the secretory pathway by its terminal tail then perhaps the cleavage of the tail that we observe is a mechanism of relieving this retention? This might not necessarily be noticeable if the N-terminal product of the tail cleavage was a very minor fraction of the total or very short-lived.

I hypothesised the RUSH system, established in section 3.4.6, could be used to synchronise trafficking of TMEM131 to make any such effects easier to see. If the full-length protein is synthesised in the ER and trapped in that location by affinity to streptavidin-BFPkdel then it should provide plenty of time for the tail to be cleaved. Once cleaved it should resemble the truncated constructs and when the block is relieved by biotin we would expect to see trafficking of the N-terminal domains of TMEM131 towards the cell surface.

Having cloned SBP-GFP, GFP-TMEM131ntd and GFP-TMEM131 before it was easy to reuse these oligos to clone SBP-GFP-TMEM131 with and without the cytoplasmic tail. I then transfected HEK cells in a microscopy slide and followed the same imaging protocol I developed for the pTCR and mTCR.

Unlike with the pre-TCR and mature TCR, there was an obvious difference between SBP-GFP-TMEM131 and the tailless SPB-GFP-TMEM131ntd (figure 4.25). The shorter construct was detectable at the cell surface about 30 minutes after biotin addition (red line) and appeared to plateau about 40mins later. There was a small increase in the number of foci without biotin (yellow line). Cells expressing SBP-GFP-TMEM131 with its cytoplasmic tail were much dimmer and there was no increase in foci compared to the no biotin control (blue line compared to green line).

I measured the correlation between the GFP-tagged proteins and the Streptavidin-BFPkdel ER hook and between the GFP-tagged proteins and the added nanobody. When biotin is added SBP-GFP-TMEM131ntd appears to diverge from the BFP after about 35 minutes, which is about the same time it started being detectable at the cell surface. Without biotin the correlation stayed roughly constant. When biotin is added SBP-GFP-TMEM131 appears to gradually diverge from the ER hook but does not appear to reach the cell surface. There is a slight increase in the correlation between the protein and the fluorescent nanobody relative to the no biotin control but this might not be significant.

If TMEM131 is beginning to diverge from the ER hook but is not present at the cell surface then where could the protein be? It's possible the protein is impeded later in the secretory pathway for instance in the trans-Golgi network. Alternatively, the protein could be being removed from the ER by ERAD for degradation in the cytosol. The lack of surface expression after synchronisation of secretory trafficking and inability to detect the HA-tagged fragments of TMEM131 tail cleavage on westerns suggests that cleavage is somehow coupled to degradation of the N-terminal product.

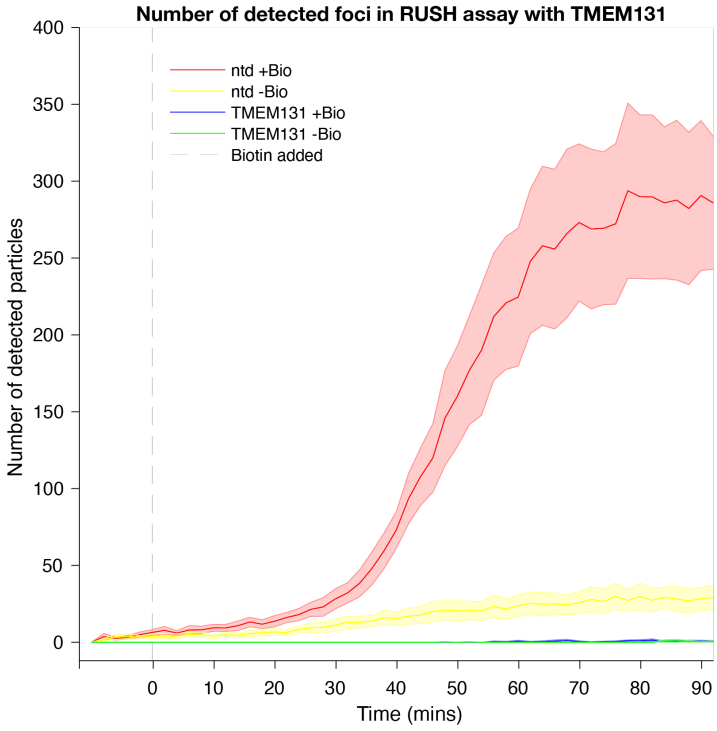

Figure 4.25: Foci detected in a RUSH assay with TMEM131.
HEK293T cells were seeded into a 4-well microscopy dish. Two wells were transfected with SBP-GFP-TMEM131 with Streptavidin-BFPkdel and two with the tailless SBP-GFP-TMEM131ntd with Streptavidin-BFPkdel. 10 cells in each were imaged simultaneously under a 40X air objective with frames every two minutes. AF647-conjugated anti-GFP nanobodies were added after 1 frame, left 10minutes to diffuse then biotin added to half the wells. The number of detected foci of the nanobody was recorded over time. Shaded areas show standard error.

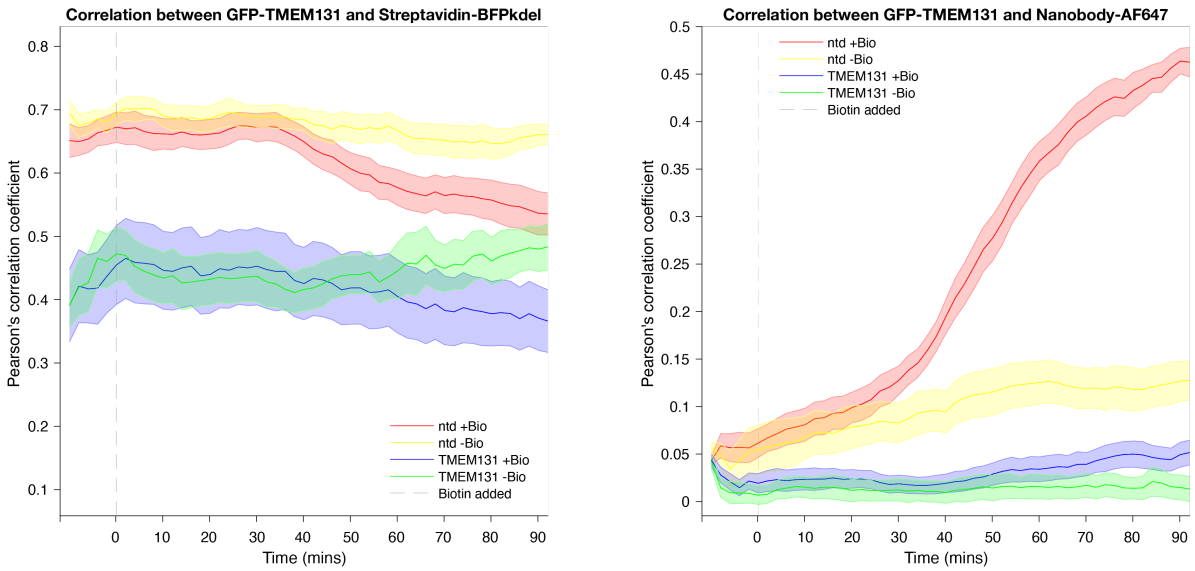

Figure 4.26: Correlation between GFP-tagged TMEM131 protein and the ER hook and externally-added nanobody in RUSH assay.
The correlation coefficient between two indicated channels was recorded for each frame of the assay. The shaded areas show standard error.

4.5 Results: the mechanism of TMEM131 tail cleavage

4.5.1 Proteomic experiments to determine TMEM131 cleavage sites

Having established via western blotting and microscopy that TMEM131 is cleaved in its cytoplasmic tail when expressed in HEK cells, the most pressing questions are where in the sequence is the protein cleaved and what protease might be responsible? As truncating the protein tail did not yield much insight into the cleavage sites, I decided to purify the protein fragments and tried to determine their sequence. This was a more difficult problem to address via proteomic methods than first anticipated.

A conventional proteomic workflow is to enrich proteins in a sample via immunoprecipitation and remove non-specific proteins via washing steps. The enriched proteins can be digested to short peptides on the beads using mass spectrometry grade trypsin (an "on-bead digest" approach). Trypsin cleaves C-terminal to K or R residues. The exact mass of these short peptides can be found using LC/MS and knowledge about trypsin sites is used to determine peptide sequence. An alternative approach (an "in-gel digest") elutes proteins from the beads using high temperature in sample buffer and these proteins are then run on a polyacrylamide gel. Non-specific protein dyes are then used to identify bands within the gel which can then be cut out. Peptides are removed from this gel band by a similar process of trypsin digestion then identified using LC/MS.

For this specific task both approaches are unsuitable. Even with optimal procedure it is rare to capture every peptide from trypsin-digestion in sample as trypsin-cleavage sites may not be optimally distributed and peptides travel differently in a mass spectrometer so are not detected in proportion to their input concentration. For identifying the cleavage sites in the TMEM131 tail this is an issue as we would need to detect the most N-terminal peptide from the natural cleavage and would not necessarily know if this had been achieved. An "on-bead digest" has the additional problem of not separating the GFP-linked fragments by length so peptides could not be assigned to each band visible via western.

An alternative method is to identify the mass of proteins without trypsin digestion ("native mass spectrometry"). In theory this could give the exact mass of each TMEM131 fragment in kDa from which the sequence could be inferred. However native mass spectrometry has much stricter requirements to generate ions and remove incompatible detergents and salts. For the spectrometer I had access to this meant an upper mass limit of 30-35kD. Previous western blots implied the mass of the TMEM131 tail fragments would be small enough on their own, however the GFP tag adds an extra ~28kD taking them over this limit.

I decided to design constructs with a specific protease cleavage site between the end of the TMEM131 tail and the GFP domain so that the fluorescent tag could be removed after GFP-linked fragments had be pulled down and contaminants washed away. I chose the HRV 3C protease sequence (LEVLFQ/GP where "/" indicates the cleavage site). This sequence can be cut by PreScission Protease, a fusion of human rhinovirus (HRV) 3C protease and glutathione S-transferase (GST). I needed a way to verify that peptides had been enriched and the GFP domain removed but having removed the GFP-tag the fragments would no longer be recognised by anti-GFP antibodies. I decided to use SYPRO Ruby, a sensitive fluorescent dye for detecting total protein.

Before taking samples to mass spectrometry, I wanted to optimise the protocol on a smaller scale which took several rounds. The final protocol is given in section 4.3.1. Briefly, cells expressing GFP-tagged TMEM131 and TMEM131dC with or without the 3C cleavage site before the GFP-tag were lysed then proteins enriched with GFP-trap agarose beads. The beads were washed extensively then incubated

overnight at 4°C with PreScission protease. After pelleting the beads, the supernatant was kept as the protease eluted fraction whilst the remaining protein on the beads was extracted more harshly by heating to 70°C in LDS sample buffer. The lysate, protease eluted and bead fractions were run on two parallel 4-12% BisTris gels then one stained with SYPRO Ruby using the rapid protocol and the other labelled with anti-HA and anti-GFP antibodies. The results from these gels are shown in figure 4.27.

As expected from prior experiments, there does not appear to be any HA-staining for the full-length protein with or without the 3C site (figure 4.27A left). There is a HA-positive band for TMEM131dC in the lane containing the cell lysate (lane 4) and this band is enriched in the bead fraction (lane 14). There is no similar band in the protease eluted fraction (lane 9). Meanwhile the TMEM131dC-3C-GFP construct has a similar HA-positive band in the lane containing cell lysate (lane 5) but there is also a band present in the protease eluted fraction (lane 15). This band is slightly smaller likely accounting for the lost mass of the GFP tag. There is a fainter band of the same size in the bead fraction (lane 15) but this could be from liquid in and around the beads, although they were washed before the bead sample was obtained.

The GFP-labelling shows a complementary picture (figure 4.27A right). There are GFP-linked bands in the lysate samples for all four constructs (lanes 2-5) though these are weaker in the TMEM131-3C-GFP lane which had seemed to have a low transfection efficiency in this run (figure 4.27C). The bands visible for TMEM131dC-GFP and TMEM131dC-3C-GFP appear strangely different despite the small mass of the 3C tag. High molecular weight bands consistent with the full construct are present for both dC variants but neither full-length construct. A nonspecific band at ~65kD is present in all samples as well as the untransfected control. In the bead fractions there are GFP-linked bands similar to the lysate fraction only in the two constructs without the 3C site. The two lanes with the 3C site only have a strong band around 28kD consistent with the mass of GFP without any of the TMEM131 protein attached. The nonspecific band is absent, apparently removed by the wash steps.

When the same fractions are labelled with SYPRO Ruby without blotting (figure 4.27B) the lysate samples are unsurprisingly strongly labelled by the non-specific protein dye. In the bead fraction there are bands for TMEM131 and TMEM131dC (lanes 12 and 15) that are consistent in mass with bands visible on the anti-GFP labelled blot. There are several bands present in all samples including the untransfected control which could represent environmental contaminants or the nanobody components of the GFP-beads. The most obvious band, in all lanes of the bead and protease eluted fractions at around 42kD, is consistent with the mass of PreScission protease. Most encouragingly there are bands in the lane for TMEM131dC-3C-GFP (lane 10) between ~24 and 35kD that are not present in the other lanes. These could be fragments of TMEM131dC missing the GFP tag. Unfortunately, these are not observed in the lane for TMEM131-3C-GFP likely due to the lower transfection efficiency. Bands for this construct were present in a prior run of this experiment though staining was worse overall (supplementary figure S17).

Having established a method for purifying the TMEM131 tail fragments and removing the GFP tag I was keen to scale up for mass spectrometry to find the peptide's sequence. I discussed this experiment with the Warwick Proteomics facility and we decided the best approach was a combined strategy using native MS and in-gel tryptic digest on the same set of samples which should give complementary information. In theory, native mass spectrometry would give me an exact list of fragment masses but these might not be easily assigned to the bands we see via Western blotting whilst the in-gel digest should give partial sequence information for each band cut out.

For native mass spectrometry, the experiment would need a change of buffer as the Tris and NP-40

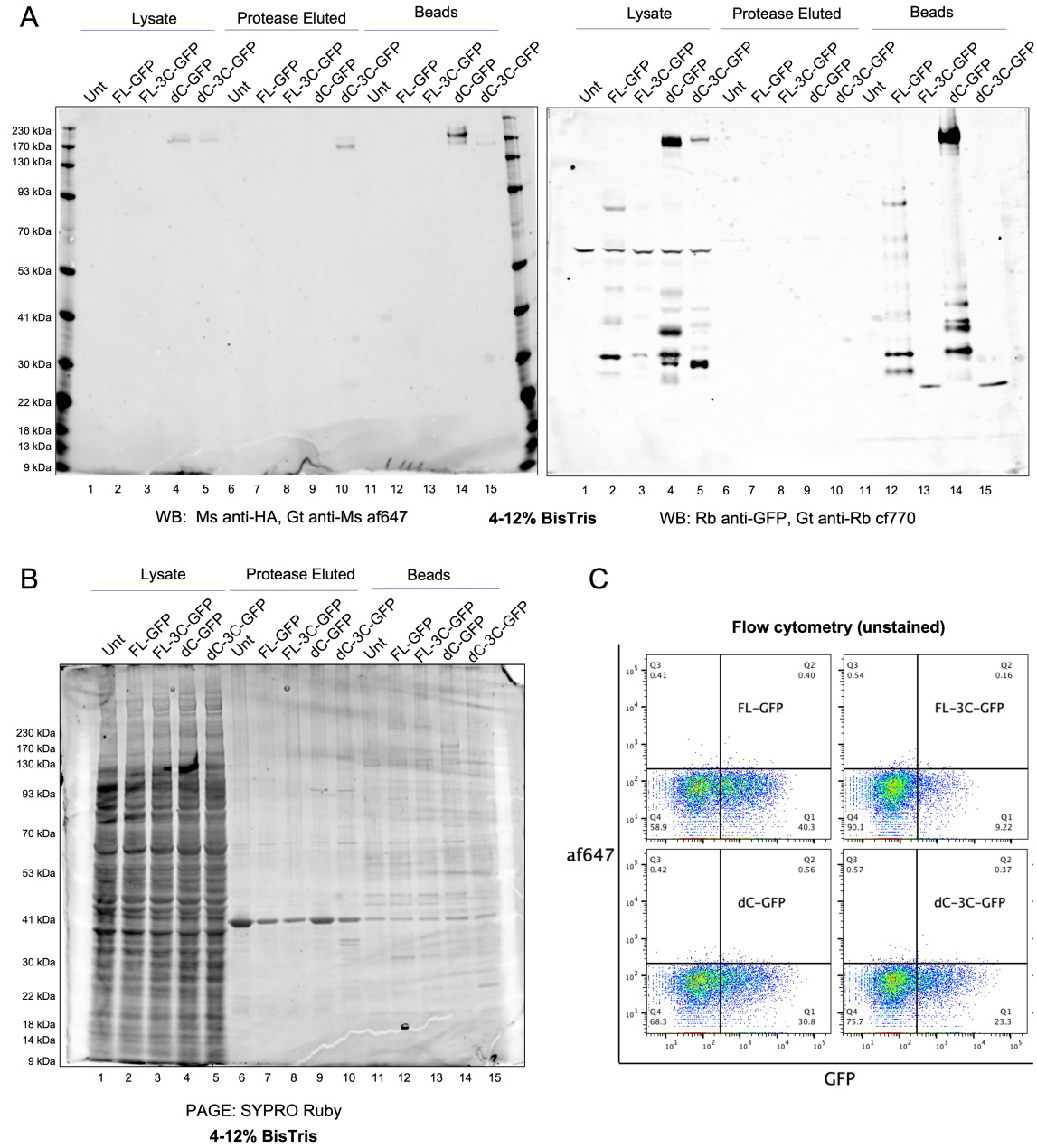

Figure 4.27: Purifying TMEM131 tail fragments via GFP-pulldown and 3C protease.
HEK cells expressing HA-TMEM131-GFP (FL-GFP), HA-TMEM-3C-GFP (FL-3C-GFP), HA-TMEM131dC-GFP (dC-GFP) and HA-TMEM131dC-3C-GFP (dC-3C-GFP) were resuspended using trypsin then lysed. GFP-tagged proteins were enriched using GFP-trap agarose beads, washed then incubated overnight at 4°C with PreScission protease. After collecting the eluted fraction, the beads were wash again then boiled in sample buffer to collect any remaining proteins.
A: Lysate, Protease Eluted and Bead fractions run on a 4-12% BisTris gel labelled with Mouse anti-HA and Rabbit anti-GFP primary antibodies and Goat anti-Mouse af647 and anti-Rabbit cf770 secondary antibodies.
B: Lysate, Protease Eluted and Bead fractions run on a 4-12% BisTris gel labelled SYPRO Ruby. The blue-coloured dyes in the protein ladder can absorb the red fluorescent emission of SYPRO Ruby dye so whilst the dye still binds to these proteins, this signal is quenched. This means the ladder has a lower signal than surrounding gel and appears here as white as the image has been inverted.
C: Flow cytometry of these cells before lysis. The af647 channel is used to separate the cells along a secondary axis, no antibodies were added.

in the lysis buffer and GTA wash buffer are not MS-compatible and there could be no C18 stage tip purification stage to remove detergents. The optimal buffer is 20-50mM Ammonium acetate which is easily vaporised during protein ionisation. This change of buffer necessitated a dialysis step. As I was unsure if the GFP-trap agarose beads or PreScission proteolysis operated well in ammonium acetate buffer, I decided the dialysis step should occur after the overnight incubation with PreScission. Another issue is that, like trypsin in tryptic-digests, the large amount of PreScission protease in the final mixture might interfere with the analysis of the rarer and more diverse TMEM131 fragments. As the PreScission protease has a GST-tag, I decided to try removing it from the eluted fraction with glutathione-magnetic beads prior to dialysis.

To scale the experiment up for mass spectrometry, I followed a similar protocol but starting from a T175 of HEK transfected with HA-TMEM131-GFP or HA-TMEM131-3C-GFP (methods section 4.3.2). After enriching GFP-tagged fragments with GFP-trap agarose beads and washing with GTA wash buffer, I did three further washes with GTA dilution buffer to dilute some of the detergent. I incubated the beads overnight with PreScission protease to cleave the tail fragments from the bound GFP-tags. I then attempted to remove the protease by incubating the supernatant with glutathione-magnetic beads then dialysed the resulting liquid into 40mM Ammonium acetate buffer using 10kD molecular weight cutoff dialysis columns. The dialysed liquid was submitted to proteomics for analysis under native mass spectrometry conditions.

The lysate, unbound, protease eluted and cleared fractions, collected during the procedure prior to dialysis, were run on a 4-12% BisTris gel, blotted and labelled with anti-HA and anti-GFP antibodies (figure 4.28A). The lysate for TMEM131-GFP shows three GFP^{+ve} bands, possibly fewer than previously due a diluted lysate. The highest molecular weight band also shows up in the unbound fraction, suggesting this is whatever protein causes the nonspecific band mentioned previously. No bands are present in the protease eluted or cleared samples but several GFP^{+ve} bands are present on the beads having been enriched by GFP-pulldown. The lane for the lysate fraction of the TMEM131-3C sample unfortunately appears distorted, likely due to being next to the high-salt and high detergent bead fraction in the previous lane. There does appear to be several GFP^{+ve} bands however that are not in the unbound fraction. The protease eluted and cleared fractions seem similarly blank as expected whilst the bead fraction shows only a single band at a consistent size for the GFP domain alone. As with previous western blots there did not appear to be any HA-positive bands with either construct so this channel was blank apart from the ladders.

30µl of the Bead and Cleared fractions for each construct were run on a parallel 4-12% BisTris gel in two lanes of 15µl then labelled overnight with SYPRO Ruby (figure 4.28B). Overnight labeling is meant to produce stronger signal but produces more speckling artefacts which are seen here. In all samples there is a band at ~70kD observed previously. There is also a strong band corresponding to molecular weight of PreScission protease. This suggests that the efforts to remove the PreScission protease with glutathione-magnetic beads were insufficient or ineffective. The bead sample for TMEM131-3C-GFP has a band at a similar size to the one in the western blot suggesting this is the GFP domain after the TMEM131 tail fragment has been cleaved off at the 3C site. A band of the same size is not present in the TMEM131-GFP bead sample but there are some bands of higher molecular weight which could correspond to the GFP with the fragment of the tail still attached. No bands are visible beneath the band for PreScission protease in the TMEM131-3C-GFP cleared sample.

I decided to cut out two bands in the TMEM131-GFP lanes and three similar sized sections of the lanes for the TMEM131-3C-GFP cleared sample (indicated by red boxes in the figure). The gel pieces were chopped up and prepared for mass spectrometry analysis with the in-gel digest protocol described in

methods section 2.6.3. Identified peptides were analysed using Scaffold with a protein threshold of 95% and a peptide threshold of 95%. Lowering these restrictions increased the numbers of peptide found but did not change the overall picture.

Unfortunately, the number of peptides identified from TMEM131 was too low to draw confident conclusions. In the 131-1 sample there were 24 unique peptides of TMEM131-GFP of which about half were in the GFP tag and half were in the TMEM131 tail. The most N-terminal peptide began at position 1260 (TSPLVLDSNTVTQGHTAGRK) relatively close to the start of the tail and close to the SpeI site used in the TMEM131dC construct. For 131-2 there were 10 unique peptides of which 7 were in the GFP tag. The most N-terminal peptide (LVDNRPPALAK) began at position 1529 about halfway through the cytoplasmic tail. In 3CG-1 a peptide from TMEM131 was identified beginning at position 1618 (GSYSSIVNSSSSSDPK). Three peptides of PreScission protease were identified in 3CG-1, one in 3CG-3, six in 131-1 and one in 131-1. Peptides from other proteins were found and are presented in supplementary table A.4.3. Many of these are common contaminants such as keratin and tubulin.

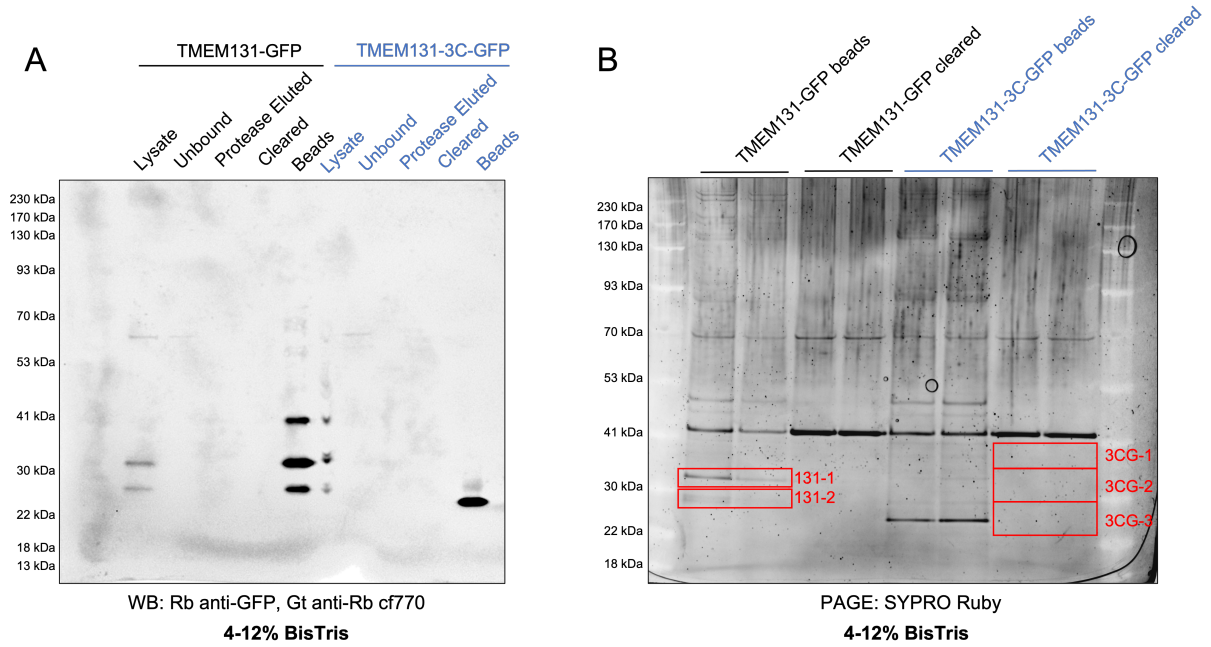

Figure 4.28: Set up for a proteomic experiment for isolating TMEM131 tail fragments.
HEK cells expressing HA-TMEM131-GFP (black) and HA-TMEM131-3C-GFP (blue) were lysed then GFP-tagged proteins enriched. The lysate sample is the diluted liquid prior to addition of GFP-trap beads whilst the unbound fraction captures the liquid after these beads were pelleted prior to washing. The Protease eluted fraction is the liquid after incubation with PreScission protease whilst the "cleared" fraction was obtained after incubation with glutathione-magnetic beads.
A: 10µl of each fraction was run on a 15-well BisTris gel, blotted then labelled with anti-HA and anti-GFP antibodies. The anti-HA/af647 channel appears blank apart from the ladder so is not shown.
B: 2x15µl of the Bead and Cleared samples were run on a parallel 10-well 4-12% BisTris gel and labelled overnight with SYPRO Ruby. The red boxes show areas cut out for mass spectrometry analysis using an in-gel tryptic digest.

The samples submitted for mass spectrometry under native conditions took many months to be analysed due to training and equipment scheduling within the proteomics facility. When analysed however no proteins could be reliably detected, likely because the protein concentration was too low. I expect the fragments of TMEM131 were at a quite low concentration after the protease step and more was likely lost on the glutathione beads. Concentration might also have been lost during dialysis. Whilst the volumes going into the dialysis columns were not measured, there seemed to be an increase in volume. Non-specific binding to the column membrane is also known to be significant for dilute samples which lowers recovery. I believe low input concentration is also likely responsible for the low peptide coverage of TMEM131 in all samples recovered from the gel by the tryptic digest.

As this experiment did not generate useful information about the TMEM131 cleavage sites I considered whether it was worth repeating with a larger input volume or any other methodological changes. One possibility would be to exchange the GFP-tag with a smaller tag such as a streptavidin-binding peptide or FLAG tag. These could also be used to enrich the protein fragments and would not have to be cleaved off for native mass spectrometry. This could however lead to other challenges such as difficulty eluting the fragments from the beads in conditions compatible for native mass spectrometry and difficulty in measuring transfection efficiency without a fluorescent tag or usable tag on the N-terminus.

Whilst the idea of performing native and tryptic-digest proteomics on the same samples was appealing so hits could be directly compared, I acknowledge that there was relatively little benefit for the 3C-cleavage site and proteolysis step to set up for the in-gel digest. Whilst removal of the GFP sequence means that all peptides identified should be in the TMEM131 tail rather than the tag, this effect is likely outweighed by the loss in yield from the extra washes to remove detergents then the incubation with PreScission and glutathione beads. Whilst waiting for the native mass spectrometry results to be returned, I decided to set up an experiment pursuing an in-gel tryptic digest without the 3C site and associated steps. I also decided to replace the SYPRO Ruby gel stain with Coomassie blue as it produces visible bands that are much easier to cut out and monitor during destaining. Whilst less sensitive than SYPRO Ruby, it should also prove a more accurate guide to if protein concentration in a band is sufficient for identification.

This experiment started with a T150 flask transfected with HA-TMEM131-GFP, a T75 flask transfected with HA-TMEM131dC-GFP and one untransfected T75 flask. After 48h, these were washed, suspended in trypsin then lysed. Diluted lysate from each flask was incubated with 50µl of conditioned GFP-trap agarose beads for 90 minutes at 4°C then washed five times. 70µl of sample buffer with DTT was used to elute the proteins from the beads. I ran two gels with these samples: one with 10µl of the lysate, unbound and bead eluted fractions and one with 2x30µl of the samples eluted from bead. The former was blotted and labelled with anti-HA and anti-GFP antibodies whilst the latter labelled with Coomassie G250 using the method described in section 2.6.3.

The blot (figure 4.29A) shows a series of GFP^{+ve} bands in the lysate lanes for TMEM131-GFP and TMEM131dC-GFP as well as the non-specific band that was also present in the lane for the untransfected control. Only the non-specific band was observed in the lanes for the unbound fraction but similar bands are present in the lanes for the fraction eluted from GFP-trap agarose beads. Although it is difficult to compare due to the distortion in lane 1, the bands in the bead elution appear to be stronger than the lysate and additional bands are apparent at higher molecular weights. For TMEM131dC, the band at approximately 230kD is stronger than in the lysate whilst lower molecular weight bands appear significantly different with a band in the lysate around 30kD appearing greatly reduced. The HA-antibody labelling shows only a band at 230kD in the bead elution. Taken together this suggests that the beads have greatly enriched GFP-tagged fragments but not done so evenly. One possibility is that smaller fragments such as the one forming the 30kD band in the TMEM131dC lysate are not eluted as efficiently as longer fragments or have been lost during blotting.

Running the bead elutions on a parallel gel labelled with Coomassie G250 stain showed bands in the lanes for TMEM131 and TMEM131dC that were not present in the lanes of the untransfected control (figure 4.29B). In the lanes for full-length TMEM131 there were six gel pieces that that could be cut out that were numbered from highest to lowest molecular weight. FL-1 consisted of two bands at the very top of the gel above the 230kD marker. These bands are not visible in the blot suggesting they have failed to transfer in the electroblotting stage. The FL-2 piece consists of two bands, though one of these also appears in the untransfected and dC lanes. This band also does not appear in the blot. Smaller molecular weight pieces numbered FL-3 to FL-6 all appear to have corresponding GFP-positive bands in

the blot. In the lanes for TMEM131dC there were five bands that could be separated and cut out as gel pieces. These were also numbered from highest to lowest molecular weight. dC-1 seemed to correspond to the bright GFP^{+ve} and HA^{+ve} band whilst dC-2 was just below. Pieces dC-3, dC-4 and dC-5 had similar apparent masses to FL-3, FL-4 and FL-5. All 11 pieces were taken to proteomics for analysis with an in-gel tryptic digest.

All gel pieces in this experiment returned more peptides from HA-TMEM131-GFP than the previous experiment with SYPRO Ruby. These were analysed using Scaffold using a protein-identification threshold of 95% and a peptide-identification threshold of 95%. The peptides of TMEM131 found for each gel piece are summarised in table 4.1 and the coverage of the HA-TMEM131-GFP construct is depicted in (figure 4.29C). The sequence of the HA-TMEM131 construct is given in supplementary figure S18.

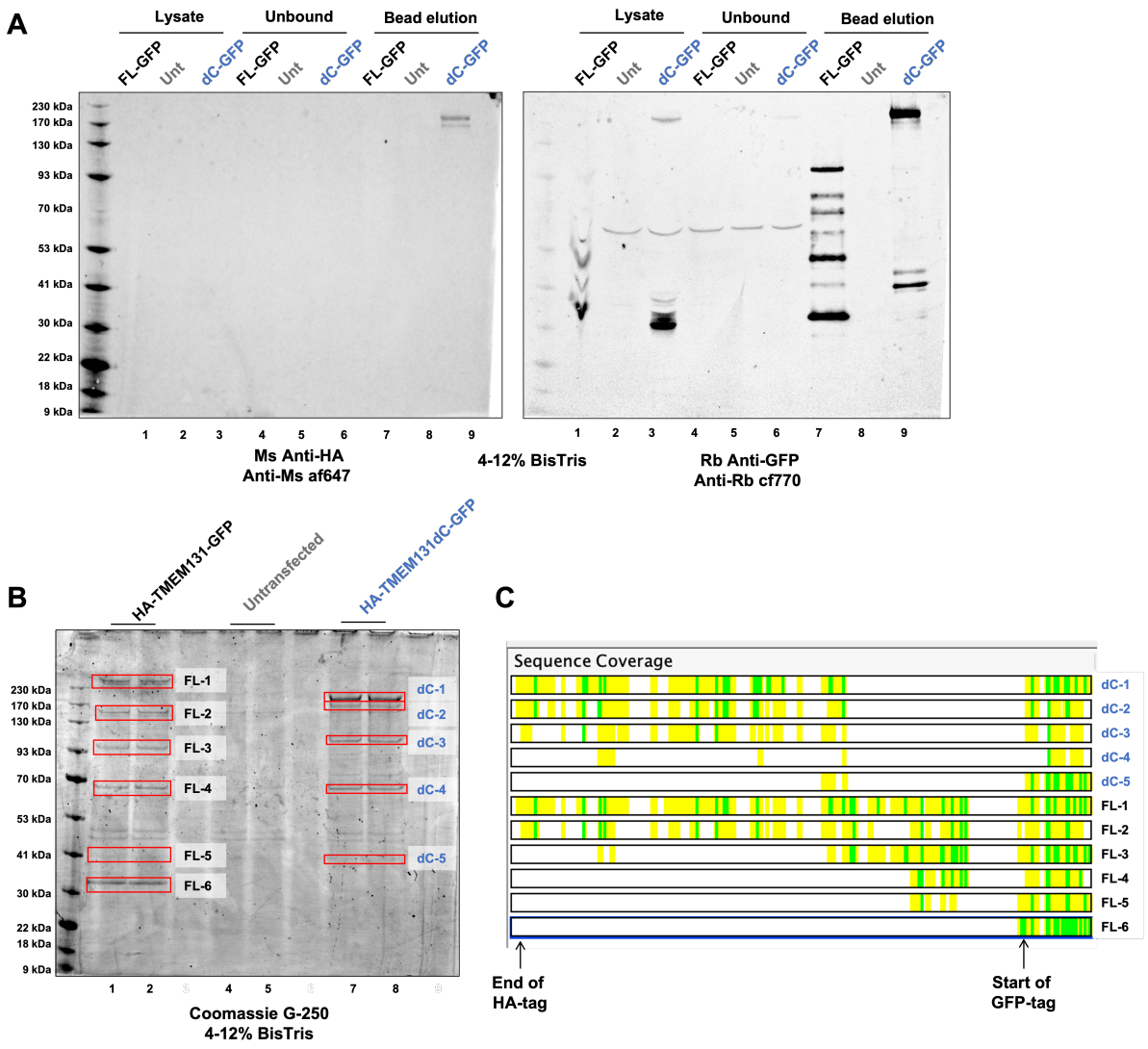

Figure 4.29: Proteomic experiment isolating TMEM131 tail fragments and mass spectrometry results.
A: Blot containing lysate (1-3), unbound (4-6) and bead eluted (7-9) fractions from cells expressing full-length HA-TMEM131-GFP (black) and HA-TMEM131dC-GFP (blue) as well as untransfected cells (grey).
B: A gel run in parallel containing two lanes of the bead eluted fraction for each sample labelled with Coomassie G250. Band that were cut out are indicated with a red box.
C: The peptide coverage of the HA-TMEM131-GFP for each of these gel pieces after an in-gel tryptic digest. Yellow indicates unmodified peptides whilst green indicates peptides with variable modifications (Deaminated N or Q, Oxidised M or an acetylated N-terminus). Arrows show the approximate positions of the end of the HA-tag and start of the GFP-tag.

As expected, the coverage of the protein decreases as the molecular weight decreases and coverage is lost

in the N-terminal regions of the protein before the C-terminal regions. FL-1 for instance has 219 unique peptides from 293 total spectra and covers 57% of the construct including the first residue of the HA-tag which becomes the mature N-terminus after removal of the signal peptide and the final residue of the GFP tag. Meanwhile FL-6 has 104 unique peptides from 221 total spectra but only covers 11% of the construct of which almost all is the GFP-tag (residue V1858 onwards). Across all FL samples, there is a 183 amino acid region between V1648 and R1830 in which no peptides are recovered. This is likely because it contains unusually few trypsin sites with only two K and two R residues.

Sample	Approx. MW on gel (kDa)	Unique Peptides from construct	Percentage Coverage of construct	Most N-terminal Residue Identified
FL-1	240	219	57	Y19
FL-2	160	93	37	E32
FL-3	95	144	29	V312
FL-4	65	56	16	E1445
FL-5	40	37	15	E1445
FL-6	35	104	11	I1831
dC-1	180	194	63	Y19
dC-2	170	93	48	Y19
dC-3	100	49	37	E32
dC-4	65	14	10	V312
dC-5	40	57	18	L1124

Table 4.1: Summary data for the peptides of TMEM131 identified in each gel piece.
Six pieces of the Coomassie gel shown in figure 4.29B were cut up in the TMEM131 lanes and five in the TMEM131dC lanes then processed with an in-gel tryptic digest. The molecular weight of the centre of each piece was estimated from the protein ladder. The values for the number of unique peptides and the percentage coverage refer to the construct expressed in that lane. The number of the most N-terminal residue found refers to the full-length HA-TMEM131-GFP construct. The HA-TMEM131dC-GFP construct is identical apart from the absence of the tail section between residues 1213 and 1854. Due to the 18 residue gaussia luciferase signal peptide and 11 residue HA epitope tag, the first residue in the expressed construct that is encoded by the human gene is E30. This corresponds to E60 in the endogenous protein which is also the mature N terminus due to the 59-residue endogenous signal peptide.

Comparing the approximate masses of the bands on the gel with the peptides found in each piece revealed some surprising inconsistencies. FL-1 appears to contain the entire peptide sequence and has a mass of approximately 240kD which is consistent with the theoretical mass of 228.3kDa for the construct minus its signal peptide. FL-2 however covers a similarly wide range of peptides from near the N-terminus to the end of the GFP tag but appears around 160kDa on the gel. A similar effect is seen for the dC-1 and dC-2 samples which appeared to have very similar peptide coverage despite a difference in apparent mass of about 10kDa. These however both appeared slightly heavier than the theoretical mass of the HA-TMEM131dC-GFP construct which is 160.2 kDa.

There also appeared to be discordance between the apparent mass of fragments and the peptide coverage in the lower molecular weight pieces. Most obviously peptides from FL-4 and FL-5 appeared to come from the same region of the construct and have near identical coverage despite a difference in apparent mass of these pieces on gel of ~25kDa. The theoretical mass from E1445, the most N-terminal residue identified in both to the end of the GFP is 70.3kDa. This more consistent with the observed mass of FL-4 (~65kDa) than FL-5 (~40kDa).

The peptide coverage of FL-3 was also strange as the coverage was strongest towards the end of the tail from I1144 onwards but two peptides were found from the N-terminal regions of the protein beginning at positions V312 and V358. The mass of the sequence from V312 to the end is 195.6 kDa which is much greater than its apparent mass of around 95kDa. The mass from position I1144 onwards is more consistent at 102.6 kDa. If these two peptides are not actually part of the same fragment as the rest of

the peptides in this gel piece then how were they enriched by the GFP-trap beads, persist through the wash steps to be present in this area of the gel? One possibility is that they are from the N-terminal product of the protein after it has been cleaved in the TMEM131 tail that has appeared in the gel due to non-covalent interactions with other fragments such as the full-length construct present in FL-1. The split Venus experiment (figure S15) suggested that the N-terminal domains of TMEM131 form oligomers which would support this idea. The two pieces would separate once denatured by the sample buffer. Something similar might also be occurring in dC-4 where peptides from the same region also appear. Again, the theoretical mass of a single peptide from V312 to the end of the GFP would produce a peptide with a mass inconsistent with the size of gel piece from which these peptides were obtained (127.6kDa vs 65kDa).

Non-covalent interactions between GFP-tagged TMEM131 and other proteins could explain the numerous other proteins identified in the gel pieces. Though I had not intended this experiment to identify TMEM131 interaction partners, the list of proteins identified did share surprising similarities with the list of interaction partners identified by BioID (section 4.6.1). The lists for both experiments are found in supplementary section A.4.4 and will be discussed later.

Whilst this experiment generated much higher peptide coverage than I expected, I am doubtful that it is sufficient alone to define the cleavage sites in the TMEM131 tail. I find it curious that the most N-terminal residue found in FL-2 was also found in dC-3 and the same was true for FL-3 and dC-4 and in FL-4 and FL-5. As these are all to the C-terminal side of a K or R residue I find it more likely that these ends were produced by trypsin digestion during the gel extraction rather than by whatever protein is cleaving the tail in HEK cells. Presumably the endogenous cleavage sites are to the N-terminal sides of these residues, though it is difficult to say where as we cannot be confident that the most N-terminal peptide of the cleaved product has been detected.

Nonetheless I believe the experiment generated useful insight. Firstly, it was interesting that peptides from the 11 amino acid HA epitope were found in FL-1 in addition to dC-1 and dC-2, given the difficulty in detecting the full-length construct using anti-HA antibodies with western blotting or flow cytometry. This supports the hypothesis that the absence of a HA-positive band for the intact HA-TMEM-GFP construct is an artefact of gel blotting. It does not explain however the apparent lack of HA-tagged fragments when the tail is cleaved.

More interesting however was the discordance between apparent mass and peptide coverage which I wanted to investigate further. The most likely explanation was a post-transcriptional modification of some of the fragments increased their apparent mass. Two common post-transcriptional modifications are large enough to explain the mass differences observed: ubiquitination and glycosylation. I decided to investigate both possibilities.

4.5.2 Discordance between apparent mass of TMEM131 fragments and the detected sequence is partially explained by glycosylation

Glycosylation is the addition of carbohydrate residues onto the extracellular domains of a protein, most commonly on the asparagine residues of Asn–X–Ser/Thr consensus sequences. This addition is initiated during translation in the ER with the transfer of a 14 sugar unit, high-mannose branched oligosaccharide onto N residues by the enzyme oligosaccharyltransferase. The terminal glucose units of this oligosaccharide core are trimmed by exoglycosidases in the ER during protein folding. This trimming acts as a quality control checkpoint before the protein is allowed to leave the ER. The oligosaccharide

can be trimmed further in the cis-Golgi before later processing in the medial and trans-Golgi which can include addition of other sugars to the chain. Different amounts of glycosylation could potentially explain the difference in apparent mass between proteins identified in gel pieces dC-1 and dC-2 in the experiment above as these have peptides from the N-terminal domains of TMEM131. However, it would not easily explain the differences in apparent mass between FL-5 and FL-6 which only have residues from the cytoplasmic tail.

To test whether the TMEM131 protein fragments visible via western blotting were glycosylated, I purified the protein fragments via the GFP tag and treated fractions of the protein with the enzymes Endoglycosidase H (EndoH) or Peptide:N-glycosidase F (PNGaseF). Full details for this protocol are provided in section 4.3.3. EndoH is able to remove most of the mannose-rich oligosaccharide core that is added to proteins in the ER but cannot remove oligosaccharides once they have been trimmed by enzymes in the Golgi. Surface glycoproteins are therefore EndoH-resistant as they have passed through the trans-Golgi network. Meanwhile the enzyme PNGaseF is able to entirely remove most N-linked glycosylations with very little selectivity for oligosaccharide or protein structure. This means it can remove glycosylations from proteins regardless of where they are in the secretory pathway. Once treated proteins can be analysed by western blotting. The downwards shift of a band present in a control lane upon treatment with PNGaseF is evidence that the band contains a glycosylation. If this band does not shift with EndoH treatment then it is evidence that the band contains protein that has been transported through the trans-Golgi network.

The treated TMEM131 proteins were run on a 4-12% gel, blotted then labelled with anti-HA and anti-GFP antibodies (figure 4.30). The pre-TCR, with a GFP-tagged TCRβ chain was included as a control (lanes 7-9). This chain appears as a single band of around 60kDa without enzyme treatment but treatment with either EndoH or PNGaseF reduces the band to around 55kDa. This EndoH-sensitivity is consistent with the majority of pre-TCR complexes being present in the ER, at least when expressed in HEK cells, with an undetectable fraction reaching the cell surface. This is consistent with multiple prior experiments discussed in the previous chapter.

The difficulty in this experiment was ensuring identical volumes of purified protein were subject to each treatment then identical amounts loaded onto the gel. This involved accurate pipetting of small volumes of rather viscous solutions. The difference in loading can be approximated by quantifying the total fluorescence in the GFP channel in each lane and calculating the ratio. Doing so suggests that more EndoH-treated TMEM131 was loaded than either the control or PNGaseF-treated protein (1.00 : 1.70 : 1.03 in the order Cnt : EndoH : PNGase as on gel). This analysis found slightly less EndoH-treated TMEM131dC was loaded (1.00 : 0.80 : 1.07) as well as an excess of PNGaseF treated pre-TCR (1.00 : 1.06 : 2.11). A previous run where the differences were more excessive is shown in supplementary figure S19. This gel also shows the comparison between the pre-TCR which appears entirely EndoH-sensitive with the mature TCR which appears partially EndoH-insensitive due to the fraction of the receptor complex that is found at the cell surface.

As in previous experiments, there are no bands consistent with the mass of the intact TMEM131-GFP construct at the top of the gel and the HA-tag on the N-terminus is not detected (lanes 1-3). The series of GFP-tagged fragments is present as normal and there does not appear to be a difference in the banding pattern between the three treatments. This implies these fragments are not glycosylated which is consistent with their cytoplasmic localisation.

As expected there are HA^{+ve}/GFP^{+ve} bands in the lanes for TMEM131dC. In the control lane (lane 4) there appears to be 3 of these bands with the middle band around 200kDa being most intense.

Treatment with EndoH (lane 5) causes the middle and lower bands shift down the gel whilst the upper band remains the same size. This implies that the majority of the intact TMEM131dC-GFP construct is EndoH-sensitive but there is a fraction of the protein that has transited through the trans-Golgi network to become EndoH-insensitive. This is consistent with previous microscopy experiments where TMEM131dC appears to localise to the ER but also forms cytoplasmic vesicles and is detectable at the cell surface with anti-HA antibodies. I am unable to conclude with certainty that this EndoH-insensitive fraction is TMEM131dC present at the cell surface but one way to confirm this would be to label cell surface proteins with a membrane-impermeable, non-specific surface marker such as biotin-NHS before the GFP-pulldown. Biotinylated protein could be detected on the gel after EndoH/PNGase treatment and a downwards shift of a band with PNGaseF but not EndoH would confirm membrane localisation of this fraction.

These high molecular weight bands are compatible with the bands cut out in the dC-1 and dC-2 gel pieces in the experiment above (figure 4.29). Both these bands were heavier than the theoretical molecular weight from the construct sequence. Detection of EndoH-sensitive and EndoH-insensitive fractions would suggest that glycosylation not only makes up the addition mass but also results in two bands of different sizes that have similar peptide coverage. Treating the protein with PNGaseF produces only a single HA^{+ve}/GFP^{+ve} band at 170kDa which is much more consistent with the theoretical molecular weight. This band is smeared in both channels likely due to the extra NP40 that is added in this enzyme treatment which could compete with the LDS detergent in the sample buffer and gel.

Much more surprising was that glycosidase treatment produced differences in the banding of GFP-tagged fragments of TMEM131dC at lower molecular weights. Two GFP^{+ve} bands are present at about 35 and 45kDa in the control lane. EndoH-treatment appears to reduce the intensity of the higher of these two bands and a new band is apparent at about 40kD. Treatment with PNGaseF completely removes the 45kDa band and strengthens this 40kD band. Overcontrasting the GFP channel reveals a similar shift in bands in the 70kD and 55kD range.

In the Coomassie-labelled gel in figure 4.29 no bands at 35kD or 45kD were detected and cut out but a band was detected at 40kD and analysed as gel piece dC-5. The most N-terminal residue detected in this piece was L1124. Prior experiments suggested that the second of two putative transmembrane helices was genuine but this ends at position 1110. If dC-5 contains glycosidase-sensitive protein then it implies that the fragment actually extends further towards the N-terminus and includes sequence in the ER lumen. The nearest glycosylation consensus sequence is a NAS sequence starting position 1066, between the two putative TM helixes. It is not unreasonable the fragment in dC-5 contains these residues but peptides from this region were not captured. This has bigger consequences for the mechanism of protein cleavage. We had previously assumed that the protein cleavage occurred in the protein's cytoplasmic tail as this was most consistent with the fragment sizes produced by the full-length TMEM131 protein. Cleavage also seemed intrinsic to the tail as it still occurred when the tail was appended to CD86. GFP-tagged fragments of TMEM131dC with ER-lumen sequence however imply that cleavage might also occur in the transmembrane region or N-terminal domains of the protein. These bands are partially EndoH-sensitive implying that at least some of this fragment has transited the trans-Golgi network. If these bands were just the result of protein misfolding and immediate degradation by Endoplasmic-reticulum-associated protein degradation (ERAD) then we would not expect this EndoH-insensitivity. We might infer that cleavage occurs after TMEM131dC-GFP has left the ER or the product of cleavage is still able to progress through the secretory pathway. The two other gel pieces from the Coomassie gel, dC-3 and dC-5, had peptides from the lumenal domains but did not appear to contain the whole sequence. These proteins in these pieces could be producing the weaker glycosidase-sensitive bands of TMEM131dC visible here

though the apparent masses do not appear to match exactly. I suspect this is caused by gel to gel variability combined with the scaling effect of gradient gels.

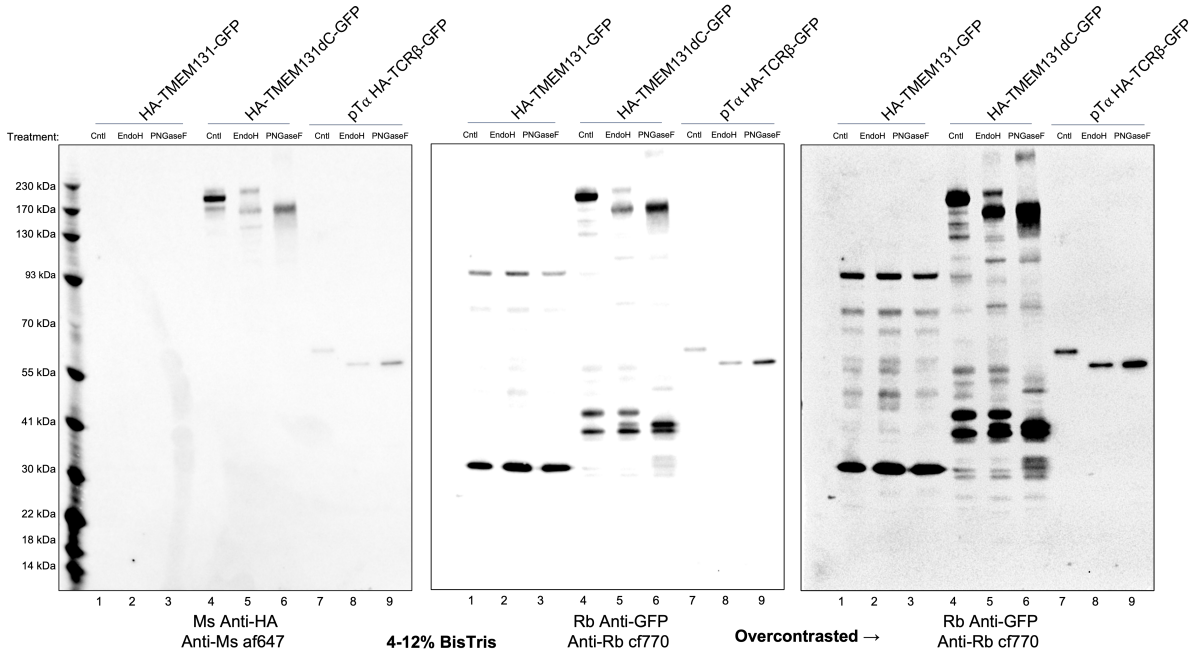

Figure 4.30: TMEM131dC is an EndoH-sensitive glycoprotein whilst the tail fragments of TMEM131 are EndoH and PNGase-insensitive.
TMEM131 and TMEM131dC proteins were enriched by GFP-pulldown then treated with EndoH, PNGaseF or left untreated (Cntl). Proteins were run on a 4-12% BisTris gel, blotted then labelled with anti-HA and anti-GFP antibodies. The picture on the right shows the same anti-GFP channel as the middle picture but has the contrast increased further to make fainter bands more noticeable.

These observations confirmed that TMEM131dC is glycosylated. The differential glycosylation observed would explain the difference in mass of the fragments forming bands in dC-1 and dC-2 that have similar peptide coverage. I would also assume that the full-length TMEM131 protein is glycosylated and the intact protein does not appear as a band because it is too big to be transferred via this blotting procedure. This means that different glycosylations could explain the similar discrepancy in apparent mass in pieces FL-1 and FL-2. An important caveat however is that full-length TMEM131 is not detectable at the cell surface via the HA-tag and it is unclear how much it progresses through the secretory pathway. ER exit and transit through the Golgi would be necessary for the glycosylations to be processed and result in mass differences. Some mobile foci of TMEM131 were observed in prior microscopy experiments with fluorescent tags on the C-terminus but these were nowhere near as frequent as in shorter constructs. Glycosylation would also not explain the difference between the proteins found in gel pieces FL-4 and FL-5. These fragments, also seemed to have different masses with near identical peptide coverage but both terminated at E1445, far from the transmembrane domains and any lumenal sequences. Whilst there are NXS/T glycosylation sequences in this region, the coverage makes it seem very likely that these fragments are entirely cytoplasmic sequence so lack glycosylation. I therefore decided to investigate protein ubiquitination which could also explain the mass discrepancy.

4.5.3 TMEM131 cleavage is related to proteasome degradation

Ubiquitination is the addition of a small 8.6kD ubiquitin (Ub) protein unit onto a target protein through an isopeptide bond between the C-terminus of ubiquitin to a lysine residue on the target protein. This is

catalysed by a diverse group of E3 ubiquitin ligase enzymes that specify protein targets. Ubiquitination can occur once (monoubiquitination) but can also be extended to form straight or branched chains (polyubiquitination). Different forms of ubiquitination can be discriminated by different antibodies, for instance the P4D1 antibody recognises monoubiquitin as efficiently as polyubiquitination whilst the FK1 and FK2 antibodies recognise only polyubiquitin chains. Different Ubiquitin modifications can have different effects on the target proteins. Typically however, polyubiquitination results in the re-targeting of a protein and its degradation by proteasomes.

The centre of the human proteasome is a 20S core particle that is shaped like a hollow barrel. Proteolytic active sites are present in the centre of this barrel but the small internal diameter prevents entry of most cytoplasmic proteins. The 20S core is flanked by 19S regulatory particles are responsible for recognising polyubiquitinated proteins and translocating these proteins into the narrow 20S core for degradation. Translocation is usually initiated at an unstructured region near the Ub tag, either one of the termini or an internal sequence, which is engaged by loops of an unfoldase domain in the 19S particle. The ATP-dependent action of motor proteins in this unfoldase domain progressively pulls the protein into the core particle. This involves the unfolding of any globular domains in the protein. The protein is deubiquitinated during translocation and the ubiquitin later reused. As the protein is translocated into the 20S core, it is usually degraded completely into short peptides 3-25 amino acids in length.[348]

Ubiquitin is attached to lysine residues of substrate proteins via its C-terminus which has the sequence LRLRGG. Trypsin digestion of ubiquitinated proteins therefore leads to two glycine residues being added onto peptides of the target protein at the ubiquitination site. After considering ubiquitination of TMEM131, I asked for the mass spectrometry spectra from the prior GFP-pulldown experiment (section 4.5.1) to be re-analysed with GG as a variable modification on Lys-residues. Using the standard 95% protein confidence threshold and 95% peptide confidence threshold there were no ubiquitinated peptides of TMEM131 detected. There were ubiquitinated peptides from other proteins with these settings including HSP70 in FL-4 and dC-4 (consistent with its MW) and ubiquitin itself in the form of polyubiquitin-B/B4DV12. If the peptide confidence threshold was lowered to 50% then the same ubiquitinated peptide of TMEM131 was found in FL-1 and FL-3 (FLPNSQELGNTSSSEGEK where the final K was modified with GG). Although the protein coverage in the experiment was surprisingly high, I don't think it improbable that even abundant ubiquitinated peptides of TMEM131 could be missed.

As an initial experiment to see if TMEM131 fragments are ubiquitinated, I expressed four TMEM131 constructs then enriched GFP-tagged proteins with GFP-trap beads. I then ran the lysates and bead elutions on a 4-12% gel. This gel was blotted and labelled with rabbit anti-GFP antibodies and the P4D1 anti-Ub antibody (figure 4.31). Unfortunately, the enrichment of GFP-tagged proteins was low and antibody staining for both channels was relatively poor. However it still appeared that the GFP-tagged bands in the TMEM131 and TMEM131dC lysate were also labelled by the Ub-antibody (lanes 1 and 2). Bands consistent with the intact TMEM131dC and the tailless TMEM131ntd construct were visible at the top of the gel in the bead elution lanes (6 and 7).

Although I had seemingly detected ubiquitinated fragments of TMEM131, I wanted more evidence that the ubiquitin-modification had a functional role in TMEM131 degradation and maybe in the cleavage of the protein itself. One way of enriching ubiquitinated proteins is treating cells with MG132: a potent, cell-permeable proteasome inhibitor. Without the proteasome to deubiquitinate and degrade proteins, the ubiquitin pathway backs up resulting in an accumulation of polyubiquitinated proteins. As the anti-Ub antibody seemed to be able to pick up GFP-linked bands in cell lysate and having a temporary shortage of GFP-trap agarose beads, I decided to just use cell lysate for further experiments.

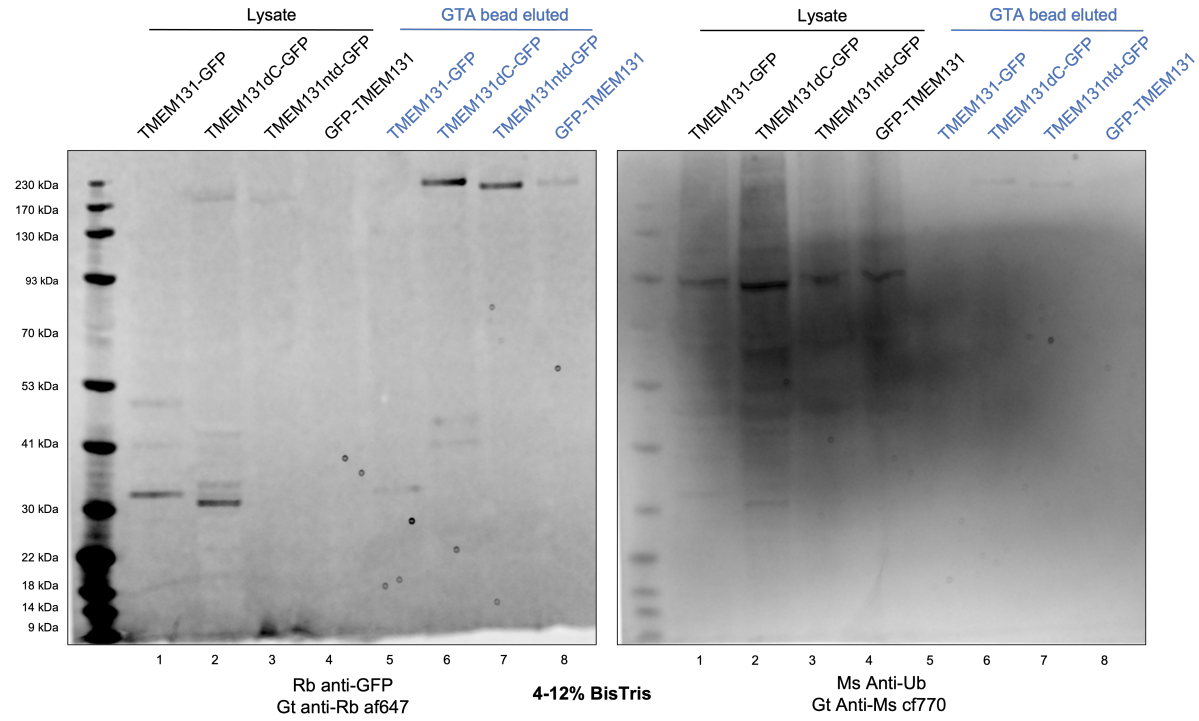

Figure 4.31: Tail fragments of TMEM131 are ubiquitinated.
HEK cells expressing TMEM131 constructs were lysed and GFP-tagged proteins enriched with GFP-trap agarose beads. Proteins were eluted from the beads by heat in sample buffer. Lysate (black) and bead elution (blue) were loaded onto a 4-12% BisTris gel then the blot was stained with Mouse anti-Ub and Rabbit anti-GFP primary antibodies then goat anti-Rabbit af647 and goat anti-Mouse cf770 secondary antibodies.

I transfected two wells with TMEM131 constructs using the same transfection mix for each. After 24h I treated one well with 50µM MG132. Cells then were lysed after another 24h. Ideally, I would have transfected a single population of cells then split the population evenly before treating one half with the drug. This would match the transfection efficiency in the treatment groups making comparisons easier. This would be feasible with a suspension cell line but would not be feasible for adherent HEK cells as the cells would need to be resuspended with trypsin to be divided and once treated with MG132 the cells are less likely to adhere back to the surface.

I ran lysate from treated and untreated cell on a 4-12% gel then labelled with blot with Total Protein normalisation (TPN) before blocking as normal and labelling with anti-GFP and anti-Ub antibodies (figure 4.32). The total protein measurement suggested that a similar amount of protein was loaded in each condition for all constructs (ratio untreated to treated between 0.97 and 1.03) with the exception of TMEM131ntd-GFP (lanes 5-6, ratio 0.91).

Treatment with 50µM MG132 appeared to increase the staining of the anti-Ub antibody at the top of the gel giving the blot a striped appearance in alternate lanes in that channel. The full-length TMEM131 appeared to be well expressed in this experiment (lanes 1 and 2) and GFP-tagged fragments of the tail were very noticeable in the GFP channel. Most of these bands, especially the bands at ~30kD and ~40kD, were also visible in the Ub channel amongst the background staining of other proteins. A faint band is present at ~28kD in the GFP channel in lanes 1 and 2 that is not present in the ubiquitin channel. None of these TMEM131 fragments were visible in the channel labelled for total protein suggesting that whilst they are a significant fraction of ubiquitinated proteins they are not a significant proportion of total proteins.

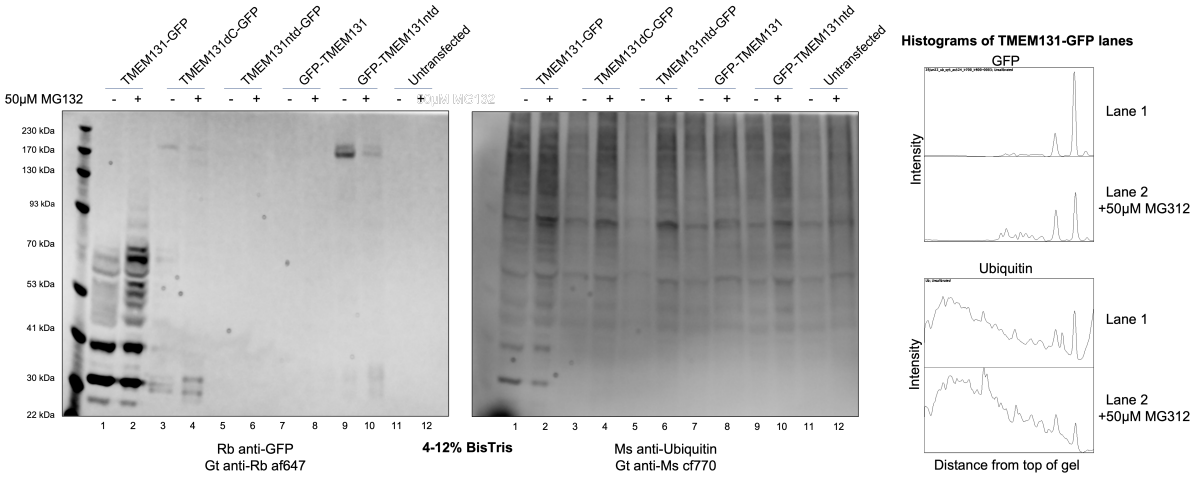

Figure 4.32: Proteasome-inhibitor MG132 increases the molecular weight of TMEM131 fragments. HEK cells expressing TMEM131 constructs were produced in duplicate with half exposed to 50μM MG132 for 24h before cells were lysed. The blot was stained with Mouse anti-Ub and Rabbit anti-GFP primary antibodies then goat anti-Rabbit af647 and goat anti-Mouse cf770 secondary antibodies. The histograms to right show the average intensity along the length of the first two lanes of the blot in the GFP and ubiquitin channels.

Comparing the lanes for TMEM131 with and without MG132 suggest there is an increase in the average mass of fragments with the inhibitor. This is strongest in the 40-70 kDa range. Though more subtle, this increase also appears to occur in the Ub-channel. Plotting the profile of the two lanes on histograms (right of the figure) makes this increase clearer. The bands in this region are more closely spaced than I have observed with TMEM131-GFP previously and seem to differ by ~5-10kD. The bands at 30kD and ~40kD are comparable in size with the bands appearing in the FL-5 and FL-6 pieces analysed by mass spectrometry. Proteins in those piece differed in apparent mass by ~10kD but had comparable peptide coverage. These might be fragments that differ by a ubiquitin monomer.

This western blot had worse anti-GFP labelling in the lanes for the other constructs. In lanes 3 and 4 containing lysate from cells expressing TMEM131dC it seems that treatment with MG132 increases the size of lower molecular weight range whilst bands at 170kDa are not affected. MG132 also did not change the size of bands of GFP-TMEM131ntd (lanes 9 and 10). I repeated this experiment with a set of TMEM131 truncated constructs with similar results (supplementary figure S20).

As the fragments of TMEM131 are ubiquitinated and stabilised by proteasome inhibition, I consider it likely that the proteasome plays a key role in the cleavage of the TMEM131 cytoplasmic tail. Degradation by the proteasome is initiated in unstructured regions near ubiquitination sites. Given the tail appears entirely unstructured and Lysine residues are abundant this could potentially occur at many points throughout this sequence. Normally after translocation is initiated in the middle of a protein it is degraded entirely in both directions. Why then is TMEM131 degraded to a series of GFP-linked fragments and corresponding HA-tagged products absent?

Partial degradation by the proteasome has been reported for a small handful of proteins.[348] These include the EBNA-1 protein produced by Epstein Barr virus which appears to avoid proteolysis to stop viral peptides being presented to the immune system by MHC-I molecules. Partial degradation of EBNA1 seems to be caused by a 239 amino acid glycine-alanine rich region as removal of this region increases its degradation rate *in vitro* and its insertion into other proteins can cause them to be partially degraded.[349] The fragments created by proteasome partial proteolysis appears to have a functional role in the case of NFκB subunit p50 that is generated from partial degradation from larger p105 precursor. The N-terminal domains of p150 appear to be protected by a gly-rich region. How these sequences lead

to partial degradation is unclear. One hypothesis is that these low-complexity regions are somewhat "slippery" to the loops of the unfoldase motor proteins so the protein dissociates during translocation. Alternatively, these low-complexity regions may not allow sufficient pulling force to generated for the required unfolding of globular domains. This effect seems to be most prominent when degradation is initiated from internal sites.[350] Perhaps something similar could be occurring in TMEM131 as its tail is relatively low in complexity and quite rich in G, A and S residues?

An alternative hypothesis which seemed a lot more credible, related to the GFP domain on the C-terminus of the TMEM131 constructs I had been using. This domain was present in all of the western blotting experiments investigating TMEM131 tail cleavage I had performed as well as experiments performed by other lab members. Using the GFP domain was convenient for measuring transfection efficiency and the domain has reliable antibodies for both flow cytometry and western blotting. However GFP is an extremely well-folding, artificial domain which has a non-standard peptide structure (the fluorophore) at its core. Incomplete degradation of GFP by the proteasome has been reported by several groups[351,352] and has used to investigate proteasome processivity.[353]

If the GFP domain was responsible for partial degradation of the TMEM131 tail then this could explain prior observations. In particular, it would explain how the GFP^{+ve} banding pattern is affected by truncation of the TMEM131 tail with effects of both sequence and distance from the TMEM131 C-terminus (section 4.4.1). I would expect that particular sequences in the tail are less easily engaged, translocated and cleaved than others which would produce bands that change size when the tail is truncated. However, I also expect that distance along the chain from the well-folded GFP-domain is an important factor in proteasome unfolding and stalling. This would produce bands persistent in length when the tail is truncated. This would also explain why the banding is intrinsic to the tail and appears when the tail is fused to CD86. I have discussed experiments that used full-length TMEM131 without a C-terminal GFP tag but these either did not investigate TMEM131 degradation directly or used mScar on the C-terminus instead which might produce the same effect.

Clearly the presence of a GFP domain is not sufficient for partial degradation as we do not see similar pattern of GFP^{+ve} bands when GFP is fused to other proteins such as the chains of the TCR. This would imply that some features of the TMEM131 tail are also important. These might be its length, sequence complexity or the distribution of Lys-residues causing proteasome initiation from an internal sequence. I was keen to investigate if partial degradation of the TMEM131 tail is an artefact of the GFP domain which seemed likely.

4.5.4 Incomplete degradation by the proteasome is not just an artefact of the GFP domain

Investigating cleavage of the protein tail without using the GFP tag would require replacing the tag with a smaller epitope. I considered using a HA tag but it was already present on the N-terminus of many of my TMEM131 constructs and would require a lot of cloning to remove and then add to the C-terminus. Instead I used PCR to clone the FLAG epitope sequence (DYKDDDDK) onto the C-terminal end of the TMEM131 tail and onto the C-terminus of GFP (and mScar) domains. These parts allowed me make TMEM131 constructs tagged with both GFP and FLAG (TMEM131-FLAG-GFP and TMEM131-GFP-FLAG) and a construct with just the flag tag (TMEM131-FLAG). If the GFP domain was causing the protein to be partially degraded by the proteasome then I would expect multiple FLAG-tagged bands with the former but not the latter as it would be degraded completely.

TMEM131 constructs with FLAG and/or GFP tags were expressed in HEK cells. After 48h, cells were collected for flow cytometry and lysed. Transfection efficiency seemed reasonable in all samples expressing a fluorescent protein. The lysate was run on a 4-12% BisTris gel that was blotted using the iBlot2 semi-dry blotting system. The blot was labelled with rabbit anti-GFP antibodies and the M2 mouse anti-FLAG antibody with goat anti-rabbit af647 and Goat anti-mouse cf770 secondary antibodies. This blot is shown in figure 4.33A.

In the anti-GFP channel, there is a pattern of low molecular weight bands for the TMEM131-GFP construct consistent with prior experiments (lane 1). The patterns for TMEM131-GFP-FLAG and TMEM131-FLAG-GFP (lanes 2 and 3) look very similar though very slightly larger consistent with the 1kDa gain in mass of the FLAG sequence. These lanes did not have any bands consistent with the masses of the intact protein. The consistency of these bands implies that the addition of the FLAG tag in either position has not affected the partial degradation. In the anti-FLAG channel there are strong bands at the same positions for the TMEM131-GFP-FLAG and TMEM131-FLAG-GFP lanes as expected but also weaker bands in the TMEM131-GFP lane which should not have the FLAG epitope. Lanes 4-6 contained the lysate for TMEM131-FLAG, GFP-TMEM131 and GFP-TMEM131-FLAG though no bands were observed in either of the GFP or FLAG channels. Faint bands were observed in the anti-GFP channel for TMEM131dC-GFP (lane 7) and TMEM131dC-GFP-FLAG (lane 8) at masses consistent for the intact protein and ~30kDa consistent with cleaved fragments. The higher molecular weight band only is visible in the anti-FLAG channel for TMEM131dC-GFP-FLAG. Lanes 9 and 10 contained lysate from TMEM131dC-mScarFLAG and TMEM131-mScar-FLAG which were only visible in the anti-FLAG channel. The former had a high molecular weight band while both contained lower molecular weight bands including a band at ~18kD.

The absence of lower molecular weight bands in the anti-FLAG channel in the lanes for TMEM131-FLAG and GFP-TMEM131-FLAG does suggest that the partial degradation does not happen without the GFP domain. However I was keen to find a way to visualise the intact protein bands for each construct to be confident that the anti-FLAG antibody would detect the epitope if it were present. Whilst I can monitor the transfection efficiency of the GFP-TMEM131-FLAG construct, I cannot for TMEM131-FLAG so bands would also be absent if the construct were poorly expressed. As high molecular weight TMEM131 proteins are poorly transferred in the blotting step, I tried two different gels using the same samples in the same order: an identical 4-12% BisTris gel that was blotted using a wet-blotting apparatus (figure 4.33B, method in section 4.3.4) and a 3-8% TrisAcetate gel blotted using the semi-dry method (figure 4.33C).

Wet blotting resulted in a higher background in both channels but managed to better resolve higher molecular weight bands. In the anti-GFP channel, bands consistent with the intact constructs are visible for in the first three lanes and the two constructs with GFP on the N-terminus. The latter appeared as two bands around 230kDa and 200kDa perhaps two different glycosylation isoforms. Unfortunately, these are not convincingly present in the anti-FLAG channel whilst a band in the TMEM131dC-GFP lane is present. The 3-8% TrisAcetate gel had better signal for higher molecular weight bands than the first 4-12% BisTris gel but there was still poor signal in the anti-FLAG channel.

Unable to confidently demonstrate expression of the constructs without the GFP-tag, I decided to fuse the TMEM131 tail to the membrane protein CD86. When the TMEM131 tail with a C-terminal GFP tag was fused to CD86 previously the tail produced the same GFP^{+ve} bands as TMEM131 implying that cleavage was intrinsic to tail sequence (figure 4.9). The CD86 fusion was better expressed than TMEM131 though its localisation was different. Since then I had cloned CD86 with an N-terminal GFP tag and I decided to add the TMEM131 tail to this as well. HEK cells expressing these constructs were

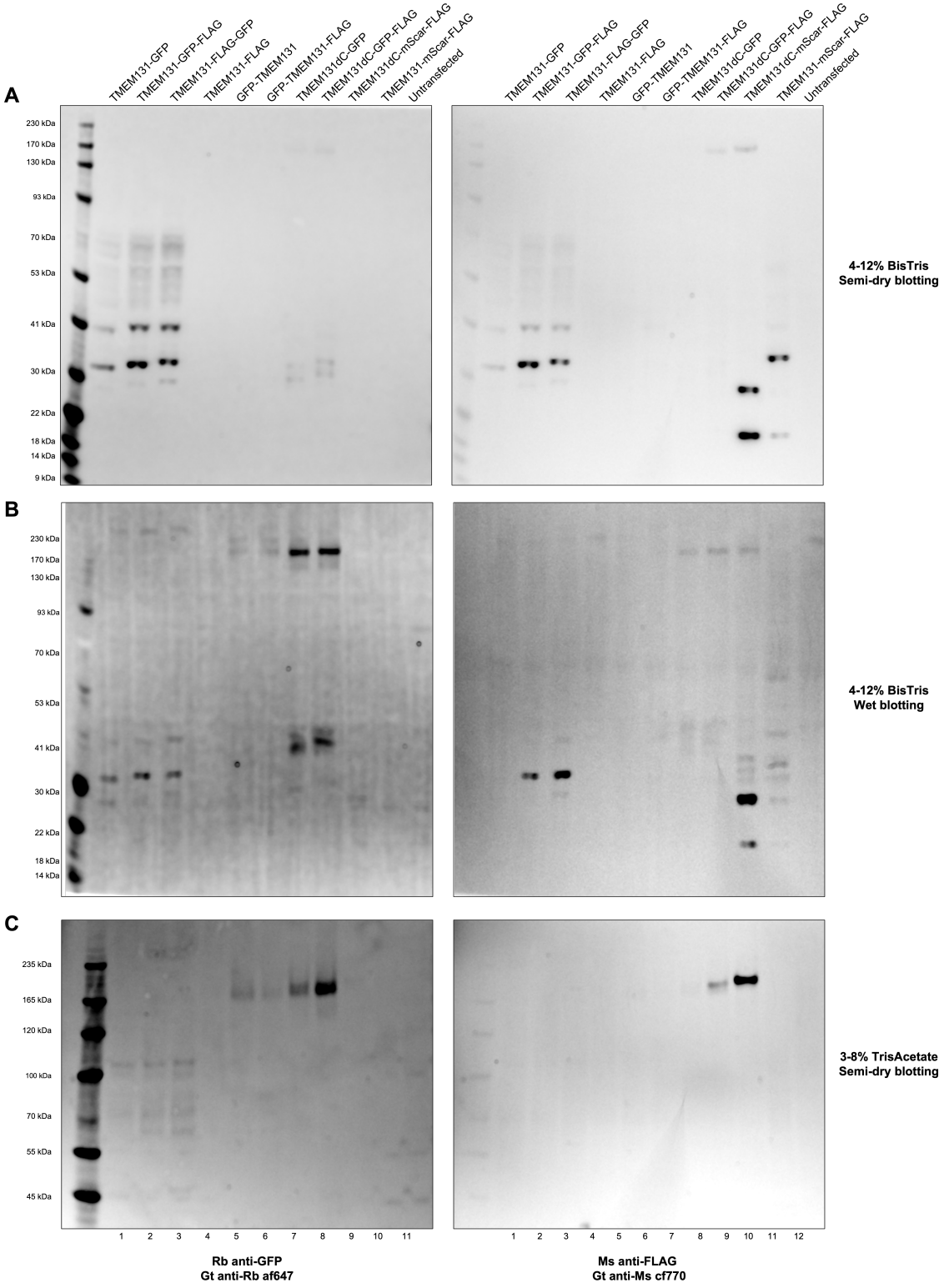

Figure 4.33: TMEM131 constructs with FLAG and/or GFP tags analysed via western blotting under three different conditions.
HEK cells were transient transfected with the constructs indicated at the top of the figure. These cells were lysed and the lysate ran on three different polyacrylamide gels:
A top: a 4-12% BisTris gel that was blotted with the iBlot2 semi-dry system for 7mins total.
B middle: a 4-12% BisTris gel that was blotted using a wet-transfer system as described in section 4.3.4.
C bottom: a 3-8% TrisAcetate gel that was blotted with the iBlot2 semi-dry system for 7mins total.
All three blots were blocked then labelled with rabbit anti-GFP antibodies and the M2 mouse anti-FLAG antibody then goat anti-rabbit af647 and goat anti-mouse cf770 secondary antibodies. Images have been cropped and aligned to aid comparison.

lysed and analysed with western blotting as previously (figure 4.34).

The first four lanes of this blot compared CD86 with an N-terminal GFP tag with four different tails: no cytoplasmic sequence (lane 1), mScar-FLAG (lane 2), the TMEM131 cytoplasmic tail with mScar-FLAG on the C-terminus (lane 3) and the TMEM131 cytoplasmic tail with just the FLAG tag on the C-terminus. High molecular weight bands, likely the intact sequences, are visible in the anti-GFP channel for all four constructs. For GFP-CD86 these were faint at ~70kDa and 90kDa, both larger than the theoretical molecular weight of 57.6kDa. In the lane for GFP-CD86-mScar-FLAG the high molecular weight bands were very strong and there is a band at ~170kDa consistent with two copies of the protein that have run together despite the reducing conditions. If the anti-GFP channel is over-contrasted then three bands are visible for GFP-CD86-TMEM131tail-mScar-FLAG and GFP-CD86-TMEM131tail-FLAG.

There were a couple of lower molecular weight bands in the anti-GFP channel in lanes with the mScar-FLAG domain for instance at ~35kDa in lanes 3 and 5. I suspect these are bleed-through from the bright bands in the FLAG channel. Lanes 10-12 compared CD86 with three different tails: the TMEM131 cytoplasmic tail with GFP on the C-terminus (lane 10), the TMEM131 cytoplasmic tail with FLAG-GFP on the C-terminus (lane 11) and the TMEM131 cytoplasmic tail with just the FLAG-tag on the C-terminus (lane 12). In the anti-GFP channel, bands consistent with partial degradation of the protein are only observed in the first two of these lanes.

In the anti-FLAG channel, the high molecular weight bands are visible in lanes 2-4 and lanes 11 and 12. In lanes 3 and 4 only the two larger molecular weight bands of the three visible in the anti-GFP channel are detected. Increasing the contrast of this channel reveals bands consistent with partial degradation of the tail. For GFP-CD86-131tail-mScar-FLAG (lane 3) the bands are similar to TMEM131-mScar-FLAG (lane 5) and CD86-131tail-FLAG-GFP (lane 11). Most surprising however was the presence of $FLAG^{+ve}$ bands in lane containing GFP-CD86-TMEM131tail-FLAG (lane 4) and CD86-TMEM131tail-FLAG (lane 12). These demonstrate that a C-terminal GFP or mScar domain is not a requirement for partial degradation, at least when attached to CD86 which has higher expression. We can conclude that there are important features of the TMEM131 tail that cause partial degradation and it is not just an artefact of the fluorescent protein.

Whilst partial degradation by the proteasome was a convincing explanation for the formation of the fragments of the TMEM131 tail I was still keen to understand the process better. In particular I was curious as to the rate this process took place. Barring the initiation of translation within the construct or premature truncation, I assume that the constructs must be synthesised as their full length complete with any tags on the N and C termini. All the blots performed so far would represent the steady state of the system, a mixture of new and old protein. I blocked *de novo* protein synthesis with cycloheximide and used western blotting to analyse fragments of GFP-CD86-131tail-FLAG in cell lysate (figure 4.35). 1h of treatment with CHX was sufficient to greatly reduce the amount of the full-length construct implying cleavage of the TMEM131 tail occurs relatively quickly compared to protein synthesis.

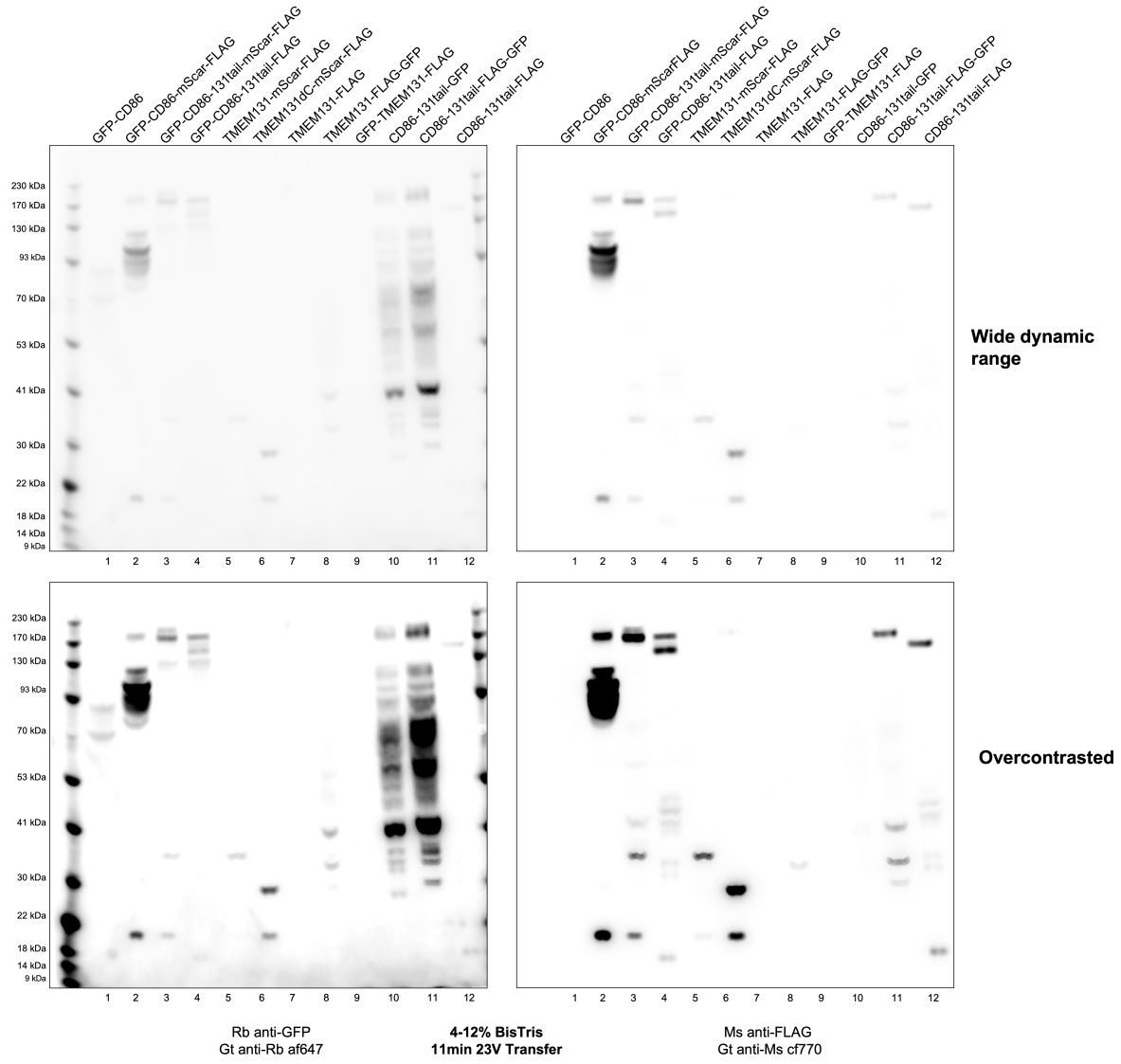

Figure 4.34: Partial degradation is observed if the TMEM131 tail is fused to CD86 without the C-terminal GFP tag.
HEK cells were transiently transfected with the indicated constructs and lysate run on a 4-12% BisTris gel. The gel was blotted for 11mins at 23V using the iBlot2 system then the blot labelled with the indicated antibodies. Both channels are presented with wide dynamic range (top) and with overcontrasting (bottom).

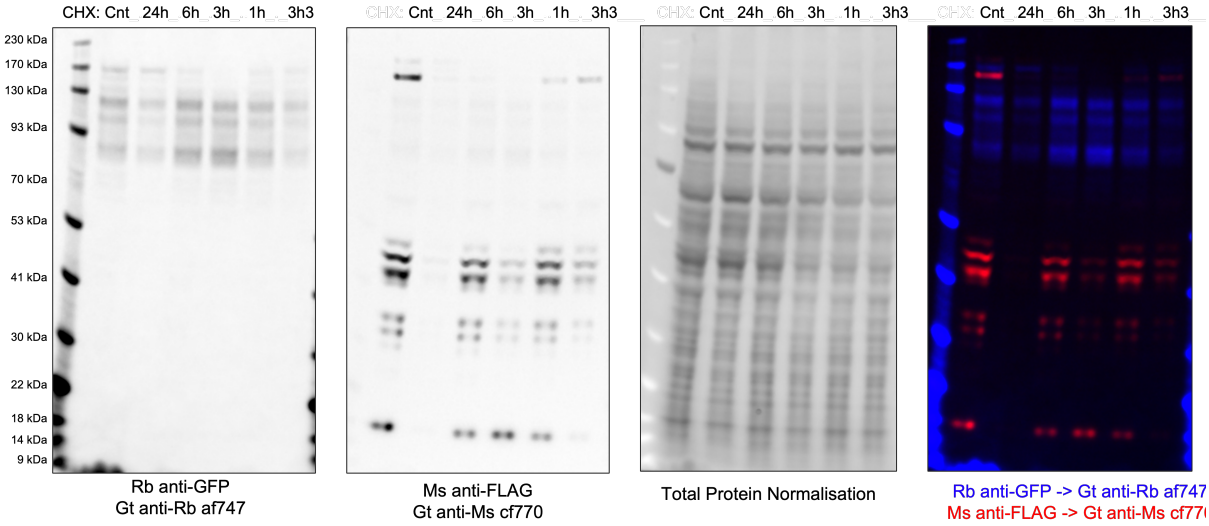

Figure 4.35: A band corresponding to the intact GFP-CD86-tail-FLAG rapidly decreases when *de novo* synthesis is inhibited.
Cells expressing GFP-CD86-131tail-FLAG were treated with 50µg/ml cyloheximide for the times indicated above the lanes then lysed at the same time. The 3h3 sample was treated for 3 hours then washed and incubated for a further 3 hours with fresh media. Samples were run on a 4-12% BisTris gel and the blot labelled with anti-GFP and anti-FLAG antibodies. Total protein normalisation showed similar amounts were loaded in each lane (1 : 0.60 : 0.94 : 0.94 : 1.13 : 0.96). The composite image shows the anti-GFP (blue) and anti-FLAG (red) channels.

4.6 Results: TMEM131 interaction partners

4.6.1 Trialling proximity labelling assays to identify TMEM131 interaction partners

The cleavage of the TMEM131 tail is an interesting feature and the tail clearly has a profound impact on the localisation and degradation of the molecule. However the "business end" of the molecule, as far any chaperone function, is likely to be the N-terminal domains located in the ER lumen. TMEM131 has sequence homology to an evolutionary-distant chaperone protein and there is published evidence for a role in the exit of proteins from the ER, particularly of collagen. The James lab's own data supports an interaction with the pre-TCR. I was curious therefore what other proteins TMEM131 might be interacting with.

Having investigated proximity labelling assays with the pre-TCR, I decided to apply these techniques to TMEM131. Before I attempted the GFP-pulldown approaches with the 3C protease site discussed in section 4.5.1, I cloned APEX2 to the C-terminus of the TMEM131 tail in place of the GFP domain. The two domains happen to be very similar in molecular weight. I made the same modification to TMEM131dC. Cells expressing these constructs were incubated with biotin phenol then labelling initiated with hydrogen peroxide. The reactions were quenched and the cells lysed. The labelling and washing procedure was performed at the same time with untransfected HEK cells and cells expressing APEX2-tagged pTCR and mTCR.

Cell lysate was analysed by western blotting using fluorescently-labelled Streptavidin and anti-HA antibodies (figure 4.36). Lanes 1 and 2 contain lysate from cells expressing TMEM131-APEX and TMEM131dC-APEX. These show a series of biotin-labelled bands in addition to bands matching the endogenously biotinylated proteins (Pyruvate Carboxylase, MCCA and PCCA) that also appear in the lane of the untransfected control. Amongst the smear of biotinylated proteins a few bands appear to be enriched. In the lower molecular weight range these could be the tail of TMEM131 with the APEX domain attached. There is a prominent band at around 93kDa in the TMEM131dC lane but this also appears to be present in the lane for the pTCR or mTCR constructs where the APEX domain is in the ER lumen rather the cytoplasm. The HA^{+ve} band in the TMEM131dC lane corresponding to the full-length construct is not particularly strongly labelled by biotinylation unlike the pTCR or mTCR. This is likely due to the lack of Tyr residues in the cytoplasmic sequence of this construct.

Whilst I considered it as potential experiment to identify the protein responsible for TMEM131 tail cleavage, this did not seem to be a viable strategy for investigating TMEM131 interaction partners. Due to cleavage of the cytoplasmic tail, the C-terminus is largely delocalised in the cytoplasm and could therefore contact a wide range of proteins and organelles through non-specific interactions. Even if it were not cleaved the length of the TMEM131 tail would give it a large sphere of influence through which other proteins would diffuse, diluting specific interaction partners. A better assay would therefore require modifications to the N-terminal side of the protein.

The AlphaFold model of TMEM131 (figure 4.4) suggested that there was significant flexibility at the N-terminus of TMEM131 and I thought an APEX2 or BioID domain could be fused to this site without disrupting the folding of the rest of the protein. This position would be optimal for detecting interactions between the chaperone domains TMEM131 and its potential target proteins in the ER lumen. I cloned both domains onto the N-terminus of full-length TMEM131-GFP and the tailless TMEM131ntd-GFP construct immediately after the signal peptide and HA-tag. I imaged the TMEM131ntd constructs

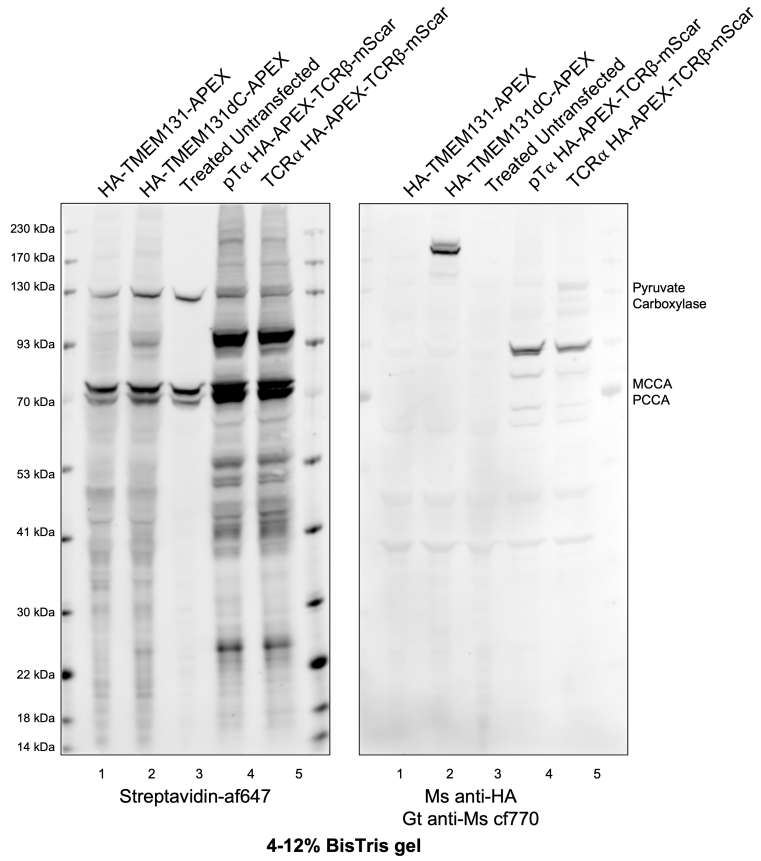

Figure 4.36: APEX2 labelling on the cytoplasmic tail of TMEM131.
APEX2-tagged constructs of TMEM131, TMEM131dC, the pre-TCR and mTCR were expressed in HEK cells. These were treated with APEX labelling assay described in 3.3.2. The lysate was run on a 4-12% BisTris gel then blotted using the iBlot2. The blot was blocked then labelled with af647-conjugated streptavidin and Ms anti-HA antibodies using 1% BSA in the buffers instead of milk powder.

with an ER marker 4.37. The localisation of the APEX and GFP-tagged construct was similar to the construct without the APEX tag with protein mostly confined to the ER with a few mobile vesicles. The BioID-TMEM131ntd-GFP construct appeared a little different with long, mobile blobs in high expressing cells. These seemed to correlate with the ER marker and did not resemble OSER artefacts (supplementary figure S13).

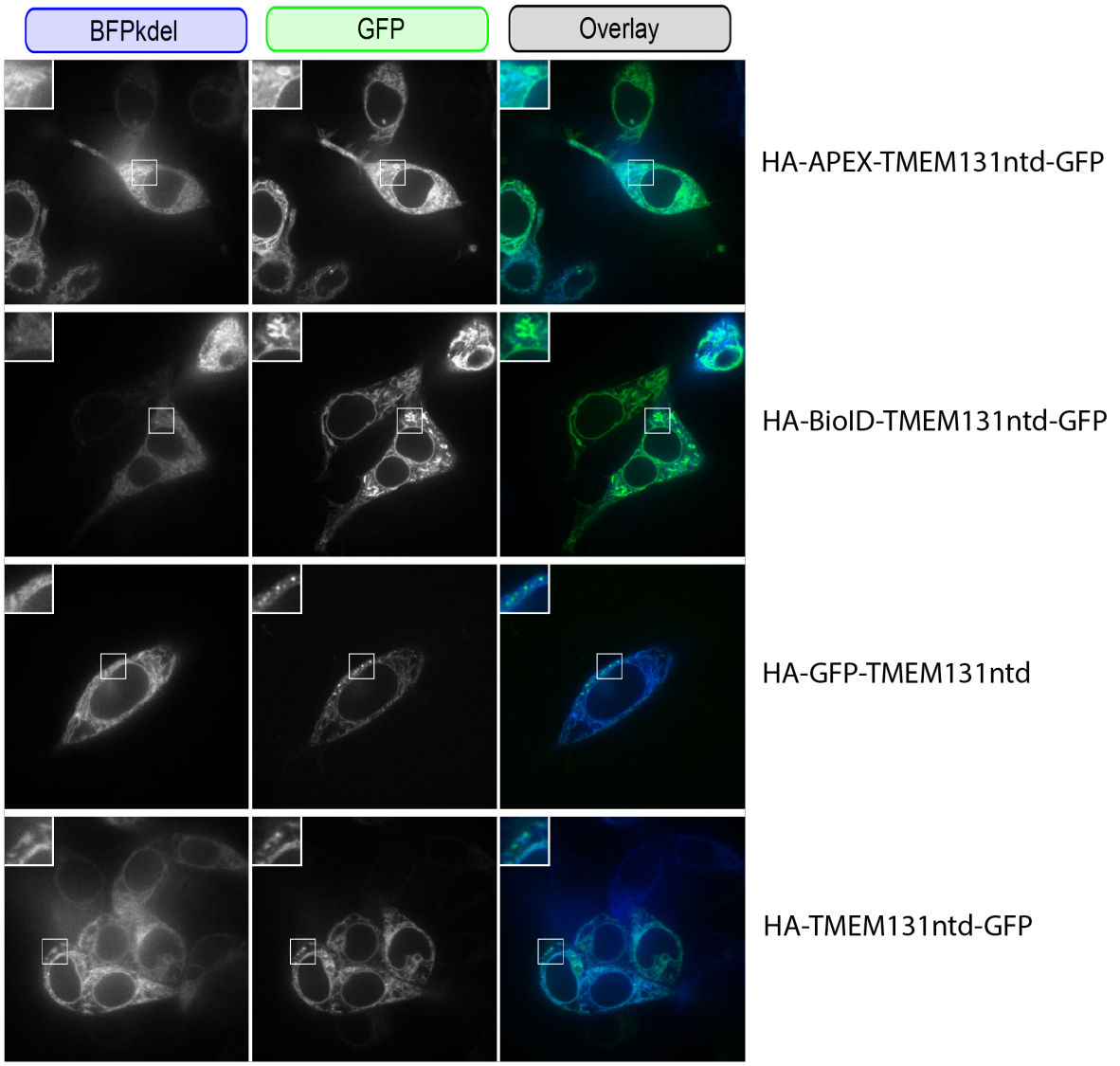

Figure 4.37: Fusion of APEX2 or BioID domain to the N-terminus of TMEM131 does not disrupt ER localisation.
The indicated constructs were expressed in HEK cells with a BFPkdel ER marker and examined by microscopy under 100X magnification. Inserts show a 2X zoom of the selected area.

All new constructs were expressed in HEK using six well plates. Cells expressing the APEX-TMEM131 constructs were labelled with biotin phenol and hydrogen peroxide as before. TMEM131-APEX and GFP-TMEM131ntd-APEX were included to allow comparisons to labelling from the cytoplasmic side of the protein. The mature TCR with APEX on the N-terminus of TCRβ was also included to demonstrate the labelling around a different protein present in the ER lumen. A well expressing GFP-tagged TMEM131ntd without the APEX domain was treated identically as a negative control. The BioID constructs were expressed and the wells treated with 50µM biotin for 16h overnight. The mature TCR

with the BioID domain on the N-terminus of TCRβ and another well with TMEM131ntd-GFP were used as positive and negative controls.

Cell lysates were run on a 4-12% BisTris gel, blotted then labelled with fluorescently-labelled Streptavidin and anti-GFP antibodies (figure 4.38). As expected there was a large range of biotinylated proteins detected in the lysates with APEX/BioID-containing constructs and only the three bands of endogenous biotinylated proteins in the negative controls. The lanes with APEX constructs (lanes 1-6) were more intensely labelled by streptavidin than the lanes with BioID constructs (lanes 7-10) implying more efficient labelling. There was a striking difference between TMEM131 with APEX on the N-terminus (lane 1) compared with on the C-terminus (lane 2) with the latter rather dim in this experiment. The TMEM131-APEX construct might not have been well expressed as its expression could not be monitored with flow cytometry. Lanes 3 and 4 show the difference between TMEM131ntd with APEX on the N-terminus and GFP on the C-terminus and with the tags swapped. The pattern of biotinylation is noticeably different, consistent with the APEX domain being present in two different environments containing different proteins. Though the two termini of the protein might be spatially close together, the biotin radicals generated in APEX-labelling are short-lived so unlikely to cross the ER membrane.

I also compared lanes 1 and 3, samples from APEX-labelling from the TMEM131 N-terminus, with lane 6, a sample from labelling from the N-terminus of the mTCR. Whilst not identical, the major bands seem to be shared between all three lanes which would imply that similar proteins have been labelled in each case. The three constructs should all be present in the ER lumen and interact with general ER proteins but we would still expect TMEM131 and the mTCR to have different interaction partners.

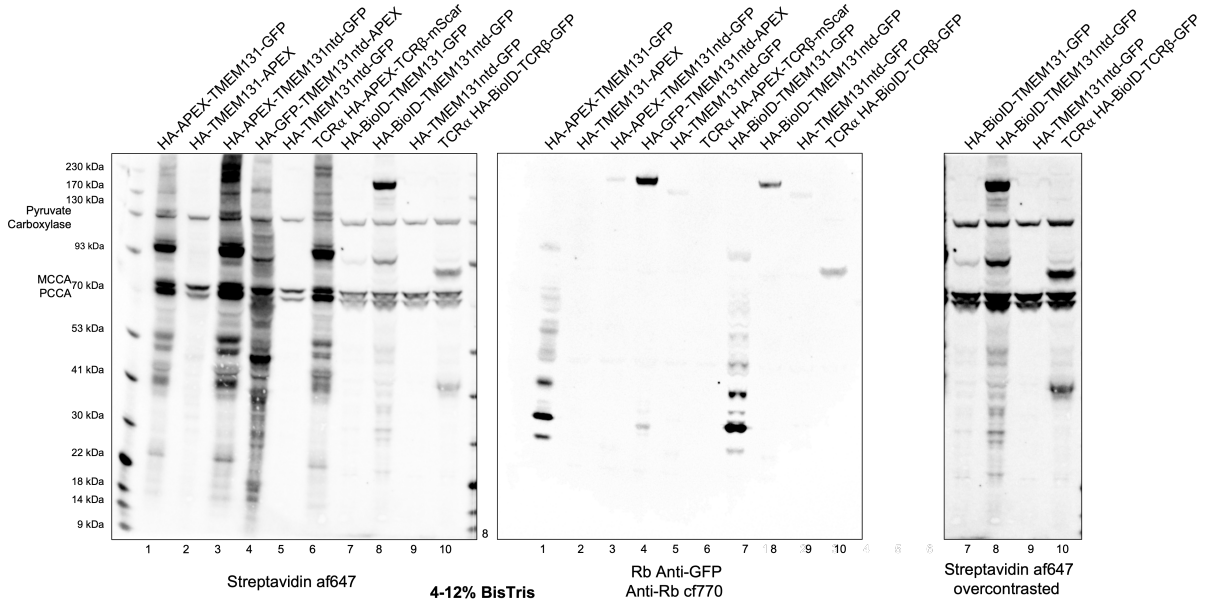

Figure 4.38: APEX2 and BioID assays with domains on the N-terminus of TMEM131.
The APEX and BioID constructs indicated at the top were expressed in HEK cells and labelled using the APEX or BioID protocol. Cells were lysed and run on a 4-12% BisTris gel that was blotted then blocked with 1% BSA in 1xTBS. The gel was labelled with Streptavidin af647 and anti-GFP antibodies. The figure to the right shows the lanes expressing BioID with additional contrasting. This shows bands not visible when presented next to the APEX constructs.

Whilst the amount of biotin labelling was less intense in the BioID samples than the APEX samples, the banding pattern appears to be more distinct between the BioID-TMEM131 samples and the mTCR control. Both TMEM131 samples have a band at ~93kDa which appear to be reduced in the mTCR lane. The TMEM131ntd lane has multiple bands across all mass ranges. In a BioID experiment, labelling occurs primarily on lysine residues which are more common and more accessible than tyrosine residues

that are labelled with APEX. This means a wider range of proteins might be labelled and this could outweigh the lower labelling efficiency.

4.6.2 TMEM131 interacts with proteins involved in protein folding and ER-associated protein degradation

Due to the distinct banding pattern and the easier experimental method, the BioID assay was scaled up to a larger number of input cells. Cells were treated with biotin then lysed. A small amount of lysate was analysed by western blotting to confirm the assay worked (shown in figure S21) whilst the rest was incubated overnight with streptavidin-agarose beads. These beads were extensively washed then treated with an on-bead trypsin digest method. Hits were searched in a database of human proteins using MaxQuant then analysed with Scaffold software. The full list of proteins found in this experiment is found in supplementary section A.4.5.

Using the recommended values of a 95% protein confidence threshold and 95% peptide confidence threshold, there were unfortunately no peptides identified in the mTCR or TMEM131ntd-GFP samples. If the peptide threshold was lowered to 50% then endogenous biotinylated proteins were detected along with other common contaminants. Whilst the mTCR sample definitely produced weaker bands than BioID-TMEM131ntd in the western blot, there was apparent labelling and endogenous biotinylated proteins should have been identified in both control samples. This implies there was an error later in the procedure perhaps in the digest or C18 clean-up stages. The data was analysed without these two controls.

There were many more proteins identified in the TMEM131ntd sample than in the full-length TMEM131 sample, consistent with generally better expression of the former. The BioID-TMEM131 construct was identified in the ntd sample but not the full-length sample. As is typical, some of the proteins identified were uninteresting. These included the five carboxylase enzymes that are naturally biotinylated in human cells (ACACA, ACACB, PCCA, MCCA, PC) as well as common contaminants such as histones, tubulin and keratin proteins. There were also proteins that are seem to be preferentially biotinylated in BioID experiments including Parp1 and EEF1A1.[289] To assess how likely each protein was to be a contaminant, I searched the Contaminant Repository for Affinity Purifications database[290] in which contaminants are more frequent.

Manual inspection of the list of hits suggested a lot of the proteins were ER chaperones or otherwise involved in protein folding. I used GO enrichment analysis[354] from the Gene Ontology Resource which found a 56-fold enrichment of genes tagged with "protein folding in the endoplasmic reticulum" compared to a reference human genome. Similar tags such as "protein refolding" and "chaperone cofactor-dependent protein refolding" were also enriched. There was also an 8.7-fold enrichment for genes tagged with "ubiquitin-dependent ERAD pathway". Proteins tagged with two quite broad GO molecular function terms "Protein Folding" and "ERAD pathway" are indicated in section A.4.5.

Endoplasmic reticulum-associated protein degradation (ERAD) is a pathway which removes misfolded proteins from the ER which are then degraded in the cytoplasm by the proteasome.[355] Misfolded proteins can be recognised by three main mechanisms: Hydrophobic regions of proteins which should normally be hidden can be recognised by members of the Heat Shock protein (HSP) 70 family such as BiP. Several members of this family were identified in this experiment though these proteins are quite abundant and frequently often observed in proteomic experiments. Unpaired cysteine residues in misfolded proteins can be recognised by proteins in the Protein disulphide isomerase (PDI) family. Two

of these proteins appear in this dataset. Misfolded protein can also be recognised by specific N-linked glycosylation chains. These chains are bound by the lectins calnexin and calreticulin which retain the protein in the ER, giving time for the protein to fold. The target can sometime be re-glycosylated by the protein UDP-glucose:glycoprotein glucosyltransferase (UGGT). If refolding does not occur then the glycosylation is trimmed and bound by the lectin EDEM with the help of OS-9. Whilst calreticulin was not identified in the dataset, peptides from calnexin, UGGT and OS-9 were found. Once misfolded proteins are identified they can be exported from the ER in a polyubiquitin-dependent process which is powered by the ATPase VCP.

TANGO1/MIA3, discussed in section 4.1.3, was also identified in this BioID dataset. This supports the prior evidence that TMEM131 and TANGO1 have similar roles in the exit of ER proteins.[319]

Comparisons to previous GFP-pulldown experiment

This BioID dataset had surprising overlap with the proteins identified by pulling down on the GFP-tagged cytoplasmic tail of TMEM131, an experiment discussed in section 4.5.1. I had planned this experiment to just identify the TMEM131 cleavage sites so ran the precipitated proteins on a gel and cut out gel pieces depending on the bands of tail fragments. Five gel pieces were cut out of lanes containing TMEM131dC-GFP and six in lanes for TMEM131-GFP. Any TMEM131 interaction proteins could potentially appear in this gel if they remained bound to GFP-tagged TMEM131 when the beads were washed. These washes were not particularly harsh, just five 1ml washes with cold GTA wash buffer (effectively 1xTBS with added EDTA and NP-40) but I anticipate this would have removed most non-specific interactions. To be identified via mass spectrometry, the masses of these interaction partners would have to coincide with the quite narrow gel pieces I cut out. Remarkably many of the protein detected seem to be genuine hits.

The full mass spectrometry results from this experiment are shown in section A.4.4. The approximate masses for each gel piece were estimated with reference to the protein ladder. The protein with the most unique peptides identified is TMEM131-GFP as expected and 9 of the next 10 hits are keratin molecules. These are present pretty evenly in each of the gel pieces regardless of the mass of the keratin protein. This is consistent with them being contaminants after the gel was cut. The other highly-identified protein is the heat shock protein HSPA1B. This appears most frequently in the pieces dC4 and FL4 which were taken from positions on the gel consistent with the molecular weight of HSPA1B. This would imply that HSPA1B is a genuine interaction partner or a contaminant prior to when the gel was run.

Many of the proteins identified in this experiment also appear to be involved in protein folding and the ERAD pathway. These include HSP70 and DNAJ chaperones, UGGT that appeared in the BioID dataset and the lectin EDEM2. This experiment also found proteins that are involved in the ERAD pathway on the cytoplasmic side including the cytoplasmic PNGase NGLY1 and the ATPase VCP that powers retrograde transport of targets to the cytoplasm. Peptides from at least nine ubiquitin-ligases are present including CHIP and SYVN1 which are both involved in ERAD.

Seven components of the regulatory subunit of the proteasome were found in various gel pieces supporting the evidence that the proteasome is involved in the cleavage of the TMEM131 tail. These include PSMC6, one of the ATPase motor subunits as well as PSMD2 and PSMD1 non-ATPase subunits. Twelve peptides from subunit PSMD14, which deubiquitinates proteins during translocation, were found in the band for FL-6. In the proteasome complex this subunit is enclosed by PSMD6 and PSMD13 which also appear.

Two peptides of ADRM1, the Ubiquitin-receptor component of the complex, appeared in FL-5[†].

TMEM39A and TMEM39B proteins also appear in the GFP pulldown. TMEM131 was proposed to cooperate with TMEM39 in the secretion of collagen proteins.[311] Both the GFP pulldown and BioID dataset also contains peptides from Serpin H1/HSP47, a collagen-specific molecular chaperone.[310,356]

Comparisons with Perturb-seq data

Across both datasets a large number of proteins identified are involved in protein folding and ER-associated degradation. It could be argued that this does not necessarily establish that TMEM131 has a role in these processes. If any large, ER-targeted protein was overexpressed in HEK cells then it would presumably interact with chaperone proteins and be subjected to degradation via the ER-pathway.

Ideally, I would want to compare these results to other GFP and BioID-tagged proteins with similar size and localisation. The pre-TCR and TCR initially seem like reasonable options. An interaction between the pre-TCR and TMEM131 was first established using a BioID assay with the BioID2-domain on the N-terminus of TCRβ. Tagged versions of the pre-TCR and mTCR were included as samples along with versions without the BioID domain as negative controls. Whilst this experiment was performed by someone else with different equipment, procedure and settings, the data might still be a valid comparison. This data featured TMEM131 (two peptides in the BioID-pTCR sample only) as well as various HSP70 and DNAJ chaperones, Calnexin, Endoplasmin, UGGT1, OS-9, PDIA6 and MIA3. A pulldown of proteins via a GFP-tag on the C-terminus of TCRβ as a component of the pre-TCR and mTCR found peptides from PDI6, endoplasmin, CHIP and several regulatory proteasome subunits.

However, I think it would be a mistake to conclude that TMEM131 is uninvolved in protein folding or ERAD just because similar hits were generated with labelling from the mTCR and pre-TCR complexes. Whilst the role of the two receptors is signal at the cell surface, both are known to interact with chaperone proteins for complex assembly. When complex assembly fails TCRα and TCRβ chains are degraded via the ERAD pathway. In fact, the TCR has been used as a model protein for ER assembly and degradation as ER export is dependent on dimerisation of the transmembrane domains. The mTCR and pre-TCR could interact with these pathways are cargos whilst TMEM131 could be involved as an active participant, either helping proteins fold or marking them for degradation.

How else could a functional role be established if not via interaction partners? One method would be to knockout TMEM131 and measure a phenotype such as failed ER export or ER stress. I made several attempts to make a HEK *tmem131* knockout cell line, as well as other edits to the endogenous locus, but was unsuccessful. These attempts are detailed in supplementary section A.1.

A public dataset that proved extremely interesting was database of genome-wide Perturb-Seq in K562 cells by Replogle *et al*.[357] These authors used CRISPR interference to disrupt every expressed protein then measured expression changes in the target and other genes using single-cell RNA sequencing. When targeted by two guides with CRISPRi TMEM131 expression was reduced by 85% and this resulted in about 50 differentially expressed genes. The data was published for the 30 top upregulated and downregulated genes along with the 10 most similar genes when targeted. This is represented in figure 4.39. A change in expression of a gene after the interference of another, does not necessarily infer a direct interaction between the two gene products, for instance genes could be upregulated if they both contribute to the same molecular pathway.

[†]In proteasome structural studies the systematic nomenclature uses the protein names from yeast. In the order mentioned in the text here these are Rpt4, Rpn1, Rpn2, Rpn11, Rpn7, Rpn9 and Rpn13

The biggest change in gene expression with *tmem131* interference is in the gene for SLC3A2 (with an extremely high z-score of 4.54). This protein is the larger chain of six heterodimeric amino acid transporters but seems to have other interaction partners as well. It is reported to be involved in processes as diverse as amino transport,[358] integrin signalling[359] and the unfolded protein response.[360] At least in HeLa cells, which are adapted for fast nutrient uptake, SLC3A2 is extremely abundant at the plasma membrane with an estimated 15 million copies. This comprises over 8% of the total protein mass for this fraction.[361] Peptides from this protein appear in both my BioID dataset and the GFP-pulldown dataset in gel pieces dC-4, dC-5 and FL-3.

Of the top 30 proteins upregulated by *tmem131* interference, four are identified by BioID (calnexin/CANX, PRDX4, PDIA6, PDIA3), one appears in the GFP pulldown (DNAJB11) and five in both (SLC3A2, TTC17, PRKCSH, Endoplasmin/HSP90B1, RPN1). Though not tagged as being involved in protein folding TTC17 contains tetratricopeptide repeats that are frequent domains in chaperones. Upregulated genes featuring in neither dataset include calreticulin, involved protein folding and ALG5, involved in protein glycosylation. Of the top 30 genes that appear downregulated by *tmem131* interference, two are found in the GFP-pulldown dataset (COPE and WASF1). COPE is involved in the retrograde transport of proteins from the Golgi to the ER, particularly those with diLeucine motifs, which are present in the TMEM131 cytoplasmic tail. Of the genes that create the most similar effect to *tmem131* when targeted, one (PPA1) appears in the GFP-pulldown dataset. The other genes include glycosyltransferases ALG2, ALG13 and ALG14 as well as SEC61B and SEC63 which are involved in the translocation of proteins across the plasma membrane.

Like the BioID and GFP-pulldown hits, many of the genes appearing in this Perturb-Seq dataset have links to protein folding or protein glycosylation. This supports the assertion that the proteins identified are genuine interaction partners and that TMEM131 is a functional protein chaperone, rather than just a target of protein folding and ERAD. I would consider interactions between these proteins and TMEM131 as potential avenues for further study.

4.7 Discussion

A number of predictions could be made about TMEM131 based solely on the amino acid sequence. The protein was predicted to have an unusual signal peptide, N-terminal domains with homology to bacterial chaperone PapD, one or two transmembrane helices and a long unstructured C-terminal tail. I have investigated all these features to see how they relate to the localisation, degradation and interaction partners of TMEM131.

Published evidence supports a role for TMEM131 in the secretion of protein complexes from the ER. The data I have acquired is consistent with this molecular function and suggests additional clients.

4.7.1 ER localisation of TMEM131

TMEM131 synthesis is directed to the ER though an unusually long, two-domain signal peptide. Images of TMEM131ntd with modified signal peptides suggests the first half of this signal peptide is dispensable for targeting. Why TMEM131 and some other ER-targeted proteins use signal peptides of this architecture is unclear. As least in my experiments, I did not notice any dual-targeting of TMEM131. The signal peptide is not especially conserved relative to the rest of the protein so its unusual architecture may not

Figure 4.39: Peturb-Seq data for TMEM131 showing the 30 top upregulated and downregulated genes along with the ten guides that gave the most similar effect.
TMEM131 was targeted by CRISPR interference as part of a genome-wide screen in K562 cells by Replogle *et al.*[357] The effect of TMEM131 interference on mRNA expression was measure by single-cell RNA sequencing and shown in column 1. The top 30 genes are the most upregulated whilst the bottom 30 genes are the most downregulated. The other columns show the guides that led to the most similar effect. Changes in expression are express by a z-score (the number of standard deviations away from mean) that is usually between 2 and -2.

be that significant to its function. Replacing the signal peptide with the gaussia luciferase signal peptide did not seem to change the behaviour of the molecule and improved expression.

Though TMEM131 was annotated in Uniprot with two transmembrane helices, a topology assay with split GFP has demonstrated that the N-terminal domains of the protein are likely present in the ER lumen and the C-terminal tail is in the cytoplasm. This is consistent with only one of these regions being a transmembrane helix. The surface expression of a truncated construct with both helices (NTD-TM1-TM2) but not a construct with only the first putative helix (NTD-TM1) would imply that it is the second of these two helixes that is transmembrane. The other could potentially lie parallel to the membrane or partially within the membrane. Without the lipids, these helices cannot be correctly modelled by the AlphaFold program. The sequence for the first of the two helices contains an Asn-Ala-Ser motif so may well be glycosylated if it is present in the ER lumen.

The cytoplasmic tail of TMEM131 has two interesting properties. Firstly the tail appears to be cleaved at multiple sites producing a series of fragments. Secondly truncation of the cytoplasmic tail, even by a small amount, greatly increases the surface expression of TMEM131. I suggest that the final 160 amino acids of the molecule are the most critical for retention in the secretory pathway. Fusing these residues to the truncated TMEM131dC construct was found to reduce surface expression though it did not eliminate it entirely. Fusing almost the entire cytoplasmic tail to CD86 was also insufficient to prevent surface expression of this protein.

Zang *et al.*, in their paper relating TMEM131 to collagen secretion, investigated the cytoplasmic tail of the protein through a Yeast-two hybrid assay. They found the cytoplasmic tail of TMEM131 interacted with TRAPPC8, part of a complex involved in ER to Golgi transport of COPII vesicles.[310] Removal of the residues 1741 to 1883 in TMEM131 abolished this interaction. The authors mutated two conserved tryptophan residues in the tail which also prevented COPII binding. I am unsure why these particular residues were chosen.

An interaction between the very end of the TMEM131 tail and the TRAPP complex that retains TMEM131 in the ER would explain the surface expression of the truncated constructs I observed. However, I did not find any peptides from the TRAPP complex in the BioID assay or GFP pulldown experiments I performed using strict protein and peptide thresholds. If these settings are lowered, making identification much less reliable, then potentially two peptides of TRAPPC8 were found in the GFP pulldown in gel piece dC-3. This band was removed from the gel at ~100kDa which is consistent with the molecular weight of TRAPPC8 (104kDa).

I also saw an increase in surface expression of TMEM131 upon removing the Lys-rich region (1375-1475) in the middle of the cytoplasmic tail. I am unsure if the sequence of this region is particularly significant. Perhaps removing any 100 amino acid sequence of the tail would lead to an increase in surface expression? This would imply that a critical feature of the TMEM131 tail is simply its length. A long tail would allow it to reach from the ER membrane to contact proteins localised to other organelles such as the Golgi. This would be consistent with the low sequence conservation throughout the most of the tail compared to the N-terminal domains. There are some patches of high conservation in the tail however, implying functionality beyond being long and disordered.

I considered whether cleavage of the protein tail was responsible for releasing the protein from ER retainment. This could potentially allow TMEM131 to accompany its interaction partners through the secretory pathway like a cargo receptor. This would be similar behaviour to TMEM131's paralog: TMEM131L which was detected at the cell surface of DN3 cells with components of the Wnt signalling pathway. This did not appear to be the case with TMEM131 however as the protein could not be detected

at the surface even if secretion was synchronised in a RUSH assay. I believe, at least for the full-length TMEM131 construct, that the fragments of the tail are produced when the protein is degraded via the ERAD pathway. This means that at no point would there be N-terminal and transmembrane domains of the molecule in the ER without the entire tail attached.

When TMEM131 missing its cytoplasmic tail was co-expressed with the pre-TCR, imaging suggested a degree of co-localisation. Establishing co-localisation rigorously is quite challenging as it is depended on expression and resolution. For these experiments, I believe my data could be significantly improved. Ideally, I would have obtained an unbiased set of images with reasonable expression in each channel. Qualitative comparison however, suggested similar co-localisation with CD86 and ICAM-1. I attribute this co-localisation to the partial mixing of cargo in the same vesicles throughout the secretory pathway, rather than a specific interaction between a chaperone and its client.

The split Venus assay suggested that the N-terminal domains of the tailless TMEM131ntd construct could form a dimer or higher order oligomers. This signal was this only detectable via flow cytometry at high expression levels and only faintly visible in the ER under microscopy. I presume the full-length TMEM131 construct also forms oligomers though this was not tested. This should be investigated further through different assays such as the NanoBiT assay or FRET. It would be interesting to know if the oligomeric state of TMEM131 is relevant to its function.

4.7.2 Cleavage and degradation of the TMEM131 cytoplasmic tail

When TMEM131 is expressed in HEK with a C-terminal GFP tag, a series of GFP-tagged fragments is produced. When analysed by western blotting these fragments appeared distinct so were not thought to be due to non-specific aggregation. These bands also appear if the GFP-tagged tail is fused to CD86 implying the cleavage is intrinsic to the tail sequence. Truncation of the cytoplasmic tail suggests there are multiple cleavage sites and these are affected by both protein sequence and position from the GFP tag. If tail cleavage were due to a cytosolic protease, such as a caspase, these factors would be very unusual.

I propose that the GFP-tagged bands observed are caused by the partial degradation of the cytoplasmic tail by the proteasome. Proteasome degradation is initiated in disordered regions of proteins near to ubiquitination sites. From this initiation site, the protein is translocated into the 20S central core of the protease by motor units in the 19S regulatory particle. This process requires the unfolding of structured domains. During this process I believe the low complexity, "slippery" tail sequence, combined with the tightly-folded GFP domain, cause the C-terminal fragments to dissociate from the proteasome during translocation. These fragments could be bound, trimmed and released again, increasing the complexity of the fragment pattern. At least for TMEM131, the N-terminal part of the protein is degraded entirely.

This model is supported by the following evidence:

- The GFP-tagged tail fragments of TMEM131 appear to be ubiquitinated, a modification associated with proteasome degradation.
- The apparent size of the GFP-tagged fragments increases upon treatment with proteasome inhibitor MG132.
- Peptides from proteins in the 19S regulatory particle were found in a GFP-pulldown of the tail fragments. In addition, the pulldown identified several E2-ubiquitin ligases and protein in the

ERAD pathway. Peptides from ERAD proteins were also identified using a BioID assay with the BioID2-domain on the N-terminal domains of the protein.

I think this model explains the sequence and distance effects observed in the bands. The proteolytic catalytic sites of the proteasome have broad specificity but I would still expect certain sequences to be cleaved preferentially. I would also expect the position of sequences relative to the GFP domain to determine how fragments dissociate.

I do not think this explanation was obvious as it is confounded by the fate of the N-terminal domains of the protein. On western blots it appeared that as the cytoplasmic tail is truncated, the stability of the intact protein increased. With the shortest constructs TMEM131ntd and TMEM131dC, GFP^{+ve}/HA^{+ve} bands are visible with sizes compatible with the theoretic mass of the construct. As size increases however these bands seem to disappear. This is likely a blotting artefact. Bands consistent with the intact size of constructs with longer tails can be observed using lower percentage weight gels or by optimising blotting conditions for larger proteins. If the blotting step is removed entirely then the intact sequence of TMEM131 can be seen via Coomassie or SYPRO ruby stains after enrichment.

I designed proteomic experiments to find the exact cleavage sites in the TMEM131 tail. As the capture rate of tryptic peptides is not 100%, the most likely method to find exact sites would have been native mass spectrometry. I think the challenges associated with this experiment could be overcome with more optimisation of the protocol. However I am not certain that knowing the exact masses of fragments would have necessarily implied the involvement of the proteasome. As the fragments appear to be ubiquitinated, it may not have been trivial to match the exact mass of a fragment to its peptide sequence due to the mass gain from the modification. If the exact sequence of the protein was discovered then the N-terminal residue might not be informative as the catalytic sites of the proteasome have broad specificity.

In all western blots of TMEM131 and its truncations, the only proteins detected were either intact with both GFP and HA tags, or C-terminal tail fragments with just the GFP-tag. No HA-tagged fragments of the N-terminal product of cleavage were observed. I do not think this is an artefact of gel blotting as the mass of the N-terminal fragments would be small enough after the cytoplasmic tail has been removed. When the TMEM131 tail was fused to CD86 however, HA-tagged fragments were observed (for instance in figure 4.9). What difference could explain why the N-terminal products persist when CD86-131tail is cleaved whilst the N-terminal products when TMEM131 is cleaved are degraded completely?

One hypothesis would be that structural differences in the N-terminal and transmembrane domains make CD86 more stable than TMEM131 after the tail is cleaved. The extent of partial degradation likely depends upon many factors including the sequence around the initiation site and the energy required to unfold flanking domains. This effect might only be noticeable with sufficient expression.

However, I think the difference may also relate to how the two proteins are trafficked through the secretory pathway. TMEM131 does not seem to leave the ER so its degradation is likely mediated by ERAD. This means it is recognised as misfolded or aggregated in the ER lumen then removed from the ER into the cytoplasm by retro-translocation. Once in the cytoplasm the N-terminal section of the protein is degraded entirely by the proteasome. This process presumably requires chaperone proteins to maintain solubility of the hydrophobic helices and folded PapD domains. CD86 meanwhile is able to leave the ER and is easily detectable at the cell surface. Once at the surface, degradation of the N-terminal domains would require input from lysosomes. If proteasome degradation was initiated on the cytoplasmic tail with the rest of the protein anchored in the cell membrane then cleavage might create stable N-terminal products.

Where in the cell is degradation initiated for the TMEM131 truncations? The GFP-tagged fragments of TMEM131dC are partially EndoH-insensitive which would imply that these fragments contain sequence that has been glycosylated in the ER lumenal but also that a fraction of the protein has been cleaved after it has passed through the trans-Golgi network during which glycosylations are modified. Crosslinking with BS3 would also suggest that a fraction of TMEM131dC is present at the cell surface with the GFP tag intact. Presumably this means the N-terminal domains of this fraction are also degraded primarily by lysosomes.

Other questions about TMEM131 degradation are still unresolved. I suspected partial degradation was an artefact of the well-folded GFP domain as partial degradation of GFP-tagged proteins has been previously reported. However multiple bands are seen when CD86 or GFP-CD86 are fused to the TMEM131 tail with a C-terminal FLAG tag. This tag is a short sequence rather than a folded domain so cannot be difficult to unfold. This demonstrates that the tail is capable of partial degradation alone. Perhaps the GFP domain just amplified the effect, making it more visible with lower expression?

What features of the tail are important for partial degradation? As the effect is seen in some of the smallest truncations, sequence in the tail near the membrane must contribute. This sequence is quite poorly conserved relative to the rest of the protein but quite serine rich. This sequence flexibility might be sufficient for slippage to dissociate the protein during translocation. It might be that any suitably long, unstructured sequence would give the same effect or perhaps there are certain structural features, as in NFκB subunit p50 or EBNA-1 where partial proteasome degradation is also observed. An interesting experiment might be to replace the middle of the TMEM131 tail with a repetitive or randomised sequence of the same length to see if the effect still occurs. An unexplored aspect is the connection between low sequence complexity and liquid phase separation.[362]

I have unfortunately not demonstrated that partial degradation also occurs with endogenously-expressed protein. I attempted to tag the final exon of the endogenous locus in HEK cells with GFP and GFP11 tags using CRISPR knock-ins. These efforts are detailed in supplementary section A.1 but I was unable to find a successful line. I suspect the expression level of the protein is too low for such a strategy. A better method might have been to tag the cytoplasmic tail with something small such as a FLAG tag then screen cell lysate on a western blot with an anti-FLAG antibody. Again however, this would require the endogenous protein to be sufficiently expressed in HEK cells to be detectable.

If partial proteasome degradation of the molecule is an inherent feature of the protein then is it coincidental or could it have a functional role? Chaperone proteins, like TMEM131, have dual roles. Firstly to help client proteins to fold and secondly to direct clients for degradation if they are unable to do so.[363] Perhaps degradation of TMEM131 via its cytoplasmic tail is involved in this process.

4.7.3 TMEM131 clients and interaction partners

My BioID and GFP-pulldown experiments returned a list of proteins containing lots of chaperones and proteins involved in the ERAD pathway. This might be expected given overexpression of a large, multi-domain protein directed to the ER. Further controls might be needed to account for this. However, I believe some of these interactions represent genuine behaviour of the protein when expressed at the endogenous level. Identified proteins include HSP90B1/endoplasmin, PDIA6, PDIA3 and calnexin which also appeared to be upregulated when TMEM131 gene expression was perturbed. If TMEM131 cooperates with these chaperones in protein folding then what might its list of clients be like?

Zhang *et al.* concluded that TMEM131 was a molecular chaperone for collagen fibers.[310] This was based

on a knockout of TMEM131 in *C. elegans* and a knockdown in U2OS cells, both of which produced defects in collagen secretion. Their yeast two hybrid assays found collagen fibers when the PapD-like domains were expressed and found TRAPPC8 when the tail was expressed. However, the lists of proteins returned by these assays were short and many functional interactions could have been missed. TMEM131 was proposed to cooperate in collagen export with TMEM39 and Serpin H1/HSP47, both of which appear in my proteomic datasets.

Roles in collagen secretion prompted comparisons between TMEM131 and the protein TANGO1.[315] TANGO1 appears to have lumenal domains to assist in folding of clients and a long cytoplasmic tail to recruit components of the COPII machinery to sites of ER export. TMEM131 appears to have similar roles for the PapD-like domains and cytoplasmic tail. TANGO1 has been reported to have a one full transmembrane helix proceeded by a hydrophobic loop that is partially inserted into the inner leaflet of the ER membrane. The sequence and predicted structure of TMEM131 could imply a similar architecture. This is consistent with my data suggesting only the second of the two helixes actually crosses the membrane. TANGO1 operates as a heterodimer with cTAGE5 whilst my split Venus assay suggests TMEM131 might have some capacity to form homodimers. TANGO1 does not get transported to the Golgi with its cargo, similar to the ER-residency of full-length TMEM131.

Whilst a role for TMEM131 in collagen secretion seem plausible, this does not exclude other clients. TMEM131 appears to be widely expressed throughout the body so would likely interact with different proteins in different cell types. The U2OS osteosarcoma cells investigated by Zhang *et al.* produce a lot of collagen so a chaperone to aid in collagen secretion might be expected. Meanwhile when TMEM131 expression was disrupted in K562 cells, the greatest change in gene expression was the up-regulation of SLC3A2. SLC3A2 is extremely abundant in HeLa cells, and presumably HEK cells, as a component of various amino acid importers that help drive cell growth. SLC3A2 appears in both my BioID and GFP pulldown datasets so seems to be a likely client of TMEM131. CRISPR screens in other cell lines, detailed in the section 4.1.3, suggest a role of TMEM131 in the secretion of MR1 and TREM2.

Section 4.1.1 in the introduction to this chapter, described evidence linking TMEM131 to T cell development. This includes an increase in gene expression during β-selection, regulation via Bcl11b and hypomethylation of the gene in lymphocytes from patients with trisomy 21. Why might TMEM131 be involved in T cell development in particular? During β-selection there is a burst of thymocyte replication and differentiation down the $\alpha\beta$ and $\gamma\delta$-lineages. This requires thymocytes changing the expression of many surface proteins. Perhaps TMEM131 is an important chaperone for these new proteins?

Unpublished data within the James lab suggests TMEM131 interacts with the pre-TCR. The pre-TCR is quite a large for a surface receptor with eight total subunits (pTαTCRβ CD3$\gamma\delta\epsilon_2\zeta_2$). TMEM131 might help to keep the pre-TCR stable whilst this complex forms. Unlike most receptors which are present in a limited number of isoforms, the pre-TCR includes the TCRβ chain which is generated through VDJ recombination. This process generates sequence diversity in the variable chain, which is required for recognising varied peptide ligands. Sequence diversity however would be expected to produce TCRβ chains that fold or assemble poorly. Extra chaperones might be beneficial in managing thymocytes survival under these circumstances.

I made several efforts to make a TMEM131 knockout in HEK cells, then expanded out to HeLa and RPE1 cells. These attempts are detailed in appendix A.1. I was ultimately unable to generate and verify a knockout cell line. The number of alleles I found with indels of multiples of 3bp, producing an in-frame protein, seemed suspicious. However, I cannot conclude that making a *tmem131* knockout line is impossible and it would be surprising if TMEM131 was essential in HEK cells. Continuing these

efforts indefinitely did not seem a productive use of time when there was no guarantee of a measurable phenotype at the end of the process. This is particularly relevant to TMEM131 because you might expect a level of redundancy with other chaperone proteins.

My original intention was to see if the pre-TCR and $\alpha\beta$TCR could still reach the cell surface in a *tmem131* knockout line. However, there is variation in the surface expression of the receptor in HEK due to many factors including variability in transfection efficiency of both the receptor and the CD3 chains as well as the age/passage number of the line. After genomic editing, single cell sorting then being grown up from a single cell, the resulting HEK cell line may not be as healthy as a result. Comparisons in surface expression might be difficult with this background. If a knockout cell line can be generated and a phenotype measured then this could be used as an assay for TMEM131 function. Transfection with structurally-informed mutants constructs, based upon the AlphaFold model, might then be tested to derive functionally important parts of the molecule.

Chapter 5

Investigating the role of TMEM131 in Zebrafish development

5.1 Introduction

As discussed in section 4.1, the molecular function of the TMEM131 is poorly defined. Nonetheless, circumstantial evidence has linked the protein to T cell development. More recent papers imply the protein is involved in the export of proteins, particularly collagen, from the ER.

After establishing an interaction between TMEM131 and the pre-TCR, the James lab commissioned a mouse knockout strain with the Sanger institute. The data for this mutant is available from the International Mouse Phenotype Consortium with the code MGI:1927110. This work found complete preweaning lethality for Tmem131-/- mice with a cross between heterozygous pairs producing 21 wild-type and 46 heterozygous pups with no homozygous pups observed. This is clearly more severe than the expected immune dysfunction phenotype (for instance that of a pTα knockout). Much more recently, data has been published for homozygous mutant mice at the E14.5-E15.5 stage of development with statistically significant observations of haemorrhage in 1/3 males and 1/3 female embryos (compared to 5/295 male and 4/308 female controls), oedema in 1/3 males and 3/3 female embryos (compared to 2/292 male and 3/299 female controls) and abnormal craniofacial morphology in 0/3 males and 2/3 female embryos (compared to 0/292 male and 0/299 female controls).

Whilst the inviability of Tmem131 -/- embryos strongly supports a critical role for the protein in an important process, it is difficult to discern molecular function without mice to study and studying the development of pups *in utero* is extremely challenging. We were therefore keen to investigate the knockout phenotype in a more tractable system where development could be studied closely.

Zebrafish (*Danio rerio*) is a small, freshwater fish, native to south east Asia. It has become an established non-mammalian, vertebrate model organism for scientific research. The fish has a short generation time (~3 months), high fecundity and can been kept relatively cheaply in high numbers. Zebrafish have long been tractable for genetical manipulation[364] and their development is well understood due their transparent embryos which are fertilised externally.[365]

Mutant Zebrafish lines can be generated using the CRISPR-cas9 system by injecting the mRNA encoding the cas9 endonuclease with a *β-globin* 3'UTR and guide RNAs into embryos at the one cell stage.[366]

These guide RNAs associate with the nuclease and target the enzyme to identical sequences in the genome. Once at a genomic locus, the nuclease can create a double strand break that is processed by the non-homologous end joining DNA repair pathway. Imperfect DNA repair causes insertion or deletion mutations (indels) that disrupt the function of the gene. The *S. Pyogenes* cas9 endonuclease can be directed to nearly any genomic locus but can only cut 3-4bp 5' of a Protospacer Adjacent Motif (PAM) with the sequence NGG. In practice however, some guide sequences are more efficient at creating indels than others. Doses of guides and cas9 mRNA need to be determined empirically.

Traditional phenotypic studies of gene mutants are done by crossing two heterozygous carriers of the mutation and studying the impact in the offspring (the F1 generation). CRISPR with efficient guides can generate bi-allelic mutations so the phenotype can be studied directly in the injected embryos (the F0 generation). However this is usually complicated by side-effects from injection, off-target effects and contributions from maternal mRNA. Instead the injected embryos, all a mosaic of somatic mutations, are raised then screened. If the frequency of editing in the germline cells is very high then these F0 fish can be in-crossed to produce F1 embryos in which the phenotype is observed. Alternatively, the edited fish are out-crossed to WT then these embryos raised and screened. These homozygous fish can then be crossed together to study the phenotype in the F2 generation.

In some cases CRISPR can cause lethality in injected embryos so fish cannot be raised. In these cases, editing can be concentrated in germline cells by fusing the *nanos* 3' UTR to cas9 mRNA rather than the *β-globin* 3'UTR.[367] These fish are raised to adulthood then screened to find ones with the highest proportion of edited gametes. The best fish can then be crossed back to wild-type producing a mixture of wild-type and homozygous offspring. These offspring can be raised and screened using fin clipping. Finally, homozygous fish can be crossed to generate homozygous fish in which the phenotype is observed.

5.1.1 Aims

This chapter aims to create a zebrafish knockout of TMEM131 using a protocol by Varshney *et al*.[368] This protocol is summarised in figure 5.1. Once mutant fish are generated, studying the mutant phenotype will provide insight into the molecular function of TMEM131 *in vivo*.

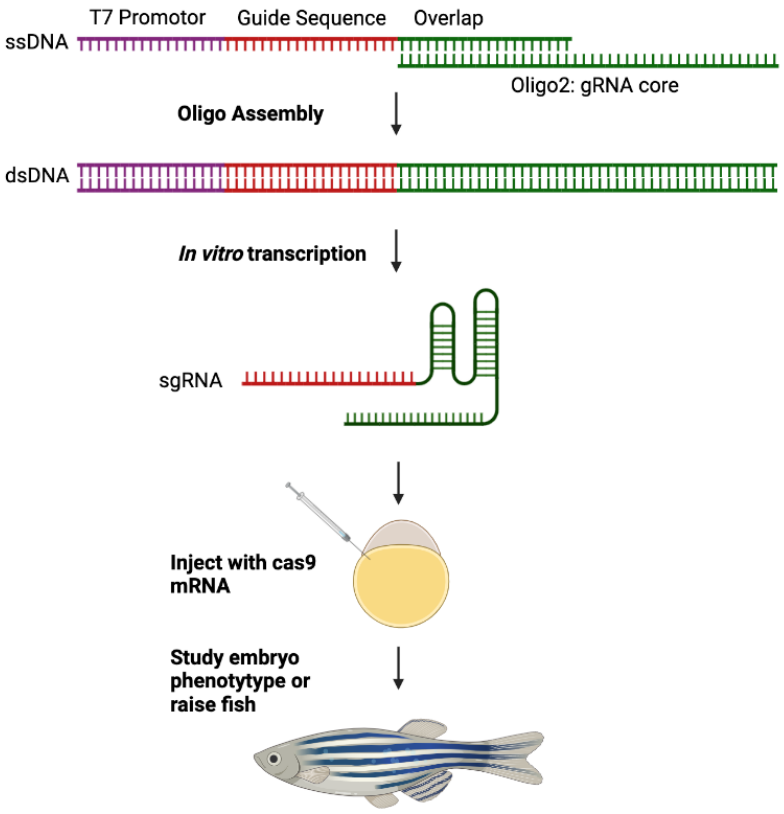

Figure 5.1: Method for generating Zebrafish mutants using CRISPR including oligo assembly, *in vitro* translation and injection steps.
Based off the protocol from Varshney *et al.*[368]

5.2 Methods

5.2.1 Design of CRISPR guide RNAs

The gene sequence for TMEM131 in *Danio rerio* was obtained from Ensembl and was annotated with exon, intron and 5' UTR sequences. The gene is located on chromosome 6 in position 29,313,220-29,377,092 on the reverse strand. Exon sequences were scanned for possible guides using the program CRISPOR[369] using the Zebrafish - UCSC May 2017 (GRCz11/danRer11) genome assembly and the setting for a 20bp-NGG PAM sequence. CRISPOR gives additional metrics including predictions for guide efficiency, the number of off-targets, predicted indels based upon microhomology surrounding the target site as well as suggested primers to amplify the on-target site.

As transcription from the T7 promotor is most efficient when starting with "GG", guides 1 and 2 had two and one extra "G" residue added to the 5' end. The 5' nucleotides are thought to have little or no effect on target specificity or guide activity. Other authors reported that transcription is efficient enough with only a single G nucleotide so guide 4 is left unmodified.

Guide	Target	Sequence	Off-targets[1]	Predicted Efficiency[2]
1	Exon1: Init. Met	*gg*CAGCAGCAGGATGGCGAGCA	0-0-0-2-41	76
2	Exon1: TSS	*g*GTTCACTCACGCGCGACACC	0-0-0-1-8	34
3	Exon 7	GGAACTCTTGCCCCTACAAA	0-0-0-0-13	32
4	Exon 9	GATGATAAACTCGTCCTCGT	0-0-0-0-17	42
5	Exon 24	GGATTGGGGTATAAAGCCAG	0-0-0-0-19	87

Table 5.1: Guide sequences for generating a Zebrafish TMEM131 knockout.
Sequences were generated using CRISPOR software using the UCSC May 2017 (GRCz11/danRer11) genome assembly and the setting for a 20bp-NGG PAM sequence. Guide 1 targets the initiator methionine (underlined) whilst guide 2 targets the transcription start site (underlined) which is a C on the opposite strand. These guides have two and one "g" residues respectively added to the 5' end (italicised) to aid transcription from the T7 RNA polymerase promotor. 1- The number of off-targets with 0-1-2-3-4 mismatches next to a PAM sequence. 2- The guide efficiency, a prediction of how likely DNA cleavage is to occur at this sequence from 0-100. This was trained on data from 1000 guides on >100 genes, from zebrafish 1-cell stage embryos injected with cas9 mRNA.[367]

Oligo Assembly

The five CRISPR guide sequences designed above were ordered as single stranded DNA oligos with the 17nt sequence for the T7 promotor (TAATACGACTCACTATA) fused to the 5' end and a 20nt overlapping sequence (GTTTTAGAGCTAGAAATAGC) on 3' end. This sequence is complementary to Oligo 2, an 80nt sequence containing the rest of the sgRNA (AAAAGCACCGACTCGGTGCCACTTTTT CAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC) that was ordered once. These RNAs can be assembled into a single ~120bp dsDNA oligo using the reaction conditions in table 5.2. Successful assembly was established by running products on a 2.5% agarose gel (figure 5.2).

Reaction Component	**Volume (µl)**
10µM Guide oligo	1
10µM Oligo 2	1
KOD polymerase	0.5
KOD mix	6.5
Water	16

Step	**Temp (°C)**	**Time (mins)**
Denature	98	2
Anneal	50	10
Extend	72	10

Table 5.2: Reaction components and conditions for guide oligo assembly

In vitro transcription

Assembled dsDNA oligos were transcribed to RNA using a HiScribe T7 High Yield RNA Synthesis Kit (NEB). These reactions were assembled in PCR tubes on ice with components added in the order of table 5.3. These tubes were incubated for 16h overnight at 37°C. The manufacturer recommends these incubation time, buffer and NTP concentrations for maximum yield of transcripts shorter than 0.3kb such as these.

The reactions were purified using a Monarch RNA cleanup kit (NEB) using the manufacturer's instructions. After measuring the concentration using a Nanodrop, these RNA's were aliquoted into PCR tubes containing 2µl of 200ng/µl and stored at -80°C.

To check the quality of the product, 1µl of each RNA or 1µl of Low Range ssRNA Ladder (NEB) was mixed with 2µl of Invitrogen Gel Loading Buffer II (95% Formamide, 18mM EDTA, 0.025% SDS, Xylene

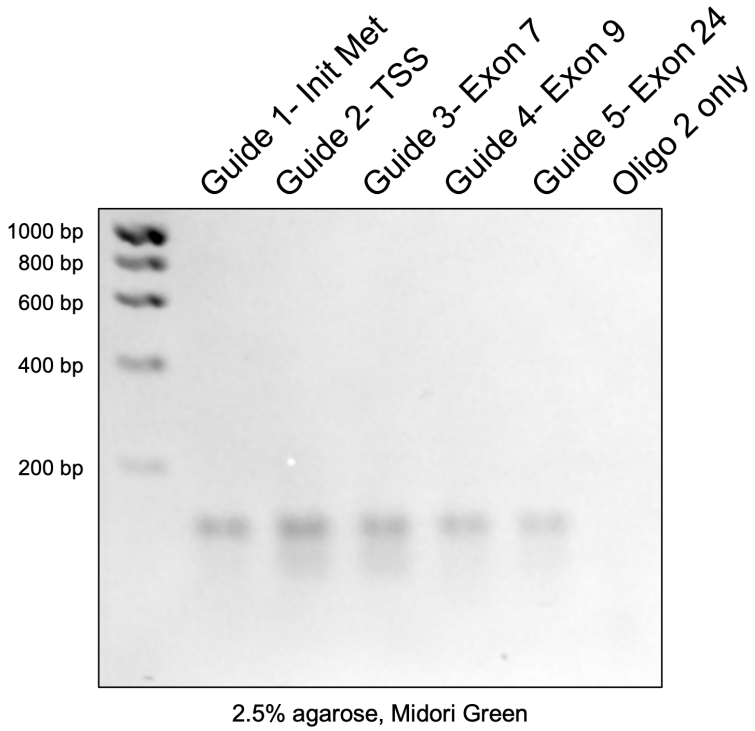

Figure 5.2: Successful assembly of dsDNA encoding sgRNAs. The dsDNA made through the oligo assembly procedure above was run on a 2.5% agarose gel with Midori direct. The correct band size is approximately 120bp.

Reaction Component	Volume	Final Conc
Nuclease-free water	6µl	
10X reaction buffer	1.5µl	0.75X
100mM NTPs	1.5µl each	7.5mM each
Template DNA	5µl	~1µg
T7 RNA polymerase mix	1.5µl	

Table 5.3: Reaction components for *in vitro* transcription using HiScribe T7 High Yield RNA Synthesis Kit.

Cyanol, and Bromophenol Blue). These mixtures were incubated at 70°C for 10 minutes before being immediately cooled on ice. These tubes were run on a 2.5% agarose gel for 20 minutes at 100V using fresh TAE 1X buffer (figure 5.3). A single band was seen for each RNA suggesting they had been transcribed correctly and had not degraded.

5.2.2 Zebrafish husbandry and injection

The Tu Zebrafish strain was used for most experiments[†]. This is a "wild-type" strain that was bred to remove embryonic lethal mutations for use as a background for mutagenesis, sequencing and screening experiments.[364]

Fish were maintained at 28.5°C at the University of Warwick Biomedical Services Unit in compliance with the University of Warwick animal welfare and ethical review board (AWERB) and the UK home office animal welfare regulations. Embryos were collected from pairwise or pooled crosses of wild-type fish using standard mating procedures.

[†]The Tu strain originated from a pet shop in the German university city of Tübingen

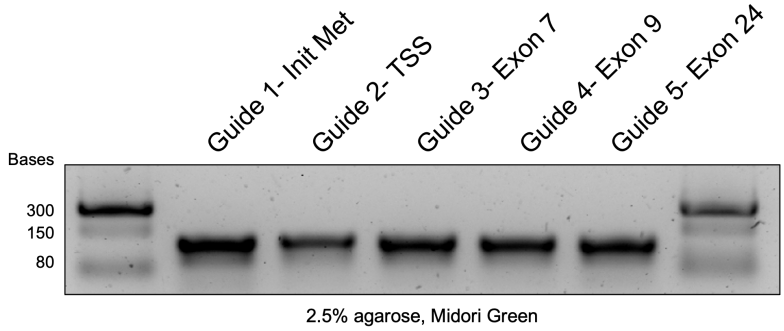

Figure 5.3: 2.5% agarose gel confirming successful *in vitro* transcription of sgRNA sequences.

Injections were performed by Professor Karuna Sampath using varying doses of guide RNAs and cas9 mRNA. The cas9:nanos used was a gift from Conor Talbot in the Nelson lab.

Embryos were grown at 28°C in egg water (60 µg/ml Instant ocean sea salt, Sera Marin) with the dye methylene blue. In most cases, embryos were raised for 24-48h before being lysed. No embryos edited with cas9:globin were kept beyond five days post-fertilisation.

5.2.3 Embryo Lysis

Embryos in egg water were placed into PCR tubes with a Pasteur pipette and as much liquid removed as possible without crushing the embryo. A mixture of 1ml of whole-mount *in situ* hybridisation (WISH) buffer and 10µl of Proteinase K was prepared and 30µl added to each sample. These were incubated at 55°C for 3 hours then heated to 95°C for 15minutes to inactivate the enzyme. 5µl of these lysates was used as template for PCR with KOD or Q5 polymerase with the number of cycles increased to 35. Lysates were stored at -20°C.

5.2.4 T7 Endonuclease1 genome editing detection assay

Genome editing events were detected in CRISPR-modified Zebrafish embryos by amplifying a ~600bp genomic region with the target site slightly off-centre. The PCR product was then heated and cooled to anneal strands from different chromosomes.

Reaction Component	Volume (µl)
PCR product	5
NEBuffer 2 10X	1
Water	4

Temperature	Condition
95°C	10min
95°C to 85°C	-2°C/s
85°C to 25°C	-0.1°C/s
4°C	hold

Table 5.4: Reaction components and conditions for oligo annealing

These annealed PCR products were then incubated with 10µl of T7E1 enzyme mix for 30 minutes at 37°C before being immediately cooled to 4°C. In control reactions the enzyme was replaced by the same volume of water.

Reaction Component	Volume (µl)
T7E1 (NEB, 10U/µl)	0.25
NEBuffer 2 10X	1
Water	8.75

Table 5.5: T7E1 enzyme mix

T7 Endonuclease I (T7E1) selectively cleaves mismatched DNA sequences. Products were run on a 3% agarose gel with 4 µl of OrangeG loading dye and 0.5µl of MidoriGreen Direct DNA stain. Cleavage of the DNA was identified by comparing treated and untreated lanes against a 100bp DNA ladder (NEB).

5.3 Results

5.3.1 Targeting TMEM131 exon 1 with guides 1 and 2

Zebrafish embryos were injected with 150pg of cas9:globin together with 150pg of guide 1 or guide 2 sgRNAs. Injected embryos were raised at 28°C. About 50% of embryos injected with guide RNA 2 seemed to have a phenotype that was similar to the loss of bmp1/tolloid metalloprotease or overexpression of chordin.[370,371] This would imply a role for TMEM131 in the modulation of BMP signalling. Some embryos appeared to have malformed hearts, fat tails and slower development that expected. However the number of injected embryos was quite low. Lethality in injected embryos was very high.

Injected embryos and uninjected controls were lysed in pools of three. The gDNA for TMEM131 exon 1 was amplified and treated with a T7E1 assay to check for genomic editing (figure 5.4). For both guide sequences, there was additional bands in the lanes treated with T7E1 implying that genomic editing had occurred. The sizes of these band were consistent with cuts at position of the guides in the amplified region. There was however also additional bands in the uninjected controls implying there was single-nucleotide polymorphisms (SNPs) in this exon sequence. These PCR parts were ligated into vectors and sequenced. For guide2-injected embryos, sequencing found a 2bp deletion immediately before the transcription start site as well as 7bp and 9bp deletions spanning transcription start site.

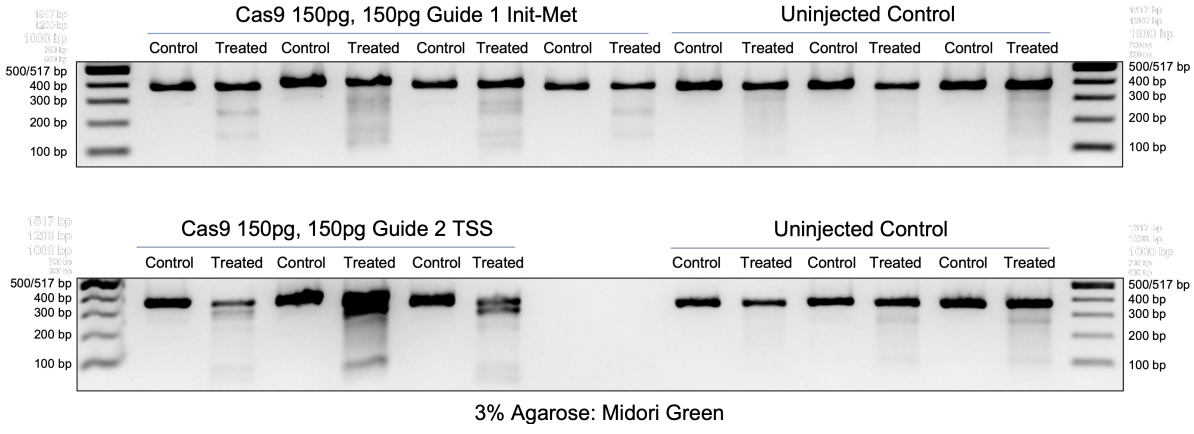

Figure 5.4: T7E1 assay of embryos injected with guides 1 and 2, both targeted to *TMEM131* exon 1 with cas9:globin
For each sample the untreated (control) lane is run before the T7E1 treated. One gel was used for both rows but the large amount of intervening space has been cut out after contrasting.

In an attempt to reduce lethality, embryos were injected again with reduced doses of cas9:globin and guide 2. The same phenotype and lethality was observed with the two higher of these doses. Lethality seemed to be slightly reduced in the lowest dose whilst the uninjected controls seemed normal. The highest dose still produced a lot of normal-looking embryos suggesting an issue with SNPs in the targeted region. A T7E1 assay with targeted embryos suggested all three doses produced genomic edits in the target sequence.

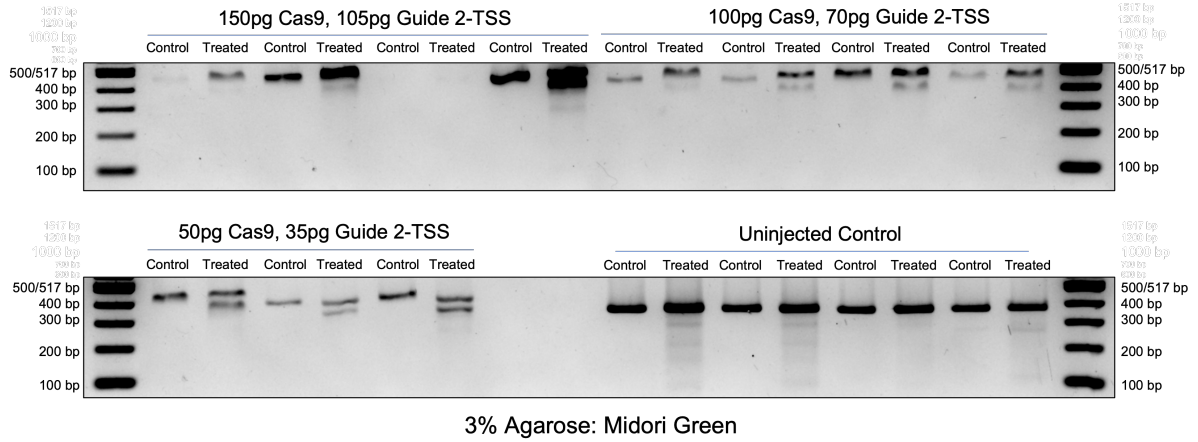

Figure 5.5: T7E1 assay of embryos injected different doses of guide 2 and cas9-globin.
One gel was used for both rows but the large amount of intervening space has been cut out after contrasting.

To see if a Zebrafish line with a different genomic background produced more consistent results, embryos from the Sg line were injected with two low doses of cas9:globin and guide 2. In this background, the lowest dose also produced an inviable phenotype. Individual embryos were lysed and amplified exon 1 DNA screened for genomic editing with T7E1 (figure 5.6). Editing looked to have taken place as before but there also seemed to be an issue with SNPs in the uninjected controls.

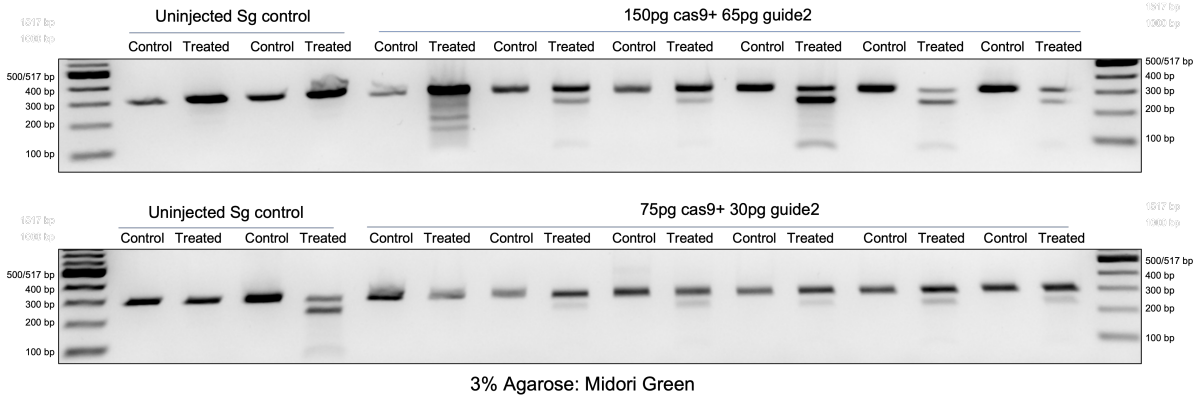

Figure 5.6: T7E1 assay for cas9:globin and guide 2 in the Sg background.
One gel was used for both rows but the large amount of intervening space has been cut out after contrasting.

5.3.2 Targeting TMEM131 exon 7 with guide 3

Whilst the phenotype with guide 2 was promising, the high lethality observed meant the effects of TMEM131 could not be studied consistently. I also could not rule out any off-target effects. We tested guide 3, targeted to TMEM131 exon 7 with the same method. Lethality was not as obvious but a T7E1

assay implied that this region has a polymorphism in the wild-type Tu background. This would make screening for indels more difficult so this guide was not used further.

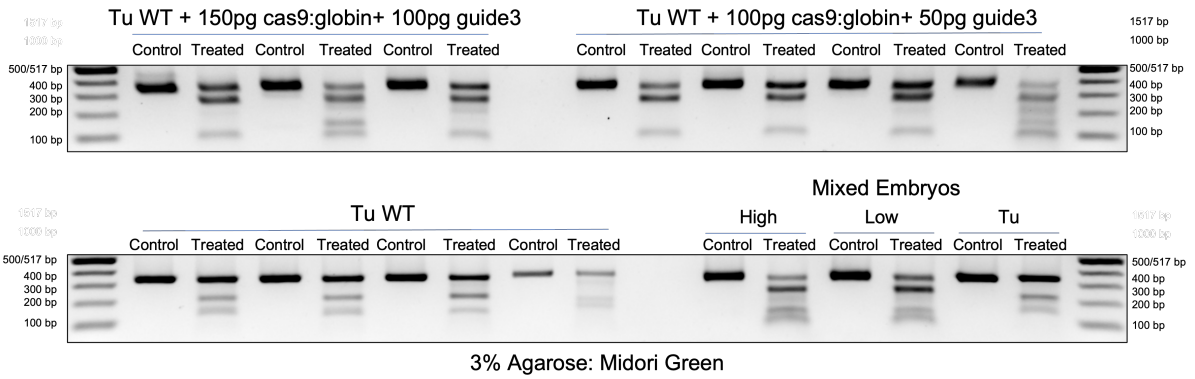

Figure 5.7: T7E1 assay for embryos injected with guide 3 targeted to *TMEM131* exon 7.
Individual embryos were lysed and TMEM131 exon 7 amplified. One gel was used for both rows but the large amount of intervening space has been cut out after contrasting.

5.3.3 Generating germline mutants using guide 4 targeted to exon 9

As the somatic mutations induced by cas9:globin and guide 2 were inviable, we decided to raise fish with germline-targeted mutations. Guide 2 and the untested guide 4 were injected with cas9:nanos mRNA and the embryos raised. Of the embryos injected with cas9:nanos and guide 4, five female fish and eight male fish survived to adulthood. Unfortunately no embryos injected with guide 2 survived this process.

As cas9 has been targeted to the germline cells, we expect the fish to have a mixture of edited and non-edited gametes. The germline cells are committed quite early in zebrafish development (before the 1000 cell stage) but depending on when editing occurred these could have a range of possible edits. To generate homozygotes without having to screen an excessive number of fish, these would need to be screened for good indels. To screen for edits in germline cells efficiently, an in-cross was set up between five pairs of fish. Embryos from these crosses were lysed in pools of five and screened for genomic edits (figure 5.8). Edits were found in the embryos of pair 1 and pair 3. The fish in pairs 2, 4 and 5 were culled.

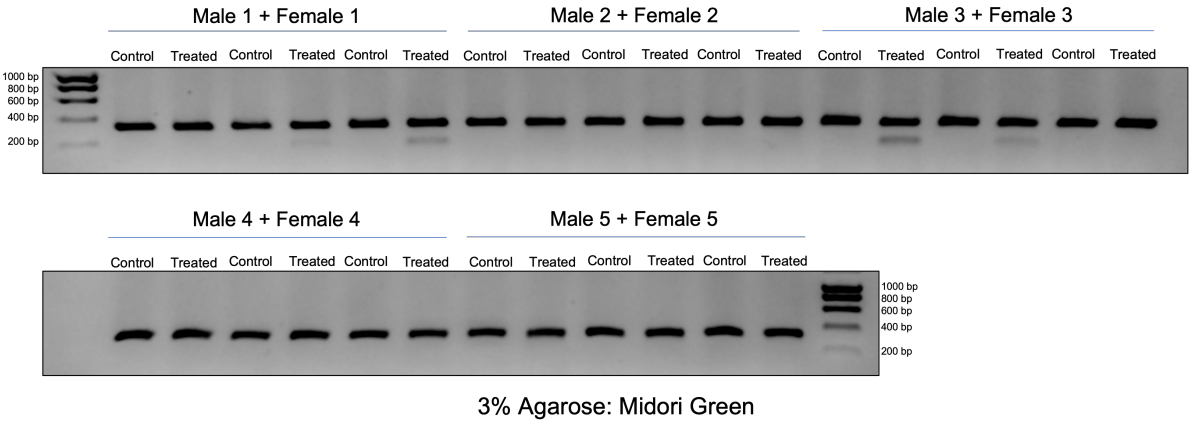

Figure 5.8: T7E1 assay of an in-cross between five pairs of fish that had cas9 targeted to TMEM131 exon 9 in germline cells.
One gel was used for both rows but the large amount of intervening space has been cut out after contrasting.

To see which of the fish in these pairs had edited gametes, they were crossed with wildtype Tu fish of

the appropriate sex. The extra males were also crossed with WT females. Embryos from these crosses were screened with T7E1 as before (figure 5.9). Both fish in the original pairs 1 and 3 seemed to have a reasonable number of edited gametes as did male 7.

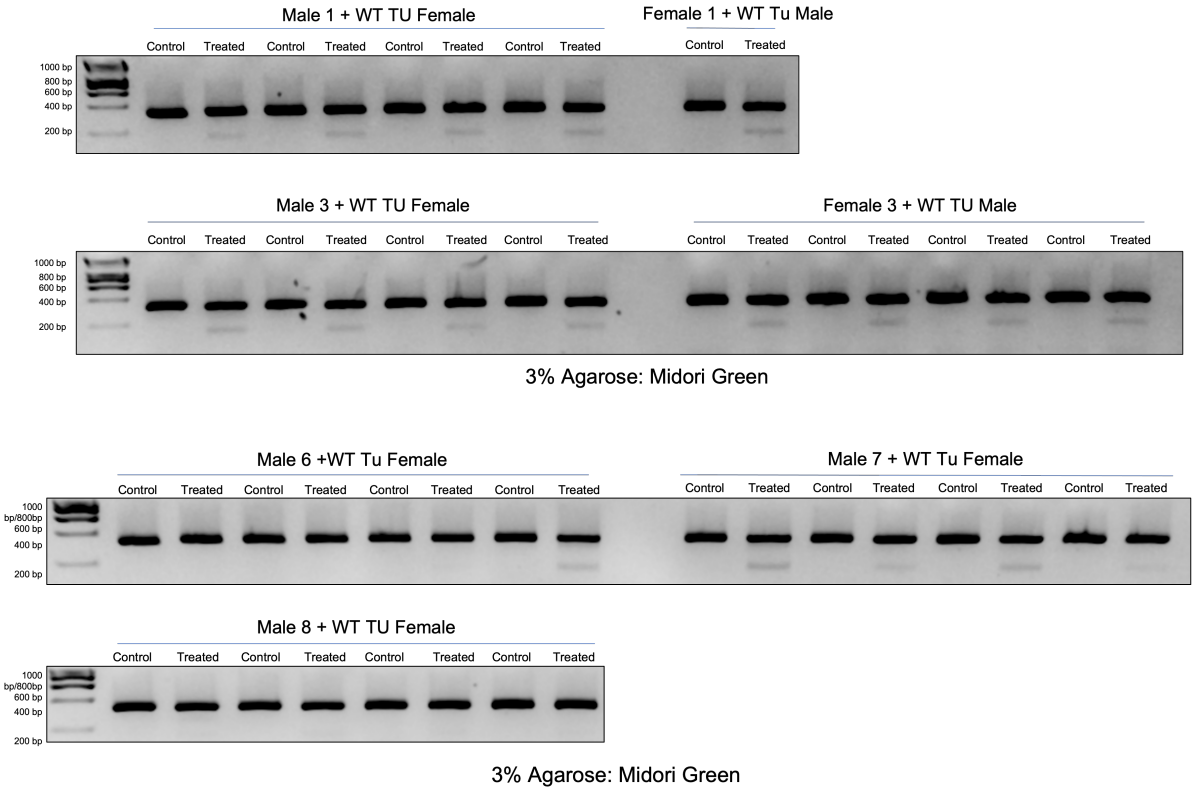

Figure 5.9: Outcross between germline-edited fish and WT TU mates.
Two agarose gels were used here with intervening space cropped out.

The amplified DNA was ligated into plasmids, transformed then several colonies sequenced for each cross. As expected from the amount of edited DNA, most colonies contained the WT sequence. Surprisingly however the same indel appeared in the embryos from four separate fish (Female 1 and Males 1,3 and 5). In fact this mutation, the loss of one base pair and insertion of 5 more, was the only indel found (figure 5.10).

CRISPR causes double strand break at sites which are repaired by non-homologous end joining. This repair pathway can produce some indels more frequently than others due to micro-homology around the break site. Whilst it seemed suspicious for an identical mutation to occur in several fish, the same indel appearing in different fish has been reported before.[372] As the indel is exactly where expected based on the guide, it is not likely a polymorphism. This indel was convenient however, as it would cause a frame shift in the protein and a stop codon to be present before the end of exon 9.

The male 1 and female 1 fish were outcrossed to wild-type Tu fish and embryos raised. As of writing, these fish are still being raised to adulthood when they will be screen for the presence of the mutation via fin clipping. Homozygous fish will then be crossed to study the homozygous phenotype in the F2 generation. If permitted the phenotypical details will be included as a supplement to this thesis.

To get an early look at the mutant phenotype two new in-crosses were set up (male 1 with female 1 and female 3 with both male 3 and male 7). About 100 embryos from each cross were collected and watched closely through development. Most of these fish would be expected to be wildtype or heterozygotes. After 24h, a few fish had obvious phenotypes (figure 5.11). One fish was found with an abnormally

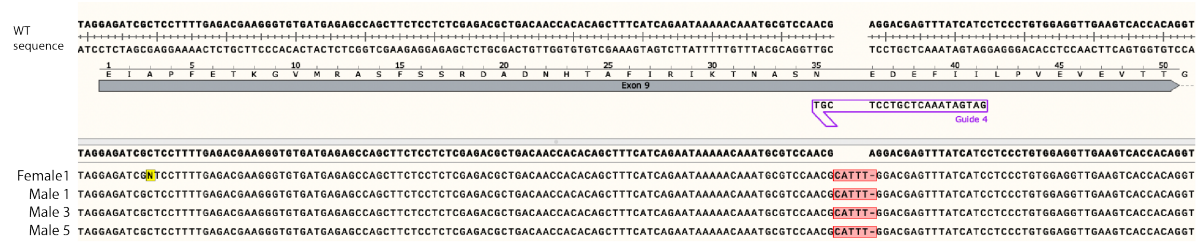

Figure 5.10: Mutant sequences found from crosses from germline edited fish.
Crosses between fish with germline edits and WT Tu mates were lysed and the gDNA for TMEM131 exon 9 amplified. This DNA was ligated into a pHR plasmid, transformed into *E. coli* and colonies sequenced. Most colonies contained the wild-type sequence.

curved tail and two appeared to have a pericardial oedema.

After 48 hours, the tail in the first abnormal embryos appeared to have partially straightened out and further embryos with the cardiac phenotype were found in the cross between female 1 and male 1 fish (figure 5.12). Eight abnormal embryos and eight normal-looking embryos were lysed and sequenced to see if the mutation correlated with the phenotype. All eight normal-looking embryos returned the WT sequence (one colony each sequenced). One allele with the same insertion identified before and one allele of the wildtype sequence was found in the first abnormal fish with the curved tail showing that this fish was heterozygous for TMEM131. The WT sequence was also found in all eight abnormal fish (2 reactions) indicating they are heterozygous or wild-type.

I conclude that the pericardial oedema is not associated with TMEM131 and is just an unfortunate background present in the inbred Tu strain. The curled tail phenotype in the heterozygous fish is concerning. If heterozygous fish have a disadvantage then raising heterozygotes to later cross might prove unsuccessful.

Figure 5.11: Abnormal embryos in a cross between TMEM131 germline mutants at approximately 24h post-fertilisation.
Images obtained on an Olympus SZ91 stereo microscope equipped with a Moticam 5 CMOS camera.

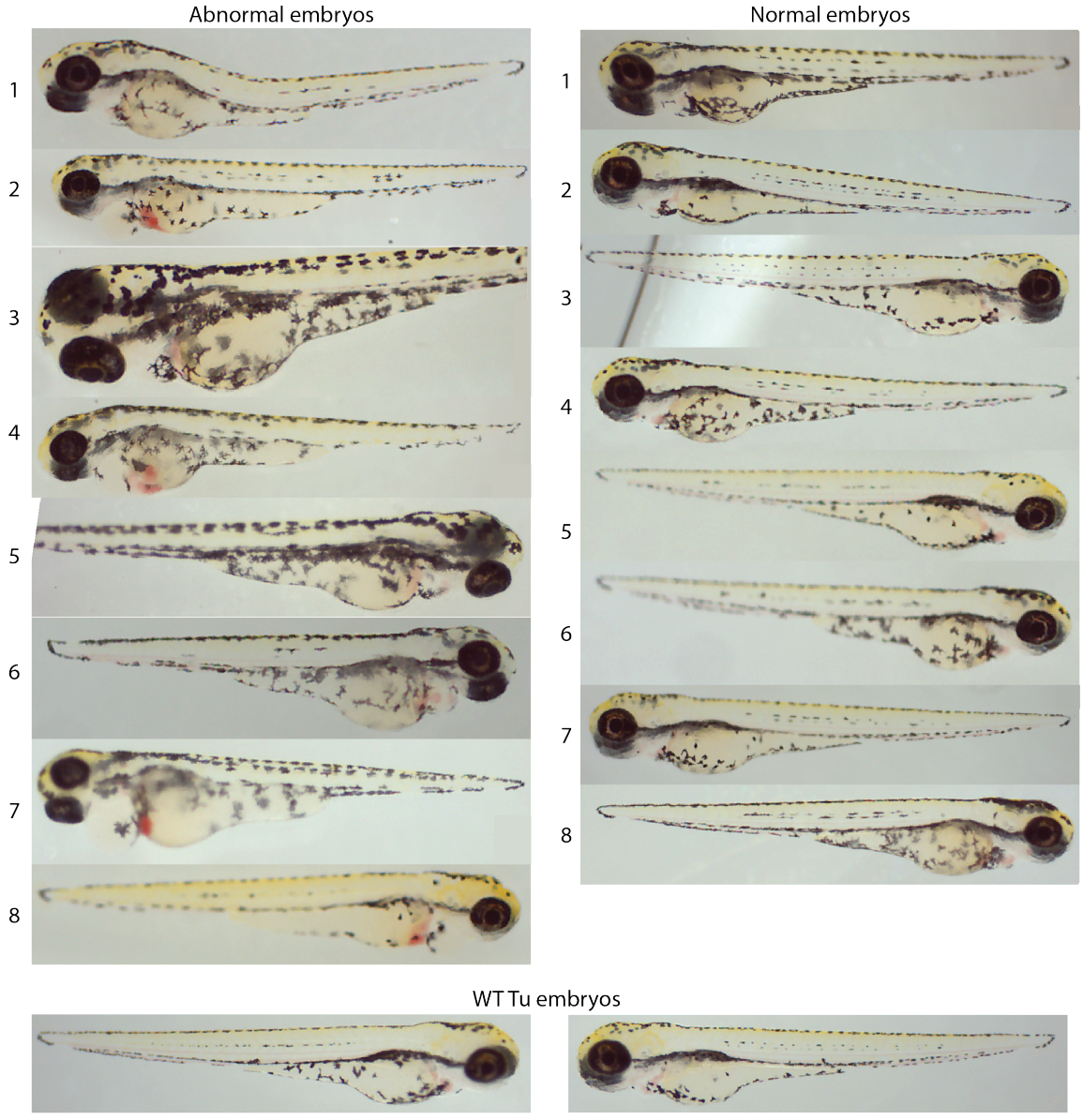

Figure 5.12: Abnormal embryos in a cross between TMEM131 germline mutants at approximately 48h post-fertilisation.
Images obtained on an Olympus SZ91 stereo microscope equipped with a Moticam 5 CMOS camera.

Chapter 6

Final Discussion

In this thesis, I have investigated the trafficking of the pre-TCR through reconstituted expression in HEK293T cells. Whilst the $\alpha\beta$TCR is readily detectable at the cell surface in this system, the pre-TCR has very low surface expression at steady state. The main aim of this project was to establish the mechanism for the difference in surface expression. Previous data in the lab implicated asymmetry in the extracellular domains of the receptor as the main structural cause.

This work searched for a membrane interactions partner for the pre-TCR that could trigger internalisation. Despite several attempts no such protein was identified. Whilst the SPPLAT assay identified the components of the pre-TCR and TCR, peptides from these proteins were not so abundant as to be confident of capturing an interaction partner if one were present. In retrospect however, it seems unlikely HEK cells would express the right protein for pre-TCR internalisation unless the mechanism was so generic that it would have been identified in other contexts. Dimerisation of the pre-TCR was investigated with various techniques which concluded the pre-TCR was primarily monomeric.

This work found that the pre-TCR and $\alpha\beta$TCR were capable of comparable rates of ER to surface trafficking when secretion was synchronised with the RUSH assay. The amount of receptor trafficked was higher with the $\alpha\beta$TCR than the pre-TCR suggesting the latter has difficulty in complex assembly. I suspect this factor makes up a large part of the difference in surface expression. Other data in the lab supports a contribution of receptor recycling. Unfortunately these are the steps that seem least amenable to experimental manipulation.[373]

Finally we demonstrated that the pre-TCR and $\alpha\beta$TCR produce distinct tonic signalling when expressed in Jurkat T cells. It would be very interesting to test the modified receptor constructs mentioned in this thesis using the same system. For instance, the pre-TCR variants missing the Vβ domain or with Vα fused to the N-terminus of pTα both had surface expression comparable to the $\alpha\beta$TCR when expressed in HEK. Would they produce tonic signalling in Jurkat cells that mimics the $\alpha\beta$TCR as well?

This thesis also investigated TMEM131 which was identified as a pre-TCR interaction partner in HEK cells. Published data links the protein to T cell development and the export of proteins from the ER. Data acquired in this thesis is consistent with TMEM131 being an ER chaperone. TMEM131 is targeted to the ER by its signal peptide and is retained by a sequence at the end of its long cytoplasmic tail. Fragments of this tail observed in western blotting are attributed to partial proteasome degradation. The functionality of this effect is unclear but I do not think it is merely an artefact of the GFP domain used in most experiments.

Work is currently ongoing to produce a knockout of TMEM131 in Zebrafish. Provisional data from embryos injected with cas9:globin and the mouse knockout imply the phenotype might be severe. Broad expression and high conservation suggest TMEM131 acts as a chaperone for a reasonably diverse set of clients. Depending on redundancy with other chaperones, this means the knockout could disrupt a wide range of developmental processes. Many suspected clients seem to be multi-protein complexes with immune system roles.

Why might TMEM131 be expressed during T cell development and interact with the pre-TCR in particular? During β-selection checkpoint, thymocytes are primed for $\alpha\beta/\gamma\delta$-lineage choice or death by apoptosis. TMEM131 might be a valuable chaperone protein for the massive changes in protein expression that accompany this progression.

Bibliography

[1] Martin F Flajnik and Masanori Kasahara. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nature Reviews Genetics*, 11(1):47–59, 2010.

[2] Daniel Ricklin, Edimara S Reis, and John D Lambris. Complement in disease: a defence system turning offensive. *Nature Reviews Nephrology*, 12(7):383, 2016.

[3] David D Chaplin. Overview of the immune response. *Journal of Allergy and Clinical Immunology*, 125(2):S3–S23, 2010.

[4] Mark D Turner, Belinda Nedjai, Tara Hurst, and Daniel J Pennington. Cytokines and chemokines: at the crossroads of cell signalling and inflammatory disease. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1843(11):2563–2582, 2014.

[5] David M Mosser and Justin P Edwards. Exploring the full spectrum of macrophage activation. *Nature reviews immunology*, 8(12):958–969, 2008.

[6] Jacqueline Parkin and Bryony Cohen. An overview of the immune system. *The Lancet*, 357(9270):1777–1789, 2001.

[7] Kiyoshi Takeda and Shizuo Akira. Toll-like receptors in innate immunity. *International immunology*, 17(1):1–14, 2005.

[8] Charles A Janeway Jr and Ruslan Medzhitov. Innate immune recognition. *Annual review of immunology*, 20(1):197–216, 2002.

[9] J Claire Hoving, Gillian J Wilson, and Gordon D Brown. Signalling c-type lectin receptors, microbial recognition and immunity. *Cellular microbiology*, 16(2):185–194, 2014.

[10] Qi Chen, Lijun Sun, and Zhijian J Chen. Regulation and function of the cgas–sting pathway of cytosolic dna sensing. *Nature immunology*, 17(10):1142, 2016.

[11] Jiaxi Wu and Zhijian J Chen. Innate immune sensing and signaling of cytosolic nucleic acids. *Annual review of immunology*, 32:461–488, 2014.

[12] Kenneth L Rock, Diego J Farfán-Arribas, Jeff D Colbert, and Alfred L Goldberg. Mhc class i-presented peptides and the drip hypothesis. *Trends in immunology*, 35(4):144, 2014.

[13] Jacques Neefjes, Marlieke LM Jongsma, Petra Paul, and Oddmund Bakke. Towards a systems understanding of mhc class i and mhc class ii antigen presentation. *Nature Reviews Immunology*, 11(12):823, 2011.

[14] Dean R Madden, Joan C Gorga, Jack L Strominger, and Don C Wiley. The three-dimensional structure of hla-b27 at 2.1 å resolution suggests a general mechanism for tight peptide binding to mhc. *Cell*, 70(6):1035–1048, 1992.

[15] Marek Wieczorek, Esam T Abualrous, Jana Sticht, Miguel Álvaro-Benito, Sebastian Stolzenberg, Frank Noé, and Christian Freund. Major histocompatibility complex (mhc) class i and mhc class ii proteins: conformational plasticity in antigen presentation. *Frontiers in immunology*, 8:292, 2017.

[16] Yi Xu and N Tony Eissa. Autophagy in innate and adaptive immunity. *Proceedings of the American Thoracic Society*, 7(1):22–28, 2010.

[17] Yuan-Hua Ding, Brian M Baker, David N Garboczi, William E Biddison, and Don C Wiley. Four a6-tcr/peptide/hla-a2 structures that generate very different t cell signals are nearly identical. *Immunity*, 11(1):45–56, 1999.

[18] Jens Hennecke, Andrea Carfi, and Don C Wiley. Structure of a covalently stabilized complex of a human $\alpha\beta$ t-cell receptor, influenza ha peptide and mhc class ii molecule, hla-dr1. *The EMBO journal*, 19(21):5611–5624, 2000.

[19] Federica Pierini and Tobias L Lenz. Divergent allele advantage at human mhc genes: signatures of past and ongoing selection. *Molecular biology and evolution*, 35(9):2145–2158, 2018.

[20] Veronique M Braud, David SJ Allan, Christopher A O'Callaghan, Kalle Söderström, Annalisa D'Andrea, Graham S Ogg, Sasha Lazetic, Neil T Young, John I Bell, Joseph H Phillips, et al. Hla-e binds to natural killer cell receptors cd94/nkg2a, b and c. *Nature*, 391(6669):795, 1998.

[21] Aura Burian, Kevin L Wang, Kathryn AK Finton, Ni Lee, Akiko Ishitani, Roland K Strong, and Daniel E Geraghty. Hla-f and mhc-i open conformers bind natural killer cell ig-like receptor kir3ds1. *PloS one*, 11(9):e0163297, 2016.

[22] Tsufit Gonen-Gross, Debra Goldman-Wohl, Berthold Huppertz, Dikla Lankry, Caryn Greenfield, Shira Natanson-Yaron, Yaron Hamani, Ronit Gilad, Simcha Yagel, and Ofer Mandelboim. Inhibitory nk receptor recognition of hla-g: regulation by contact residues and by cell specific expression at the fetal-maternal interface. *PLoS One*, 5(1):e8941, 2010.

[23] José A Villadangos and Petra Schnorrer. Intrinsic and cooperative antigen-presenting functions of dendritic-cell subsets in vivo. *Nature Reviews Immunology*, 7(7):543–555, 2007.

[24] Alexander Mildner and Steffen Jung. Development and function of dendritic cell subsets. *Immunity*, 40(5):642–656, 2014.

[25] Leslie Summers Deluca and Jennifer L Gommerman. Fine-tuning of dendritic cell biology by the tnf superfamily. *Nature Reviews Immunology*, 12(5):339–351, 2012.

[26] Stefan Nierkens, Jurjen Tel, Edith Janssen, and Gosse J Adema. Antigen cross-presentation by dendritic cell subsets: one general or all sergeants? *Trends in immunology*, 34(8):361–370, 2013.

[27] Christian Kurts, Bruce WS Robinson, and Percy A Knolle. Cross-priming in health and disease. *Nature Reviews Immunology*, 10(6):403–414, 2010.

[28] Andrew Smith, Paula Stanley, Kristian Jones, Lena Svensson, Alison McDowall, and Nancy Hogg. The role of the integrin lfa-1 in t-lymphocyte migration. *Immunological reviews*, 218(1):135–146, 2007.

[29] Michael L Dustin and Kaushik Choudhuri. Signaling and polarized communication across the t cell immunological synapse. *Annual review of cell and developmental biology*, 32:303–325, 2016.

[30] P Anton van der Merwe, Simon J Davis, Andrey S Shaw, and Michael L Dustin. Cytoskeletal polarization and redistribution of cell-surface molecules during t cell antigen recognition. In *Seminars in immunology*, volume 12, pages 5–21. Elsevier, 2000.

[31] George F Gao, Zihe Rao, and John I Bell. Molecular coordination of $\alpha\beta$ t-cell receptors and coreceptors cd8 and cd4 in their recognition of peptide-mhc ligands. *Trends in immunology*, 23(8):408–413, 2002.

[32] Rishi Vishal Luckheeram, Rui Zhou, Asha Devi Verma, and Bing Xia. Cd4+ t cells: differentiation and functions. *Clinical and developmental immunology*, 2012, 2012.

[33] Jinfang Zhu and William E Paul. Cd4 t cells: fates, functions, and faults. *Blood, The Journal of the American Society of Hematology*, 112(5):1557–1569, 2008.

[34] Foo Y Liew. Th 1 and th 2 cells: a historical perspective. *Nature Reviews Immunology*, 2(1):55–60, 2002.

[35] Selma Tuzlak, Anne S Dejean, Matteo Iannacone, Francisco J Quintana, Ari Waisman, Florent Ginhoux, Thomas Korn, and Burkhard Becher. Repositioning th cell polarization from single cytokines to complex help. *Nature immunology*, pages 1–8, 2021.

[36] Susan L Swain, K Kai McKinstry, and Tara M Strutt. Expanding roles for cd4+ t cells in immunity to viruses. *Nature Reviews Immunology*, 12(2):136–148, 2012.

[37] Christopher E Rudd, Alison Taylor, and Helga Schneider. Cd28 and ctla-4 coreceptor expression and signal transduction. *Immunological reviews*, 229(1):12–26, 2009.

[38] Shane Crotty. Follicular helper cd4 t cells (tfh). *Annual review of immunology*, 29:621–663, 2011.

[39] Matthew A Williams and Michael J Bevan. Effector and memory ctl differentiation. *Annu. Rev. Immunol.*, 25:171–192, 2007.

[40] Neil Q Tay, Debbie CP Lee, Yen Leong Chua, Nayana Prabhu, Nicholas RJ Gascoigne, and David M Kemeny. Cd40l expression allows cd8+ t cells to promote their own expansion and differentiation through dendritic cells. *Frontiers in immunology*, 8:1484, 2017.

[41] TWH Flinsenberg, L Spel, M Jansen, D Koning, C de Haar, M Plantinga, R Scholman, MM van Loenen, S Nierkens, L Boon, et al. Cognate cd4 t-cell licensing of dendritic cells heralds anti-cytomegalovirus cd8 t-cell immunity after human allogeneic umbilical cord blood transplantation. *Journal of virology*, 89(2):1058–1069, 2015.

[42] Stefanie Hoyer, Sabrina Prommersberger, Isabell A Pfeiffer, Beatrice Schuler-Thurner, Gerold Schuler, Jan Dörrie, and Niels Schaft. Concurrent interaction of dcs with cd4+ and cd8+ t cells improves secondary ctl expansion: It takes three to tango. *European journal of immunology*, 44(12):3543–3559, 2014.

[43] Stephan Halle, Olga Halle, and Reinhold Förster. Mechanisms and dynamics of t cell-mediated cytotoxicity in vivo. *Trends in immunology*, 38(6):432–443, 2017.

[44] De Dong, Lvqin Zheng, Jianquan Lin, Bailing Zhang, Yuwei Zhu, Ningning Li, Shuangyu Xie, Yuhang Wang, Ning Gao, and Zhiwei Huang. Structural basis of assembly of the human t cell receptor–cd3 complex. *Nature*, 573(7775):546–552, 2019.

[45] Emil H Palacios and Arthur Weiss. Function of the src-family kinases, lck and fyn, in t-cell development and activation. *Oncogene*, 23(48):7990–8000, 2004.

[46] Javier Casas, Joanna Brzostek, Veronika I Zarnitsyna, Jin-sung Hong, Qianru Wei, John AH Hoerter, Guo Fu, Jeanette Ampudia, Rose Zamoyska, Cheng Zhu, et al. Ligand-engaged tcr is triggered by lck not associated with cd8 coreceptor. *Nature communications*, 5(1):1–11, 2014.

[47] Rebecca J Brownlie and Rose Zamoyska. T cell receptor signalling networks: branched, diversified and bounded. *Nature Reviews Immunology*, 13(4):257, 2013.

[48] Adam H Courtney, Wan-Lin Lo, and Arthur Weiss. Tcr signaling: mechanisms of initiation and propagation. *Trends in biochemical sciences*, 43(2):108–123, 2018.

[49] Oreste Acuto and Frédérique Michel. Cd28-mediated co-stimulation: a quantitative support for tcr signalling. *Nature Reviews Immunology*, 3(12):939–951, 2003.

[50] Ardiyanto Liaunardy-Jopeace, Ben L Murton, Mohan Mahesh, Jason W Chin, and John R James. Encoding optical control in lck kinase to quantitatively investigate its activity in live cells. *Nature structural & molecular biology*, 24(12):1155, 2017.

[51] Ricardo Fernandes, Jiandong Huo, Yuan Lui, James Felce, and Simon John Davis. On the control of tcr phosphorylation. *Frontiers in immunology*, 3:92, 2012.

[52] Jeanne E Baker, Ravindra Majeti, Stuart G Tangye, and Arthur Weiss. Protein tyrosine phosphatase cd148-mediated inhibition of t-cell receptor signal transduction is associated with reduced lat and phospholipase cγ1 phosphorylation. *Molecular and cellular biology*, 21(7):2393–2403, 2001.

[53] Michelle L Hermiston, Zheng Xu, and Arthur Weiss. Cd45: a critical regulator of signaling thresholds in immune cells. *Annual review of immunology*, 21(1):107–137, 2003.

[54] Nick Holmes. Cd45: all is not yet crystal clear. *Immunology*, 117(2):145–155, 2006.

[55] Diana Gil, Wolfgang WA Schamel, Maria Montoya, Francisco Sánchez-Madrid, and Balbino Alarcón. Recruitment of nck by cd3ϵ reveals a ligand-induced conformational change essential for t cell receptor signaling and synapse formation. *Cell*, 109(7):901–912, 2002.

[56] Lisa A Pitcher and Nicolai SC Van Oers. T-cell receptor signal transmission: who gives an itam? *Trends in immunology*, 24(10):554–560, 2003.

[57] Jatuporn Ngoenkam, Wolfgang W Schamel, and Sutatip Pongcharoen. Selected signalling proteins recruited to the t-cell receptor–cd3 complex. *Immunology*, 153(1):42–50, 2018.

[58] Matthew L Bettini, Po-Chein Chou, Clifford S Guy, Thomas Lee, Kate M Vignali, and Dario AA Vignali. Cutting edge: Cd3 itam diversity is required for optimal tcr signaling and thymocyte development. *The Journal of Immunology*, 199(5):1555–1560, 2017.

[59] Steven C Neier, Alejandro Ferrer, Katelynn M Wilton, Stephen EP Smith, April MH Kelcher, Kevin D Pavelko, Jenna M Canfield, Tessa R Davis, Robert J Stiles, Zhenjun Chen, et al. The

early proximal $\alpha\beta$ tcr signalosome specifies thymic selection outcome through a quantitative protein interaction network. *Science immunology*, 4(32):eaal2201, 2019.

[60] John R James. Tuning itam multiplicity on t cell receptors can control potency and selectivity to ligand density. *Science signaling*, 11(531), 2018.

[61] P Anton Van Der Merwe and Omer Dushek. Mechanisms for t cell receptor triggering. *Nature Reviews Immunology*, 11(1):47–55, 2011.

[62] Roy A Mariuzza, Pragati Agnihotri, and John Orban. The structural basis of t-cell receptor (tcr) activation: An enduring enigma. *Journal of Biological Chemistry*, 295(4):914–925, 2020.

[63] Wolfgang W Schamel, Balbino Alarcon, and Susana Minguet. The tcr is an allosterically regulated macromolecular machinery changing its conformation while working. *Immunological reviews*, 291(1):8–25, 2019.

[64] Lars Kjer-Nielsen, Craig S Clements, Anthony W Purcell, Andrew G Brooks, James C Whisstock, Scott R Burrows, James McCluskey, and Jamie Rossjohn. A structural basis for the selection of dominant $\alpha\beta$ t cell receptors in antiviral immunity. *Immunity*, 18(1):53–64, 2003.

[65] Travis Beddoe, Zhenjun Chen, Craig S Clements, Lauren K Ely, Simon R Bushell, Julian P Vivian, Lars Kjer-Nielsen, Siew Siew Pang, Michelle A Dunstone, Yu Chih Liu, et al. Antigen ligation triggers a conformational change within the constant domain of the $\alpha\beta$ t cell receptor. *Immunity*, 30(6):777–788, 2009.

[66] Sneha Rangarajan, Yanan He, Yihong Chen, Melissa C Kerzic, Buyong Ma, Ragul Gowthaman, Brian G Pierce, Ruth Nussinov, Roy A Mariuzza, and John Orban. Peptide–mhc (pmhc) binding to a human antiviral t cell receptor induces long-range allosteric communication between pmhc-and cd3-binding sites. *Journal of Biological Chemistry*, 293(41):15991–16005, 2018.

[67] Michael S Kuhns, Mark M Davis, and K Christopher Garcia. Deconstructing the form and function of the tcr/cd3 complex. *Immunity*, 24(2):133–139, 2006.

[68] Lukas Sušac, Mai T Vuong, Christoph Thomas, Sören von Bülow, Caitlin O'Brien-Ball, Ana Mafalda Santos, Ricardo A Fernandes, Gerhard Hummer, Robert Tampé, and Simon J Davis. Structure of a fully assembled tumor-specific t cell receptor ligated by pmhc. *Cell*, 185(17):3201–3213, 2022.

[69] Susana Minguet, Mahima Swamy, Balbino Alarcón, Immanuel F Luescher, and Wolfgang WA Schamel. Full activation of the t cell receptor requires both clustering and conformational changes at cd3. *Immunity*, 26(1):43–54, 2007.

[70] Martin F Bachmann and Pamela S Ohashi. The role of t-cell receptor dimerization in t-cell activation. *Immunology today*, 20(12):568–576, 1999.

[71] Wolfgang WA Schamel, Ignacio Arechaga, Ruth M Risueño, Hisse M van Santen, Pilar Cabezas, Cristina Risco, José M Valpuesta, and Balbino Alarcón. Coexistence of multivalent and monovalent tcrs explains high sensitivity and wide range of response. *The Journal of experimental medicine*, 202(4):493–503, 2005.

[72] Pauline M Rudd, Mark R Wormald, Robyn L Stanfield, Mingdong Huang, Niklas Mattsson, Jeffrey A Speir, Jeannine A DiGennaro, Jacquelyn S Fetrow, Raymond A Dwek, and Ian A Wilson.

Roles for glycosylation of cell surface receptors involved in cellular immune recognition. *Journal of molecular biology*, 293(2):351–366, 1999.

[73] Jamey D Marth and Prabhjit K Grewal. Mammalian glycosylation in immunity. *Nature Reviews Immunology*, 8(11):874–887, 2008.

[74] Jia-Huai Wang and Ellis L Reinherz. Revisiting the putative tcr cα dimerization model through structural analysis. *Frontiers in immunology*, 4:16, 2013.

[75] Bozidar Purtic, Lisa A Pitcher, Nicolai SC Van Oers, and Christoph Wülfing. T cell receptor (tcr) clustering in the immunological synapse integrates tcr and costimulatory signaling in selected t cells. *Proceedings of the National Academy of Sciences*, 102(8):2904–2909, 2005.

[76] Wolfgang WA Schamel and Balbino Alarcón. Organization of the resting tcr in nanoscale oligomers. *Immunological reviews*, 251(1):13–20, 2013.

[77] Sophie V Pageon, Thibault Tabarin, Yui Yamamoto, Yuanqing Ma, Philip R Nicovich, John S Bridgeman, André Cohnen, Carola Benzing, Yijun Gao, Michael D Crowther, et al. Functional role of t-cell receptor nanoclusters in signal initiation and antigen discrimination. *Proceedings of the National Academy of Sciences*, 113(37):E5454–E5463, 2016.

[78] Wanli Liu, Haopeng Wang, and Chenqi Xu. Antigen receptor nanoclusters: small units with big functions. *Trends in immunology*, 37(10):680–689, 2016.

[79] Björn F Lillemeier, Manuel A Mörtelmaier, Martin B Forstner, Johannes B Huppa, Jay T Groves, and Mark M Davis. Tcr and lat are expressed on separate protein islands on t cell membranes and concatenate during activation. *Nature immunology*, 11(1):90, 2010.

[80] Katharina Beck-García, Esmeralda Beck-García, Sheila Bohler, Carina Zorzin, Erdinc Sezgin, Ilya Levental, Balbino Alarcón, and Wolfgang WA Schamel. Nanoclusters of the resting t cell antigen receptor (tcr) localize to non-raft domains. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1853(4):802–809, 2015.

[81] Peter W Janes, Steven C Ley, Anthony I Magee, and Panagiotis S Kabouridis. The role of lipid rafts in t cell antigen receptor (tcr) signalling. In *Seminars in immunology*, volume 12, pages 23–34. Elsevier, 2000.

[82] Pallavi Varshney, Vikas Yadav, and Neeru Saini. Lipid rafts in immune signalling: current progress and future perspective. *Immunology*, 149(1):13–24, 2016.

[83] Kristen Feher, James M Halstead, Jesse Goyette, and Katharina Gaus. Can single molecule localization microscopy detect nanoclusters in t cells? *Current opinion in chemical biology*, 51:130–137, 2019.

[84] Jesse Goyette, Daniel J Nieves, Yuanqing Ma, and Katharina Gaus. How does t cell receptor clustering impact on signal transduction? *J Cell Sci*, 132(4):jcs226423, 2019.

[85] John R James, James McColl, Marta I Oliveira, Paul D Dunne, Elizabeth Huang, Andreas Jansson, Patric Nilsson, David L Sleep, Carine M Gonçalves, Sara H Morgan, et al. The t cell receptor triggering apparatus is composed of monovalent or monomeric proteins. *Journal of Biological Chemistry*, 286(37):31993–32001, 2011.

[86] Mario Brameshuber, Florian Kellner, Benedikt K Rossboth, Haisen Ta, Kevin Alge, Eva Sevcsik, Janett Göhring, Markus Axmann, Florian Baumgart, Nicholas RJ Gascoigne, et al. Monomeric tcrs drive t cell antigen recognition. *Nature immunology*, 19(5):487–496, 2018.

[87] Benedikt Rossboth, Andreas M Arnold, Haisen Ta, René Platzer, Florian Kellner, Johannes B Huppa, Mario Brameshuber, Florian Baumgart, and Gerhard J Schütz. Tcrs are randomly distributed on the plasma membrane of resting antigen-experienced t cells. *Nature immunology*, 19(8):821–827, 2018.

[88] Darrell J Irvine, Marco A Purbhoo, Michelle Krogsgaard, and Mark M Davis. Direct observation of ligand recognition by t cells. *Nature*, 419(6909):845–849, 2002.

[89] Yuri Sykulev, Michael Joo, Irina Vturina, Theodore J Tsomides, and Herman N Eisen. Evidence that a single peptide–mhc complex on a target cell can elicit a cytolytic t cell response. *Immunity*, 4(6):565–571, 1996.

[90] P Anton van der Merwe. Do t cell receptors do it alone? *Nature immunology*, 3(12):1122–1123, 2002.

[91] Simon J Davis, Shinji Ikemizu, Edward J Evans, Lars Fugger, Talitha R Bakker, and P Anton van der Merwe. The nature of molecular recognition by t cells. *Nature immunology*, 4(3):217–224, 2003.

[92] Simon J Davis and P Anton van der Merwe. Tcr triggering: co-receptor-dependent or-independent? *Trends in immunology*, 24(12):624–626, 2003.

[93] Veronica T Chang, Ricardo A Fernandes, Kristina A Ganzinger, Steven F Lee, Christian Siebold, James McColl, Peter Jönsson, Matthieu Palayret, Karl Harlos, Charlotte H Coles, et al. Initiation of t cell signaling by cd45 segregation at'close contacts'. *Nature immunology*, 17(5):574, 2016.

[94] Simon J Davis and P Anton Van Der Merwe. The kinetic-segregation model: Tcr triggering and beyond. *Nature immunology*, 7(8):803–809, 2006.

[95] Shaun-Paul Cordoba, Kaushik Choudhuri, Hao Zhang, Marcus Bridge, Alp Bugra Basat, Michael L Dustin, and P Anton van der Merwe. The large ectodomains of cd45 and cd148 regulate their segregation from and inhibition of ligated t-cell receptor. *Blood, The Journal of the American Society of Hematology*, 121(21):4295–4302, 2013.

[96] Kaushik Choudhuri, David Wiseman, Marion H Brown, Keith Gould, and P Anton van der Merwe. T-cell receptor triggering is critically dependent on the dimensions of its peptide-mhc ligand. *Nature*, 436(7050):578–582, 2005.

[97] Jenny J Lin, Geoff P O'Donoghue, Kiera B Wilhelm, Michael P Coyle, Shalini T Low-Nam, Nicole C Fay, Katherine N Alfieri, and Jay T Groves. Membrane association transforms an inert anti-tcrβ fab'ligand into a potent t cell receptor agonist. *Biophysical Journal*, 2020.

[98] Fenglei Li and Kaushik Choudhuri. Small size of membrane-tethered ligands regulates phosphatase segregation in $\gamma\delta$ tcr triggering, 2017.

[99] James H Felce, Erdinc Sezgin, Madina Wane, Heather Brouwer, Michael L Dustin, Christian Eggeling, and Simon J Davis. Cd45 exclusion–and cross-linking–based receptor signaling together broaden fcεri reactivity. *Sci. Signal.*, 11(561):eaat0756, 2018.

[100] Matthew H Bakalar, Aaron M Joffe, Eva M Schmid, Sungmin Son, Marija Podolski, and Daniel A Fletcher. Size-dependent segregation controls macrophage phagocytosis of antibody-opsonized targets. *Cell*, 174(1):131–142, 2018.

[101] John R James and Ronald D Vale. Biophysical mechanism of t-cell receptor triggering in a reconstituted system. *Nature*, 487(7405):64, 2012.

[102] Wendell A Lim and Carl H June. The principles of engineering immune cells to treat cancer. *Cell*, 168(4):724–740, 2017.

[103] Gianpietro Dotti, Stephen Gottschalk, Barbara Savoldo, and Malcolm K Brenner. Design and development of therapies using chimeric antigen receptor-expressing t cells. *Immunological reviews*, 257(1):107–126, 2014.

[104] Kyeong-Hee Lee, Amy D Holdorf, Michael L Dustin, Andrew C Chan, Paul M Allen, and Andrey S Shaw. T cell receptor signaling precedes immunological synapse formation. *Science*, 295(5559):1539–1542, 2002.

[105] Balbino Alarcón, David Mestre, and Nuria Martínez-Martín. The immunological synapse: a cause or consequence of t-cell receptor triggering? *Immunology*, 133(4):420–425, 2011.

[106] Nele MG Dieckmann, Gordon L Frazer, Yukako Asano, Jane C Stinchcombe, and Gillian M Griffiths. The cytotoxic t lymphocyte immune synapse at a glance. *J Cell Sci*, 129(15):2881–2886, 2016.

[107] Sašo Čemerski, Jayajit Das, Emanuele Giurisato, Mary A Markiewicz, Paul M Allen, Arup K Chakraborty, and Andrey S Shaw. The balance between t cell receptor signaling and degradation at the center of the immunological synapse is determined by antigen quality. *Immunity*, 29(3):414–422, 2008.

[108] Alexandra Zanin-Zhorov, Michael L Dustin, and Bruce R Blazar. Pkc-θ and the immunological synapse: mechanisms and implications. *Trends in immunology*, 32(8):358, 2011.

[109] Enrique Calleja, Balbino Alarcón, and Clara L Oeste. Studying the dynamics of tcr internalization at the immune synapse. In *The Immune Synapse*, pages 89–99. Springer, 2017.

[110] Andrea L Szymczak and Dario AA Vignali. Plasticity and rigidity in adaptor protein-2-mediated internalization of the tcr: Cd3 complex. *The Journal of Immunology*, 174(7):4153–4160, 2005.

[111] Ugo D'Oro, Ivana Munitic, George Chacko, Tatiana Karpova, James McNally, and Jonathan D Ashwell. Regulation of constitutive tcr internalization by the ζ-chain. *The Journal of Immunology*, 169(11):6269–6278, 2002.

[112] Jens Peter H Lauritsen, Charlotte Menné Bonefeld, Marina von Essen, Martin Weiss Nielsen, Anette Bødker Rasmussen, Niels Ødum, Jes Dietrich, and Carsten Geisler. Masking of the cd3γ di-leucine-based motif by ζ is required for efficient t-cell receptor expression. *Traffic*, 5(9):672–684, 2004.

[113] Nuria Martínez-Martín, Elena Fernández-Arenas, Saso Cemerski, Pilar Delgado, Martin Turner, John Heuser, Darrell J Irvine, Bonnie Huang, Xosé R Bustelo, Andrey Shaw, et al. T cell receptor internalization from the immunological synapse is mediated by tc21 and rhog gtpase-dependent phagocytosis. *Immunity*, 35(2):208–222, 2011.

[114] Laura Patrussi and Cosima T Baldari. Analysis of tcr/cd3 recycling at the immune synapse. In *The Immune Synapse*, pages 143–155. Springer, 2017.

[115] Kathrin Pieper, Bodo Grimbacher, and Hermann Eibel. B-cell biology and development. *Journal of Allergy and Clinical Immunology*, 131(4):959–971, 2013.

[116] Min Zhang, Gopesh Srivastava, and Liwei Lu. The pre-b cell receptor and its function during b cell development. *Cell Mol Immunol*, 1(2):89–94, 2004.

[117] James B Chung, Michael Silverman, and John G Monroe. Transitional b cells: step by step towards immune competence. *Trends in immunology*, 24(6):342–348, 2003.

[118] By Florienne Loder, Bettina Mutschler, Robert J Ray, Christopher J Paige, Paschalis Sideras, Raul Torres, Marinus C Lamers, and Rita Carsetti. B cell development in the spleen takes place in discrete steps and is determined by the quality of b cell receptor–derived signals. *The Journal of experimental medicine*, 190(1):75–90, 1999.

[119] Andrea Cerutti, Montserrat Cols, and Irene Puga. Marginal zone b cells: virtues of innate-like antibody-producing lymphocytes. *Nature Reviews Immunology*, 13(2):118–132, 2013.

[120] Mark J Shlomchik and Florian Weisel. Germinal center selection and the development of memory b and plasma cells. *Immunological reviews*, 247(1):52–63, 2012.

[121] Willi K Born, M Kemal Aydintug, and Rebecca L O'brien. Diversity of $\gamma\delta$ t-cell antigens. *Cellular & molecular immunology*, 10(1):13–20, 2013.

[122] Myriam Lawand, Julie Déchanet-Merville, and Marie-Caroline Dieu-Nosjean. Key features of gamma-delta t-cell subsets in human diseases and their immunotherapeutic implications. *Frontiers in immunology*, 8:761, 2017.

[123] Bin Xu, Juan C Pizarro, Margaret A Holmes, Christine McBeth, Veronika Groh, Thomas Spies, and Roland K Strong. Crystal structure of a $\gamma\delta$ t-cell receptor specific for the human mhc class i homolog mica. *Proceedings of the National Academy of Sciences*, 108(6):2414–2419, 2011.

[124] Adrienne M Luoma, Caitlin D Castro, and Erin J Adams. $\gamma\delta$ t cell surveillance via cd1 molecules. *Trends in immunology*, 35(12):613–621, 2014.

[125] Bruno Silva-Santos, Karine Serre, and Håkan Norell. $\gamma\delta$ t cells in cancer. *Nature Reviews Immunology*, 15(11):683–691, 2015.

[126] Stefano Vavassori, Anil Kumar, Gan Siok Wan, Gundimeda S Ramanjaneyulu, Marco Cavallari, Sary El Daker, Travis Beddoe, Alex Theodossis, Neal K Williams, Emma Gostick, et al. Butyrophilin 3a1 binds phosphorylated antigens and stimulates human $\gamma\delta$ t cells. *Nature immunology*, 14(9):908, 2013.

[127] Benjamin E Willcox and Carrie R Willcox. $\gamma\delta$ tcr ligands: The quest to solve a 500-million-year-old mystery. *Nature immunology*, 20(2):121–128, 2019.

[128] Sourav Paul and Girdhari Lal. Regulatory and effector functions of gamma–delta ($\gamma\delta$) t cells and their therapeutic potential in adoptive cellular therapy for cancer. *International journal of cancer*, 139(5):976–985, 2016.

[129] Alex M Abel, Chao Yang, Monica Thakar, and Subramaniam Malarkannan. Nk cells: development, maturation, and clinical utilization. *Frontiers in immunology*, 9:1869, 2018.

[130] Michael A Caligiuri. Human natural killer cells. *Blood*, 112(3):461–469, 2008.

[131] Lucia Mori, Marco Lepore, and Gennaro De Libero. The immunology of cd1-and mr1-restricted t cells. *Annual review of immunology*, 34:479–510, 2016.

[132] Daniel G Pellicci, Hui-Fern Koay, and Stuart P Berzins. Thymic development of unconventional t cells: how nkt cells, mait cells and $\gamma\delta$ t cells emerge. *Nature Reviews Immunology*, pages 1–15, 2020.

[133] Marianthi Gioulbasani, Alexandros Galaras, Sofia Grammenoudi, Panagiotis Moulos, Alexander L Dent, Mikael Sigvardsson, Pantelis Hatzis, Barbara L Kee, and Mihalis Verykokakis. The transcription factor bcl-6 controls early development of innate-like t cells. *Nature Immunology*, pages 1–12, 2020.

[134] Melanie S Vacchio, Thomas Ciucci, and Rémy Bosselut. 200 million thymocytes and i: a beginner's survival guide to t cell development. In *T-Cell Development*, pages 3–21. Springer, 2016.

[135] Jacques FAP Miller. The golden anniversary of the thymus. *Nature Reviews Immunology*, 11(7):489–495, 2011.

[136] Izumi Ohigashi and Yousuke Takahama. Flow cytometry analysis of thymic epithelial cells and their subpopulations. In *T-Cell Development*, pages 65–73. Springer, 2016.

[137] Jan YM Lee and Paul E Love. Assessment of t cell development by flow cytometry. In *T-Cell Development*, pages 47–64. Springer, 2016.

[138] Ellen V Rothenberg. T cell lineage commitment: identity and renunciation. *The Journal of Immunology*, 186(12):6649–6655, 2011.

[139] Wooseok Seo and Ichiro Taniuchi. Transcriptional regulation of early t-cell development in the thymus. *European journal of immunology*, 46(3):531–538, 2016.

[140] Inge Van de Walle, Anne-Catherine Dolens, Kaat Durinck, Katrien De Mulder, Wouter Van Loocke, Sagar Damle, Els Waegemans, Jelle De Medts, Imke Velghe, Magda De Smedt, et al. Gata3 induces human t-cell commitment by restraining notch activity and repressing nk-cell fate. *Nature communications*, 7(1):1–14, 2016.

[141] Hao Yuan Kueh, Mary A Yui, Kenneth KH Ng, Shirley S Pease, Jingli A Zhang, Sagar S Damle, George Freedman, Sharmayne Siu, Irwin D Bernstein, Michael B Elowitz, et al. Asynchronous combinatorial action of four regulatory factors activates bcl11b for t cell commitment. *Nature immunology*, 17(8):956, 2016.

[142] Scott K Durum. Bcl11: sibling rivalry in lymphoid development. *Nature immunology*, 4(6):512–514, 2003.

[143] Dorina Avram and Danielle Califano. The multifaceted roles of bcl11b in thymic and peripheral t cells: impact on immune diseases. *The Journal of Immunology*, 193(5):2059–2065, 2014.

[144] Hiroyuki Hosokawa, Maile Romero-Wolf, Mary A Yui, Jonas Ungerbäck, Maria LG Quiloan, Masaki Matsumoto, Keiichi I Nakayama, Tomoaki Tanaka, and Ellen V Rothenberg. Bcl11b sets pro-t cell fate by site-specific cofactor recruitment and by repressing id2 and zbtb16. *Nature immunology*, 19(12):1427–1440, 2018.

[145] Renée F De Pooter and Barbara L Kee. E proteins and the regulation of early lymphocyte development. *Immunological reviews*, 238(1):93–109, 2010.

[146] Mary Elizabeth Jones and Yuan Zhuang. Stage-specific functions of e-proteins at the β-selection and t-cell receptor checkpoints during thymocyte development. *Immunologic research*, 49(1-3):202–215, 2011.

[147] Marie-Paule Lefranc and Gérard Lefranc. *The T cell receptor FactsBook*. Elsevier, 2001.

[148] Isaac Engel and Cornelis Murre. The function of e-and id proteins in lymphocyte development. *Nature reviews Immunology*, 1(3):193–199, 2001.

[149] Reetta Vanhanen, Nelli Heikkilä, Kunal Aggarwal, David Hamm, Heikki Tarkkila, Tommi Pätilä, T Sakari Jokiranta, Jari Saramäki, and T Petteri Arstila. T cell receptor diversity in the human thymus. *Molecular immunology*, 76:116–122, 2016.

[150] Janko Nikolich-Žugich, Mark K Slifka, and Ilhem Messaoudi. The many important facets of t-cell repertoire diversity. *Nature Reviews Immunology*, 4(2):123–132, 2004.

[151] Lars Kjer-Nielsen, Craig S Clements, Andrew G Brooks, Anthony W Purcell, James McCluskey, and Jamie Rossjohn. The 1.5 å crystal structure of a highly selected antiviral t cell receptor provides evidence for a structural basis of immunodominance. *Structure*, 10(11):1521–1532, 2002.

[152] Isaac Engel and Cornelis Murre. E2a proteins enforce a proliferation checkpoint in developing thymocytes. *The EMBO journal*, 23(1):202–211, 2004.

[153] Jason Wojciechowski, Anne Lai, Motonari Kondo, and Yuan Zhuang. E2a and heb are required to block thymocyte proliferation prior to pre-tcr expression. *The Journal of Immunology*, 178(9):5717–5726, 2007.

[154] Alfredo Pagliuca, Pasquale Gallo, Pasquale De Luca, and Luigi Lania. Class a helix-loop-helix proteins are positive regulators of several cyclin-dependent kinase inhibitors' promoter activity and negatively affect cell growth. *Cancer Research*, 60(5):1376–1382, 2000.

[155] Michael S Krangel. Mechanics of t cell receptor gene rearrangement. *Current opinion in immunology*, 21(2):133–139, 2009.

[156] Yuichi Wakabayashi, Hisami Watanabe, Jun Inoue, Naoki Takeda, Jun Sakata, Yukio Mishima, Jiro Hitomi, Takashi Yamamoto, Masanori Utsuyama, Ohtsura Niwa, et al. Bcl11b is required for differentiation and survival of $\alpha\beta$ t lymphocytes. *Nature immunology*, 4(6):533–539, 2003.

[157] Adrian C Hayday. $\gamma\delta$ cells: a right time and a right place for a conserved third way of protection. *Annual review of immunology*, 18(1):975–1026, 2000.

[158] Immo Prinz, Amandine Sansoni, Adrien Kissenpfennig, Laurence Ardouin, Marie Malissen, and Bernard Malissen. Visualization of the earliest steps of $\gamma\delta$ t cell development in the adult thymus. *Nature immunology*, 7(9):995–1003, 2006.

[159] Adrian C Hayday and Daniel J Pennington. Key factors in the organized chaos of early t cell development. *Nature immunology*, 8(2):137–144, 2007.

[160] Taras Kreslavsky and Harald von Boehmer. $\gamma\delta$tcr ligands and lineage commitment. In *Seminars in immunology*, volume 22(4), pages 214–221. Elsevier, 2010.

[161] Maria Ciofani and Juan Carlos Zúñiga-Pflücker. Determining $\gamma\delta$ versus $\alpha\beta$ t cell development. *Nature Reviews Immunology*, 10(9):657, 2010.

[162] Helen E Porritt, Lynn L Rumfelt, Sahba Tabrizifard, Thomas M Schmitt, Juan Carlos Zúñiga-Pflücker, and Howard T Petrie. Heterogeneity among dn1 prothymocytes reveals multiple progenitors with different capacities to generate t cell and non-t cell lineages. *Immunity*, 20(6):735–745, 2004.

[163] Joonsoo Kang, Ariane Volkmann, and David H Raulet. Evidence that $\gamma\delta$ versus $\alpha\beta$ t cell fate determination is initiated independently of t cell receptor signaling. *The Journal of experimental medicine*, 193(6):689–698, 2001.

[164] Mark S Schlissel, Scott D Durum, and Kathrin Muegge. The interleukin 7 receptor is required for t cell receptor γ locus accessibility to the v (d) j recombinase. *The Journal of experimental medicine*, 191(6):1045–1050, 2000.

[165] Heather J Melichar, Kavitha Narayan, Sandy D Der, Yoshiki Hiraoka, Noemie Gardiol, Gregoire Jeannet, Werner Held, Cynthia A Chambers, and Joonsoo Kang. Regulation of $\gamma\delta$ versus αß t lymphocyte differentiation by the transcription factor sox13. *Science*, 315(5809):230–233, 2007.

[166] Nicholas A Spidale, Katelyn Sylvia, Kavitha Narayan, Bing Miu, Michela Frascoli, Heather J Melichar, Wu Zhihao, Jan Kisielow, Amy Palin, Thomas Serwold, et al. Interleukin-17-producing $\gamma\delta$ t cells originate from sox13+ progenitors that are independent of $\gamma\delta$tcr signaling. *Immunity*, 49(5):857–872, 2018.

[167] Morgan E Parker and Maria Ciofani. Regulation of $\gamma\delta$ t cell effector diversification in the thymus. *Frontiers in Immunology*, 11, 2020.

[168] Gleb Turchinovich and Daniel J Pennington. T cell receptor signalling in $\gamma\delta$ cell development: strength isn't everything. *Trends in immunology*, 32(12):567–573, 2011.

[169] Kavitha Narayan and Joonsoo Kang. Disorderly conduct in $\gamma\delta$ versus $\alpha\beta$ t cell lineage commitment. In *Seminars in immunology*, volume 22-4, pages 222–227. Elsevier, 2010.

[170] Renée de Pooter and Juan Carlos Zúñiga-Pflücker. T-cell potential and development in vitro: the op9-dl1 approach. *Current opinion in immunology*, 19(2):163–168, 2007.

[171] Maria Ciofani, Gisèle C Knowles, David L Wiest, Harald von Boehmer, and Juan Carlos Zúñiga-Pflücker. Stage-specific and differential notch dependency at the $\alpha\beta$ and $\gamma\delta$ t lineage bifurcation. *Immunity*, 25(1):105–116, 2006.

[172] Taras Kreslavsky, Annette I Garbe, Andreas Krueger, and Harald von Boehmer. T cell receptor–instructed $\alpha\beta$ versus $\gamma\delta$ lineage commitment revealed by single-cell analysis. *The Journal of experimental medicine*, 205(5):1173–1186, 2008.

[173] Jens Peter H Lauritsen, Mariëlle C Haks, Juliette M Lefebvre, Dietmar J Kappes, and David L Wiest. Recent insights into the signals that control $\alpha\beta/\gamma\delta$-lineage fate. *Immunological reviews*, 209(1):176–190, 2006.

[174] Kathleen Terrence, Christian P Pavlovich, Errin O Matechak, and BJ Fowlkes. Premature expression of t cell receptor (tcr) $\alpha\beta$ suppresses tcr$\gamma\delta$ gene rearrangement but permits development of $\gamma\delta$ lineage t cells. *Journal of Experimental Medicine*, 192(4):537–548, 2000.

[175] Sandra M Hayes, LiQi Li, and Paul E Love. Tcr signal strength influences $\alpha\beta/\gamma\delta$ lineage fate. *Immunity*, 22(5):583–593, 2005.

[176] Mariëlle C Haks, Juliette M Lefebvre, Jens Peter H Lauritsen, Michael Carleton, Michele Rhodes, Toru Miyazaki, Dietmar J Kappes, and David L Wiest. Attenuation of $\gamma\delta$tcr signaling efficiently diverts thymocytes to the $\alpha\beta$ lineage. *Immunity*, 22(5):595–606, 2005.

[177] Payam Zarin, Gladys W Wong, Mahmood Mohtashami, David L Wiest, and Juan Carlos Zúñiga-Pflücker. Enforcement of $\gamma\delta$-lineage commitment by the pre–t-cell receptor in precursors with weak $\gamma\delta$-tcr signals. *Proceedings of the National Academy of Sciences*, 111(15):5658–5663, 2014.

[178] Seungyoul Oh, Xin Liu, Sara Tomei, Shalin H Naik, Daniel HD Gray, and Mark MW Chong. The decision to develop into $\alpha\beta$ or $\gamma\delta$ t cells is pre-programmed in distinct subpopulations of dn1 thymocytes. *bioRxiv*, 2022.

[179] Sara Scaramuzzino, Delphine Potier, Robin Ordioni, Pierre Grenot, Dominique Payet-Bornet, Hervé Luche, and Bernard Malissen. Single-cell transcriptomics uncovers an instructive t-cell receptor role in adult $\gamma\delta$ t-cell lineage commitment. *The EMBO journal*, 41(5):e110023, 2022.

[180] David L Wiest. Development of $\gamma\delta$ t cells, the special-force soldiers of the immune system. In *T-Cell Development*, pages 23–32. Springer, 2016.

[181] Andrea C Carpenter and Rémy Bosselut. Decision checkpoints in the thymus. *Nature immunology*, 11(8):666, 2010.

[182] Jens Peter Holst Lauritsen, Gladys W Wong, Sang-Yun Lee, Juliette M Lefebvre, Maria Ciofani, Michele Rhodes, Dietmar J Kappes, Juan Carlos Zúñiga-Pflücker, and David L Wiest. Marked induction of the helix-loop-helix protein id3 promotes the $\gamma\delta$ t cell fate and renders their functional maturation notch independent. *Immunity*, 31(4):565–575, 2009.

[183] Mai Xu, Archna Sharma, David L Wiest, and Jyoti Misra Sen. Pre-tcr-induced β-catenin facilitates traversal through β-selection. *The Journal of Immunology*, 182(2):751–758, 2009.

[184] Nicola J Rowbotham, Ariadne L Hager-Theodorides, Anna L Furmanski, Susan E Ross, Susan V Outram, Johannes T Dessens, and Tessa Crompton. Sonic hedgehog negatively regulates pre-tcr–induced differentiation by a gli2-dependent mechanism. *Blood, The Journal of the American Society of Hematology*, 113(21):5144–5156, 2009.

[185] Konstantinos Mengrelis, Ching-In Lau, Jasmine Rowell, Anisha Solanki, Sonia Norris, Susan Ross, Masahiro Ono, Susan Outram, and Tessa Crompton. Sonic hedgehog is a determinant of gamma delta t-cell differentiation in the thymus. *Frontiers in immunology*, 10:1629, 2019.

[186] Paul C Trampont, Annie-Carole Tosello-Trampont, Yuelei Shen, Amanda K Duley, Ann E Sutherland, Timothy P Bender, Dan R Littman, and Kodi S Ravichandran. Cxcr4 acts as a costimulator during thymic β-selection. *Nature immunology*, 11(2):162, 2010.

[187] Michelle L Janas and Martin Turner. Stromal cell-derived factor 1α and cxcr4: newly defined requirements for efficient thymic β-selection. *Trends in immunology*, 31(10):370–376, 2010.

[188] Amr H Allam, Mirren Charnley, Kim Pham, and Sarah M Russell. Developing t cells form an immunological synapse for passage through the β-selection checkpoint. *Journal of Cell Biology*, 220(3), 2021.

[189] Tom Taghon, Els Waegemans, and Inge Van de Walle. Notch signaling during human t cell development. In *Notch Regulation of the Immune System*, pages 75–97. Springer, 2012.

[190] Inge Van de Walle, Els Waegemans, Jelle De Medts, Greet De Smet, Magda De Smedt, Sylvia Snauwaert, Bart Vandekerckhove, Tessa Kerre, Georges Leclercq, Jean Plum, et al. Specific notch receptor–ligand interactions control human tcr-$\alpha\beta/\gamma\delta$ development by inducing differential notch signal strength. *Journal of Experimental Medicine*, 210(4):683–697, 2013.

[191] Taras Kreslavsky, Michael Gleimer, Masaki Miyazaki, Yoon Choi, Etienne Gagnon, Cornelis Murre, Piotr Sicinski, and Harald von Boehmer. β-selection-induced proliferation is required for $\alpha\beta$ t cell differentiation. *Immunity*, 37(5):840–853, 2012.

[192] Bin Zhao, Kogulan Yoganathan, LiQi Li, Jan Y Lee, Juan Carlos Zúñiga-Pflücker, and Paul E Love. Notch and the pre-tcr coordinate thymocyte proliferation by induction of the scf subunits fbxl1 and fbxl12. *Nature immunology*, 20(10):1381–1392, 2019.

[193] Sébastien Trop, Michele Rhodes, David L Wiest, Patrice Hugo, and Juan Carlos Zúñiga-Pflücker. Competitive displacement of ptα by tcr-α during tcr assembly prevents surface coexpression of pre-tcr and $\alpha\beta$ tcr. *The Journal of Immunology*, 165(10):5566–5572, 2000.

[194] Bruno Kyewski and Ludger Klein. A central role for central tolerance. *Annu. Rev. Immunol.*, 24:571–606, 2006.

[195] Hiroyuki Takaba and Hiroshi Takayanagi. The mechanisms of t cell selection in the thymus. *Trends in immunology*, 38(11):805–816, 2017.

[196] Shigeo Murata, Katsuhiro Sasaki, Toshihiko Kishimoto, Shin-ichiro Niwa, Hidemi Hayashi, Yousuke Takahama, and Keiji Tanaka. Regulation of cd8+ t cell development by thymus-specific proteasomes. *Science*, 316(5829):1349–1353, 2007.

[197] Jelena Nedjic, Martin Aichinger, Noboru Mizushima, and Ludger Klein. Macroautophagy, endogenous mhc ii loading and t cell selection: the benefits of breaking the rules. *Current opinion in immunology*, 21(1):92–97, 2009.

[198] Ludger Klein, Bruno Kyewski, Paul M Allen, and Kristin A Hogquist. Positive and negative selection of the t cell repertoire: what thymocytes see (and don't see). *Nature Reviews Immunology*, 14(6):377–391, 2014.

[199] Elise R Breed, S Thera Lee, and Kristin A Hogquist. Directing t cell fate: How thymic antigen presenting cells coordinate thymocyte selection. In *Seminars in cell & developmental biology*, volume 84, pages 2–10. Elsevier, 2018.

[200] Ronald N Germain. T-cell development and the cd4–cd8 lineage decision. *Nature reviews immunology*, 2(5):309–322, 2002.

[201] Andrea Itano and Ellen Robey. Highly efficient selection of cd4 and cd8 lineage thymocytes supports an instructive model of lineage commitment. *Immunity*, 12(4):383–389, 2000.

[202] Andrea Itano, Patrick Salmon, Dimitris Kioussis, Mauro Tolaini, Paola Corbella, and Ellen Robey. The cytoplasmic domain of cd4 promotes the development of cd4 lineage t cells. *The Journal of experimental medicine*, 183(3):731–741, 1996.

[203] Batu Erman, Amala S Alag, Oyvind Dahle, François van Laethem, Sophia D Sarafova, Terry I Guinter, Susan O Sharrow, Alexander Grinberg, Paul E Love, and Alfred Singer. Coreceptor signal

strength regulates positive selection but does not determine cd4/cd8 lineage choice in a physiologic in vivo model. *The journal of immunology*, 177(10):6613–6625, 2006.

[204] Paul E Love, Jan Lee, and Elizabeth W Shores. Critical relationship between tcr signaling potential and tcr affinity during thymocyte selection. *The Journal of Immunology*, 165(6):3080–3087, 2000.

[205] Alfred Singer, Stanley Adoro, and Jung-Hyun Park. Lineage fate and intense debate: myths, models and mechanisms of cd4-versus cd8-lineage choice. *Nature Reviews Immunology*, 8(10):788, 2008.

[206] Satoshi Kojo, Hirokazu Tanaka, Takaho A Endo, Sawako Muroi, Ye Liu, Wooseok Seo, Mari Tenno, Kiyokazu Kakugawa, Yoshinori Naoe, Krutula Nair, et al. Priming of lineage-specifying genes by bcl11b is required for lineage choice in post-selection thymocytes. *Nature communications*, 8(1):1–14, 2017.

[207] Reinhold Förster, Ana Clara Davalos-Misslitz, and Antal Rot. Ccr7 and its ligands: balancing immunity and tolerance. *Nature Reviews Immunology*, 8(5):362–371, 2008.

[208] Tejas Kadakia, Xuguang Tai, Michael Kruhlak, Jan Wisniewski, Il-Young Hwang, Sumedha Roy, Terry I Guinter, Amala Alag, John H Kehrl, Yuan Zhuang, et al. E-protein–regulated expression of cxcr4 adheres preselection thymocytes to the thymic cortex. *Journal of Experimental Medicine*, 216(8):1749–1761, 2019.

[209] Ivan Dzhagalov and Hyewon Phee. How to find your way through the thymus: a practical guide for aspiring t cells. *Cellular and Molecular Life Sciences*, 69(5):663–682, 2012.

[210] Geraldo A Passos, Cesar A Speck-Hernandez, Amanda F Assis, and Daniella A Mendes-da Cruz. Update on aire and thymic negative selection. *Immunology*, 153(1):10–20, 2018.

[211] Ming O Li and Alexander Y Rudensky. T cell receptor signalling in the control of regulatory t cell differentiation and function. *Nature Reviews Immunology*, 16(4):220–233, 2016.

[212] Timothy K Starr, Stephen C Jameson, and Kristin A Hogquist. Positive and negative selection of t cells. *Annual review of immunology*, 21(1):139–176, 2003.

[213] Tom M McCaughtry, Matthew S Wilken, and Kristin A Hogquist. Thymic emigration revisited. *Journal of Experimental Medicine*, 204(11):2513–2520, 2007.

[214] Julie D Saba. The low down on sphingosine-1-phosphate lyase as a regulator of thymic egress. *Journal of immunological sciences*, 1(1):1, 2017.

[215] Alexey I Nesvizhskii, Andrew Keller, Eugene Kolker, and Ruedi Aebersold. A statistical model for identifying proteins by tandem mass spectrometry. *Analytical chemistry*, 75(17):4646–4658, 2003.

[216] Michael S Kuhns and Hemant B Badgandi. Piecing together the family portrait of tcr-cd 3 complexes. *Immunological reviews*, 250(1):120–143, 2012.

[217] Marcus Groettrup, Katharina Ungewiss, Orly Azogui, Ronald Palacios, Michael J Owen, Adrian C Hayday, and Harald von Boehmer. A novel disulfide-linked heterodimer on pre—t cells consists of the t cell receptor β chain and a 33 kd glycoprotein. *Cell*, 75(2):283–294, 1993.

[218] Claude Saint-Ruf, Katharina Ungewiss, Marcus Groettrup, Ludovica Bruno, Hans Joerg Fehling, and Harald von Boehmer. Analysis and expression of a cloned pre-t cell receptor gene. *Science*, 266(5188):1208–1212, 1994.

[219] Domingo F Barber, Lorena Passoni, Li Wen, Liping Geng, and Adrian C Hayday. Cutting edge: the expression in vivo of a second isoform of ptα: implications for the mechanism of ptα action. *The Journal of Immunology*, 161(1):11–16, 1998.

[220] Claude Saint-Ruf, Oskar Lechner, Jacqueline Feinberg, and Harald von Boehmer. Genomic structure of the human pre-t cell receptor α chain and expression of two mrna isoforms. *European journal of immunology*, 28(11):3824–3831, 1998.

[221] Antonio F Campese, Paola Grazioli, Sara Colantoni, Emanuela Anastasi, Marco Mecarozzi, Saula Checquolo, Gabriele De Luca, Diana Bellavia, Luigi Frati, Alberto Gulino, et al. Notch3 and ptα/pre-tcr sustain the in vivo function of naturally occurring regulatory t cells. *International immunology*, 21(6):727–743, 2009.

[222] Juliet L Mahtani-Patching. *Investigating the roles of the alternative Isoforms of the preTCR alpha (pTα) chain in T cell development.* PhD thesis, University of London, 2010.

[223] Hans Jörg Fehling, Anna Krotkova, Claude Saint-Ruf, and Harald von Boehmer. Crucial role of the pre-t-cell receptor α gene in development of ap but not $\gamma\delta$ t cells. *Nature*, 375(6534):795–798, 1995.

[224] Peter Mombaerts, Alan R Clarke, Michael A Rudnicki, John Iacomini, Shigeyoshi Itohara, Juan J Lafaille, Lili Wang, Yoshiaki Ichikawa, Rudolf Jaenisch, Martin L Hooper, et al. Mutations in t-cell antigen receptor genes α and β block thymocyte development at different stages. *Nature*, 360(6401):225–231, 1992.

[225] Marie Malissen, Anne Gillet, Laurence Ardouin, Gaelle Bouvier, Jeannine Trucy, Pierre Ferrier, Eric Vivier, and Bernard Malissen. Altered t cell development in mice with a targeted mutation of the cd3-epsilon gene. *The EMBO journal*, 14(19):4641–4653, 1995.

[226] Peter Mombaerts, John Iacomini, Randall S Johnson, Karl Herrup, Susumu Tonegawa, and Virginia E Papaioannou. Rag-1-deficient mice have no mature b and t lymphocytes. *Cell*, 68(5):869–877, 1992.

[227] Jan Buer, Iannis Aifantis, James P DiSanto, Hans Joerg Fehling, and Harald von Boehmer. Role of different t cell receptors in the development of pre–t cells. *The Journal of experimental medicine*, 185(9):1541–1548, 1997.

[228] Iannis Aifantis, Jan Buer, Harald von Boehmer, and Orly Azogui. Essential role of the pre-t cell receptor in allelic exclusion of the t cell receptor β locus. *Immunity*, 7(5):601–607, 1997.

[229] HJ Fehling, BM Iritani, A Krotkova, KA Forbush, C Laplace, RM Perlmutter, and H von Boehmer. Restoration of thymopoiesis in ptaą/ą mice by anti-cd3e antibody treatment or with transgenes encoding activated lck or tailless pta. *Immunity*, 6, 1997.

[230] Deena Gibbons, Nataki C Douglas, Domingo F Barber, Qiang Liu, Renee Sullo, Liping Geng, Hans-Joerg Fehling, Harald von Boehmer, and Adrian C Hayday. The biological activity of natural and mutant ptα alleles. *Journal of Experimental Medicine*, 194(5):695–704, 2001.

[231] Christine Borowski, Xiaoyan Li, Iannis Aifantis, Fotini Gounari, and Harald Von Boehmer. Pre-tcrα and tcrα are not interchangeable partners of tcrβ during t lymphocyte development. *Journal of Experimental Medicine*, 199(5):607–615, 2004.

[232] Harald Von Boehmer. Unique features of the pre-t-cell receptor α-chain: not just a surrogate. *Nature Reviews Immunology*, 5(7):571, 2005.

[233] Harald von Boehmer and Hans Jörg Fehling. Structure and function of the pre-t cell receptor. *Annual review of immunology*, 15(1):433–452, 1997.

[234] Nigel Killeen. A new spin on an unusual tail. *Nature immunology*, 3(5):419–421, 2002.

[235] Philippe Smelty, Céline Marchal, Romain Renard, Ludivine Sinzelle, Nicolas Pollet, Dominique Dunon, Thierry Jaffredo, Jean-Yves Sire, and Julien S Fellah. Identification of the pre–t-cell receptor α chain in nonmammalian vertebrates challenges the structure–function of the molecule. *Proceedings of the National Academy of Sciences*, 107(46):19991–19996, 2010.

[236] Iannis Aifantis, Christine Borowski, Fotini Gounari, H Daniel Lacorazza, Janko Nikolich-Zugich, and Harald von Boehmer. A critical role for the cytoplasmic tail of ptα in t lymphocyte development. *Nature immunology*, 3(5):483, 2002.

[237] Yolanda R Carrasco, Almudena R Ramiro, César Trigueros, Virginia G de Yébenes, Marina García-Peydró, and María L Toribio. An endoplasmic reticulum retention function for the cytoplasmic tail of the human pre–t cell receptor (tcr) α chain: potential role in the regulation of cell surface pre-tcr expression levels. *Journal of Experimental Medicine*, 193(9):1045–1058, 2001.

[238] Yolanda R Carrasco, Maria N Navarro, and María L Toribio. A role for the cytoplasmic tail of the pre-t cell receptor (tcr) α chain in promoting constitutive internalization and degradation of the pre-tcr. *Journal of Biological Chemistry*, 278(16):14507–14513, 2003.

[239] Ivan Dikic. Cin85/cms family of adaptor molecules. *FEBS letters*, 529(1):110–115, 2002.

[240] Baerbel Keller, Moneef Shoukier, Kathrin Schulz, Arshiya Bhatt, Ines Heine, Valentina Strohmeier, Carsten Speckmann, Niklas Engels, Klaus Warnatz, and Jürgen Wienands. Germline deletion of cin85 in humans with x chromosome–linked antibody deficiencycin85 as novel candidate gene for x-linked antibody deficiency. *The Journal of experimental medicine*, 215(5):1327–1336, 2018.

[241] María N Navarro, Gretel Nusspaumer, Patricia Fuentes, Sara González-García, Juan Alcain, and María L Toribio. Identification of cms as a cytosolic adaptor of the human ptα chain involved in pre-tcr function. *Blood*, 110(13):4331–4340, 2007.

[242] Julyun Oh, Dora Toledo Warshaviak, Mikayel Mkrtichyan, Melanie Lisette Munguia, Abby Lin, Falene Chai, Craig Pigott, Jaspal Kang, Michael Gallo, and Alexander Kamb. Single variable domains from the t cell receptor β chain function as mono-and bifunctional cars and tcrs. *Scientific reports*, 9(1):1–12, 2019.

[243] Siew Siew Pang, Richard Berry, Zhenjun Chen, Lars Kjer-Nielsen, Matthew A Perugini, Glenn F King, Christina Wang, Sock Hui Chew, Nicole L La Gruta, Neal K Williams, et al. The structural basis for autonomous dimerization of the pre-t-cell antigen receptor. *Nature*, 467(7317):844, 2010.

[244] Christiaan N Levelt and Klaus Eichmann. Receptors and signals in early thymic selection. *Immunity*, 3(6):667–672, 1995.

[245] Bernard Malissen and Marie Malissen. Functions of tcr and pre-tcr subunits: lessons from gene ablation. *Current opinion in immunology*, 8(3):383–393, 1996.

[246] David L Wiest and Michael Carleton. Control of early thymocyte development by the pre-t cell receptor complex: A receptor without a ligand? In *Seminars in immunology*, volume 11-4, pages 251–262. Elsevier, 1999.

[247] Yolanda R Carrasco, Maria N Navarro, Virginia G de Yébenes, Almudena R Ramiro, and María L Toribio. Regulation of surface expression of the human pre-t cell receptor complex. In *Seminars in immunology*, volume 14-5, pages 325–334. Elsevier, 2002.

[248] Marc A Berger, Michael Carleton, Michele Rhodes, J Michael Sauder, Sébastien Trop, Roland L Dunbrack, Patrice Hugo, and David L Wiest. Identification of a novel pre-tcr isoform in which the accessibility of the tcrβ subunit is determined by occupancy of themissing'v domain of pre-tα. *International immunology*, 12(11):1579–1591, 2000.

[249] Elizabeth Chi-Fang Huang. *Organisation of, and ligand-independent signalling by the TCR, with a special emphasis on the pre-TCR.* PhD thesis, University of Oxford, 2017.

[250] Bo Zhou, Qiang Chen, Robert Mallis, Hongmin Zhang, Jin-huan Liu, Ellis L Reinherz, and Jia-huai Wang. A conserved hydrophobic patch on vβ domains revealed by tcrβ chain crystal structures: Implications for pre-tcr dimerization. *Frontiers in immunology*, 2:5, 2011.

[251] Avik Dutta, Bin Zhao, and Paul E Love. New insights into tcr β-selection. *Trends in Immunology*, 2021.

[252] Robert J Mallis, Ke Bai, Haribabu Arthanari, Rebecca E Hussey, Maris Handley, Zhenhai Li, Loice Chingozha, Jonathan S Duke-Cohan, Hang Lu, Jia-Huai Wang, et al. Pre-tcr ligand binding impacts thymocyte development before $\alpha\beta$tcr expression. *Proceedings of the National Academy of Sciences*, 112(27):8373–8378, 2015.

[253] Robert J Mallis, Haribabu Arthanari, Matthew J Lang, Ellis L Reinherz, and Gerhard Wagner. Nmr-directed design of pre-tcrβ and pmhc molecules implies a distinct geometry for pre-tcr relative to $\alpha\beta$tcr recognition of pmhc. *Journal of Biological Chemistry*, 293(3):754–766, 2018.

[254] Dibyendu Kumar Das, Robert J Mallis, Jonathan S Duke-Cohan, Rebecca E Hussey, Paul W Tetteh, Mark Hilton, Gerhard Wagner, Matthew J Lang, and Ellis L Reinherz. Pre-t cell receptors (pre-tcrs) leverage vβ complementarity determining regions (cdrs) and hydrophobic patch in mechanosensing thymic self-ligands. *Journal of Biological Chemistry*, 291(49):25292–25305, 2016.

[255] Xiaolong Li, Réka Mizsei, Kemin Tan, Robert J Mallis, Jonathan S Duke-Cohan, Aoi Akitsu, Paul W Tetteh, Abhinav Dubey, Wonmuk Hwang, Gerhard Wagner, et al. Pre–t cell receptors topologically sample self-ligands during thymocyte β-selection. *Science*, 371(6525):181–185, 2021.

[256] K Christopher Garcia, Massimo Degano, Larry R Pease, Mingdong Huang, Per A Peterson, Luc Teyton, and Ian A Wilson. Structural basis of plasticity in t cell receptor recognition of a self peptide-mhc antigen. *Science*, 279(5354):1166–1172, 1998.

[257] J Mahtani-Patching, JF Neves, DJ Pang, KV Stoenchev, AM Aguirre-Blanco, B Silva-Santos, and DJ Pennington. Pretcr and tcrgd signal initiation in thymocyte progenitors does not require domains implicated in receptor oligomerization. *Sci. Signal*, 4, 2011.

[258] Jannie Borst, Heinz Jacobs, and Gaby Brouns. Composition and function of t-cell receptor and b-cell receptor complexes on precursor lymphocytes. *Current opinion in immunology*, 8(2):181–190, 1996.

[259] Marc A Berger, Vibhuti Davé, Michele R Rhodes, Gayle C Bosma, Melvin J Bosma, Dietmar J Kappes, and David L Wiest. Subunit composition of pre–t cell receptor complexes expressed by primary thymocytes: Cd3δ is physically associated but not functionally required. *The Journal of experimental medicine*, 186(9):1461–1467, 1997.

[260] Vibhuti P Dave, Zhensheng Cao, Carol Browne, Balbino Alarcon, Gemma Fernandez-Miguel, Juan Lafaille, Antonio de la Hera, Susumu Tonegawa, and Dietmar J Kappes. Cd3δ deficiency arrests development of the $\alpha\beta$ but not the $\gamma\delta$ t cell lineage. *The EMBO Journal*, 16(6):1360–1370, 1997.

[261] Mariëlle C Haks, Paul Krimpenfort, Jannie Borst, and Ada M Kruisbeek. The cd3γ chain is essential for development of both the tcr$\alpha\beta$ and tcr$\gamma\delta$ lineages. *The EMBO journal*, 17(7):1871–1882, 1998.

[262] Maddalena Panigada, Simona Porcellini, Eliane Barbier, Sonja Hoeflinger, Pierre-André Cazenave, Hua Gu, Hamid Band, Harald Von Boehmer, and Fabio Grassi. Constitutive endocytosis and degradation of the pre-t cell receptor. *Journal of Experimental Medicine*, 195(12):1585–1597, 2002.

[263] Alison M Michie and Juan Carlos Zúñiga-Pflücker. Regulation of thymocyte differentiation: pre-tcr signals and β-selection. In *Seminars in immunology*, volume 14-5, pages 311–323. Elsevier, 2002.

[264] Claude Saint-Ruf, Maddalena Panigada, Orly Azogui, Pascale Debey, Harald Von Boehmer, and Fabio Grassi. Different initiation of pre-tcr and $\gamma\delta$tcr signalling. *Nature*, 406(6795):524–527, 2000.

[265] Mariëlle C Haks, Stanley M Belkowski, Maria Ciofani, Michele Rhodes, Juliette M Lefebvre, Sebastién Trop, Patrice Hugo, Juan Carlos Zúñiga-Pflücker, and David L Wiest. Low activation threshold as a mechanism for ligand-independent signaling in pre-t cells. *The Journal of Immunology*, 170(6):2853–2861, 2003.

[266] Sho Yamasaki, Eri Ishikawa, Machie Sakuma, Koji Ogata, Kumiko Sakata-Sogawa, Michio Hiroshima, David L Wiest, Makio Tokunaga, and Takashi Saito. Mechanistic basis of pre–t cell receptor–mediated autonomous signaling critical for thymocyte development. *Nature immunology*, 7(1):67–75, 2006.

[267] Sho Yamasaki and Takashi Saito. Molecular basis for pre-tcr-mediated autonomous signaling. *Trends in immunology*, 28(1):39–43, 2007.

[268] Eri Ishikawa, Yasunobu Miyake, Hiromitsu Hara, Takashi Saito, and Sho Yamasaki. Germ-line elimination of electric charge on pre–t-cell receptor (tcr) impairs autonomous signaling for β-selection and tcr repertoire formation. *Proceedings of the National Academy of Sciences*, 107(46):19979–19984, 2010.

[269] Richard Berry, Zhenjun Chen, James McCluskey, and Jamie Rossjohn. Insight into the basis of autonomous immunoreceptor activation. *Trends in immunology*, 32(4):165–170, 2011.

[270] Emil H Palacios and Arthur Weiss. Distinct roles for syk and zap-70 during early thymocyte development. *The Journal of experimental medicine*, 204(7):1703–1715, 2007.

[271] Andrei Iosif Smid. *Study of the intracellular trafficking of the pre-T-cell receptor and its role in T-cell lineage commitment*. PhD thesis, University of Cambridge, 2019.

[272] Emily Crowley, Euphemia Leung, Jóhannes Reynisson, and Alan Richardson. Rapid changes in the atg5-atg16l1 complex following nutrient deprivation measured using nanoluc binary technology (nanobit). *The FEBS Journal*, 2020.

[273] Yuan-Li Tsai, Dat P Ha, He Zhao, Anthony J Carlos, Shan Wei, Tsam Kiu Pun, Kaijin Wu, Ebrahim Zandi, Kevin Kelly, and Amy S Lee. Endoplasmic reticulum stress activates src, relocating chaperones to the cell surface where grp78/cd109 blocks tgf-β signaling. *Proceedings of the National Academy of Sciences*, 115(18):E4245–E4254, 2018.

[274] David L Wiest, Avinash Bhandoola, Jennifer Punt, Gert Kreibich, David McKean, and Alfred Singer. Incomplete endoplasmic reticulum (er) retention in immature thymocytes as revealed by surface expression of "er-resident" molecular chaperones. *Proceedings of the National Academy of Sciences*, 94(5):1884–1889, 1997.

[275] Johanna S Rees, Xue-Wen Li, Sarah Perrett, Kathryn S Lilley, and Antony P Jackson. Protein neighbors and proximity proteomics. *Molecular & Cellular Proteomics*, 14(11):2848–2856, 2015.

[276] Kyle J Roux, Dae In Kim, and Brian Burke. Bioid: a screen for protein-protein interactions. *Current protocols in protein science*, 74(1):19–23, 2013.

[277] Laura Trinkle-Mulcahy. Recent advances in proximity-based labeling methods for interactome mapping. *F1000Research*, 8, 2019.

[278] Dae In Kim, Samuel C Jensen, Kyle A Noble, Birendra KC, Kenneth H Roux, Khatereh Motamedchaboki, and Kyle J Roux. An improved smaller biotin ligase for bioid proximity labeling. *Molecular biology of the cell*, 27(8):1188–1196, 2016.

[279] Chiao-Lin Chen and Norbert Perrimon. Proximity-dependent labeling methods for proteomic profiling in living cells. *Wiley Interdisciplinary Reviews: Developmental Biology*, 6(4):e272, 2017.

[280] Victoria Hung, Namrata D Udeshi, Stephanie S Lam, Ken H Loh, Kurt J Cox, Kayvon Pedram, Steven A Carr, and Alice Y Ting. Spatially resolved proteomic mapping in living cells with the engineered peroxidase apex2. *Nature protocols*, 11(3):456, 2016.

[281] Hyun-Woo Rhee, Peng Zou, Namrata D Udeshi, Jeffrey D Martell, Vamsi K Mootha, Steven A Carr, and Alice Y Ting. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*, 339(6125):1328–1331, 2013.

[282] Stephanie S Lam, Jeffrey D Martell, Kimberli J Kamer, Thomas J Deerinck, Mark H Ellisman, Vamsi K Mootha, and Alice Y Ting. Directed evolution of apex2 for electron microscopy and proximity labeling. *Nature methods*, 12(1):51–54, 2015.

[283] Braden T Lobingier, Ruth Hüttenhain, Kelsie Eichel, Kenneth B Miller, Alice Y Ting, Mark von Zastrow, and Nevan J Krogan. An approach to spatiotemporally resolve protein interaction networks in living cells. *Cell*, 169(2):350–360, 2017.

[284] Johanna Susan Rees, Xue-Wen Li, Sarah Perrett, Kathryn Susan Lilley, and Antony Philip Jackson. Selective proteomic proximity labeling assay using tyramide (spplat): a quantitative method for the proteomic analysis of localized membrane-bound protein clusters. *Current protocols in protein science*, 80(1):19–27, 2015.

[285] Eszter Zavodszky and Ramanujan S Hegde. Misfolded gpi-anchored proteins are escorted through the secretory pathway by er-derived factors. *Elife*, 8:e46740, 2019.

[286] Ziqing Liu, Olivia Chen, J Wall, Michael Zheng, Yang Zhou, Li Wang, Haley Ruth Vaseghi, Li Qian, and Jiandong Liu. Systematic comparison of 2a peptides for cloning multi-genes in a polycistronic vector. *Scientific reports*, 7(1):1–9, 2017.

[287] Jean K Lee, Guillaume Stewart-Jones, Tao Dong, Karl Harlos, Kati Di Gleria, Lucy Dorrell, Daniel C Douek, P Anton Van Der Merwe, E Yvonne Jones, and Andrew J McMichael. T cell cross-reactivity and conformational changes during tcr engagement. *The Journal of experimental medicine*, 200(11):1455–1466, 2004.

[288] Yi Li, Ruth Moysey, Peter E Molloy, Anne-Lise Vuidepot, Tara Mahon, Emma Baston, Steven Dunn, Nathaniel Liddy, Jansen Jacob, Bent K Jakobsen, et al. Directed evolution of human t-cell receptors with picomolar affinities by phage display. *Nature biotechnology*, 23(3):349–354, 2005.

[289] Kyle J Roux, Dae In Kim, Brian Burke, and Danielle G May. Bioid: a screen for protein-protein interactions. *Current protocols in protein science*, 91(1):19–23, 2018.

[290] Dattatreya Mellacheruvu, Zachary Wright, Amber L Couzens, Jean-Philippe Lambert, Nicole A St-Denis, Tuo Li, Yana V Miteva, Simon Hauri, Mihaela E Sardiu, Teck Yew Low, et al. The crapome: a contaminant repository for affinity purification–mass spectrometry data. *Nature methods*, 10(8):730–736, 2013.

[291] Wenjuan Wang, Tianqi Li, Klara Felsovalyi, Chunlai Chen, Timothy Cardozo, and Michelle Krogsgaard. Quantitative analysis of t cell receptor complex interaction sites using genetically encoded photo-cross-linkers. *ACS chemical biology*, 9(9):2165–2172, 2014.

[292] Matthew E Call and Kai W Wucherpfennig. Molecular mechanisms for the assembly of the t cell receptor–cd3 complex. *Molecular immunology*, 40(18):1295–1305, 2004.

[293] Matthias J Feige, Julia Behnke, Tanja Mittag, and Linda M Hendershot. Dimerization-dependent folding underlies assembly control of the clonotypic $\alpha\beta$t cell receptor chains. *Journal of Biological Chemistry*, 290(44):26821–26831, 2015.

[294] Gaelle Boncompain, Severine Divoux, Nelly Gareil, Helene De Forges, Aurianne Lescure, Lynda Latreche, Valentina Mercanti, Florence Jollivet, Graça Raposo, and Franck Perez. Synchronization of secretory protein traffic in populations of cells. *Nature methods*, 9(5):493–498, 2012.

[295] Joel Eggert and Byron B Au-Yeung. Functional heterogeneity and adaptation of naive t cells in response to tonic tcr signals. *Current Opinion in Immunology*, 73:43–49, 2021.

[296] Luke Gabriel, Zachary Stevens, and Haley Melikian. Measuring plasma membrane protein endocytic rates by reversible biotinylation. *JoVE (Journal of Visualized Experiments)*, 34:e1669, 2009.

[297] Timothy J Allison, Christine C Winter, Jean-Jacques Fournie, Marc Bonneville, and David N Garboczi. Structure of a human $\gamma\delta$ t-cell antigen receptor. *Nature*, 411(6839):820–824, 2001.

[298] Frederik Lindberg, JM Tennent, SJ Hultgren, B Lund, and STAFFAN Normark. Papd, a periplasmic transport protein in p-pilus biogenesis. *Journal of bacteriology*, 171(11):6052–6058, 1989.

[299] Frederic G Sauer, Klaus Fütterer, Jerome S Pinkner, Karen W Dodson, Scott J Hultgren, and Gabriel Waksman. Structural basis of chaperone function and pilus biogenesis. *Science*, 285(5430):1058–1061, 1999.

[300] Michelle M Barnhart, Jerome S Pinkner, Gabriel E Soto, Frederic G Sauer, Sol Langermann, Gabriel Waksman, Carl Frieden, and Scott J Hultgren. Papd-like chaperones provide the missing

information for folding of pilin proteins. *Proceedings of the National Academy of Sciences*, 97(14):7709–7714, 2000.

[301] Jue Ruan, Heng Li, Zhongzhong Chen, Avril Coghlan, Lachlan James M Coin, Yiran Guo, Jean-Karim Heriche, Yafeng Hu, Karsten Kristiansen, Ruiqiang Li, et al. Treefam: 2008 update. *Nucleic acids research*, 36(suppl_1):D735–D740, 2007.

[302] Stéphane Guindon, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of phyml 3.0. *Systematic biology*, 59(3):307–321, 2010.

[303] Michael Mingueneau, Taras Kreslavsky, Daniel Gray, Tracy Heng, Richard Cruse, Jeffrey Ericson, Sean Bendall, Matthew H Spitzer, Garry P Nolan, Koichi Kobayashi, et al. The transcriptional landscape of $\alpha\beta$ t cell differentiation. *Nature immunology*, 14(6):619–632, 2013.

[304] Stephanie Vargas Aguilar, Oscar Aguilar, Rhys Allan, El Ad David Amir, Veronique Angeli, Maxim N Artyomov, Natasha Asinovski, Jilian Astarita, K Frank Austen, Geetika Bajpai, et al. Immgen at 15. *NATURE IMMUNOLOGY*, 21(7):700–703, 2020.

[305] C Chace Tydell, Elizabeth-Sharon David-Fung, Jonathan E Moore, Lee Rowen, Tom Taghon, and Ellen V Rothenberg. Molecular dissection of prethymic progenitor entry into the t lymphocyte developmental pathway. *The Journal of Immunology*, 179(1):421–438, 2007.

[306] Hemant Sahni, Susan Ross, Alessandro Barbarulo, Anisha Solanki, Ching-In Lau, Anna Furmanski, José Ignacio Saldaña, Masahiro Ono, Mike Hubank, Martino Barenco, et al. A genome wide transcriptional model of the complex response to pre-tcr signalling during thymocyte differentiation. *Oncotarget*, 6(30):28646, 2015.

[307] Kristi Kerkel, Nicole Schupf, Kota Hatta, Deborah Pang, Martha Salas, Alexander Kratz, Mark Minden, Vundavalli Murty, Warren B Zigman, Richard P Mayeux, et al. Altered dna methylation in leukocytes with trisomy 21. *PLoS Genet*, 6(11):e1001212, 2010.

[308] Nesrine Maharzi, Véronique Parietti, Elisabeth Nelson, Simona Denti, Macarena Robledo-Sarmiento, Niclas Setterblad, Aude Parcelier, Marika Pla, François Sigaux, Jean Claude Gluckman, et al. Identification of tmem131l as a novel regulator of thymocyte proliferation in humans. *The Journal of Immunology*, 190(12):6187–6197, 2013.

[309] Sébastien Szuplewski, Nesrine Maharzi, Elisabeth Nelson, Kutaiba Alhaj Hussen, Bernard Mignotte, Isabelle Guénal, and Bruno Canque. Evolutionary conservation of notch signaling inhibition by tmem131l overexpression. *Biochemical and Biophysical Research Communications*, 486(4):909–915, 2017.

[310] Zhe Zhang, Meirong Bai, Guilherme Oliveira Barbosa, Andrew Chen, Yuehua Wei, Shuo Luo, Xin Wang, Bingying Wang, Tatsuya Tsukui, Hao Li, et al. Broadly conserved roles of tmem131 family proteins in intracellular collagen assembly and secretory cargo trafficking. *Science advances*, 6(7):eaay7667, 2020.

[311] Zhe Zhang, Shuo Luo, Guilherme Oliveira Barbosa, Meirong Bai, Thomas B Kornberg, and Dengke K Ma. The conserved transmembrane protein tmem-39 coordinates with copii to promote collagen secretion and regulate er stress response. *PLoS genetics*, 17(2):e1009317, 2021.

[312] Ishier Raote, Maria Ortega-Bellido, António JM Santos, Ombretta Foresti, Chong Zhang, Maria F Garcia-Parajo, Felix Campelo, and Vivek Malhotra. Tango1 builds a machine for collagen export by recruiting and spatially organizing copii, tethers and membranes. *Elife*, 7:e32723, 2018.

[313] Giulia Zanetti, Kanika Bajaj Pahuja, Sean Studer, Soomin Shim, and Randy Schekman. Copii and the regulation of protein sorting in mammals. *Nature cell biology*, 14(1):20–28, 2012.

[314] Meir Aridor. A tango for coats and membranes: New insights into er-to-golgi traffic. *Cell Reports*, 38(3):110258, 2022.

[315] Ishier Raote and Vivek Malhotra. Tunnels for protein export from the endoplasmic reticulum. *Annual Review of Biochemistry*, 2021.

[316] Aubrey V Weigel, Chi-Lun Chang, Gleb Shtengel, C Shan Xu, David P Hoffman, Melanie Freeman, Nirmala Iyer, Jesse Aaron, Satya Khuon, John Bogovic, et al. Er-to-golgi protein delivery through an interwoven, tubular network extending from er. *Cell*, 184(9):2412–2429, 2021.

[317] Ishier Raote, Morgan Chabanon, Nikhil Walani, Marino Arroyo, Maria F Garcia-Parajo, Vivek Malhotra, and Felix Campelo. A physical mechanism of tango1-mediated bulky cargo export. *Elife*, 9:e59426, 2020.

[318] Ishier Raote, Andreas M Ernst, Felix Campelo, James E Rothman, Frederic Pincet, and Vivek Malhotra. Tango1 membrane helices create a lipid diffusion barrier at curved membranes. *Elife*, 9, 2020.

[319] Janine McCaughey, Nicola L Stevenson, Judith M Mantell, Chris R Neal, Alex Paterson, Kate Heesom, and David J Stephens. A general role for tango1, encoded by mia3, in secretory pathway organization and function. *Journal of cell science*, 134(17):jcs259075, 2021.

[320] Luis Daniel Rios-Barrera, S Sigurbjörnsdóttir, Magdalena Baer, and Maria Leptin. Dual function for tango1 in secretion of bulky cargo and in er-golgi morphology. *Proceedings of the National Academy of Sciences*, 114(48):E10389–E10398, 2017.

[321] António JM Santos, Cristina Nogueira, Maria Ortega-Bellido, and Vivek Malhotra. Tango1 and mia2/ctage5 (tali) cooperate to export bulky pre-chylomicrons/vldls from the endoplasmic reticulum. *Journal of Cell Biology*, 213(3):343–354, 2016.

[322] Anja-Katrin Bosserhoff, M Moser, and R Buettner. Characterization and expression pattern of the novel mia homolog tango. *Gene expression patterns*, 4(4):473–479, 2004.

[323] Corinna A Kulicke, Erica De Zan, Zeynep Hein, Claudia Gonzalez-Lopez, Swapnil Ghanwat, Natacha Veerapen, Gurdyal S Besra, Paul Klenerman, John C Christianson, Sebastian Springer, et al. The p5-type atpase atp13a1 modulates major histocompatibility complex i-related protein 1 (mr1)-mediated antigen presentation. *Journal of Biological Chemistry*, 298(2), 2022.

[324] Wilbur Madison Song. *Regulation and Function of TREM2-dependent Pathways in Neurodegeneration*. PhD thesis, Washington University In St. Louis, 2020.

[325] Maulik Upadhyay, Elisabeth Kunz, Edson Sandoval-Castellanos, Andreas Hauser, Stefan Krebs, Alexander Graf, Helmut Blum, Arsen Dotsev, Innokentiy Okhlopkov, Alexey Shakhin, et al. Whole genome sequencing reveals a complex introgression history and the basis of adaptation to subarctic climate in wild sheep. *Molecular Ecology*, 30(24):6701–6717, 2021.

[326] Torsten Günther, Helena Malmström, Emma M Svensson, Ayça Omrak, Federico Sánchez-Quinto, Gülşah M Kılınç, Maja Krzewińska, Gunilla Eriksson, Magdalena Fraser, Hanna Edlund, et al. Population genomics of mesolithic scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS biology*, 16(1):e2003703, 2018.

[327] Stephanie M Yan, Rachel M Sherman, Dylan J Taylor, Divya R Nair, Andrew N Bortvin, Michael C Schatz, and Rajiv C McCoy. Local adaptation and archaic introgression shape global diversity at human structural variant loci. *Elife*, 10:e67615, 2021.

[328] Lee Roth, Swati Srivastava, Moshit Lindzen, Aldema Sas-Chen, Michal Sheffer, Mattia Lauriola, Yehoshua Enuka, Ashish Noronha, Maicol Mancini, Sara Lavi, et al. Silac identifies lad1 as a filamin-binding regulator of actin dynamics in response to egf and a marker of aggressive breast tumors. *Science signaling*, 11(515):eaan0949, 2018.

[329] Diogo Pellegrina, Alexander T Bahcheli, Michal Krassowski, and Jüri Reimand. Human phospho-signaling networks of sars-cov-2 infection are rewired by population genetic variants. *Molecular Systems Biology*, 18(5):e10823, 2022.

[330] Mehran Piran, Neda Sepahi, Afagh Moattari, Amir Rahimi, and Ali Ghanbariasad. Systems biomedicine of primary and metastatic colorectal cancer reveals potential therapeutic targets. *Frontiers in Oncology*, page 2314, 2021.

[331] Samuel Mathieu, Mewen Briend, Erik Abner, Christian Couture, Zhonglin Li, Yohan Bossé, Sébastien Thériault, Tõnu Esko, Benoit J Arsenault, and Patrick Mathieu. Genetic association and mendelian randomization for hypothyroidism highlight immune molecular mechanisms. *Iscience*, 25(9):104992, 2022.

[332] Maxat Kulmanov and Robert Hoehndorf. Deeppheno: Predicting single gene loss-of-function phenotypes using an ontology-aware hierarchical classifier. *PLoS computational biology*, 16(11):e1008453, 2020.

[333] Victoria Hung, Stephanie S Lam, Namrata D Udeshi, Tanya Svinkina, Gaelen Guzman, Vamsi K Mootha, Steven A Carr, and Alice Y Ting. Proteomic mapping of cytosol-facing outer mitochondrial and er membranes in living human cells by proximity biotinylation. *elife*, 6:e24463, 2017.

[334] Haim Ashkenazy, Shiran Abadi, Eric Martz, Ofer Chay, Itay Mayrose, Tal Pupko, and Nir Ben-Tal. Consurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, 44(W1):W344–W350, 2016.

[335] Gershon Celniker, Guy Nimrod, Haim Ashkenazy, Fabian Glaser, Eric Martz, Itay Mayrose, Tal Pupko, and Nir Ben-Tal. Consurf: using evolutionary data to raise testable hypotheses about protein function. *Israel Journal of Chemistry*, 53(3-4):199–206, 2013.

[336] Haim Ashkenazy, Elana Erez, Eric Martz, Tal Pupko, and Nir Ben-Tal. Consurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic acids research*, 38(suppl_2):W529–W533, 2010.

[337] Carine Berezin, Fabian Glaser, Josef Rosenberg, Inbal Paz, Tal Pupko, Piero Fariselli, Rita Casadio, and Nir Ben-Tal. Conseq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics*, 20(8):1322–1324, 2004.

[338] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, pages 1–11, 2021.

[339] Kathryn Tunyasuvunakool, Jonas Adler, Zachary Wu, Tim Green, Michal Zielinski, Augustin Žídek, Alex Bridgland, Andrew Cowie, Clemens Meyer, Agata Laydon, et al. Highly accurate protein structure prediction for the human proteome. *Nature*, 596(7873):590–596, 2021.

[340] Seong-In Hyun, Liliana Maruri-Avidal, and Bernard Moss. Topology of endoplasmic reticulum-associated cellular and viral proteins determined with split-gfp. *Traffic*, 16(7):787–795, 2015.

[341] Stéphanie Cabantous, Thomas C Terwilliger, and Geoffrey S Waldo. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature biotechnology*, 23(1):102–107, 2005.

[342] Hajar Owji, Navid Nezafat, Manica Negahdaripour, Ali Hajiebrahimi, and Younes Ghasemi. A comprehensive review of signal peptides: Structure, roles, and applications. *European journal of cell biology*, 97(6):422–441, 2018.

[343] Jan A Hiss, Eduard Resch, Alexander Schreiner, Michael Meissner, Anna Starzinski-Powitz, and Gisbert Schneider. Domain organization of long signal peptides of single-pass integral membrane proteins reveals multiple functional capacity. *Plos one*, 3(7):e2767, 2008.

[344] Jan A Hiss and Gisbert Schneider. Architecture, function and prediction of long signal peptides. *Briefings in bioinformatics*, 10(5):569–578, 2009.

[345] Eduard Resch, Jan A Hiss, Alexander Schreiner, Gisbert Schneider, and Anna Starzinski-Powitz. Long signal peptides of rgma and dcbld2 are dissectible into subdomains according to the ntrac model. *Molecular BioSystems*, 7(3):942–951, 2011.

[346] Corinna Meyer, Luis Barniol, Jan A Hiss, and Jude M Przyborski. The n-terminal extension of the p. falciparum gbp130 signal peptide is irrelevant for signal sequence function. *International Journal of Medical Microbiology*, 308(1):3–12, 2018.

[347] Yutaka Kodama and Chang-Deng Hu. An improved bimolecular fluorescence complementation assay with a high signal-to-noise ratio. *Biotechniques*, 49(5):793–805, 2010.

[348] Daniel Finley. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annual review of biochemistry*, 78:477–513, 2009.

[349] Nicholas D Nassif, Samantha E Cambray, and Daniel A Kraut. Slipping up: Partial substrate degradation by atp-dependent proteases. *IUBMB life*, 66(5):309–317, 2014.

[350] Daniel A Kraut and Andreas Matouschek. Proteasomal degradation from internal sites favors partial proteolysis via remote domain stabilization. *ACS chemical biology*, 6(10):1087–1095, 2011.

[351] Eden L Reichard, Giavanna G Chirico, William J Dewey, Nicholas D Nassif, Katelyn E Bard, Nickolas E Millas, and Daniel A Kraut. Substrate ubiquitination controls the unfolding ability of the proteasome. *Journal of Biological Chemistry*, 291(35):18547–18561, 2016.

[352] Anton Khmelinskii, Matthias Meurer, Chi-Ting Ho, Birgit Besenbeck, Julia Füller, Marius K Lemberg, Bernd Bukau, Axel Mogk, and Michael Knop. Incomplete proteasomal degradation of

green fluorescent proteins in the context of tandem fluorescent protein timers. *Molecular biology of the cell*, 27(2):360–370, 2016.

[353] Christopher Eric Bragança and Daniel Adam Kraut. Mode of targeting to the proteasome determines gfp fate. *Journal of Biological Chemistry*, 295(47):15892–15901, 2020.

[354] Huaiyu Mi, Anushya Muruganujan, Dustin Ebert, Xiaosong Huang, and Paul D Thomas. Panther version 14: more genomes, a new panther go-slim and improvements in enrichment analysis tools. *Nucleic acids research*, 47(D1):D419–D426, 2019.

[355] Birgit Meusser, Christian Hirsch, Ernst Jarosch, and Thomas Sommer. Erad: the long road to destruction. *Nature cell biology*, 7(8):766–772, 2005.

[356] Shinya Ito and Kazuhiro Nagata. Biology of hsp47 (serpin h1), a collagen-specific molecular chaperone. In *Seminars in cell & developmental biology*, volume 62, pages 142–151. Elsevier, 2017.

[357] Joseph M Replogle, Reuben A Saunders, Angela N Pogson, Jeffrey A Hussmann, Alexander Lenail, Alina Guna, Lauren Mascibroda, Eric J Wagner, Karen Adelman, Gila Lithwick-Yanai, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale perturb-seq. *Cell*, 2022.

[358] Kayo Ikeda, Makoto Kinoshita, Hisako Kayama, Shushi Nagamori, Pornparn Kongpracha, Eiji Umemoto, Ryu Okumura, Takashi Kurakawa, Mari Murakami, Norihisa Mikami, et al. Slc3a2 mediates branched-chain amino-acid-dependent maintenance of regulatory t cells. *Cell reports*, 21(7):1824–1838, 2017.

[359] Chloe C Feral, Naoyuki Nishiya, Csilla A Fenczik, Heidi Stuhlmann, Marina Slepak, and Mark H Ginsberg. Cd98hc (slc3a2) mediates integrin signaling. *Proceedings of the National Academy of Sciences*, 102(2):355–360, 2005.

[360] Chunlei Liu, Xin Li, Chen Li, Zeyu Zhang, XiaoJian Gao, Zhilong Jia, HaiXu Chen, Qian Jia, Xiaojing Zhao, Jixuan Liu, et al. Slc3a2 is a novel endoplasmic reticulum stress-related signaling protein that regulates the unfolded protein response and apoptosis. *PLoS One*, 13(12):e0208993, 2018.

[361] Daniel N Itzhak, Stefka Tyanova, Jürgen Cox, and Georg HH Borner. Global, quantitative and dynamic mapping of protein subcellular localization. *elife*, 5:e16950, 2016.

[362] Qian Xiao, Ceara K McAtee, and Xiaolei Su. Phase separation in immune signalling. *Nature Reviews Immunology*, pages 1–12, 2021.

[363] Shuh-ichi Nishikawa, Jeffrey L Brodsky, and Kunio Nakatsukasa. Roles of molecular chaperones in endoplasmic reticulum (er) quality control and er-associated degradation (erad). *Journal of biochemistry*, 137(5):551–555, 2005.

[364] Pascal Haffter, Michael Granato, Michael Brand, Mary C Mullins, Matthias Hammerschmidt, Donald A Kane, Jörg Odenthal, FJ Van Eeden, Yun-Jin Jiang, Carl-Philipp Heisenberg, et al. The identification of genes with unique and essential functions in the development of the zebrafish, danio rerio. *Development*, 123(1):1–36, 1996.

[365] Charles B Kimmel, William W Ballard, Seth R Kimmel, Bonnie Ullmann, and Thomas F Schilling. Stages of embryonic development of the zebrafish. *Developmental dynamics*, 203(3):253–310, 1995.

[366] Woong Y Hwang, Yanfang Fu, Deepak Reyon, Morgan L Maeder, Shengdar Q Tsai, Jeffry D Sander, Randall T Peterson, JR Yeh, and J Keith Joung. Efficient genome editing in zebrafish using a crispr-cas system. *Nature biotechnology*, 31(3):227–229, 2013.

[367] Miguel A Moreno-Mateos, Charles E Vejnar, Jean-Denis Beaudoin, Juan P Fernandez, Emily K Mis, Mustafa K Khokha, and Antonio J Giraldez. Crisprscan: designing highly efficient sgrnas for crispr-cas9 targeting in vivo. *Nature methods*, 12(10):982–988, 2015.

[368] Gaurav K Varshney, Wuhong Pei, Matthew C LaFave, Jennifer Idol, Lisha Xu, Viviana Gallardo, Blake Carrington, Kevin Bishop, MaryPat Jones, Mingyu Li, et al. High-throughput gene targeting and phenotyping in zebrafish using crispr/cas9. *Genome research*, 25(7):1030–1042, 2015.

[369] Jean-Paul Concordet and Maximilian Haeussler. Crispor: intuitive guide selection for crispr/cas9 genome editing experiments and screens. *Nucleic acids research*, 46(W1):W242–W245, 2018.

[370] Stephanie A Connors, Jamie Trout, Marc Ekker, and Mary C Mullins. The role of tolloid/mini fin in dorsoventral pattern formation of the zebrafish embryo. *Development*, 126(14):3119–3130, 1999.

[371] Reema Jasuja, Nikolas Voss, Gaoxiang Ge, Guy G Hoffman, Jamie Lyman-Gingerich, Francisco Pelegri, and Daniel S Greenspan. bmp1 and mini fin are functionally redundant in regulating formation of the zebrafish dorsoventral axis. *Mechanisms of development*, 123(7):548–558, 2006.

[372] François Kroll, Gareth T Powell, Marcus Ghosh, Gaia Gestri, Paride Antinucci, Timothy J Hearn, Hande Tunbak, Sumi Lim, Harvey W Dennis, Joseph M Fernandez, et al. A simple and effective f0 knockout method for rapid screening of behaviour and other complex phenotypes. *Elife*, 10, 2021.

[373] Naava Naslavsky and Steve Caplan. The enigmatic endosome–sorting the ins and outs of endocytic trafficking. *Journal of cell science*, 131(13):jcs216499, 2018.

[374] Manuel D Leonetti, Sayaka Sekine, Daichi Kamiyama, Jonathan S Weissman, and Bo Huang. A scalable strategy for high-throughput gfp tagging of endogenous human proteins. *Proceedings of the National Academy of Sciences*, 113(25):E3501–E3508, 2016.

[375] Lindsey M Costantini, Matteo Fossati, Maura Francolini, and Erik Lee Snapp. Assessing the tendency of fluorescent proteins to oligomerize under physiologic conditions. *Traffic*, 13(5):643–649, 2012.

Appendix A

Supplementary Information

A.1 Efforts to make a *tmem131* knockout cell line and CRISPR knock-ins

CRISPR knock-out using transient transfection

Whilst investigating the cell biology of TMEM131 through transient expression, I made several attempts to construct a *tmem131* knockout cell line. We speculated this might reveal a measurable phenotype, perhaps the inability of the pre-TCR or TCR to be transported to the cell surface. HEK293T, like other common cancer cell lines such as HeLa, are hypotriploid so I would have to disrupt three copies of the gene. TMEM131 is not expressed at the cell surface and we currently lack a reliable antibody. This meant the only way to screen potential knockout lines was to PCR amplify the genomic locus, check for genomic editing via a T7E1 assay then sequence exhaustively to capture all three alleles.

I designed CRISPR guides to five different exons in the gene shown in table S1. These were ordered as two complementary ssDNA oligos with the sequence "CACC" preceding the forward sequences and "AAAC" preceding the reverse sequence. The two oligos were annealed then ligated into a plasmid encoding pHR_Cas9[X]-BFP where X is a stuffer containing two copies of the recognition sequence for the restriction enzyme BsmBI. This enzyme creates cuts in the plasmid immediately after a U6 promotor sequence and prior to the rest of the sgRNA sequence, removing the stuffer and allowing correct insertion of the guide sequences. Cas9-P2A-BFP is expressed from a SFFV promotor elsewhere in the plasmid. Efficient transcription from the U6 promoter requires the first nucleotide to be G. This could either be added to the 5' end of a 20nt guide or the first nt of the guide replaced with G. For this first round I had so many guide options that I just chose guides beginning with G with reasonable predicted specificity and efficiency. I targeted five different exons as some exons prove easier to edit than others. All exons were prior to the putative transmembrane domain so should produce a dysfunctional product if disrupted.

Plasmids encoding these guides were transiently expressed in HEK cells. A BD Aria Fusion was used sort individual BFP-positive cells into a 96 well plate. Cells that survived took at least two weeks to grow, producing wells with all cells descended from a single clone. These populations could not be maintained in 96 well plates so surviving wells were moved into 6 well plates. Once they had grown, genomic DNA was extracted and cells split depending on their confluency while screening took place. To screen cell

Target Exon	Forward Oligo	Reverse Oligo
Exon 1	caccGATGACCCTCGTAGTGGCTG	aaacCAGCCACTACGAGGGTCATC
Exon 7	caccGTGAGAAACTGCTATTCACA	aaacTGTGAATAGCAGTTTCTCAC
Exon 8	caccGAACTCCCAACGGGTCAACA	aaacTGTTGACCCGTTGGGAGTTC
Exon 14	caccGTAGCAACACATCGTGAATG	aaacCATTCACGATGTGTTGCTAC
Exon 19	caccGACCTGCTTCCCTAAGCACG	aaacCGTGCTTAGGGAAGCAGGTC

Table S1: Guide sequences for making a *tmem131* knockout cell line.
Guides against five different exons in human *tmem131* were designed using CRISPOR[369] and ordered as ssDNA oligos. The forward sequence without the cacc prefix is the 20nt guide sequence.

lines I PCR amplified the genomic locus surrounding the edit site and treated the PCR part with a T7E1 assay (figure S1) using the same method as section 5.2.4. In this first round the survival rate was relatively low but still the number of different lines, all of which grew at different rates, quickly became difficult to manage.

In this attempt, no edits were found in the lines transfected with guides specific to exon 8, exon 19 or exon 14. There did seem to be some edits in lines transfected with the guide to exon 1 but these also seemed to be faintly present in DNA from untransfected cells so these might have been single nucleotide polymorphisms that are present in the gene normally. The DNA flanking exon 1 was very troublesome to PCR amplify due to the high GC-content (77%) in this exon. More success was observed with cells transfected with guides to exon 7 as it looked like there was genomic editing in lines Ex7-H7 and Ex7-F5. The PCR amplified DNA from these lines was ligated into a plasmid then this plasmid was transformed into *E. coli*. Eight colonies for each line were selected and sequenced.

For the Ex7-H7 cell line, six of the sequencing reactions returned the WT sequence and two had a large deletion encompassing the second half of exon 7. For the Ex7-F5 cell line, three sequencing reactions returned a one base pair addition (a G nt that 5bp 5' to the PAM) and two sequencing reactions returned the same 42bp insertion that would result in two in-frame STOP codons in exon 7. The final three sequencing reactions returned a 12bp deletion which would result in the loss of four amino acids (NSSF). Although the conservation scores from the Consurf model suggests this is a conserved part of the molecule, the absence of a frame shift mutation means that the protein may still be functional.

Given there are three alleles of *tmem131* that needed disrupting, it is not seem especially unlikely to get one indel that does not cause a change in reading frame of the protein. I decided to target this remaining allele with three new guides all of which should be specific to the new sequence (table S2). Plasmids containing these guides were expressed in the Ex7-F5 line, then BFP-positive cells selected via FACS and surviving lines grown up as before.

Target Exon	Forward Oligo	Reverse Oligo
AltF5g1	caccGGTTTATTATAGGTGAGACA	aaacTGTCTCACCTATAATAAACC
AltF5g2	caccGTGTTTATTATAGGTGAGAC	aaacGTCTCACCTATAATAAACAC
AltF5g3	caccGATAGGTGAGACAGGGACTC	aaacGAGTCCCTGTCTCACCTATC

Table S2: Guide sequences to alter the remaining in-frame allele of Ex7-F5.
Guides against the remaining allele were designed using CRISPOR[369] and ordered as ssDNA oligos. The forward sequence without the cacc prefix is the 20nt guide sequence. All three were 19nt guides with G (underlined) added to the 5'end for transcription from the U6 promotor.

As two of the alleles of F5 were already different, I would not be able to use a T7E1 assay to screen these lines and sequencing every line until I found a new allele would be prohibitively expensive. Instead I tried to screen lines with via a multiplex PCR. This PCR used two pairs of primers: one set should

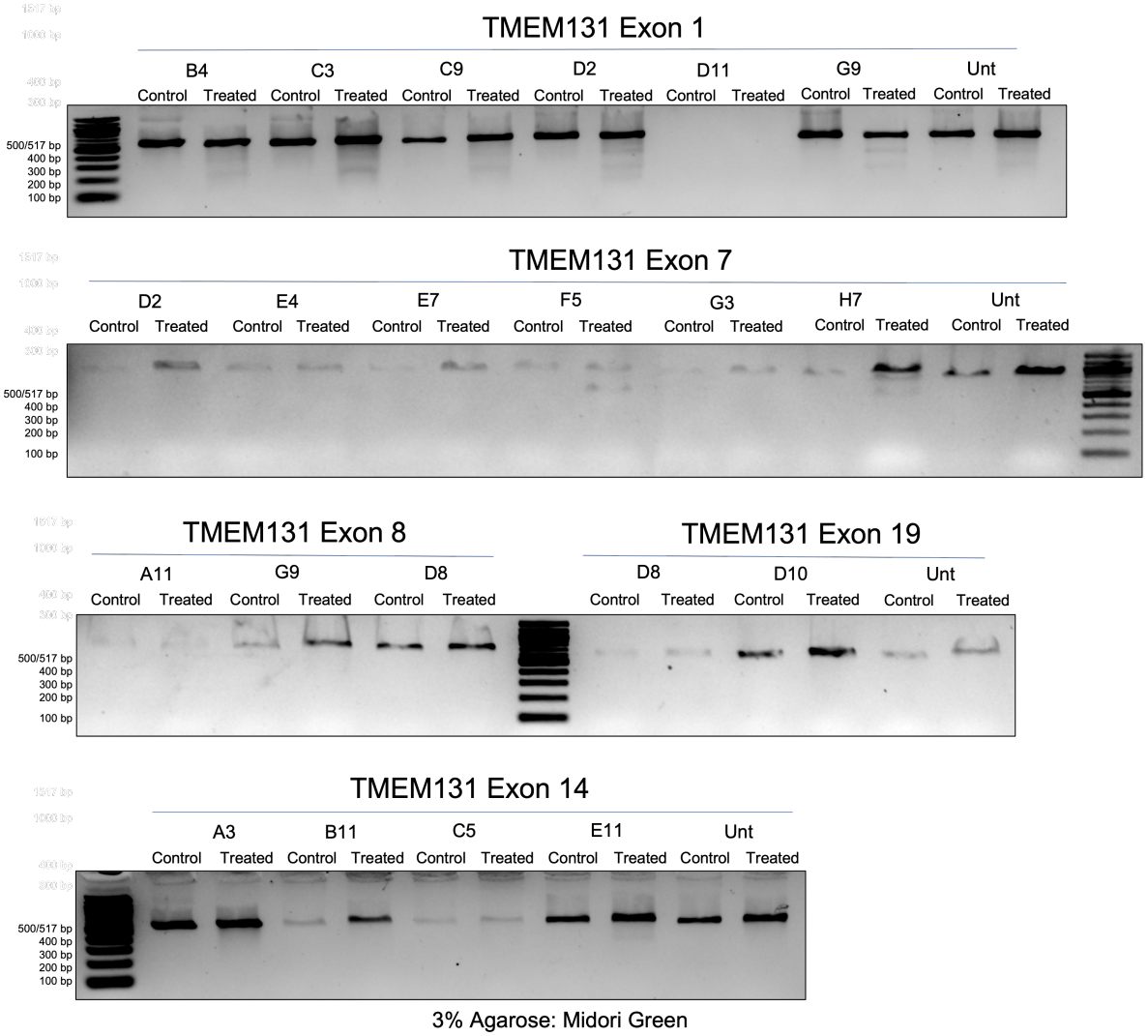

Figure S1: T7E1 assay for HEK cell lines transfected with CRISPR plasmids targeting *tmem131* gene. HEK293T cells were transfected with pHR_cas9[*tmem131*]-BFP plasmids. After 48h cells were suspended then 96 BFP-positive cells selected using FACS with the BP Aria Fusion. Surviving cell lines were identified by the location in the 96 well plate (e.g. row F, column 5). These cells were grown up, genomic DNA extracted then the genomic locus of *tmem131* amplified. Amplified sequences were screened for genomic edits with a T7E1 assay. For some exons untransfected HEK (Unt) were used as a negative control.

amplify a 566bp section of *tmem131* exon 8 and the other should only amplify a 400bp sequence from the last in-frame allele of Ex7-F5. The 3' end of the reverse primer for this set was positioned so it should be disrupted if any of the three new guides had made an edit. These PCR were run on a 3% agarose gel (figure S2). Unfortunately both bands were present in all new lines suggesting that editing of the final allele had not been successful. The lower band seemed faintly present when genomic DNA of untransfected HEK was amplified suggesting primer choice may not have been optimal.

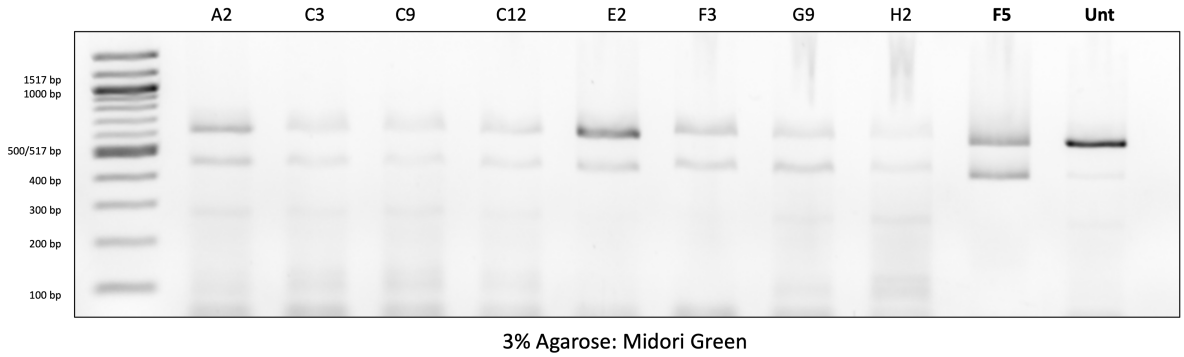

Figure S2: A multiplex PCR to screen for edits in the final in-frame allele of the Ex7-F5 line. The genomic DNA for the new lines was amplified using a set of primers for exon 8 and a set that should amplify only the in frame allele of Ex7-F5. This line and untransfected HEK are used as controls (right).

As well as trying to edit the final allele of Ex7-F5, I tried again with the guides to exon 8 and exon 19. Once again there was no editing in the only survivors from cells transfected with guides to exon 8 but editing was detected in lines transfected with guides to exon 19 (figure S3). These lines were sequenced. Ex19-C2 had an allele with a 1bp insertion, an allele with a 3bp insertion and an allele with a 6bp deletion. Ex19-D9 had an allele with a 1bp insertion and an allele with a 6bp deletion. Ex19-G9 had a WT allele, an allele with a 1bp insertion and an allele with a 21bp deletion. The number of alleles with indels of multiples of 3 resulting in an in-frame protein seemed suspicious but I cannot conclude that making a TMEM131-KO line is not possible.

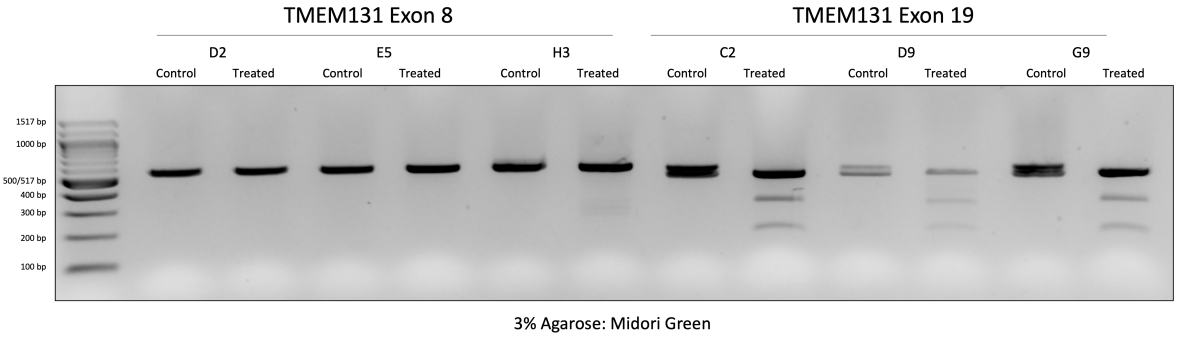

Figure S3: T7E1 assay for HEK cell lines transfected with guides to exon 8 and exon 19.

CRISPR knock-in mutations

I also made several attempts to make CRISPR knock-in mutations at the *tmem131* endogenous locus. The double strand breaks induced by cas9 can be repaired by the non-homologous end joining pathway of DNA repair, often causing indels, or via the Homology-directed repair (HDR) pathway. By supplying a DNA template with homology arms flanking the insertion site HDR can incorporate desired mutations. The homology arms need to be modified to avoid recognition by cas9 either by changing sequences bound

by the guides or removing the PAMs. Whilst the efficiency of HDR is quite low I thought this would be outweighed by the need to only alter one allele rather than all three.

I decided I wanted to tag *tmem131* with GFP11 to confirm the localisation of the endogenous protein. This tag is small enough (60nt) to order as a single stranded oligonucleotide donor (ssODN) with 70bp homology arms.[374] HDR is more efficient with ssODNs than using dsDNA templates. I designed and ordered ssODNs and CRISPR guides for exon 3, exon 31 and exon 41. Exon 3 was as close I could get to the N-terminus due to issues with GC content in exon 1 and the lack of PAM sequences in exon 2. Exon 31 is the position in the gene where the TMEM131dC construct ends and this ssODN included a stop codon after the GFP11 tag to try to create this construct from the endogenous allele. Exon 41 includes the protein C-terminus.

I ordered ssODNs for both forward and reverse strands for all three points in the gene. Cells were transfected with 500ng of ssODN along with 500ng of plasmids encoding guide sequences to these sites. After a week for editing to occur these were transfected with GFP(1-10) or ER-targeted GFP(1-10)kdel. The 96 most GFP-positive cells were selected into plates via FACS. Unfortunately once they had grown up and were transfected again with GFP(1-10) these were not noticeably GFP-positive compared to a untransfected HEK control. I suspect that the endogenous expression of TMEM131 is too low for GFP11-tagging to work. This might be improved in future by including multiple copies of the GFP11 tag in the ssODN to increase the signal from each protein. This would however exceed the length of ssDNA that I could easily order.

I also tried to tag exon 41 with GFP. This required a plasmid template with 500bp homology arms. I amplified these arms from the genomic DNA and cloned them into a PC1 plasmid flanking the sequence for the GFP domain. The intention was to transient transfect HEK cells with the guides to exon 41 and the HDR template then sort on the basis of GFP expression. Unfortunately even without promotor elements there seemed to be enough GFP expression from the HDR template to mask any effect from tagging the endogenous protein. In an attempt to remove this I transfected with the HDR template and guide plasmids then waiting two weeks before sorting for the transfected plasmids to be diluted away. Unfortunately like with the attempt at GFP11-tagging, after the sorted cells had grown up they were too dim to be distinguished from untransfected HEK. Given the time required to maintain and screen so many cell lines, I did not consider it productive to continue with either method.

CRISPR knock-out using lentiviral transfection

I wanted to make one last attempt to make a TMEM131 knockout cell line. To increase my chances of success I cloned three additional guides in TMEM131 exon 7, one that was 1bp shifted compared to my original guide as well as two overlapping guides earlier in the exon. I though the use of multiple guides would increase the chance of an edit occurring on every allele. As well as using HEK cells I acquired HeLa cells and hTERT-RPE1 cells. Retinal pigment epithelial-1 (RPE1) cells are an adherent but non-cancerous human cell line that was immortalised by expression of human telomerase reverse transcriptase (hTERT). These cells are near-diploid so it should be much easier to edit both copies of TMEM131 then screen. RPE1 cells are not as amenable to transient transfection as HEK so I generated lentivirus for all four guides and transduced all three cells lines with a mixture of guides. After 4 days to recover, BFP-positive cells were then sorted. I increased survivability of the HEK cells compared to previous attempt by sorting cells into a mixture of fresh and conditioned media. Survival of HeLa cells was still low so these were not used further.

I grew up cells and extracted genomic DNA as before however screening cells was interrupted when I had to isolate with COVID-19. Lines from which genomic DNA had been extracted were frozen down to -80°C whilst the rest had to be discarded.

Two RPE1 lines from this selection looked promising and alleles were sequenced. In both case more than two alleles were found. For the RPE1-F9 line the most common allele found by sequencing had a 16bp deletion at a site consistent with editing by either of the overlapping pair of guides in the 3' part of the exon. However an allele was also found with the same deletion as well as a 6bp deletion from the overlapping pair of guide in the 5' half of the exon. Alleles with 16bp and 36bp deletions in the 3' half of the exon were also found. The sequencing in the RPE-F11 line returned a mixture of 7 different alleles including two with the same insertion in the 5' half of the exon but different edits in the 3' half of the exon. One of the two viral-transduced HEK cell lines that was sequenced had a 36bp deletion, an 18bp deletion and two deletions that would produce a frameshift. The other had a 16bp deletion and poor sequencing for the other reads.

How did sequencing find more than two alleles from screening diploid cell lines that were supposed to be descended from a single cell? As the CRISPR plasmids were integrated into the genome by lentiviral transduction, the cas9 and guides could continue to be expressed after the cells were sorted. This means when the individual cells grow and divide there is the potential for different edits to then occur meaning the cells are no longer clonal. If I had used only one guide or one set of overlapping guides then this might not have been as much of an issue as editing could only occur twice maximum. Using two pairs of guides in the same exon however leaves the possibility of an edit correcting a frame shift caused by another edit. Eventually editing would be expected to stop as both sequences would be edited or expression of cas9 falls due to methylation however the population of cells would not be homogenous.

I could fix this issue by sorting these cells again or through generating new lines by transducing with only a single guide then sorting after a longer interval. This was not thought to be best use of time especially given there was no guarantee of a measurable phenotype at the end of the process. I had hoped to use the rate of export of a marker in the RUSH assay as potential assay for TMEM131 function. I had enough difficulties getting this assay working in HEK cells and don't think it would have been trivial to perform in RPE1 cells whilst also dealing with lower and less consistent expression from lentiviral transduction.

A.2 Supplementary figures for experiments investigating the pre-TCR, chapter 3

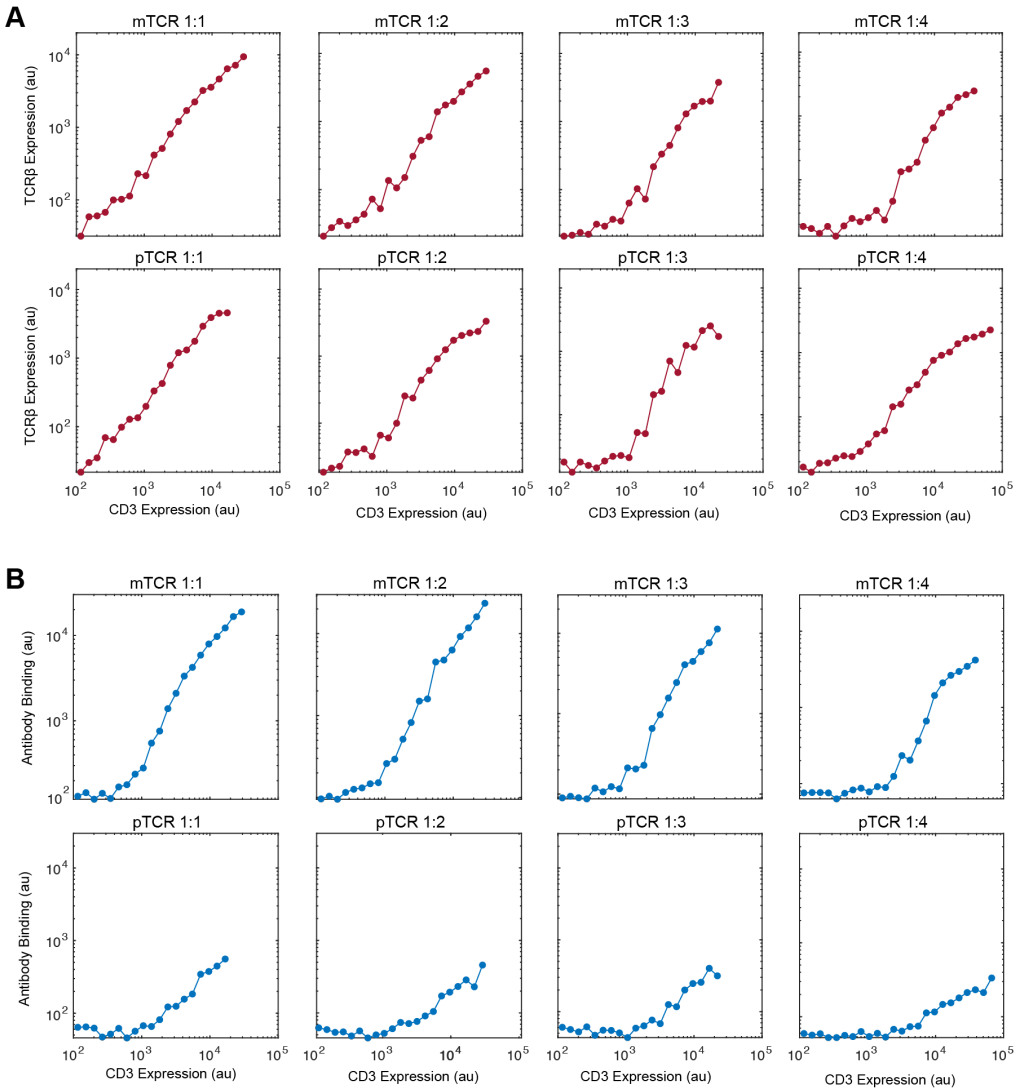

Figure S4: An internalisation assay finds no increase in receptor expression or surface trafficking by increasing the ratio of TCR:CD3 plasmids.
HEK cells were transiently transfected with a plasmid encoding the pTCR/mTCR with mScar tagged on the C-terminus of TCRβ and a plasmid encoding CD3$\gamma\delta\epsilon\zeta$-IRES-BFP in a 1:1, 1:2, 1:3 or 1:4 ratio keeping the total amount of DNA constant. Cells were resuspended then labelled with anti-TCRβ V5.1 antibody conjugated to APC for 30min in 37°C media. Cells were washed three times, fixed then analysed by flow cytometry. The data was sorted into 25 log-distributed bins between 10^3 and 10^5 along the "V 450/50" axis.
A: The mean value in the "YG 610/20" channel for each bin was plotted on the y-axis in the middle value of each bin.
B: The mean value in the "R 685/35" channel for each bin was plotted on the y-axis in the middle value of each bin. Bins with less than 10 events were excluded.

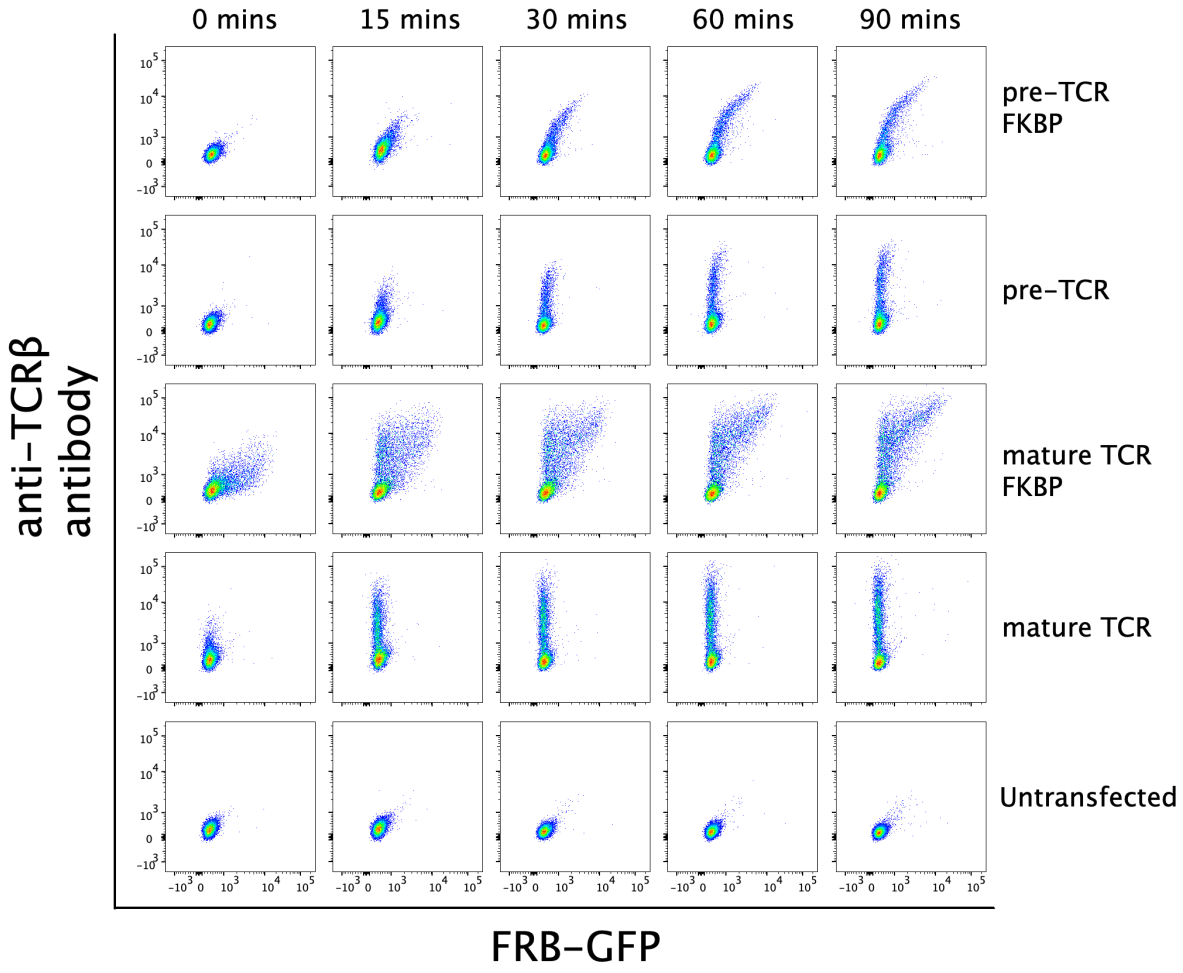

Figure S5: Timecourse of FRB-GFP binding to FKBP-TCRβ with rapamycin.
HEK cells were transfected with versions of the pTCR and mTCR with and without the FKBP domain on the N-terminus of TCRβ. After 24h, cells were suspended with trypsin, then incubated at 37°C in 1ml of medium with 1:200 APC-conjugated anti-TCRβ antibody, 0.5µM FRB-GFP and 200nM rapamycin. 200µl aliquots were removed, washed and fixed after 0, 15, 30, 60 and 90 minutes. Samples were analysed by flow cytometry.

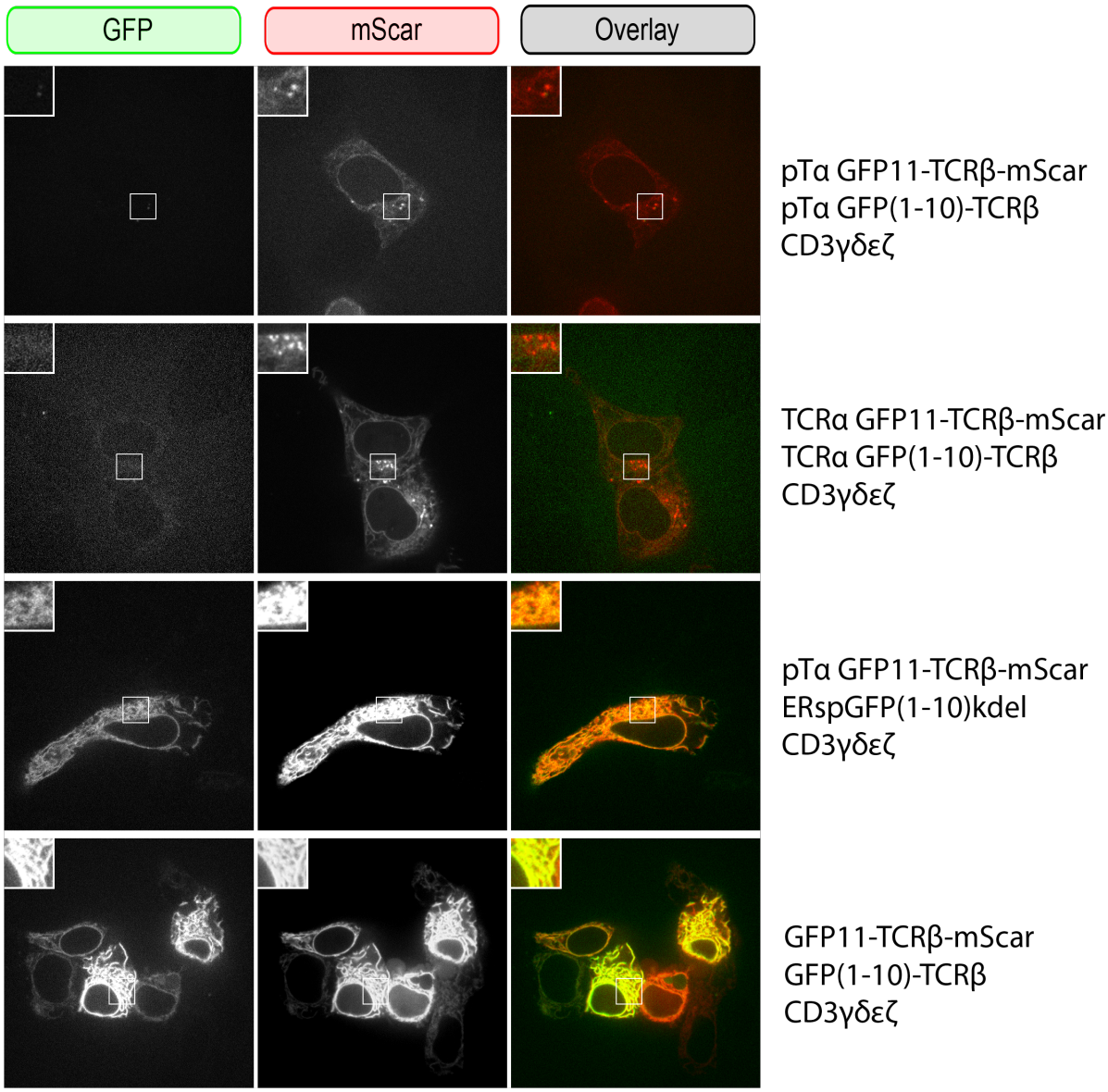

Figure S6: A split GFP assay shows oligomers of TCRβ in the ER.
HEK cells were transfected with the indicated constructs then imaged under 100X magnification. The inserts show a 2X zoom of the indicated areas.

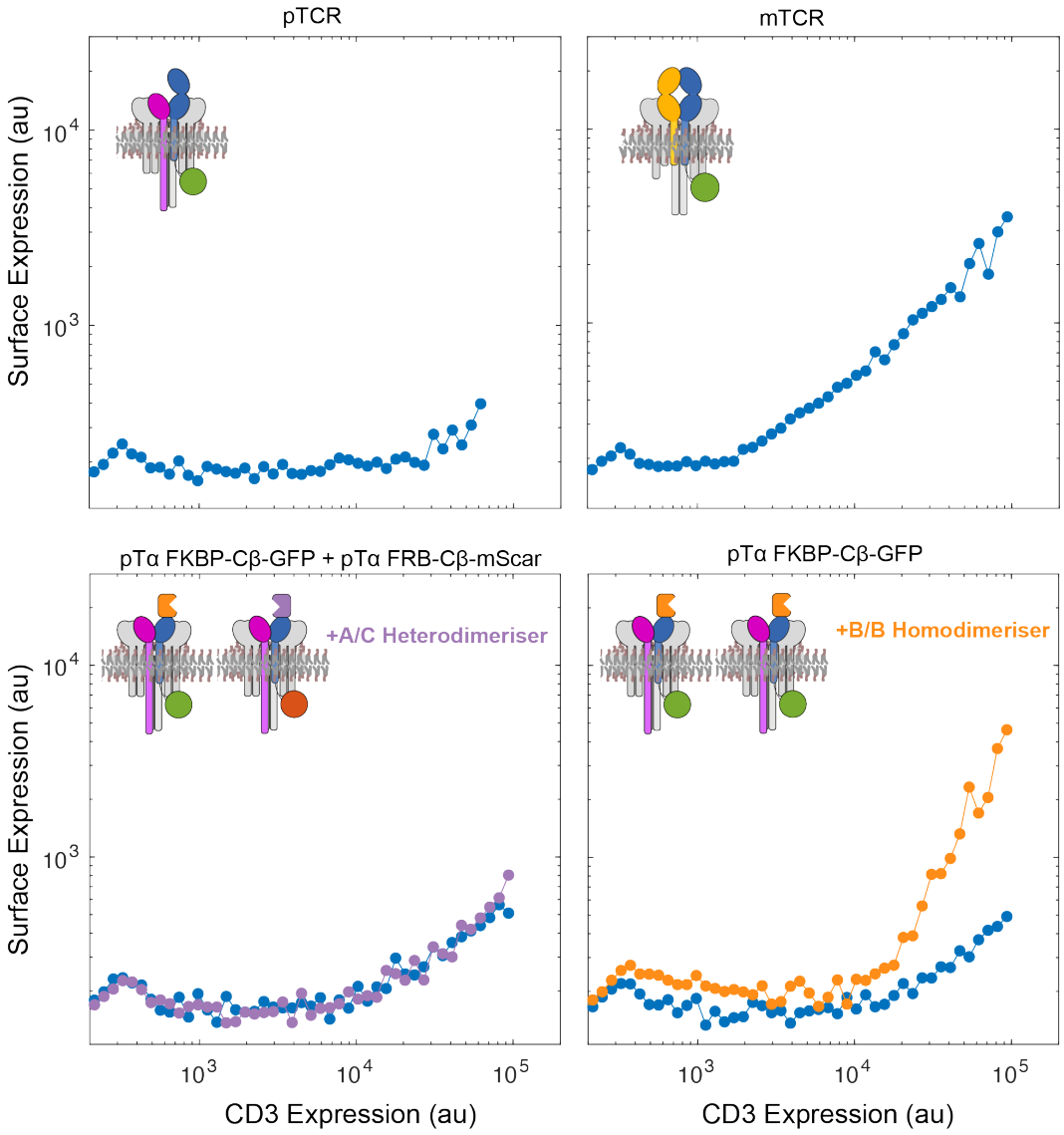

Figure S7: Surface expression of Cβ is increased by artificial dimerisation.
HEK cells were transfected with CD3$\gamma\delta\epsilon\zeta$ with pTα HA-FKBP-Cβ-GFP $\pm$ pTα FRB-Cβ-mScar. These were treated with 10nM of B/B homodimerizer (orange) and A/C heterodimerizer (purple, both Takara Bio) for 30mins. Cells were then suspended and labelled on ice with 1:200 anti-HA antibody for 30mins. HA and GFP-tagged versions of the pTCR and mTCR constructs were label simultaneous as controls. Cells were washed, fixed then analysed by flow cytometry. The data was sorted into 50 log-distributed bins between 10^2 and 10^5 along the "V 450/50" axis. The mean value in the "R 685/35" channel for each bin was plotted on the y-axis in the middle value of each bin. Bins with less than 10 events were excluded.

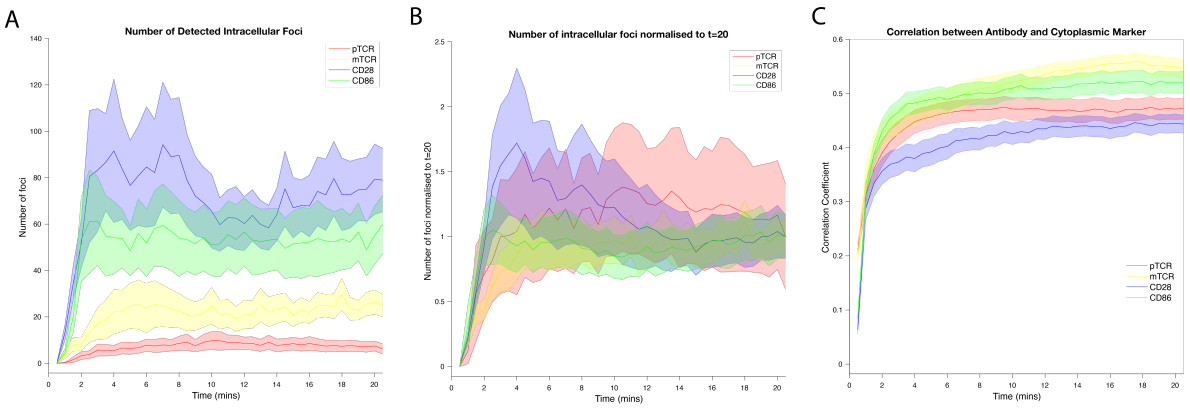

Figure S8: The rate of internalisation of the pTCR and mTCR is comparable to that of CD28 and CD86. HEK cells were transfected with GFP-tagged CD28, CD86, pTCR or mTCR with an mScar cytoplasmic marker. After 48h cells were imaged at 100X magnification every thirty seconds for 20 minutes. After frame 1, af647-conjugated anti-GFP nanobody was added to the external medium. The accumulation of antibody over time was analysed in MATLAB. The experiment was performed four times. In each replicate, 10 positions in a well were imaged every frame.
A: The average number of intracellular foci identified over time. Shaded areas show the standard error of the measurement.
B: The average number of intracellular foci normalised to the amount present after 20 minutes.
C: The correlation between the nanobody and the mScar cytoplasmic marker over time.

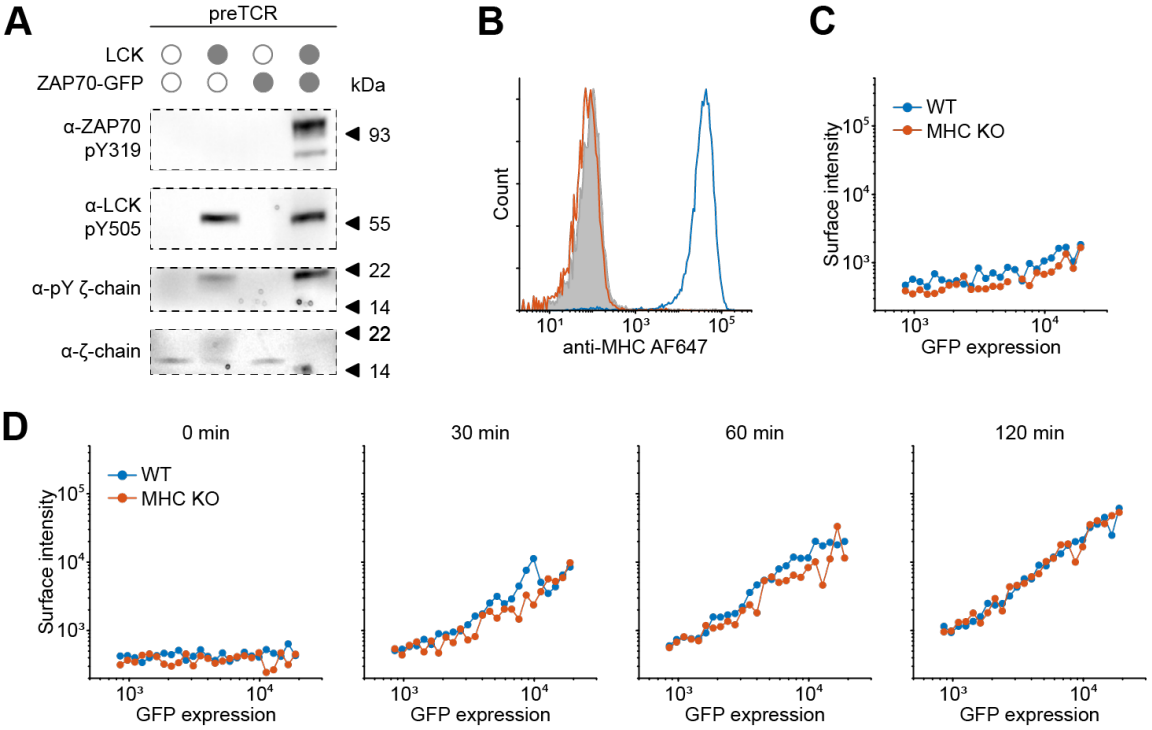

Figure S9: Phosphorylation of the pre-TCR is not detectable in HEK in the absence of signalling kinases. MHC-I expression is not required for pre-TCR surface expression.
A: The mScar-tagged pre-TCR was expressed in HEK cells with the LCK and ZAP70 kinases as indicated. Cell lysate was analysed by western blotting. The blot was cut after blocking with reference to the molecular weight markers and labelled with the indicated antibodies. Phosphorylated ζ-chain was detectable only with the expression of exogenous kinases.
B: Wildtype HEK cells (blue) were stained with af647-conjugated anti-MHC antibody showing high expression of MHC. This is completely absent on B2M-knockout cells (red) that are indistinguishable from isotype control staining (filled grey).
C: There is no significant difference in pre-TCR surface staining between WT HEK and a MHC KO line using an APC-conjugated anti-TCRβ V5.1 antibody.
D: A pre-TCR internalisation assay with WT and B2M KO HEK cells does not show substantial differences at any timepoint.
Data acquired by the author with additional analysis by John James.

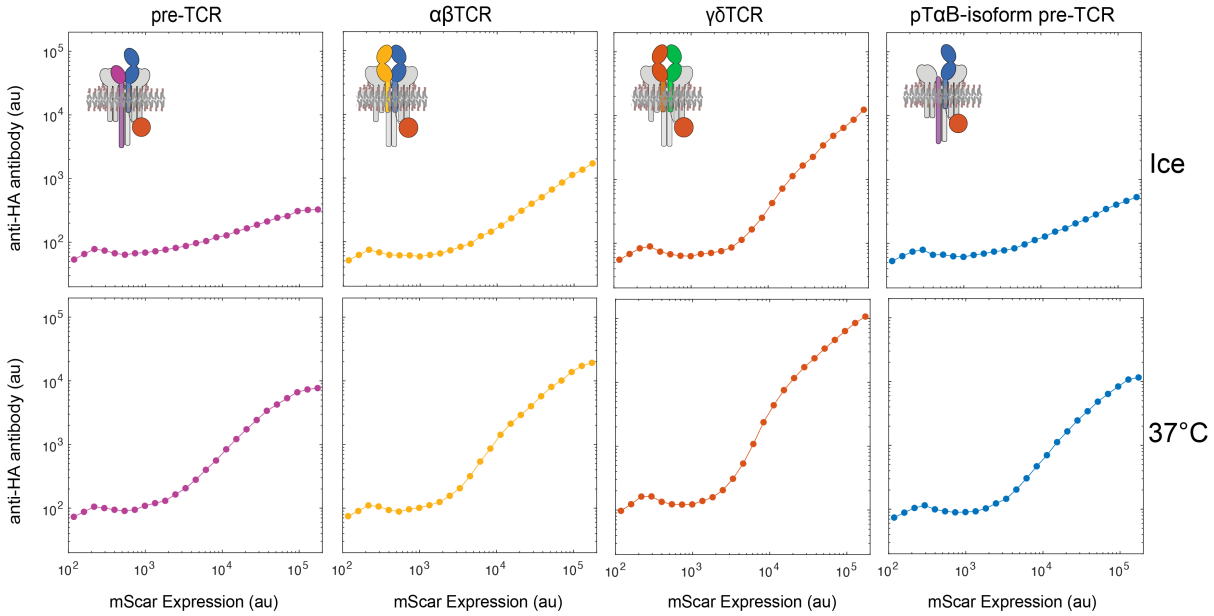

Figure S10: The $\gamma\delta$TCR is well expressed at the cell surface. The pre-TCR with the pTαB isoform has similar surface expression to the normal pre-TCR.
The G115 clone of the $\gamma\delta$TCR was expressed with a HA-tag on the N-terminus of TCRδ and mScar-fluorophore on the C-terminus. The G10 clone of the $\alpha\beta$-TCR, pre-TCR and pre-TCR using the pTα^{b} isoform were expressed with the same tags on TCRβ. All were co-expressed with the same CD3$\gamma\delta\epsilon\zeta$-BFP plasmid. Cells were labelled with anti-HA antibodies for 30 minutes on ice (top) or at 37°C in growth medium. Cells were washed then analysed via flow cytometry. The data was sorted into 25 log-distributed bins between 10^2 and 2x10^5 along the "YG 610/20" axis. The mean value in the "R 685/35" channel for each bin was plotted on the y-axis in the middle value of each bin. Bins with less than 10 events were excluded.

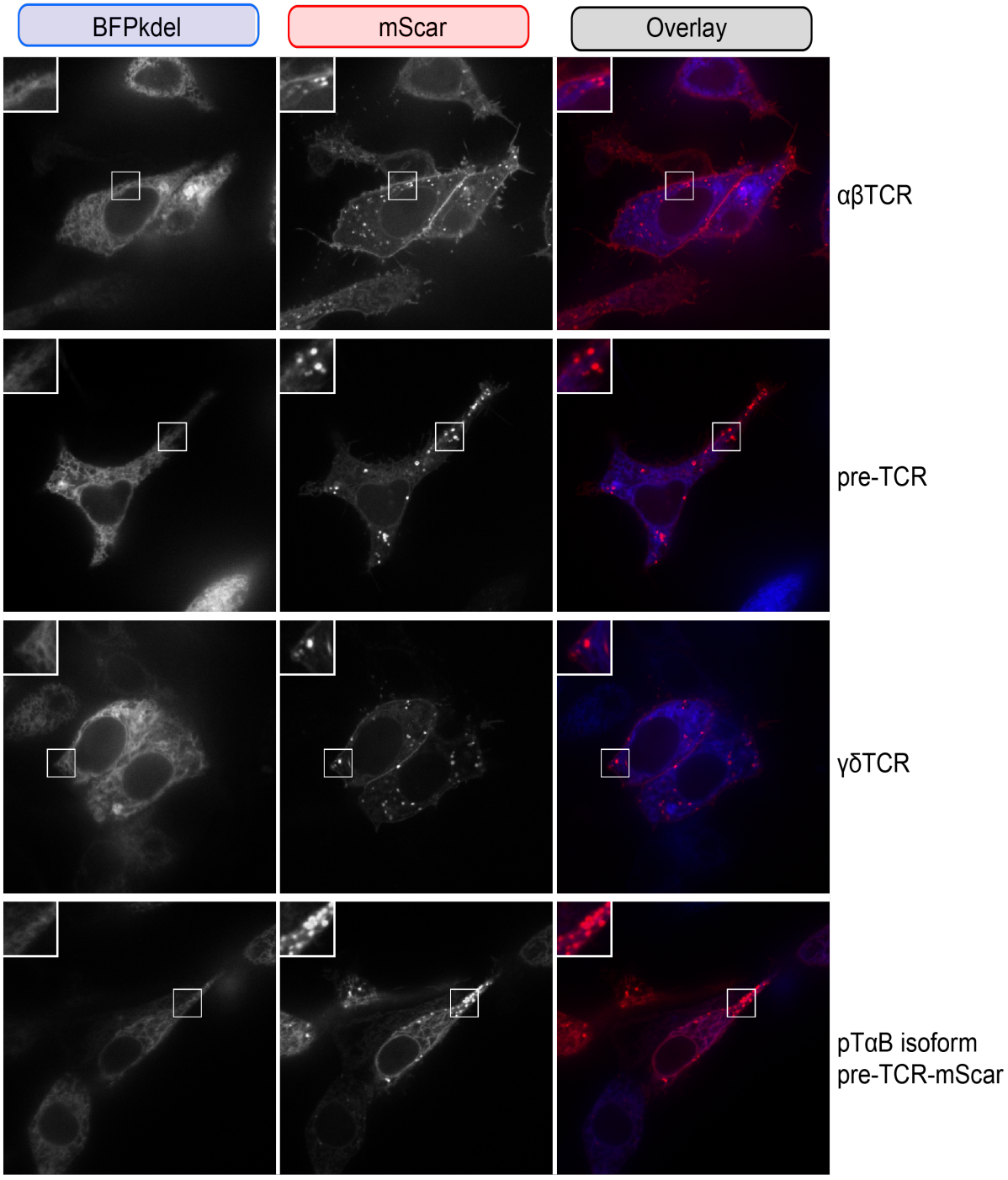

Figure S11: Microscopy of cells expressing the $\gamma\delta$TCR and pre-TCR with the pTαB isoform.
The four receptors discussed in the figure above were imaged under 100X magnification with an BFPkdel ER marker. Inserts show a 2x zoom of the selected area.

A.3 Supplementary figures for experiments investigating TMEM131, chapter 4

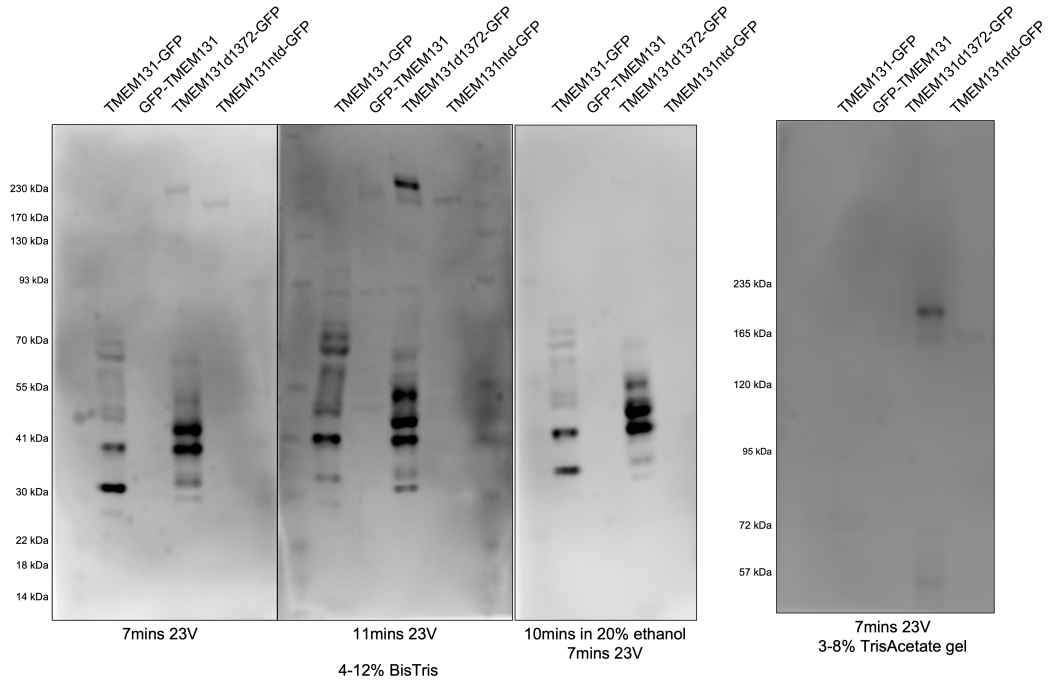

Figure S12: Higher molecular weight bands of TMEM131 can be resolved with longer blotting time. The same 4 samples were loaded into 3 wells each of a 17-well 4-12% BisTris gel separated by two lanes loaded with protein ladder. After running for 55 minutes at 200V the gel was cut between the ladders and blotted using the iBlot2 with different procedures. One third was blotted with 7mins at 23V, one third with 11mins at 23V and a third with 7mins at 23V after soaking in 20% ethanol for 10mins. Ethanol reportedly removes contaminating buffer salts and shrinks the gel to its final size before blotting. A 3-8% gel of the same samples was run at the same time and blotted with 7mins at 23V. All blots were stained with Mouse anti-HA and Rabbit anti-GFP primary antibodies then goat anti-Mouse af647 and goat anti-Rabbit cf770 secondary antibodies. Only the cf770 channel shown here as the af647 channel appeared blank apart from the ladder.

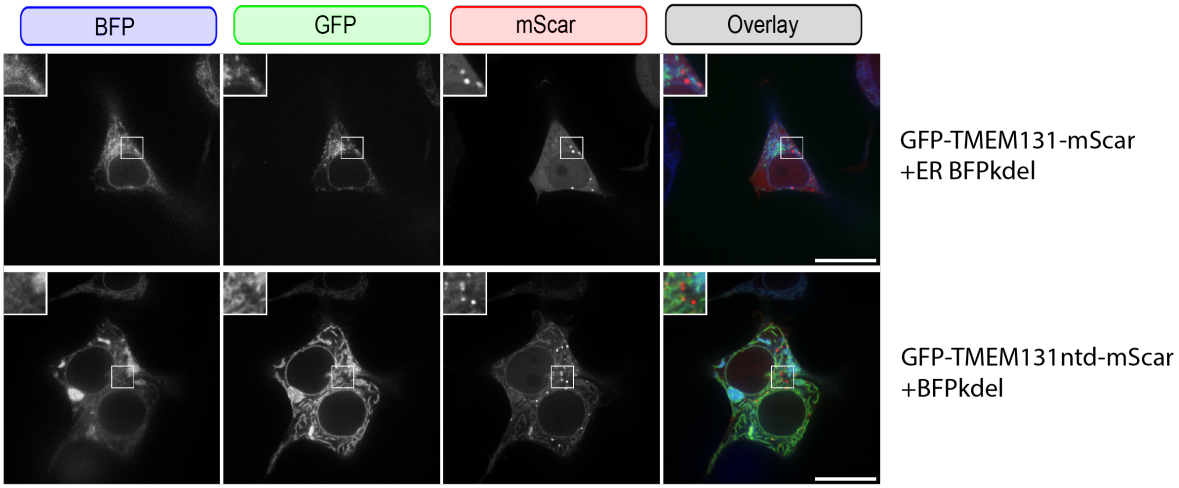

Figure S13: Images of TMEM131 and tailless TMEM131ntd with N-terminal GFP and C-terminal mScar tags.
GFP-TMEM131-mScar and GFP-TMEM131ntd-mScar were expressed in HEK cells with a BFPkdel ER marker and cells imaged under 100X magnification. Inserts show a 2x zoom of the selected area. There is some discordance between the two tags with the ntd construct despite the construct appearing intact on gel. Similar but more minor discordance was observed for TCRβ with the same tags as a component of the TCR. I suspect this is caused by misfolding of the EGFP domain in the ER lumen due to inappropriate disulphide bond formation. Cells transfected with the dual-labelled ntd construct had frequent organised smooth ER (OSER) whorl structures (the large BFP-positive structures). This is caused by fluorescent proteins on opposite membranes forming homo-oligomers.[375]

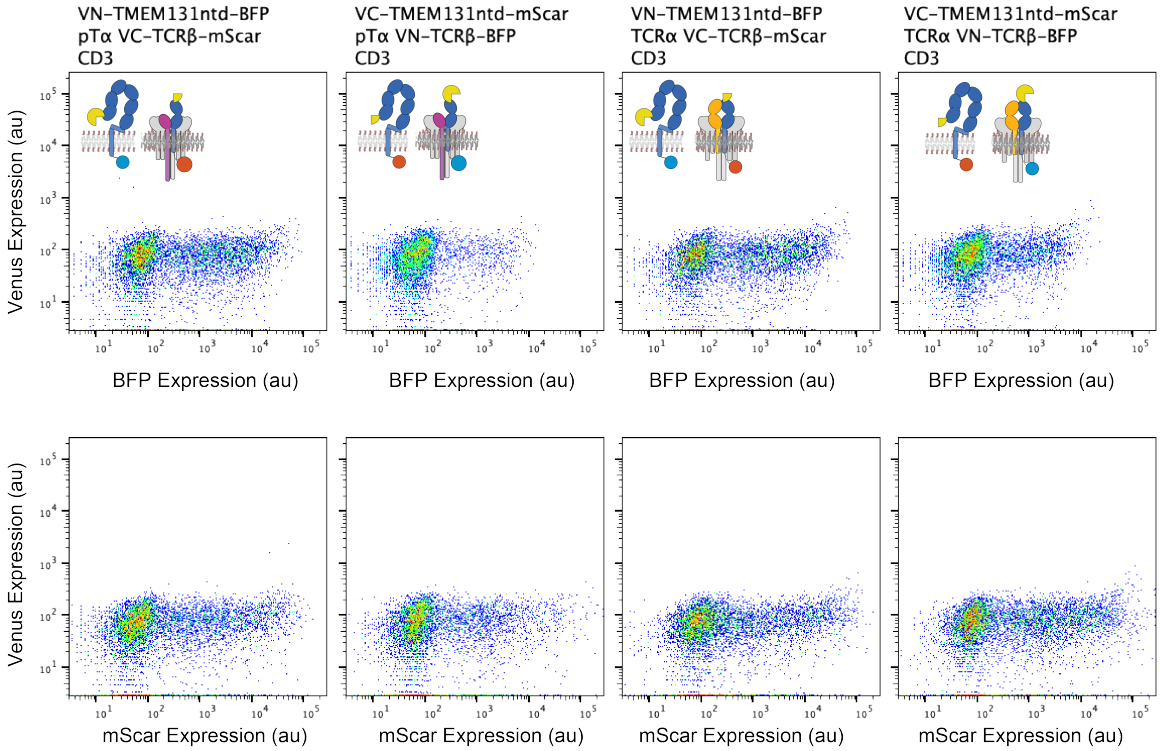

Figure S14: The split Venus bimolecular fluorescence complementation assay does not detect an interaction between TMEM131ntd and the pre-TCR or mTCR.
HEK cells were transfected with TMEM131ntd and the pre-TCR or mature TCR with the VN and VC components of Venus on the N-terminus and BFP/mScar fluorophores on the C-terminus. A plasmid without fluorophores encoding CD3$\gamma\delta\epsilon\zeta$ was included in all samples with volumes of all three plasmids being equal. Presented as dot plots of Venus against BFP (top) and mScar (bottom).

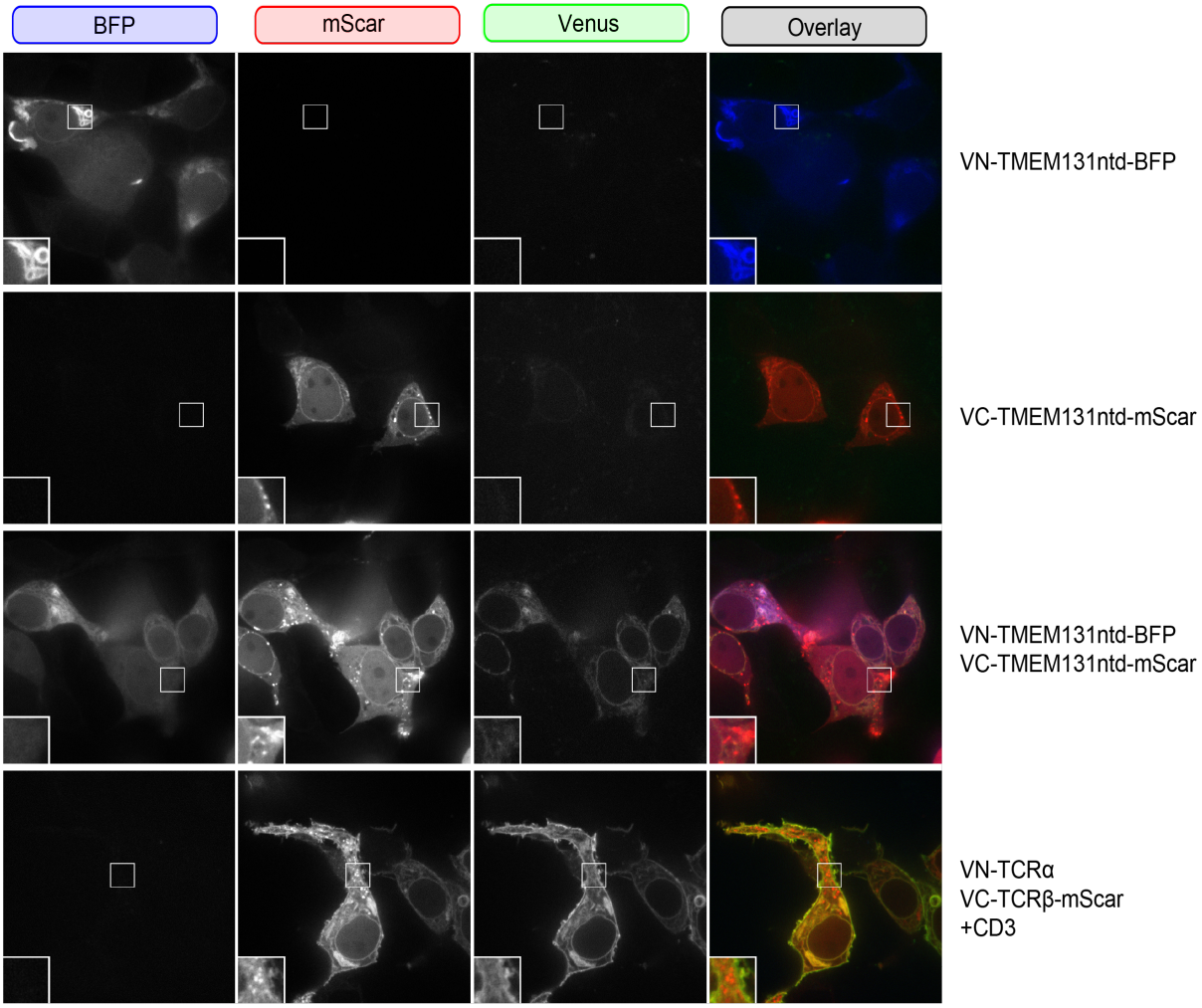

Figure S15: The tailless TMEM131 construct appears to form oligomers in the ER.
Proteins tagged with VN/VC components of split Venus were expressed in HEK cells and imaged under 100X. Inserts show a 2x zoom of the selected area.

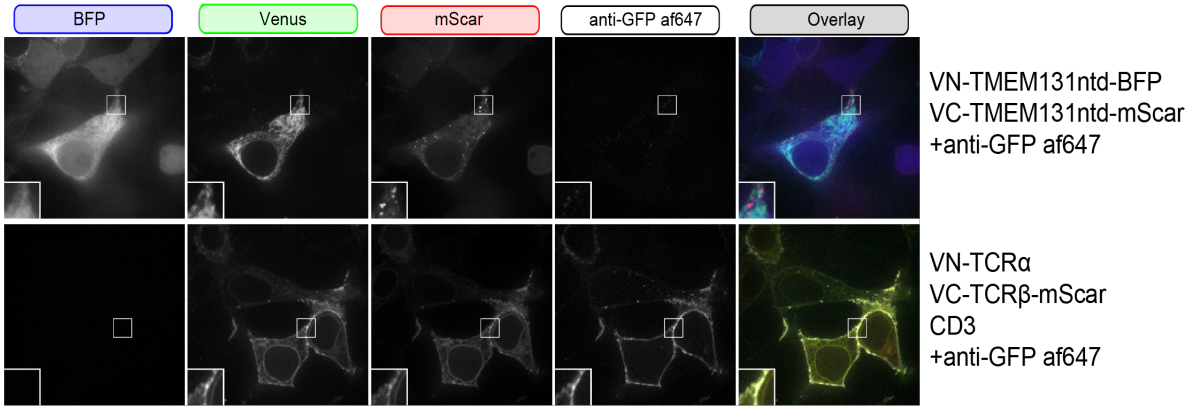

Figure S16: Oligomerised TMEM131ntd is not detected at the cell surface.
Proteins tagged with VN/VC components of split Venus were expressed in HEK cells and imaged under 100X with af647-conjugated anti-GFP nanobody added to the external medium. Inserts show a 2x zoom of the selected area.

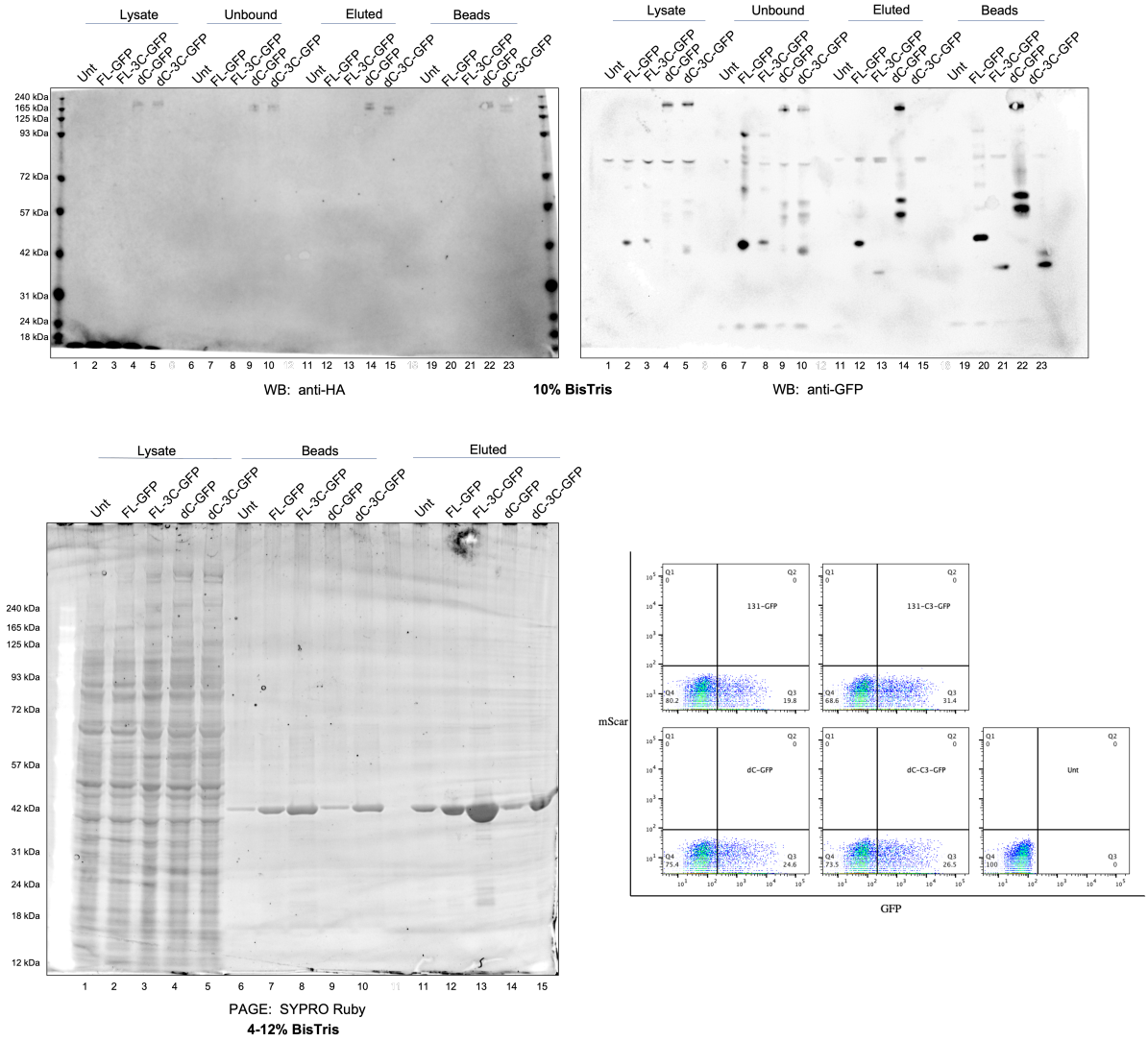

Figure S17: SYPRO-Ruby staining shows possible TMEM131 tail fragments.
HEK cells expressing HA-TMEM131-GFP (FL-GFP), HA-TMEM-3C-GFP (FL-3C-GFP), TMEM131dC-GFP (dC-GFP) and TMEM131dC-3C-GFP (dC-3C-GFP) were resuspended using trypsin then lysed and diluted. After a sample of the diluted lysate was taken, GFP-tagged proteins were enriched using GFP-trap agarose beads. A sample of the first supernatant after the beads were pelleted was taken as the unbound fraction then the beads washed extensively. The beads were then incubated overnight at 4°C with PreScission protease. This liquid was taken as the eluted fraction. The beads were then boiled in sample buffer to collect any remaining proteins.
Top: Lysate, Protease Eluted and Bead fractions run on a 10% BisTris gel labelled with Ms anti-HA and Rb anti-GFP primary antibodies and Gt anti-Ms af647 and Gt anti-Rb cf770 secondary antibodies. In this run a lot of GFP-tagged proteins were not bound by the GTA beads and the eluted fraction too closely resembled the bead fraction possibly as some beads were drawn up when the protease-treated liquid was removed.
Bottom left: Lysate, Protease Eluted and Bead fractions run on a 4-12% BisTris gel labelled SYPRO Ruby. Unlike later attempts no protein eluted from the beads using heat in sample buffer are detectable despite obvious bands in the western blot. The protease eluted fraction of TMEM131-3C-GFP (lane 13) seems to show bands consistent with the tail fragment of TMEM131 after GFP has been removed.
Bottom right: Flow cytometry of these cells before lysis. No antibody was used but the channels is used to separate the cells along a secondary axis.

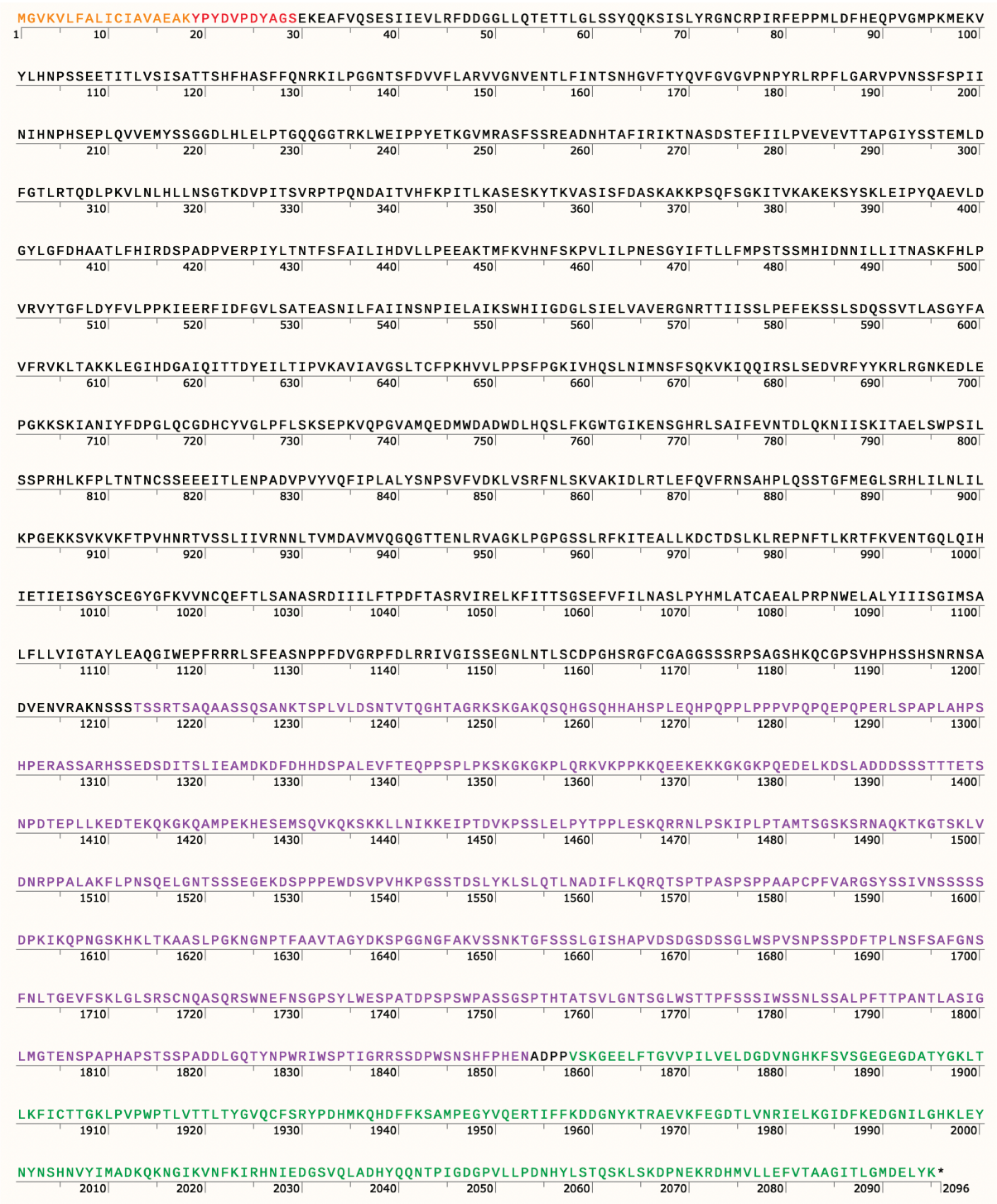

Figure S18: Annotated Protein Sequence of HA-TMEM131-GFP construct.
The gaussia luciferase signal peptide is shown in orange and the HA-tag shown in red. The sequence shown in purple is present in the TMEM131 full length version of this construct but not in the TMEM131dC construct. The sequence in green is the GFP tag on the C-terminus which is preceded by the residues ADPP which are encoded by the BamHI restriction site.

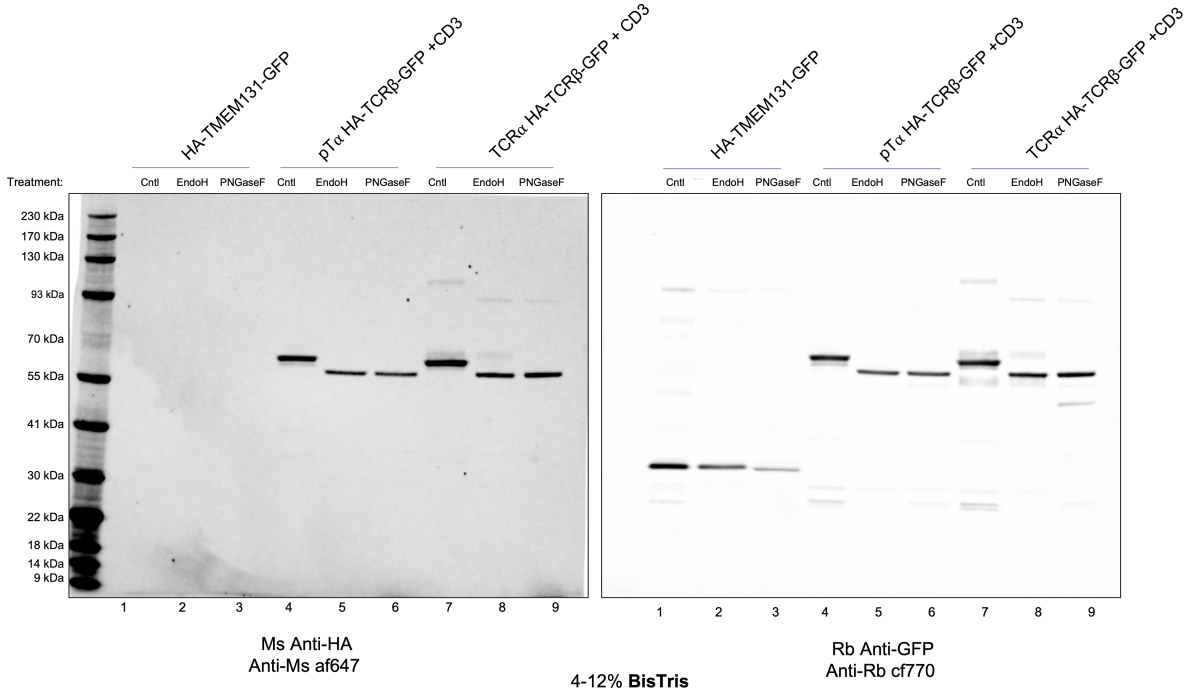

Figure S19: A larger fraction of the mTCR than pTCR is EndoH-insensitive.
TMEM131-GFP, as well as TCRβ-GFP as part of the pTCR and mTCR was enriched by GFP-pulldown then treated with EndoH, PNGaseF or untreated (Cntl). Proteins were run on a 4-12% BisTris gel, blotted then labelled with anti-HA and anti-GFP antibodies. An EndoH-insensitive band is apparent for the mTCR but not the pTCR. Too little TMEM131 seems to have been loaded into lanes 2 and 3 compared to lane 1 making it seem that bands have disappeared upon enzyme treatment rather than shifted.

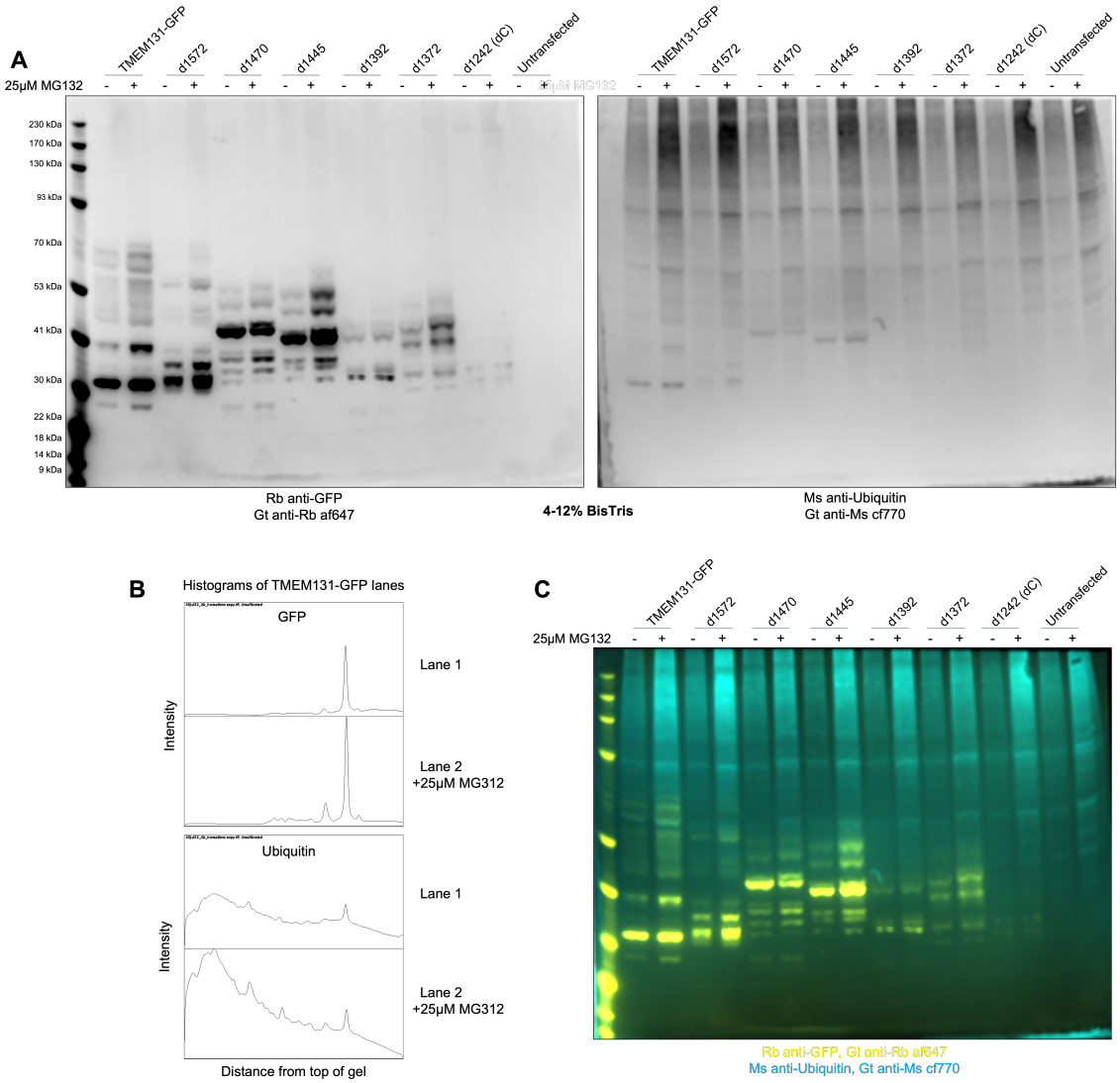

Figure S20: MG132 increases the molecular weight of fragments for a range of TMEM131 truncations.
A: HEK cells expressing TMEM131 constructs were produced in duplicate with half exposed to 50μM MG132 for 24h before cells were lysed. The blot was stained with Mouse anti-Ub and Rabbit anti-GFP primary antibodies then goat anti-Rabbit af647 and goat anti-Mouse cf770 secondary antibodies.
B: The histograms showing the average intensity along the length of the first two lanes of the blot in the GFP and ubiquitin channels.
C: Composite image with the anti-GFP channel shown in yellow and anti-Ub shown in cyan.

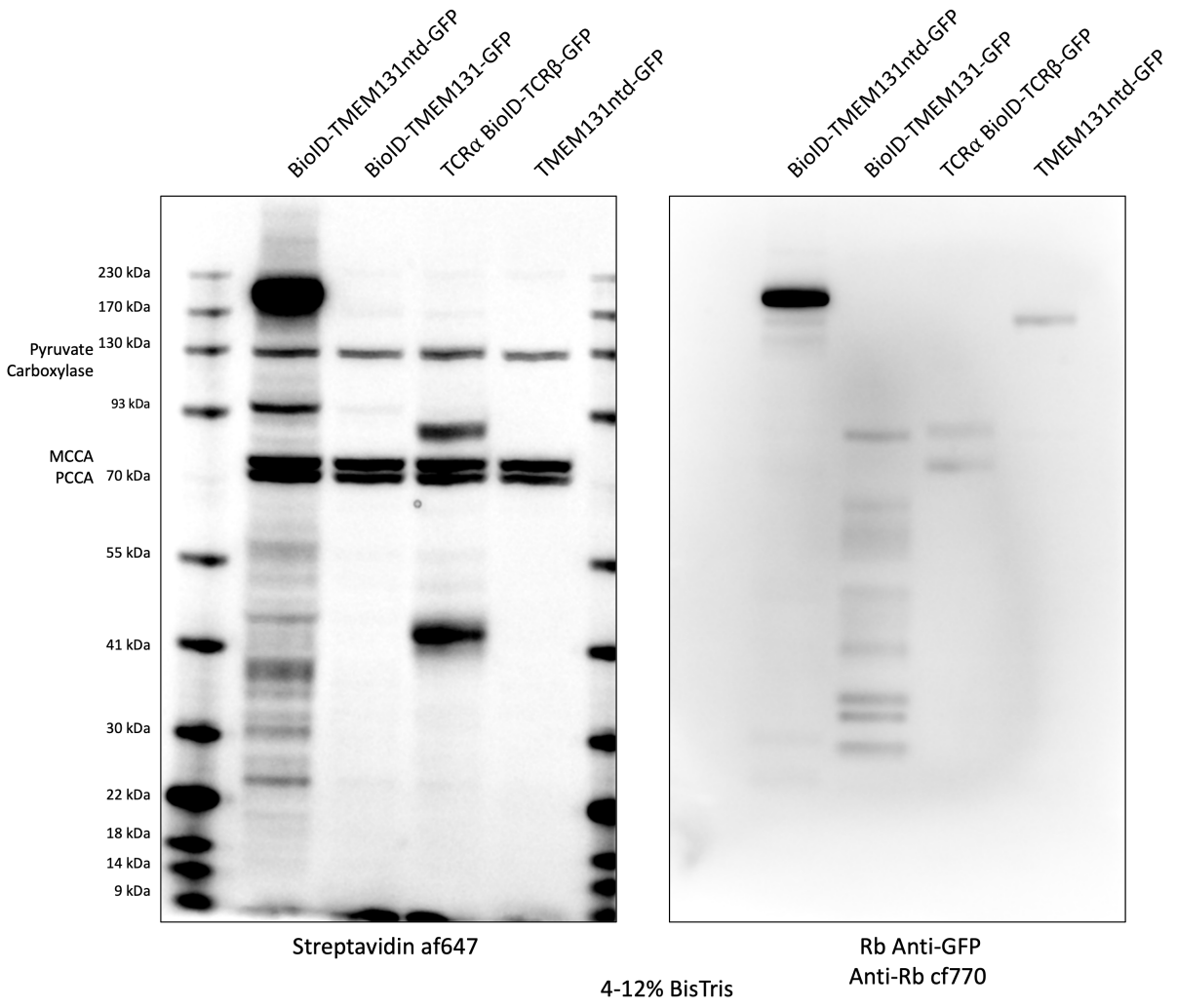

Figure S21: Western blot of the TMEM131 BioID assay samples submitted for analysis by mass spectrometry using an on-bead tryptic digest.
The indicated constructs were expressed in HEK cells and treated with 50µM biotin overnight. Cells were lysed and a sample of this lysate was run on a 4-12% BisTris gel. The gel was blotted then labelled with af647 streptavidin and anti-GFP antibodies. The remaining lysate was incubated with streptavidin-agarose beads and analysed by mass spectrometry using an on bead digest protocol. The results are discussed in section 4.6.2.

A.4 Mass spectrometry results

A.4.1 Proteins identified in a Nano-APEX assay with the pTCR and mTCR

Three T75 flasks were transfected with pTα GFP-TCRβ, pTα TCRβ-GFP and TCRα GFP-TCRβ with CD3$\gamma\delta\epsilon\zeta$-IRES-BFP using eight times the normal volume of GeneJuice, plasmid and serum free media. After 48h, the media in each flask was replaced with 10ml of warm media with 10µl of Nano-APEX. After 30 minutes for the Nanobody to bind and be internalised, the cells were washed three times in warm PBS by inverting the flask, removing the medium and inverting the flask again. Each flask was then incubated for 5 minutes with 5ml of medium containing 500µM biotin phenol. Labelling was initiated with 1mM hydrogen peroxide then cells washed twice with 10mls of Quenching solution and twice with 10ml of PBS. A final wash in quench solution was used to blast cells off the surface. Cells were lysed with 800µl of lysis buffer and clarified lysate frozen over a weekend.

The defrosted lysate was incubated with 40µl of streptavidin-agarose beads for 1h then the beads were washed five times in 500µl of lysis buffer. The beads were processed with the on-bead tryptic digest and C18 stage tip clean-up protocols. Peptides were identified by LC-MS then analysed with Scaffold using a protein confidence threshold of 95% and a peptide confidence threshold of 95%. A minimum of two peptides were required for identification. The values presented here show the number of unique peptides of each protein found in that sample. The following lane shows the number of occurrences that protein appears in the Contaminant Repository for Affinity Purification (CRAPome) database out of 716 experiments that followed a Proximity Dependent Biotinylation protocol.

Identified proteins	Alternate ID	MW (kDa)	GFP-mTCR	pTCR-GFP control	GFP-pTCR	CRAPome Exp Found/716
Keratin, type II cytoskeletal 1	KRT1	66	52	56	59	671
Keratin, type I cytoskeletal 9	KRT9	62	39	38	48	577
Keratin, type II cytoskeletal 2 epidermal	KRT2	65	42	50	54	628
Keratin, type I cytoskeletal 10	KRT10	59	36	42	44	616
Keratin, type II cytoskeletal 6B	KRT6B	60	34	49	55	570
Keratin, type II cytoskeletal 6C	KRT6C	60	0	48	57	515
Keratin, type II cytoskeletal 6A	KRT6A	60	35	50	57	515
Desmoplakin	DSP	332	35	60	75	328
Keratin, type I cytoskeletal 16	KRT16	51	33	41	45	516
Keratin, type II cytoskeletal 5	KRT5	62	27	38	47	508
Keratin, type I cytoskeletal 14	KRT14	52	29	40	46	523
Macrophage migration inhibitory factor	MIF	12	8	8	7	211
Pyruvate carboxylase, mitochondrial	PC	130	21	31	35	306
Heat shock 70 kDa protein 1B	HSPA1B	70	28	25	27	698
Poly [ADP-ribose] polymerase 1	PARP1	113	34	18	30	438
Peptidyl-prolyl cis-trans isomerase A	PPIA	18	17	14	18	439
Keratin, type I cytoskeletal 17	KRT17	48	11	27	34	465
Tubulin beta chain	TUBB	50	26	19	25	685
Peroxiredoxin-1	PRDX1	22	15	19	20	549
Tubulin beta-4B chain	TUBB4B	50	24	18	23	678
Propionyl-CoA carboxylase alpha chain, mitochondrial	PCCA	80	16	22	27	284
60S ribosomal protein L4	RPL4	48	32	12	22	444
Vimentin	VIM	54	27	15	22	543
Acetyl-CoA carboxylase 1	ACACA	266	13	23	30	296
Keratin, type II cytoskeletal 75	KRT75	60	0	0	16	522
Fatty acid synthase	FASN	273	40	11	17	436
Myosin-9	MYH9	227	40	9	10	448
Heterogeneous nuclear ribonucleoprotein M	HNRNPM	78	32	8	14	520
Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial	MCCC1	80	16	18	24	280
Protein RCC2	RCC2	56	18	15	19	336
Trifunctional purine biosynthetic protein adenosine-3	GART	108	14	16	23	336
Myosin-10	MYH10	229	45	5	7	379
Keratin, type II cytoskeletal 3	KRT3	64	0	12	14	360
Keratin, type II cytoskeletal 79	KRT79	58	0	12	13	470
Tubulin beta-2B chain	TUBB2B	50	18	12	17	675
Insulin-like growth factor 2 mRNA-binding protein 1	IGF2BP1	63	29	7	10	343
Keratin, type II cytoskeletal 2 oral	KRT76	66	0	0	13	426
Polyadenylate-binding protein	PABPC1	59	29	13	12	408
60S ribosomal protein L7a	RPL7A	30	23	11	15	398
Nucleolin	NCL	77	30	4	7	527
60S ribosomal protein L6	RPL6	33	20	11	13	432
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	88	18	10	11	578
Elongation factor 2	EEF2	95	25	12	14	488
Elongation factor 1-alpha 1	EEF1A1	50	14	15	15	653
40S ribosomal protein S4, X isoform	RPS4X	30	20	5	7	451
Probable ATP-dependent RNA helicase DDX17	DDX17	80	26	10	12	523
Heterogeneous nuclear ribonucleoprotein K	HNRNPK	51	19	12	14	581
40S ribosomal protein S8	RPS8	24	17	10	12	498
Polyadenylate-binding protein	PABPC4	68	25	8	11	376
40S ribosomal protein S3a	RPS3A	30	24	7	11	434
60S ribosomal protein L3	RPL3	46	19	7	8	402
Keratinocyte proline-rich protein	KPRP	64	8	12	12	81
40S ribosomal protein S11	RPS11	18	19	8	11	340
60S ribosomal protein L8	RPL8	28	16	8	9	390
Keratin, type II cytoskeletal 8	KRT8	54	0	6	0	545
Pyruvate kinase PKM	PKM	58	14	13	17	536
Ubiquitin-like modifier-activating enzyme 1	UBA1	118	8	14	23	301
Keratin, type II cuticular Hb4	KRT84	65	0	11	16	322
Splicing factor, proline- and glutamine-rich	SFPQ	76	19	7	10	484
40S ribosomal protein S6	RPS6	29	16	6	9	466
Histone H2A	hCG_2039566	18	8	6	6	-
60S ribosomal protein L13a (Fragment)	RPL13A	24	15	6	10	294
Polyadenylate-binding protein (Fragment)	PABPC4	60	20	0	0	376
Actin, cytoplasmic 1	ACTB	42	12	9	11	667
40S ribosomal protein S13	RPS13	17	16	7	8	311
Nucleolin (Fragment)	NCL	32	21	0	0	527
Heat shock cognate 71 kDa protein	HSPA8	71	13	13	12	703
ATP-dependent RNA helicase A	DHX9	141	27	3	9	470
Tubulin alpha-1B chain (Fragment)	TUBA1B	27	9	6	7	694
Keratin, type I cytoskeletal 13	KRT13	50	5	11	17	441
60S ribosomal protein L7	RPL7	29	15	10	9	335
60S ribosomal protein L13	RPL13	24	17	7	11	474
Polypyrimidine tract-binding protein 1	PTBP1	57	21	5	6	438
60S ribosomal protein L13a	RPL13A	24	15	6	9	294
60 kDa heat shock protein, mitochondrial	HSPD1	61	12	10	16	520
Junction plakoglobin	JUP	82	7	14	16	244
ATP-dependent RNA helicase DDX3X	DDX3X	81	17	6	10	498
Tubulin alpha-4A chain	TUBA4A	50	11	7	7	679
Keratin, type II cytoskeletal 4	KRT4	57	0	7	11	391
ATP-citrate synthase	ACLY	121	19	5	8	296
60S ribosomal protein L17 (Fragment)	RPL17	20	14	4	8	424
Filamin-A	FLNA	281	23	3	11	508
Probable ATP-dependent RNA helicase DDX5	DDX5	69	24	5	9	528
Profilin	PFN2	21	10	10	11	186
Histone H2A type 1-B/E	HIST1H2AB	14	7	5	5	550
Desmoglein-1	DSG1	114	4	10	14	116
Non-POU domain-containing octamer-binding protein	NONO	54	16	3	6	505
60S ribosomal protein L10	RPL10	25	18	4	7	369
Poly(rC)-binding protein 1	PCBP1	37	13	6	9	472
40S ribosomal protein S14 (Fragment)	RPS14	16	9	7	7	514
Heterogeneous nuclear ribonucleoprotein H	HNRNPH1	47	17	6	6	558
40S ribosomal protein S2	RPS2	31	15	5	8	447
Histone H2B type 1-L	HIST1H2BL	14	7	8	8	515
Tubulin alpha-3E chain	TUBA3E	50	10	5	7	686
Interleukin enhancer-binding factor 3	ILF3	95	21	0	4	379
Insulin-like growth factor 2 mRNA-binding protein 3	IGF2BP3	64	20	3	4	288
Keratin, type I cytoskeletal 15	KRT15	49	0	0	12	452
Endoplasmic reticulum chaperone BiP	HSPA5	72	15	6	8	661
40S ribosomal protein S3	RPS3	27	17	5	8	491
Keratin, type II cytoskeletal 1b	KRT77	62	0	7	6	592
Eukaryotic translation initiation factor 3 subunit D	EIF3D	64	16	3	8	249
Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1	37	14	3	6	536
Histone H2B type 1-O	HIST1H2BO	14	7	6	8	513
Peroxiredoxin-2	PRDX2	22	6	11	11	482
40S ribosomal protein S18	RPS18	18	15	5	7	399
60S ribosomal protein L24	RPL24	18	10	6	7	447
Cysteine and glycine-rich protein 2	CSRP2	21	7	8	9	217
60S ribosomal protein L18	RPL18	19	11	6	6	399
ATP synthase subunit alpha, mitochondrial	ATP5F1A	60	9	9	9	-
40S ribosomal protein S20	RPS20	13	7	4	5	383
60S ribosomal protein L10	RPL10	20	15	0	6	369
Far upstream element-binding protein 2	KHSRP	73	14	7	12	403
C-1-tetrahydrofolate synthase, cytoplasmic	MTHFD1	105	10	5	9	311
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	36	10	8	10	458
60S ribosomal protein L14	RPL14	23	10	5	6	333
60S ribosomal protein L23	RPL23	15	12	4	5	507
60S ribosomal protein L10a	RPL10A	25	13	5	7	280

Ras GTPase-activating protein-binding protein 1	G3BP1	52	10	3	4	353
Serine/arginine repetitive matrix protein 2	SRRM2	300	20	3	5	324
Ribosomal protein L19	RPL19	23	10	4	6	458
Nucleophosmin	NPM1	33	11	3	4	554
Histone H4	H4C1	11	9	6	8	-
Heterogeneous nuclear ribonucleoproteins C1/C2	HNRNPC	32	10	7	9	444
Histone H1.3	HIST1H1D	22	10	8	7	600
Keratin, type II cuticular Hb6	KRT86	53	0	8	15	132
L-lactate dehydrogenase B chain	LDHB	37	8	7	12	340
T-complex protein 1 subunit beta	CCT2	57	5	7	11	428
Cofilin-1	CFL1	23	7	6	9	496
60S ribosomal protein L12	RPL12	18	10	5	6	397
Receptor of activated protein C kinase 1	RACK1	35	7	8	10	352
Heterogeneous nuclear ribonucleoprotein D0 (Fragment)	HNRNPD	30	11	2	4	489
Poly(rC)-binding protein 2	PCBP2	34	12	0	9	475
60S ribosomal protein L18 (Fragment)	RPL18	19	10	0	0	399
Histone H1.2	HIST1H1C	21	9	0	0	614
L-lactate dehydrogenase A chain	LDHA	37	12	7	6	304
40S ribosomal protein S9	RPS9	23	14	2	5	388
Heterogeneous nuclear ribonucleoproteins C1/C2	HNRNPC	34	10	0	8	444
Ribosome-binding protein 1	RRBP1	152	21	0	0	172
Profilin-2	PFN2	15	0	0	7	186
Keratin, type I cuticular Ha1	KRT31	47	0	9	12	356
Homeobox protein cut-like 1	CUX1	164	9	4	9	98
60S ribosomal protein L27	RPL27	16	8	4	6	324
60S ribosomal protein L18a	RPL18A	21	10	3	6	300
Isochorismatase domain-containing protein 2	ISOC2	22	6	7	7	90
Keratin, type II cuticular Hb5	KRT85	56	0	9	15	130
D-3-phosphoglycerate dehydrogenase	PHGDH	56	6	9	9	394
Serum albumin	ALB	69	3	6	8	239
40S ribosomal protein S19	RPS19	16	10	6	7	334
GMP synthase [glutamine-hydrolyzing]	GMPS	77	7	8	8	229
X-ray repair cross-complementing protein 6	XRCC6	64	11	2	4	388
Eukaryotic translation initiation factor 2 subunit 3	EIF2S3	51	15	2	5	270
ATP-dependent RNA helicase DHX30	DHX30	131	20	0	5	188
Poly(rC)-binding protein 2 (Fragment)	PCBP2	32	11	0	7	475
Heterogeneous nuclear ribonucleoprotein M (Fragment)	HNRNPM	40	12	2	0	520
Keratin, type II cytoskeletal 80	KRT80	51	0	3	4	281
60S ribosomal protein L30 (Fragment)	RPL30	13	10	4	7	255
60S ribosomal protein L9	RPL9	22	12	3	4	390
40S ribosomal protein S15a	RPS15A	15	8	5	4	401
Protein disulfide-isomerase A6	PDIA6	48	8	7	7	357
Keratin, type II cytoskeletal 78	KRT78	57	5	7	7	191
Transcription intermediary factor 1-beta	TRIM28	89	12	2	4	466
Heterogeneous nuclear ribonucleoprotein F	HNRNPF	46	11	4	4	528
40S ribosomal protein S16	RPS16	16	9	2	6	411
Stress-70 protein, mitochondrial	HSPA9	74	10	6	8	552
60S ribosomal protein L27a	RPL27A	17	7	6	5	430
60S ribosomal protein L36	RPL36	12	10	0	7	221
Tubulin beta-6 chain	TUBB6	50	8	0	0	617
Lamina-associated polypeptide 2, isoforms beta/gamma	TMPO	46	10	5	5	386
Ribonuclease inhibitor	RNH1	50	9	4	5	70
T-complex protein 1 subunit alpha	TCP1	60	8	4	7	432
40S ribosomal protein S12	RPS12	15	7	5	6	382
Desmocollin-1	DSC1	100	3	6	9	71
Structural maintenance of chromosomes protein 1A	SMC1A	143	16	2	0	248
60S ribosomal protein L21	RPL21	19	9	0	4	329
Glial fibrillary acidic protein (Fragment)	GFAP	8	2	2	3	367
Plakophilin-1	PKP1	83	0	7	12	41
RNA-binding motif protein, X chromosome	RBMX	32	7	4	4	377
60S acidic ribosomal protein P0	RPLP0	34	11	3	6	436
Lamin-B1	LMNB1	66	5	8	11	284
Ubiquitin-40S ribosomal protein S27a	RPS27A	18	8	3	3	478
Chromobox protein homolog 3	CBX3	21	7	5	7	309
Interleukin enhancer-binding factor 2	ILF2	43	10	3	2	313
Insulin-like growth factor 2 mRNA-binding protein 2	IGF2BP2	67	11	3	3	257
Thioredoxin	TXN	12	6	4	4	432
40S ribosomal protein S23	RPS23	16	8	4	5	413
Leucine-rich repeat-containing protein 59	LRRC59	35	10	4	3	175
BTB/POZ domain-containing protein KCTD12	KCTD12	36	10	4	4	74
Matrin-3	MATR3	95	8	2	4	441
tRNA-splicing ligase RtcB homolog	RTCB	55	12	2	2	290
Eukaryotic initiation factor 4A-I	EIF4A1	46	9	3	6	479
40S ribosomal protein S7	RPS7	22	9	2	2	368
Elongation factor Tu, mitochondrial	TUFM	50	10	0	4	408
Histone H2AX	H2AFX	15	5	3	4	484
CAD protein	CAD	236	15	0	2	387
Eukaryotic translation initiation factor 4 gamma 1	EIF4G1	172	13	0	3	293
Serine/arginine-rich-splicing factor 1	SRSF1	28	7	3	6	280
60S ribosomal protein L11	RPL11	20	7	4	4	495
DNA ligase 3	LIG3	113	8	2	6	101
60S ribosomal protein L28	RPL28	16	7	2	3	316
Alpha-enolase	ENO1	47	4	6	8	478
40S ribosomal protein S25	RPS25	14	9	3	4	356
Heterogeneous nuclear ribonucleoprotein U-like protein 1	HNRNPUL1	86	11	3	2	254
Y-box-binding protein 1	YBX1	36	12	2	5	448
THO complex subunit 4	ALYREF	27	8	2	5	390
La-related protein 1	LARP1	124	9	2	3	251
Structural maintenance of chromosomes protein 3	SMC3	142	9	0	3	194
NanobodyAPEX	**NanoAPEX**	45	10	2	0	-
40S ribosomal protein S24	RPS24	15	10	2	0	353
Heterogeneous nuclear ribonucleoprotein A1	HNRNPA1	39	7	2	3	552
Fructose-bisphosphate aldolase	ALDOA	40	5	6	8	301
40S ribosomal protein S5 (Fragment)	RPS5	22	7	3	4	321
60S ribosomal protein L15	RPL15	24	6	0	3	365
60S ribosomal protein L35	RPL35	15	6	0	4	302
Nucleolar transcription factor 1	UBTF	87	7	2	5	68
Histone H2A.Z	H2AFZ	14	3	2	3	491
Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	52	9	0	4	459
60S ribosomal protein L23a (Fragment)	RPL23A	18	9	3	3	418
60S ribosomal protein L23a (Fragment)	RPL23A	19	9	0	0	418
Serine/arginine-rich-splicing factor 3	SRSF3	10	4	3	2	392
Core histone macro-H2A.1	H2AFY	40	7	4	6	148
Elongation factor 1-gamma	EEF1G	50	5	4	6	461
Nucleolar RNA helicase 2	DDX21	87	8	3	4	438
Heterogeneous nuclear ribonucleoprotein L (Fragment)	HNRNPL	59	8	2	5	460
Cytoskeleton-associated protein 4	CKAP4	66	11	2	3	168
Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	MCCC2	61	6	0	5	251
Galectin-7	LGALS7	15	0	5	6	43
Heterogeneous nuclear ribonucleoprotein D-like	HNRNPDL	40	7	0	3	369
60S acidic ribosomal protein P2	RPLP2	12	9	2	0	440
Eukaryotic translation initiation factor 5B	EIF5B	139	8	0	3	200
Protein PRRC2A	PRRC2A	229	12	0	3	212
Ataxin-2-like protein	ATXN2L	113	12	0	2	305
Barrier-to-autointegration factor	BANF1	10	7	0	2	128
Fragile X mental retardation syndrome-related protein 1	FXR1	68	10	0	4	140
Histone H1.1	H1-1	22	4	4	0	-
Coatomer subunit alpha	COPA	136	10	0	0	202
Heterogeneous nuclear ribonucleoprotein R	HNRNPR	71	13	0	4	422

Histone H3	H3F3A	14	3	3	4	384
Protein-L-isoaspartate O-methyltransferase	PCMT1	30	4	4	4	299
T-complex protein 1 subunit eta	CCT7	59	3	3	6	360
40S ribosomal protein S28	RPS28	8	4	2	6	404
40S ribosomal protein S10	RPS10	19	4	4	2	336
Caprin-1	CAPRIN1	78	11	0	3	282
Heat shock protein HSP 90-beta	HSP90AB1	83	7	2	6	573
RNA-binding protein 39	RBM39	59	8	3	4	366
Double-stranded RNA-binding protein Staufen homolog 1	STAU1	55	13	0	2	228
ATPase family AAA domain-containing protein 3A	ATAD3A	71	9	0	3	229
Keratin, type I cytoskeletal 27	KRT27	50	0	0	3	449
Mitochondrial-processing peptidase subunit beta	PMPCB	54	3	3	4	109
Sodium/potassium-transporting ATPase subunit alpha-1	ATP1A1	113	8	3	6	335
T-complex protein 1 subunit delta	CCT4	58	4	4	7	433
Scaffold attachment factor B1	SAFB	103	6	4	4	206
3-hydroxyacyl-CoA dehydrogenase type-2	HSD17B10	27	5	3	4	290
RNA-binding protein 14	RBM14	69	7	3	6	293
60S ribosomal protein L38	RPL38	8	6	2	4	300
Filamin-B	FLNB	278	6	2	4	417
X-ray repair cross-complementing protein 5	XRCC5	83	7	2	3	428
ATP-dependent RNA helicase DDX1	DDX1	82	6	0	2	331
ATPase family AAA domain-containing protein 3B	ATAD3B	73	7	0	3	172
Microtubule-associated protein	MAP4	245	8	0	4	312
Kinesin-1 heavy chain	KIF5B	110	15	0	0	245
Acetyl-CoA carboxylase 2	ACACB	277	0	0	7	265
Myosin light polypeptide 6	MYL6	17	7	0	0	411
Immunoglobulin heavy constant alpha 1 (Fragment)	IGHA1	43	0	5	7	51
Myosin light polypeptide 6	MYL6	16	7	0	0	411
Helicase MOV-10	MOV10	114	13	0	0	87
Lactotransferrin	LTF	78	0	4	10	61
Keratin, type I cuticular Ha4	KRT34	49	0	5	8	42
F-actin-capping protein subunit beta	CAPZB	29	4	3	6	366
PHD finger-like domain-containing protein 5A	PHF5A	12	2	3	4	133
ADP/ATP translocase 2	SLC25A5	33	3	3	5	525
60S ribosomal protein L32 (Fragment)	RPL32	16	7	2	3	184
Bifunctional glutamate/proline--tRNA ligase	EPRS	171	7	2	4	393
Pre-mRNA-processing factor 19	PRPF19	55	7	0	2	262
Cell division cycle 5-like protein	CDC5L	92	13	0	0	209
Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	DHX15	91	9	0	0	432
Heterogeneous nuclear ribonucleoprotein H2	HNRNPH2	49	7	0	0	540
BolA-like protein 2	BOLA2B	17	3	3	5	250
Propionyl-CoA carboxylase beta chain, mitochondrial	PCCB	56	5	4	6	263
T-complex protein 1 subunit gamma	CCT3	61	4	4	5	367
T-complex protein 1 subunit theta	CCT8	60	4	3	8	451
ATP synthase subunit O, mitochondrial	ATP5PO	23	6	3	4	-
Annexin (Fragment)	ANXA2	20	2	4	6	398
CTP synthase	CTPS1	64	5	2	4	312
Low molecular weight phosphotyrosine protein phosphatase	ACP1	18	0	5	5	95
DNA replication licensing factor MCM3	MCM3	91	4	0	4	322
Leucine-rich repeat-containing protein 47	LRRC47	63	8	2	3	82
60S ribosomal protein L29	RPL29	19	6	0	3	399
FACT complex subunit SPT16	SUPT16H	120	6	0	2	175
60S ribosomal protein L22	RPL22	15	6	0	3	429
ELAV-like protein 1	ELAVL1	36	9	0	3	322
U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	245	9	0	2	333
Polymerase delta-interacting protein 3	POLDIP3	48	10	0	0	247
Y-box-binding protein 3	YBX3	40	8	0	0	363
60S ribosomal protein L26	RPL26	17	6	0	3	312
Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	SDHA	73	4	3	6	163
Zinc finger CCCH-type antiviral protein 1-like	ZC3HAV1L	33	4	3	4	21
Dermcidin	DCD	11	2	3	4	279
Creatine kinase B-type	CKB	43	7	2	3	308
Putative RNA-binding protein Luc7-like 2	LUC7L2	47	6	2	2	332
Splicing factor U2AF 65 kDa subunit	U2AF2	54	6	2	2	231
60S ribosomal protein L31	RPL31	14	6	2	0	335
Staphylococcal nuclease domain-containing protein 1	SND1	102	8	0	0	265
60S ribosomal protein L5 (Fragment)	RPL5	27	6	0	2	395
Annexin A1	ANXA1	39	0	4	5	68
116 kDa U5 small nuclear ribonucleoprotein component	EFTUD2	109	11	0	0	381
Ataxin-2-like protein	ATXN2L	113	9	0	0	305
Putative transferase CAF17, mitochondrial	IBA57	38	3	4	4	46
Thioredoxin-dependent peroxide reductase, mitochondrial	PRDX3	28	2	5	5	315
Guanine nucleotide-binding protein subunit beta-like protein 1	GNB1L	36	2	6	4	35
40S ribosomal protein S30	FAU	11	4	2	3	320
Fatty acid-binding protein 5	FABP5	15	2	5	6	106
T-complex protein 1 subunit zeta	CCT6A	58	3	2	5	413
Cancer-related nucleoside-triphosphatase	NTPCR	25	4	2	2	101
Serine/arginine-rich splicing factor 6	SRSF6	40	4	2	2	326
ATP-dependent 6-phosphofructokinase, platelet type	PFKP	86	2	0	4	201
RNA-binding protein FUS	FUS	53	5	0	2	428
Ubiquitin carboxyl-terminal hydrolase 10	USP10	87	7	0	0	114
Clathrin heavy chain	CLTC	192	8	0	0	406
Serine-threonine kinase receptor-associated protein	STRAP	38	7	0	3	319
Synaptic functional regulator FMR1	FMR1	71	5	0	0	131
ATP-binding cassette sub-family F member 1	ABCF1	96	9	0	2	288
Cold shock domain-containing protein E1	CSDE1	89	9	0	0	191
Ribosomal RNA processing protein 1 homolog B	RRP1B	84	10	0	0	214
5'-3' exoribonuclease 2	XRN2	109	7	0	0	276
Immunoglobulin heavy constant alpha 2 (Fragment)	IGHA2	37	0	0	5	49
Double-stranded RNA-specific adenosine deaminase	ADAR	136	8	0	0	235
Serine/arginine-rich-splicing factor 7	SRSF7	16	6	2	3	351
Profilin-1	PFN1	15	2	3	4	308
60S ribosomal protein L34	RPL34	13	4	2	2	247
NEDD8-conjugating enzyme Ubc12	UBE2M	21	2	4	2	140
E3 ubiquitin-protein ligase HUWE1	HUWE1	482	4	2	3	204
Ribose-phosphate pyrophosphokinase 1	PRPS1	31	3	2	2	225
Protein/nucleic acid deglycase DJ-1	PARK7	18	3	0	2	236
Heat shock protein beta-1	HSPB1	23	0	3	4	128
Plastin-3	PLS3	69	0	4	3	208
40S ribosomal protein S26	RPS26	13	5	0	2	375
Glutaredoxin-related protein 5, mitochondrial	GLRX5	17	0	3	4	20
Enhancer of mRNA-decapping protein 4	EDC4	152	5	0	2	163
DNA replication licensing factor MCM7	MCM7	81	7	0	2	328
RNA-binding protein Raly	RALY	32	5	0	3	116
39S ribosomal protein L12, mitochondrial	MRPL12	21	7	0	0	175
Guanine nucleotide-binding protein-like 3	GNL3	62	8	0	0	256
DNA topoisomerase 1	TOP1	91	7	0	2	285
Zinc finger CCCH-type antiviral protein 1	ZC3HAV1	101	7	0	2	210
Splicing factor 3B subunit 2	SF3B2	100	10	0	0	317
Hydroxysteroid dehydrogenase-like protein 2	HSDL2	45	3	3	4	79
Eukaryotic translation initiation factor 6	EIF6	27	3	3	4	216
40S ribosomal protein SA	RPSA	33	5	2	3	335
Phosphoribosylformylglycinamidine synthase	PFAS	145	4	2	3	166
Importin subunit alpha-1	KPNA2	58	3	2	4	367
Signal recognition particle 14 kDa protein	SRP14	15	3	2	2	338
Peroxiredoxin-6	PRDX6	25	3	2	3	388
S-methyl-5'-thioadenosine phosphorylase	MTAP	33	2	3	3	50
Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	45	3	3	4	161
Proteasome activator complex subunit 3	PSME3	27	2	3	5	198

60S ribosomal protein L36a	RPL36A	16	4	0	0	220
KH domain-containing, RNA-binding, signal transduction-associated protein 1	KHDRBS1	48	4	0	0	348
Eukaryotic translation initiation factor 2 subunit 1	EIF2S1	36	6	0	0	240
39S ribosomal protein L11, mitochondrial	MRPL11	21	8	0	0	38
Poly(U)-binding-splicing factor PUF60 (Fragment)	PUF60	57	3	0	0	307
Replication factor C subunit 1	RFC1	128	7	0	0	129
SNW domain-containing protein 1	SNW1	65	8	0	0	191
Transitional endoplasmic reticulum ATPase	VCP	89	5	0	0	314
Heterogeneous nuclear ribonucleoprotein A1 (Fragment)	HNRNPA1	19	4	0	0	552
Protein LYRIC	MTDH	64	8	0	0	179
Peroxisomal multifunctional enzyme type 2	HSD17B4	77	2	2	4	139
Transgelin-2	TAGLN2	22	2	2	4	410
Neutral alpha-glucosidase AB	GANAB	107	5	2	3	216
Aconitate hydratase, mitochondrial	ACO2	88	2	2	4	89
DNA helicase	MCM5	78	4	0	2	208
Transcription factor E3	TFE3	62	0	3	4	8
T-complex protein 1 subunit epsilon	CCT5	55	3	0	3	457
39S ribosomal protein L27, mitochondrial (Fragment)	MRPL27	15	4	0	2	41
Zinc-alpha-2-glycoprotein	AZGP1	34	0	3	4	58
Proliferation-associated protein 2G4	PA2G4	44	5	0	2	260
CDSN	CDSN	52	0	4	4	50
Heterogeneous nuclear ribonucleoprotein C-like 1	HNRNPCL1	32	4	0	3	377
ADP/ATP translocase 3	SLC25A6	33	3	2	0	495
Protein POF1B	POF1B	68	0	2	6	28
Transcription factor A, mitochondrial (Fragment)	TFAM	26	4	0	2	107
40S ribosomal protein S29	RPS29	7	3	0	3	223
Cell cycle and apoptosis regulator protein 2	CCAR2	103	8	0	2	297
Bifunctional polynucleotide phosphatase/kinase	PNKP	54	5	0	3	3
Interferon-induced, double-stranded RNA-activated protein kinase	EIF2AK2	62	8	0	0	76
39S ribosomal protein L15, mitochondrial	MRPL15	33	8	0	0	56
Slit homolog 2 protein	SLIT2	170	8	0	0	3
cDNA FLJ60124, highly similar to Mitochondrial dicarboxylate carrier		48	7	0	0	-
Myosin-14	MYH14	228	7	0	0	318
Voltage-dependent anion-selective channel protein 2 (Fragment)	VDAC2	30	2	2	3	256
Nucleoside diphosphate kinase A	NME1	15	2	2	3	320
Serine/arginine-rich splicing factor 10	SRSF10	31	2	2	2	150
Protein quaking	QKI	38	4	2	2	25
Protein S100-A14	S100A14	12	2	2	3	16
Adenosylhomocysteinase	AHCY	48	3	0	4	232
Protein disulfide-isomerase A3	PDIA3	57	0	2	4	281
Filaggrin-2	FLG2	248	0	2	3	153
Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	SDHB	32	0	3	3	87
Heat shock protein 75 kDa, mitochondrial	TRAP1	80	2	2	3	478
Nuclear autoantigenic sperm protein	NASP	85	0	2	3	327
Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	PPP2R2A	52	5	0	0	304
Programmed cell death 6-interacting protein	PDCD6IP	96	0	3	3	227
Histone H1x	H1FX	22	4	0	0	257
60S ribosomal protein L35a	RPL35A	13	4	0	3	220
RNA-binding protein 4	RBM4	40	6	0	0	82
Treacle protein	TCOF1	152	4	0	0	332
Tyrosine-protein kinase CSK	CSK	51	0	0	5	53
Protein transport protein Sec16A	SEC16A	252	7	0	0	179
Polypyrimidine tract-binding protein 3	PTBP3	60	3	0	0	182
Plasminogen activator inhibitor 1 RNA-binding protein	SERBP1	45	3	0	0	468
Golgin subfamily B member 1	GOLGB1	376	5	0	0	30
Serine/arginine-rich splicing factor 5	SRSF5	31	5	0	0	260
Thyroid hormone receptor-associated protein 3	THRAP3	104	8	0	0	382
RNA cytosine C(5)-methyltransferase NSUN2	NSUN2	86	9	0	0	254
Protein PRRC2C	PRRC2C	309	5	0	0	223
Ribonucleases P/MRP protein subunit POP1	POP1	115	7	0	0	73
Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	37	3	2	3	239
Electron transfer flavoprotein subunit beta	ETFB	28	2	2	2	74
Jupiter microtubule-associated homolog 2	JPT2	19	3	2	3	242
60S ribosomal protein L37a	RPL37A	8	3	2	2	200
Isochorismatase domain-containing protein 1	ISOC1	32	2	3	3	40
Crk-like protein	CRKL	34	2	2	3	218
S-adenosylmethionine synthase isoform type-2	MAT2A	44	2	2	3	228
Phenylalanine--tRNA ligase beta subunit	FARSB	66	4	2	0	139
Proliferating cell nuclear antigen	PCNA	29	0	2	2	268
Cystatin-B	CSTB	11	2	0	3	164
39S ribosomal protein L4, mitochondrial (Fragment)	MRPL4	34	2	0	2	72
Threonine--tRNA ligase, mitochondrial	TARS2	81	2	0	3	18
HCG1984214, isoform CRA_a	hCG_1984214	26	4	0	3	64
Multifunctional protein ADE2 (Fragment)	PAICS	46	3	0	2	322
60S ribosomal protein L39	RPL39	6	2	0	0	162
Forkhead box protein C1	FOXC1	57	0	0	2	51
Signal recognition particle subunit SRP68	SRP68	71	5	0	0	192
Deoxyuridine 5'-triphosphate nucleotidohydrolase, mitochondrial	DUT	27	0	2	4	236
Cleavage stimulation factor subunit 1	CSTF1	48	0	2	2	46
Heterogeneous nuclear ribonucleoprotein A0	HNRNPA0	31	4	0	0	405
Ribosomal L1 domain-containing protein 1	RSL1D1	55	4	0	0	277
Chromodomain-helicase-DNA-binding protein 4	CHD4	220	2	0	2	256
ATP synthase subunit beta, mitochondrial	ATP5F1B	57	0	0	2	-
Protein S100-A9	S100A9	13	0	0	2	116
Histone-binding protein RBBP7	RBBP7	47	3	0	0	370
Protein-glutamine gamma-glutamyltransferase E	TGM3	77	0	3	3	47
40S ribosomal protein S27	RPS27L	11	2	0	2	372
Myb-binding protein 1A	MYBBP1A	149	5	0	0	301
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	RPN1	69	5	0	2	246
Phosphoglycerate kinase 1	PGK1	45	0	3	2	200
Tropomyosin alpha-3 chain	TPM3	26	0	0	3	354
28S ribosomal protein S11, mitochondrial	MRPS11	21	4	0	0	60
DnaJ homolog subfamily C member 9	DNAJC9	30	4	0	2	86
RNA-binding protein 28	RBM28	86	5	0	0	158
40S ribosomal protein S15	RPS15	14	3	0	0	239
Cleavage and polyadenylation-specificity factor subunit 6	CPSF6	52	5	0	0	299
Keratin, type I cuticular Ha8	KRT38	50	0	2	2	349
AP-1 complex subunit mu-1	AP1M1	40	4	0	0	21
DNA helicase	MCM4	101	4	0	0	293
Prohibitin	PHB	30	5	0	0	269
RNA-binding protein 10	RBM10	104	5	0	0	327
Ras GTPase-activating protein-binding protein 2	G3BP2	54	5	0	0	214
Eukaryotic initiation factor 4A-III	EIF4A3	45	3	0	0	418
Zinc finger CCCH domain-containing protein 11A	ZC3H11A	89	7	0	0	161
pre-rRNA 2'-O-ribose RNA methyltransferase FTSJ3	FTSJ3	97	6	0	0	184
Negative elongation factor E	NELFE	43	6	0	0	152
Endoribonuclease Dicer	DICER1	219	7	0	0	70
G-rich sequence factor 1 (Fragment)	GRSF1	51	7	0	0	114
Heterogeneous nuclear ribonucleoprotein A/B	HNRNPAB	30	6	0	0	442
Myotrophin	MTPN	6	2	2	3	124
Vesicle-fusing ATPase	NSF	82	0	3	2	98
Skin-specific protein 32	XP32	26	0	2	3	54
Destrin	DSTN	19	2	2	2	131
Eukaryotic translation initiation factor 5A (Fragment)	EIF5A2	13	2	0	2	314
Histone-binding protein RBBP4	RBBP4	48	3	0	2	393
Mitotic checkpoint protein BUB3 (Fragment)	BUB3	32	3	0	0	298
Multifunctional methyltransferase subunit TRM112-like protein	TRMT112	12	2	0	3	49
Nucleosome assembly protein 1-like 1	NAP1L1	41	4	0	0	374
Ataxin-2 (Fragment)	ATXN2	117	2	0	2	116

Heterogeneous nuclear ribonucleoprotein A3	HNRNPA3	40	3	0	0	447
RuvB-like 1	RUVBL1	50	4	0	0	378
Enoyl-CoA delta isomerase 2, mitochondrial	ECI2	40	2	0	2	25
Eukaryotic translation initiation factor 3 subunit C-like protein	EIF3CL	105	3	0	0	274
Heterochromatin protein 1-binding protein 3	HP1BP3	61	3	0	0	214
Small proline-rich protein 2E	SPRR2E	8	0	2	3	28
Adenylate kinase 2, mitochondrial	AK2	26	0	2	3	142
Desmocollin-3	DSC3	100	0	2	3	15
Chromobox protein homolog 5	CBX5	22	3	0	3	100
Ankyrin repeat domain-containing protein 17 (Fragment)	ANKRD17	263	5	0	0	129
Chromatin target of PRMT1 protein	CHTOP	26	3	0	2	135
Splicing factor U2AF 35 kDa subunit-like protein	U2AF1L5	28	2	0	2	351
60S acidic ribosomal protein P1	RPLP1	12	3	0	0	378
28S ribosomal protein S31, mitochondrial	MRPS31	45	5	0	0	92
Pre-rRNA-processing protein TSR1 homolog	TSR1	92	2	0	0	148
High density lipoprotein binding protein (Vigilin), isoform CRA_a	HDLBP	141	3	0	0	218
Cleavage and polyadenylation specificity factor subunit 5	NUDT21	26	4	0	0	295
14-3-3 protein theta (Fragment)	YWHAQ	17	2	0	0	410
28S ribosomal protein S7, mitochondrial	MRPS7	32	4	0	0	76
E3 ubiquitin/ISG15 ligase TRIM25	TRIM25	51	6	0	0	136
Cellular tumor antigen p53	TP53	44	3	0	2	131
Nuclear mitotic apparatus protein 1	NUMA1	236	5	0	0	268
Regulator of nonsense transcripts 1	UPF1	124	6	0	0	117
Fragile X mental retardation syndrome-related protein 2	FXR2	74	4	0	0	139
Liprin-alpha-1 (Fragment)	PPFIA1	139	6	0	0	14
Apoptotic chromatin condensation inducer in the nucleus	ACIN1	147	4	0	0	206
Kinesin-like protein KIF2A	KIF2A	80	5	0	0	105
UPF0488 protein C8orf33	C8orf33	25	5	0	0	59
ADP-ribose glycohydrolase MACROD1	MACROD1	36	2	2	2	13
ATP synthase subunit gamma, mitochondrial	ATP5F1C	33	2	0	2	-
Elongation factor 1-beta	EEF1B2	25	0	2	2	379
DNA-dependent protein kinase catalytic subunit	PRKDC	469	2	0	3	406
EBNA1 binding protein 2, isoform CRA_d	EBNA1BP2	41	3	0	2	137
F-actin-capping protein subunit alpha-1	CAPZA1	33	2	2	2	318
39S ribosomal protein L13, mitochondrial	MRPL13	21	2	0	3	49
Growth arrest and DNA damage-inducible proteins-interacting protein 1	GADD45GIP1	25	3	2	0	57
28S ribosomal protein S22, mitochondrial	MRPS22	41	4	0	0	90
Exosome complex exonuclease RRP44	DIS3	109	4	0	0	114
GTP-binding protein 1	GTPBP1	72	3	0	0	47
Serpin H1	SERPINH1	36	3	0	0	281
Heat shock protein 105 kDa	HSPH1	97	0	2	2	347
Electron transfer flavoprotein subunit alpha, mitochondrial	ETFA	24	2	0	0	148
Density-regulated protein	DENR	18	0	0	3	52
39S ribosomal protein L45, mitochondrial	MRPL45	35	3	0	0	41
Lysozyme	LYZ	15	0	2	2	74
Host cell factor 1	HCFC1	209	2	0	0	278
28S ribosomal protein S29, mitochondrial	DAP3	46	3	0	0	96
39S ribosomal protein L39, mitochondrial (Fragment)	MRPL39	34	2	0	0	77
Probable ATP-dependent RNA helicase DDX6	DDX6	54	3	0	0	303
Serine/arginine-rich splicing factor 11	SRSF11	54	5	0	0	243
Microtubule-associated protein 1B	MAP1B	271	2	0	0	260
14-3-3 protein zeta/delta (Fragment)	YWHAZ	16	0	0	2	410
Catalase	CAT	60	0	3	3	57
Ubiquitin-conjugating enzyme E2 D3	UBE2D3	17	0	3	3	96
GTP-binding nuclear protein Ran	RAN	15	3	0	2	440
Ubiquitin-fold modifier-conjugating enzyme 1	UFC1	19	0	3	2	12
Unconventional myosin-VI	MYO6	145	4	0	2	148
Squamous cell carcinoma antigen recognized by T-cells 3	SART3	110	3	0	0	198
Glutathione reductase, mitochondrial	GSR	56	0	2	3	109
BAG family molecular chaperone regulator 2	BAG2	24	5	0	0	247
39S ribosomal protein L28, mitochondrial	MRPL28	30	4	0	0	42
Probable ATP-dependent RNA helicase DDX23	DDX23	96	4	0	0	144
Developmentally-regulated GTP-binding protein 1	DRG1	41	5	0	0	172
Replication protein A 70 kDa DNA-binding subunit	RPA1	68	4	0	0	242
DNA mismatch repair protein Msh6	MSH6	153	4	0	0	222
Zinc finger CCCH domain-containing protein 4 (Fragment)	ZC3H4	96	3	0	0	134
ATP-dependent RNA helicase DDX39A	DDX39A	49	2	0	0	404
RNA-binding protein 3	RBM3	17	2	0	0	176
U4/U6.U5 tri-snRNP-associated protein 1	SART1	90	5	0	0	233
Exosome component 10	EXOSC10	101	3	0	0	210
Signal recognition particle subunit SRP72	SRP72	75	5	0	0	168
Cytoplasmic dynein 1 heavy chain 1	DYNC1H1	532	4	0	0	345
Nuclear fragile X mental retardation-interacting protein 2	NUFIP2	76	3	0	0	223
Protein flightless-1 homolog	FLII	145	5	0	0	81
mRNA turnover protein 4 homolog	MRTO4	28	4	0	0	128
HA-TCRb-GFP	**HA-TCRb-GFP**	62	4	0	0	-
Structural maintenance of chromosomes protein 2	SMC2	136	5	0	0	194
ATP synthase subunit d, mitochondrial	ATP5PD	9	0	0	2	-
Protein S100-A8	S100A8	11	0	2	2	79
Glutathione S-transferase LANCL1 (Fragment)	LANCL1	22	0	2	2	94
Inosine-5'-monophosphate dehydrogenase 2 (Fragment)	IMPDH2	51	2	0	0	195
Lipocalin-1	LCN1	19	0	0	3	36
Haloacid dehalogenase-like hydrolase domain-containing protein 3	HDHD3	28	0	0	2	11
TAR DNA-binding protein 43	TARDBP	45	2	0	0	324
TBC1 domain family member 4	TBC1D4	147	2	0	0	113
Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial	PDHA1	43	0	0	2	126
Importin subunit beta-1	KPNB1	97	2	0	0	418
Mothers against decapentaplegic homolog	SMAD2	49	2	2	0	10
3'(2'),5'-bisphosphate nucleotidase 1	BPNT1	32	0	2	0	31
Transcription factor AP-4	TFAP4	39	0	0	2	11
Phosphoglycerate mutase 1	PGAM1	29	0	0	2	239
Fascin	FSCN1	55	0	0	3	100
Serpin B5	SERPINB5	42	0	2	0	3
Insulin receptor substrate 2	IRS2	137	4	0	0	21
40S ribosomal protein S27	RPS27	9	2	0	0	411
Serine/threonine-protein phosphatase (Fragment)	PPP5C	55	0	2	2	39
Malonate--CoA ligase ACSF3, mitochondrial (Fragment)	ACSF3	22	0	0	2	4
SRA stem-loop-interacting RNA-binding protein, mitochondrial	SLIRP	10	3	0	0	113
U1 small nuclear ribonucleoprotein 70 kDa	SNRNP70	52	2	0	0	266
Glutathione S-transferase P	GSTP1	19	0	2	0	164
rRNA 2'-O-methyltransferase fibrillarin (Fragment)	FBL	27	3	0	0	387
39S ribosomal protein L53, mitochondrial	MRPL53	12	4	0	0	42
39S ribosomal protein L23, mitochondrial	MRPL23	18	4	0	0	38
Calcium-binding mitochondrial carrier protein Aralar2	SLC25A13	74	3	0	0	160
Eukaryotic translation initiation factor 4 gamma 3	EIF4G3	181	3	0	0	181
Pre-mRNA-splicing factor SPF27	BCAS2	26	4	0	0	104
TATA-binding protein-associated factor 2N	TAF15	49	4	0	0	383
Transferrin receptor protein 1	TFRC	85	3	0	0	87
Melanoma-associated antigen D2	MAGED2	65	3	0	0	133
Probable ATP-dependent RNA helicase DDX46	DDX46	117	4	0	0	302
ATP-dependent 6-phosphofructokinase, liver type	PFKL	85	3	0	0	178
Ras GTPase-activating-like protein IQGAP1	IQGAP1	189	3	0	0	197
DnaJ homolog subfamily A member 3, mitochondrial	DNAJA3	52	3	0	0	116
Parafibromin	CDC73	44	2	0	0	162
Afadin	AFDN	202	2	0	0	116
Serine/threonine-protein kinase PRP4 homolog	PRPF4B	117	4	0	0	131
Protein lin-28 homolog B	LIN28B	28	2	0	0	58
DNA replication licensing factor MCM2	MCM2	102	4	0	0	173
Serine/arginine-rich-splicing factor 4 (Fragment)	SRSF4	44	2	0	0	317

14-3-3 protein sigma	SFN	28	0	0	2	314
Pinin	PNN	82	4	0	0	166
Serine--tRNA ligase, mitochondrial	SARS2	58	3	0	0	75
cAMP-dependent protein kinase type II-alpha regulatory subunit	PRKAR2A	46	5	0	0	140
Eukaryotic translation initiation factor 2 subunit 2	EIF2S2	38	4	0	0	176
28S ribosomal protein S6, mitochondrial	MRPS6	14	4	0	0	38
La-related protein 4	LARP4	81	3	0	0	131
ATP-dependent RNA helicase DDX54	DDX54	99	5	0	0	148
Serine/threonine-protein phosphatase PGAM5, mitochondrial	PGAM5	32	4	0	0	223
Coatomer protein complex, subunit epsilon, isoform CRA_g	COPE	37	5	0	0	93
RNA-binding protein EWS	EWSR1	65	2	0	0	301
Ribosome biogenesis protein BRX1 homolog	BRIX1	41	2	0	0	178
DNA polymerase delta subunit 2	POLD2	55	0	0	2	14
Chromatin assembly factor 1 subunit A	CHAF1A	107	2	0	0	71
Myosin regulatory light chain 12A	MYL12A	20	2	0	0	294
Talin-1	TLN1	270	2	0	0	217
Alpha-globin transcription factor CP2 (Fragment)	TFCP2	44	0	0	2	87
RNA 3'-terminal phosphate cyclase	RTCA	39	0	2	0	31
Complement component 1 Q subcomponent-binding protein, mitochondrial	C1QBP	20	2	0	0	420
39S ribosomal protein L18, mitochondrial	MRPL18	21	2	0	0	37
Chromodomain-helicase-DNA-binding protein 1-like	CHD1L	101	2	0	0	4
Actin-related protein 3	ACTR3	47	2	0	0	137
Notchless protein homolog 1	NLE1	49	0	0	2	33
CDK4 protein	CDK4	12	0	0	2	163
ADP-ribosylation factor-like protein 3	ARL3	20	0	0	2	15
Protein-glutamine gamma-glutamyltransferase K	TGM1	90	0	2	2	35
Ferrochelatase, mitochondrial	FECH	48	2	0	0	15
E3 ubiquitin-protein ligase RBBP6	RBBP6	202	2	0	0	147
Cyclin-dependent kinase inhibitor 2A	CDKN2A	15	0	2	2	81
39S ribosomal protein L2, mitochondrial	MRPL2	33	2	0	2	43
Nicotinamide phosphoribosyltransferase	NAMPT	56	0	0	2	135
HCG2044799	HNRNPUL2-BSCL	85	2	0	0	135
Oxygen-dependent coproporphyrinogen-III oxidase, mitochondrial	CPOX	50	0	0	2	40
Transformer-2 protein homolog alpha	TRA2A	33	2	0	0	116
Methionine-R-sulfoxide reductase B2, mitochondrial	MSRB2	20	0	0	2	24
Helicase-like transcription factor	HLTF	114	2	0	0	45
39S ribosomal protein L43, mitochondrial	MRPL43	21	2	0	0	47
Polyadenylate-binding protein 2	PABPN1	33	2	0	0	206
Dimethyladenosine transferase 1, mitochondrial	TFB1M	40	2	0	0	8
Mitochondrial import inner membrane translocase subunit TIM50	TIMM50	40	2	0	0	354
Loricrin	LOR	26	0	2	0	5
3-ketoacyl-CoA thiolase, mitochondrial	ACAA2	42	0	0	2	49
Delta(14)-sterol reductase LBR	LBR	71	2	0	0	241
Spermidine synthase	SRM	34	2	0	0	85
Cytosolic Fe-S cluster assembly factor NUBP2	NUBP2	29	2	0	0	66
28S ribosomal protein S5, mitochondrial	MRPS5	48	3	0	0	66
NF-kappa-B-repressing factor	NKRF	78	3	0	0	168
Serine/arginine-rich splicing factor 9	SRSF9	26	2	0	0	151
Protein LTV1 homolog	LTV1	55	2	0	0	137
Nuclear factor of activated T-cells 5	NFAT5	166	0	0	3	23
DNA topoisomerase 2-beta	TOP2B	183	3	0	0	146
Melanoma-associated antigen B2	MAGEB2	35	2	0	0	8
26S proteasome regulatory subunit 6A (Fragment)	PSMC3	35	3	0	0	225
Cystathionine beta-synthase-like protein	CBSL	61	0	0	2	120
G patch domain-containing protein 4	GPATCH4	50	3	0	0	114
DNA topoisomerase 2-alpha	TOP2A	174	2	0	0	179
Prolactin-inducible protein	PIP	17	0	0	3	59
Dihydrolipoyl dehydrogenase	DLD	52	2	0	0	231
Protein TFG	TFG	43	3	0	0	126
Protein arginine N-methyltransferase 5 (Fragment)	PRMT5	21	2	0	0	285
Methyl-CpG-binding domain protein 3	MBD3	33	2	0	0	133
F-actin-capping protein subunit alpha	CAPZA2	20	2	0	0	232
NF-X1-type zinc finger protein NFXL1	NFXL1	101	3	0	0	9
Small nuclear ribonucleoprotein Sm D1	SNRPD1	8	3	0	0	356
ATP-dependent RNA helicase DDX50	DDX50	82	2	0	0	344
Eukaryotic translation initiation factor 2A	EIF2A	65	2	0	0	90
Zinc finger CCCH domain-containing protein 15	ZC3H15	49	4	0	0	149
ATP-dependent DNA/RNA helicase DHX36	DHX36	115	2	0	0	78
3'-5' RNA helicase YTHDC2	YTHDC2	160	2	0	0	133
DEAD (Asp-Glu-Ala-Asp) box polypeptide 56, isoform CRA_a	DDX56	42	3	0	0	115
Protein PRRC2B	PRRC2B	243	3	0	0	140
28S ribosomal protein S9, mitochondrial	MRPS9	46	3	0	0	72
Rho guanine nucleotide exchange factor 2	ARHGEF2	116	4	0	0	78
Pericentrin	PCNT	378	4	0	0	35
Histone lysine demethylase PHF8 (Fragment)	PHF8	106	3	0	0	54
14-3-3 protein eta	YWHAH	28	0	0	2	313
Glutamate-rich WD repeat-containing protein 1	GRWD1	49	2	0	0	160
U4/U6 small nuclear ribonucleoprotein Prp4	PRPF4	58	4	0	0	148
ATP-dependent RNA helicase DDX24	DDX24	69	2	0	0	138
Alpha-taxilin	TXLNA	62	3	0	0	53
Core histone macro-H2A.2	H2AFY2	40	3	0	0	78
Emerin	EMD	29	3	0	0	270
La-related protein 7	LARP7	67	4	0	0	92
Neurofilament light polypeptide	NEFL	62	0	0	2	187
E3 ubiquitin-protein ligase UHRF1	UHRF1	97	3	0	0	89
AP-2 complex subunit mu	AP2M1	50	3	0	0	48
Condensin complex subunit 2	NCAPH	82	4	0	0	212
2',5'-phosphodiesterase 12	PDE12	52	3	0	0	71
Small nuclear ribonucleoprotein Sm D3	SNRPD3	14	4	0	0	377
Paraspeckle component 1	PSPC1	59	4	0	0	217
39S ribosomal protein L16, mitochondrial	MRPL16	28	3	0	0	57
General transcription factor 3C polypeptide 2	GTF3C2	101	3	0	0	23
Serine protease inhibitor Kazal-type 5	SPINK5	121	0	0	3	0
La-related protein 4B	LARP4B	81	3	0	0	86
Biliverdin reductase A	BLVRA	33	2	0	0	98
Keratin, type II cuticular Hb2	KRT82	57	0	0	3	38
Transcriptional repressor p66-alpha	GATAD2A	68	2	0	0	164
Asparagine--tRNA ligase, cytoplasmic	NARS	63	2	0	0	64
glucosidase (Fragment)			2	0	0	-
5'-nucleotidase domain-containing protein 1	NT5DC1	52	2	0	0	40
Septin-2	SEPTIN2	37	2	0	0	-
Putative ATP-dependent RNA helicase DHX57	DHX57	156	2	0	0	26
Disks large-associated protein 5	DLGAP5	95	2	0	0	79
Prelamin-A/C	LMNA	56	0	0	2	240
Exosome complex component CSL4 (Fragment)	EXOSC1	18	0	2	0	55
Ubiquitin carboxyl-terminal hydrolase 5	USP5	96	0	0	2	148
Calmodulin-like protein 3	CALML3	17	0	0	2	81
WD repeat domain phosphoinositide-interacting protein 3	WDR45B	38	0	0	2	0
Pyridoxal phosphate phosphatase	PDXP	32	2	0	0	24
Cornifin-B	SPRR1B	10	0	2	0	19
Keratin-associated protein 13-1	KRTAP13-1	18	0	0	2	5
28S ribosomal protein S28, mitochondrial	MRPS28	21	2	0	0	65
Coronin	CORO7-PAM16	114	2	0	0	25
Cornifin-A	SPRR1A	10	0	2	0	19
Arginine and glutamate-rich protein 1	ARGLU1	33	0	0	2	150
Methylosome subunit pICln	CLNS1A	20	2	0	0	205
Cleavage and polyadenylation-specificity factor subunit 7 (Fragment)	CPSF7	41	2	0	0	154
Gamma-glutamylcyclotransferase	GGCT	21	0	0	2	96
Ubiquitin carboxyl-terminal hydrolase	UCHL1	23	0	0	2	167

NAD-dependent protein deacylase sirtuin-5, mitochondrial	SIRT5	34	0	0	2	0
SRSF protein kinase 1 (Fragment)	SRPK1	76	2	0	0	157
(E3-independent) E2 ubiquitin-conjugating enzyme (Fragment)	UBE2O	89	2	0	0	127
Serine hydroxymethyltransferase, mitochondrial	SHMT2	56	2	0	0	221
Protein FAM207A	FAM207A	25	2	0	0	102
Nuclear pore complex protein Nup153	NUP153	154	3	0	0	200
V-type proton ATPase subunit B, brain isoform	ATP6V1B2	57	2	0	0	90
Archain 1, isoform CRA_a	ARCN1	62	2	0	0	263
ATP-binding cassette sub-family F member 2	ABCF2	71	2	0	0	137
Spectrin beta chain	SPTBN1	275	2	0	0	250
Structural maintenance of chromosomes flexible hinge domain-containing protein 1	SMCHD1	226	2	0	0	155
General transcription factor 3C polypeptide 1	GTF3C1	239	2	0	0	134
Histone deacetylase 1	HDAC1	55	3	0	0	306
Acyl-protein thioesterase 2	LYPLA2	25	0	0	2	128
Trifunctional enzyme subunit alpha, mitochondrial	HADHA	54	3	0	0	154
Probable 28S rRNA (cytosine(4447)-C(5))-methyltransferase	NOP2	89	3	0	0	276
Pre-mRNA-processing-splicing factor 8	PRPF8	274	3	0	0	339
Double-stranded RNA-binding protein Staufen homolog 2	STAU2	54	2	0	0	121
Nuclear export mediator factor NEMF	NEMF	123	2	0	0	11
Uncharacterized protein C7orf50	C7orf50	22	2	0	0	62
KRR1 small subunit processome component homolog	KRR1	44	2	0	0	76
E3 ubiquitin-protein ligase RING2	RNF2	38	3	0	0	67
Regulator of nonsense transcripts 3B	UPF3B	58	3	0	0	42
Nucleolar protein 14	NOP14	98	3	0	0	46
Constitutive coactivator of PPAR-gamma-like protein 1	FAM120A	122	3	0	0	159
Chromobox protein homolog 1 (Fragment)	CBX1	20	2	0	0	204
Transcriptional activator protein Pur-alpha	PURA	35	3	0	0	55
39S ribosomal protein L47, mitochondrial	MRPL47	29	2	0	0	40
Cell growth-regulating nucleolar protein	LYAR	44	3	0	0	94
Interferon-inducible double-stranded RNA-dependent protein kinase activator A	PRKRA	34	3	0	0	58
Integrin beta-1	ITGB1	88	3	0	0	38
Protein FAM98B	FAM98B	46	3	0	0	113
CD166 antigen	ALCAM	60	3	0	0	1
Probable ATP-dependent RNA helicase DDX28	DDX28	60	2	0	0	17
28S ribosomal protein S18b, mitochondrial	MRPS18B	29	2	0	0	67
YLP motif-containing protein 1 (Fragment)	YLPM1	180	3	0	0	162
Zinc finger CCCH domain-containing protein 14 (Fragment)	ZC3H14	73	2	0	0	145
5'-3' exoribonuclease 1 (Fragment)	XRN1	130	3	0	0	105
Polymeric immunoglobulin receptor	PIGR	83	0	0	2	12
UMP-CMP kinase	CMPK1	22	2	0	0	18
H/ACA ribonucleoprotein complex subunit 1	GAR1	22	2	0	0	89
G10a (273 aa)	**TCRa**	30	**2**	**0**	**0**	0
28S ribosomal protein S34, mitochondrial	MRPS34	26	2	0	0	58
39S ribosomal protein L37, mitochondrial	MRPL37	48	2	0	0	61
Peptidyl-prolyl cis-trans isomerase-like 4	PPIL4	57	2	0	0	213
39S ribosomal protein L22, mitochondrial	MRPL22	24	2	0	0	72
39S ribosomal protein L33, mitochondrial	MRPL33	8	2	0	0	27
Transcription activator BRG1	SMARCA4	189	2	0	0	183
39S ribosomal protein L17, mitochondrial	MRPL17	20	2	0	0	39
Pericentriolar material 1 protein	PCM1	229	2	0	0	127
Partitioning defective 3 homolog	PARD3	82	2	0	0	46
Triosephosphate isomerase	TPI1	31	0	0	2	177
Phosphate carrier protein, mitochondrial	SLC25A3	36	2	0	0	362
E3 ubiquitin protein ligase	RNF40	109	2	0	0	50
DnaJ homolog subfamily A member 1	DNAJA1	45	2	0	0	243
Lupus La protein (Fragment)	SSB	21	2	0	0	296
Protein KRI1 homolog	KRI1	83	2	0	0	131
Leukotriene A-4 hydrolase	LTA4H	69	2	0	0	66
Metastasis-associated protein MTA1	MTA1	79	2	0	0	192
Nucleosome-remodeling factor subunit BPTF (Fragment)	BPTF	272	2	0	0	114
U4/U6 small nuclear ribonucleoprotein Prp3	PRPF3	78	2	0	0	168
Transducin beta-like protein 2	TBL2	46	2	0	0	47
28S ribosomal protein S2, mitochondrial	MRPS2	33	2	0	0	65
eEF1A lysine and N-terminal methyltransferase	EEF1AKNMT	79	2	0	0	-
N-alpha-acetyltransferase 15, NatA auxiliary subunit	NAA15	101	2	0	0	68
Unconventional myosin-Ib	MYO1B	128	2	0	0	159
Cold inducible RNA binding protein, isoform CRA_c	CIRBP	32	2	0	0	208
28S ribosomal protein S16, mitochondrial	MRPS16	15	2	0	0	26
Filamin-C	FLNC	291	2	0	0	373
DNA-directed RNA polymerase II subunit RPB1	POLR2A	217	2	0	0	175
tRNA (adenine(58)-N(1))-methyltransferase catalytic subunit TRMT61A (Fragment)	TRMT61A	21	2	0	0	16
39S ribosomal protein L44, mitochondrial	MRPL44	38	2	0	0	59
RNA-binding protein NOB1	NOB1	47	2	0	0	32
SRSF protein kinase 2	SRPK2	78	2	0	0	122
DNA-directed RNA polymerase I subunit RPA34	CD3EAP	55	2	0	0	31
Protein kinase domain-containing protein		95	2	0	0	-
Sorcin	SRI	18	0	0	2	70
Adenylosuccinate synthetase isozyme 2	ADSS2	50	2	0	0	-
Splicing factor 1	SF1	68	2	0	0	256
Zinc finger CCHC domain-containing protein 8	ZCCHC8	79	2	0	0	138
39S ribosomal protein L55, mitochondrial	MRPL55	15	2	0	0	31
Probable ATP-dependent RNA helicase DHX40	DHX40	89	2	0	0	54
Transcription factor 25	TCF25	77	2	0	0	1
Regulator of nonsense transcripts 2	UPF2	148	2	0	0	14
RNA transcription, translation and transport factor protein	RTRAF	28	2	0	0	-
WD repeat-containing protein 5	WDR5	37	2	0	0	99
WD repeat-containing protein 6	WDR6	125	2	0	0	12
Reticulocalbin-2	RCN2	37	2	0	0	208
28S ribosomal protein S35, mitochondrial	MRPS35	37	2	0	0	74
Cleavage and polyadenylation specificity factor subunit 2	CPSF2	88	2	0	0	166
Ribosome biogenesis protein NSA2 homolog	NSA2	23	2	0	0	36
E3 ubiquitin-protein ligase MIB1	MIB1	110	2	0	0	15
28S ribosomal protein S23, mitochondrial	MRPS23	18	2	0	0	85
Transcription elongation regulator 1	TCERG1	124	2	0	0	183
TBC1 domain family member 10B	TBC1D10B	87	2	0	0	60
TRMT1-like protein	TRMT1L	82	2	0	0	162
NADH dehydrogenase [ubiquinone] iron-sulfur protein 7, mitochondrial	NDUFS7	24	2	0	0	30
Trifunctional enzyme subunit beta, mitochondrial	HADHB	38	2	0	0	123
Dihydropyrimidinase-related protein 2	DPYSL2	74	0	0	2	199
Ribosome biogenesis regulatory protein homolog	RRS1	41	2	0	0	93
Pre-mRNA-splicing factor 38B	PRPF38B	64	2	0	0	159
Serine/threonine-protein kinase TAO2	TAOK2	138	2	0	0	1
Zinc finger CCHC domain-containing protein 3	ZCCHC3	44	2	0	0	37
Actin-related protein 2/3 complex subunit 1A	ARPC1A	42	2	0	0	62
Protein Red	IK	66	2	0	0	134
rRNA methyltransferase 3, mitochondrial	MRM3	47	2	0	0	39
39S ribosomal protein L3, mitochondrial (Fragment)	MRPL3	41	2	0	0	39
Annexin A7	ANXA7	53	0	0	2	53
Plexin-B2	PLXNB2	205	2	0	0	11
Ribonuclease P protein subunit p25-like protein	RPP25L	18	2	0	0	19
Cleavage and polyadenylation specificity factor subunit 1	CPSF1	161	2	0	0	139
Semenogelin-1	SEMG1	52	2	0	0	8

Protein threshold 95%, Peptide threshold 95%, 2 peptides minimum

A.4.2 Proteins identified by a SPPLAT assay with the pTCR and mTCR

Three T75 flasks were transfected with pTα HA-G10β-mScar, G10α HA-G10β-mScar and pTα HA-1G4β -GFP with CD3$\gamma\delta\epsilon\zeta$-IRES-BFP using eight times the normal volume of GeneJuice, plasmid and serum free media.

After 48h, the cells were suspended with 0.25% trypsin then made up to 10ml with fresh media. 0.5ml of this cell suspension was labelled with 1:200 APC-conjugated anti-TCRβ V5.1 antibody for 30 minutes in media at 37°C, washed three times in cold PBS and analysed by flow cytometry. The remainder was labelled with 1:100 HRP-conjugated anti-TCRβ V5.1 antibody for 30min in media at 37°C before being washed three times in 12ml of PBS. After the final wash, the cells were incubated in PBS for 5 mins with 500µM biotin phenol. Hydrogen peroxide was then added to 1mM final concentration to initiate labelling. After 1min the cells were washed three times in 5ml of cold quench buffer. Cells were lysed then clarified lysate incubated overnight at 4°C with 40µl of streptavidin-sepharose beads.

These beads were then washed once with lysis buffer, once with 1M KCl, once with 0.1M Na_2CO_3, once with 2M urea in 10mM Tris-HCl pH 8.0 then twice with lysis buffer without detergent. Each wash involved addition of 1ml of ice-cold solution and a short centrifugation to pellet the beads before the solution was removed. The washed beads transferred to a fresh tube with a final wash in detergent-free lysis buffer then processed with the on-bead digest protocol described in section 2.6.1. Peptides were identified by LC-MS then analysed with Scaffold using a protein confidence threshold of 95% and a peptide confidence threshold of 95%. A minimum of one peptide was required for identification. The values presented here show the number of unique peptides of each protein found in that sample. The following lane shows the number of occurrences that protein appears in the Contaminant Repository for Affinity Purification (CRAPome) database out of 716 experiments that followed a Proximity Dependent Biotinylation protocol.

Identified Proteins	Alternate ID	MW (kDa)	Total Unique Peptide count			CRAPome Exp.
			Control	pTCR	mTCR	Found/716 total
Heterogeneous nuclear ribonucleoprotein M	HNRNPM	78	7	8	17	520
Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial	MCCC1	80	6	20	19	280
Histone H2B type 1-H	HIST1H2BH	14	3	8	13	515
Histone H2B type 1-D	HIST1H2BD	14	3	8	13	515
Propionyl-CoA carboxylase alpha chain, mitochondrial (Fragment)	PCCA	67	5	13	12	284
Histone H2B type 1-L	HIST1H2BL	14	0	0	12	515
Heat shock 70 kDa protein 1B	HSPA1B	70	2	11	15	698
Histone H2B type 3-B	HIST3H2BB	14	0	0	8	513
ADP/ATP translocase 2	SLC25A5	33	0	4	7	525
HA-G10b-mScar	**TCRb**	60	0	1	8	0
Actin, cytoplasmic 1	ACTB	42	0	0	6	667
Lamina-associated polypeptide 2, isoforms beta/gamma	TMPO	46	0	1	6	386
Histone H2AX	H2AFX	15	0	0	6	484
Endoplasmin	HSP90B1	92	3	5	3	460
Stress-70 protein, mitochondrial	HSPA9	74	2	4	5	552
Far upstream element-binding protein 2	KHSRP	73	0	2	8	403
Poly [ADP-ribose] polymerase 1	PARP1	113	0	2	6	438
SERPINE1 mRNA binding protein 1	SERBP1	42	0	1	4	468
39S ribosomal protein L27, mitochondrial	MRPL27	10	1	0	0	41
Lactotransferrin (Fragment)	LTF	77	0	7	0	61
60S ribosomal protein L8	RPL8	28	1	2	3	390
Tubulin beta chain	TUBB	50	1	0	3	685
Putative elongation factor 1-alpha-like 3	EEF1A1P5	50	0	1	2	653
Tubulin beta-4B chain	TUBB4B	50	0	2	3	678
T-cell surface glycoprotein CD3 epsilon chain	**CD3E**	23	0	1	2	0
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	RPN1	69	0	3	2	246
Vimentin	VIM	54	0	1	5	543
T-cell surface glycoprotein CD3 gamma chain	**CD3G**	20	0	2	4	0
Ribosomal protein L19	RPL19	23	1	0	4	458
Splicing factor 3B subunit 2	SF3B2	100	0	2	3	317
14-3-3 protein theta (Fragment)	YWHAQ	17	0	3	4	410
Heat shock cognate 71 kDa protein	HSPA8	69	0	2	3	703
60S ribosomal protein L29	RPL29	19	0	1	4	399
Myristoylated alanine-rich C-kinase substrate	MARCKS	32	0	2	3	283
60S ribosomal protein L32 (Fragment)	RPL32	16	1	2	2	184
Microtubule-associated protein	MAP4	88	0	1	1	312
Keratin, type II cytoskeletal 1	KRT1	66	1	0	4	671
60S ribosomal protein L7a	RPL7A	30	0	1	4	398
Matrin-3	MATR3	95	0	1	6	441
40S ribosomal protein S6	RPS6	25	1	1	3	466
Trypsin-2	PRSS2	26	1	1	1	57
60S ribosomal protein L4	RPL4	48	2	1	4	444
Golgi integral membrane protein 4	GOLIM4	79	0	1	2	10
ADP/ATP translocase 3	SLC25A6	33	1	0	0	495
Tyrosine-protein kinase receptor	IGF1R	155	0	2	2	2
40S ribosomal protein S11	RPS11	18	0	0	4	340
60S ribosomal protein L6	RPL6	33	0	0	3	432
Serum albumin	ALB	69	0	0	2	239
Propionyl-CoA carboxylase beta chain, mitochondrial	PCCB	61	1	1	2	263
Alpha-enolase	ENO1	47	2	0	1	478
Delta(14)-sterol reductase LBR	LBR	47	0	1	2	241
Very-long-chain enoyl-CoA reductase	TECR	12	0	1	2	192
Keratin, type II cytoskeletal 5 (Fragment)	KRT5	12	0	3	0	508
60S ribosomal protein L14	RPL14	23	1	1	2	333
Heme oxygenase 2 (Fragment)	HMOX2	27	1	2	2	20
Keratin, type I cytoskeletal 9	KRT9	62	1	1	2	577
Mitochondrial 2-oxoglutarate/malate carrier protein (Fragment)	SLC25A11	32	1	2	2	290
Glucosidase 2 subunit beta	PRKCSH	60	1	1	1	174
HIG1 domain-containing protein		9	0	1	2	-
7-dehydrocholesterol reductase (Fragment)	DHCR7	12	0	1	1	60
60S ribosomal protein L13a	RPL13A	24	1	0	1	294
D-3-phosphoglycerate dehydrogenase	PHGDH	56	0	1	2	394
Elongation factor Tu, mitochondrial	TUFM	50	0	1	4	408
Large neutral amino acids transporter small subunit 1	SLC7A5	55	0	1	3	73
Minor histocompatibility antigen H13	HM13	46	0	1	1	52
Enhancer of rudimentary homolog	ERH	8	0	1	1	454
Liprin-beta-1	PPFIBP1	32	0	1	1	23
Cationic amino acid transporter 3	SLC7A3	67	0	0	1	0
14-3-3 protein zeta/delta (Fragment)	YWHAZ	8	0	3	0	410
14-3-3 protein beta/alpha	YWHAB	17	0	3	0	368
4F2 cell-surface antigen heavy chain	SLC3A2	65	0	0	3	189
Integrin alpha-2	ITGA2	89	0	0	2	1
EEF1E1-BLOC1S5 readthrough (NMD candidate)	EEF1E1-BLOC1S5	17	1	1	1	214
Nucleoplasmin-3	NPM3	19	1	1	1	209
Heterogeneous nuclear ribonucleoproteins C1/C2	HNRNPC	32	1	1	1	444
Host cell factor 1	HCFC1	213	1	1	1	278
Dolichol-phosphate mannosyltransferase subunit 1	DPM1	30	0	1	1	107
TMEM256-PLSCR3 readthrough (NMD candidate)	TMEM256-PLSCR3	4	0	0	1	6

Cofilin-1	CFL1	19	1	0	2	496
60S ribosomal protein L34	RPL34	13	0	1	1	247
MICOS complex subunit	CHCHD3	27	0	1	2	61
Prohibitin-2	PHB2	33	0	2	1	243
Importin subunit alpha-1 (Fragment)	KPNA2	16	0	1	2	367
Protein transport protein Sec61 subunit beta	SEC61B	10	0	1	1	166
Probable ATP-dependent RNA helicase DDX17	DDX17	80	0	1	1	523
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	88	0	1	3	578
60S ribosomal protein L18a	RPL18A	18	0	0	1	300
C-1-tetrahydrofolate synthase, cytoplasmic	MTHFD1	102	0	0	3	311
60S ribosomal protein L10a	RPL10A	25	0	0	2	280
EPH receptor B4, isoform CRA_b	EPHB4	103	0	1	0	20
Prostaglandin F2 receptor negative regulator	PTGFRN	99	0	0	1	2
Pyrroline-5-carboxylate reductase 2	PYCR2	26	0	0	3	98
CD3 delta	**CD3D**	11	0	0	3	0
Cell division cycle 2, G1 to S and G2 to M, isoform CRA_a	CDK1	34	0	0	1	282
Signal peptidase complex subunit 2	SPCS2	25	1	1	1	17
Serine/threonine-protein kinase PLK1	PLK1	19	1	1	1	54
Calcium-binding mitochondrial carrier protein Aralar2	SLC25A13	8	1	1	1	160
ATP5MF-PTCD1 readthrough	ATP5MF-PTCD1	6	1	1	1	-
L-lactate dehydrogenase A chain (Fragment)	LDHA	12	1	1	1	304
Polyubiquitin-B	UBB	17	1	0	1	457
Cathepsin B	CTSB	13	0	0	1	47
ATP-dependent 6-phosphofructokinase, liver type	PFKL	4	1	0	0	178
Etoposide-induced protein 2.4 homolog	EI24	5	0	1	1	4
Titin	TTN	3994	0	0	1	15
40S ribosomal protein S27	RPS27	9	0	1	2	411
Wolframin	WFS1	100	0	1	1	1
ATP synthase subunit e, mitochondrial	ATP5ME	8	0	1	0	-
G2/mitotic-specific cyclin-B1 (Fragment)	CCNB1	27	0	0	1	33
Matrix-remodeling-associated protein 7	MXRA7	2	0	1	2	41
Endoplasmic reticulum chaperone BiP	HSPA5	72	0	1	2	661
Catenin beta-1	CTNNB1	78	0	1	1	78
Protein disulfide-isomerase A3 (Fragment)	PDIA3	14	1	0	1	281
60S ribosomal protein L24	RPL24	18	1	0	0	447
Phosphate carrier protein, mitochondrial	SLC25A3	36	0	1	1	362
BTB/POZ domain-containing protein KCTD12	KCTD12	36	1	0	2	74
Pyruvate kinase	PKM	31	2	0	1	536
60S ribosomal protein L38	RPL38	8	0	1	0	300
Elongation factor 2	EEF2	95	0	0	1	488
Cell adhesion molecule 1	CADM1	46	0	0	1	14
Solute carrier family 35 member F6	SLC35F6	40	0	0	1	8
Galactokinase (Fragment)	GALK1	20	0	0	1	51
40S ribosomal protein S16	RPS16	14	0	0	1	411
Ataxin-2-like protein	ATXN2L	113	0	0	1	305
60S ribosomal protein L5	RPL5	20	0	0	2	395
Transferrin receptor protein 1	TFRC	85	0	0	3	87
G10a	**TCRa**	30	0	0	2	0
Desmoglein-2	DSG2	122	0	0	1	113
HLA class I histocompatibility antigen, A-3 alpha chain	**HLA-A**	34	0	0	1	114
Extended synaptotagmin-2	ESYT2	94	0	0	1	77
40S ribosomal protein S14 (Fragment)	RPS14	16	0	0	1	514
Amino acid transporter	SLC1A5	39	0	1	0	236
Thyroid hormone receptor-associated protein 3	THRAP3	104	0	1	1	382
SWI/SNF complex subunit SMARCC2	SMARCC2	136	0	1	1	164
Inosine-5'-monophosphate dehydrogenase 2 (Fragment)	IMPDH2	51	0	1	1	195
ATP synthase subunit O, mitochondrial	ATP5PO	17	0	1	1	-
ATP-dependent 6-phosphofructokinase, platelet type	PFKP	63	0	1	1	201
Multifunctional protein ADE2 (Fragment)	PAICS	46	0	1	1	322
Cytochrome b-c1 complex subunit 8	UQCRQ	10	0	1	1	5
m-AAA protease-interacting protein 1, mitochondrial	MAIP1	33	0	1	1	25
BolA-like protein 2	BOLA2B	17	0	1	1	250
Lysine-specific demethylase 9	RSBN1	85	0	1	1	50
Peroxiredoxin-4	PRDX4	21	0	1	0	394
Poly(rC)-binding protein 2 (Fragment)	PCBP2	17	0	1	1	475
U5 small nuclear ribonucleoprotein 200 kDa helicase (Fragment)	SNRNP200	158	0	1	1	333
Protein Shroom3	SHROOM3	208	0	1	1	125
Protein spinster homolog 1	SPNS1	61	0	0	1	14
Endoplasmic reticulum metallopeptidase 1	ERMP1	93	0	0	1	1
Pyrroline-5-carboxylate reductase	PYCR3	30	0	0	1	10
Translocation protein SEC62	SEC62	12	0	0	1	2
mRNA cap guanine-N7 methyltransferase (Fragment)	RNMT	11	0	0	1	81
Plexin-B2	PLXNB2	205	0	0	1	11
MICOS complex subunit MIC60	IMMT	73	0	0	2	93
Sulfate transporter (Fragment)	SLC26A2	3	0	0	1	0
Keratin, type II cytoskeletal 6A	KRT6A	60	0	1	0	515
40S ribosomal protein S2	RPS2	25	0	0	1	447
Tropomyosin alpha-3 chain	TPM3	26	0	0	1	354
Neurogenic locus notch homolog protein 2	**NOTCH2**	265	0	0	2	21
A-kinase anchor protein 8	AKAP8	76	0	0	1	111

Keratin, type II cytoskeletal 6B	KRT6B	60	0	0	2	570
Heterogeneous nuclear ribonucleoprotein A1	HNRNPA1	33	0	0	1	552
Bcl-2-associated transcription factor 1	BCLAF1	81	0	0	1	390
Mitochondrial import inner membrane translocase subunit TIM44 (Fragment)	TIMM44	31	0	0	1	112
Pre T-cell antigen receptor alpha	**PTCRA**	31	0	1	0	0
Beta-2-microglobulin	**B2M**	14	0	0	1	14
Myb-binding protein 1A (Fragment)	MYBBP1A	140	0	0	1	301
Emerin	EMD	29	0	0	1	270
FMR1-interacting protein 1	CYFIP1	149	0	0	1	217
40S ribosomal protein S20	RPS20	5	0	0	1	383
eIF-2-alpha kinase activator GCN1	GCN1	293	0	0	1	246
T-complex protein 1 subunit beta	CCT2	57	0	0	1	428
Adhesion G protein-coupled receptor L2	ADGRL2	156	0	0	1	1
60S ribosomal protein L7	RPL7	29	1	0	0	335
Importin subunit alpha-4	KPNA3	58	0	0	1	143
Adenylate kinase 2, mitochondrial	AK2	21	0	0	1	142
Acetyl-coenzyme A transporter 1	SLC33A1	49	0	0	1	7
HEAT repeat-containing protein 6	HEATR6	117	1	0	0	5
Activator of 90 kDa heat shock protein ATPase homolog 1 (Fragment)	AHSA1	15	0	0	1	230
Junctional adhesion molecule A	F11R	30	0	0	1	0
Chromodomain-helicase-DNA-binding protein 4	CHD4	220	0	0	1	256
WD repeat-containing protein 5	WDR5	37	0	0	1	99
39S ribosomal protein L14, mitochondrial	MRPL14	16	0	0	1	74
40S ribosomal protein S3	RPS3	25	1	0	0	491
40S ribosomal protein S18	RPS18	18	0	0	1	399
Profilin-1	PFN1	15	1	0	0	308
Neutrophil defensin 1	DEFA1	10	0	1	0	21
Apolipoprotein C-III	APOC3	13	0	1	0	0
Polypyrimidine tract binding protein 1, isoform CRA_e	PTBP1	22	0	0	1	438
60S ribosomal protein L18	RPL18	16	0	0	1	399
Plasmolipin	PLLP	18	0	0	1	0
Nucleolar and coiled-body phosphoprotein 1 (Fragment)	NOLC1	75	0	0	1	297
DNA-directed RNA polymerases I and III subunit RPAC1	POLR1C	39	0	0	1	84
Ras-related protein Rab-34, isoform NARR (Fragment)	RAB34	24	0	0	1	37
THO complex subunit 4	ALYREF	28	0	0	1	390
HCLS1-associated protein X-1	HAX1	17	0	0	1	63
Sec1 family domain-containing protein 1	SCFD1	5	0	0	1	81
60S ribosomal protein L27 (Fragment)	RPL27	16	0	0	1	324
Leucine-rich repeat-containing protein 47	LRRC47	63	0	0	1	82
Torsin-1A-interacting protein 1	TOR1AIP1	68	0	0	1	69
Lysine--tRNA ligase	KARS	8	0	0	1	216
6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 3	PFKFB3	64	0	0	1	23
Chromobox homolog 3 (HP1 gamma homolog, Drosophila), isoform CRA_b	CBX3	12	0	0	1	309
Afadin	AFDN	202	0	0	1	116
Vacuolar protein sorting-associated protein 18 homolog	VPS18	15	0	0	1	2
Peptidyl-prolyl cis-trans isomerase FKBP8	FKBP8	27	0	0	1	82
Glutamine--tRNA ligase	QARS	7	0	0	1	291
Programmed cell death protein 2-like	PDCD2L	39	0	0	1	107
SRA stem-loop-interacting RNA-binding protein, mitochondrial	SLIRP	10	0	0	1	113
Homeobox protein Hox-A5	HOXA5	29	0	0	1	10
N-acetyltransferase 14	NAT14	5	0	0	1	0
Motile sperm domain-containing protein 2	MOSPD2	56	0	0	1	1
Cytochrome c oxidase subunit 4 isoform 1, mitochondrial	COX4I1	14	0	0	1	90
CKLF-like MARVEL transmembrane domain-containing protein 7	CMTM7	16	0	0	1	0
Leucine-rich repeat and fibronectin type III domain-containing protein 1	LRFN1	82	0	0	1	0
DNA mismatch repair protein Msh6	MSH6	11	0	0	1	222
Pyruvate dehydrogenase E1 component subunit beta	PDHB	38	0	0	1	141
Ribonucleoprotein PTB-binding 1	RAVER1	78	0	1	0	129
Very-long-chain 3-oxoacyl-CoA reductase	HSD17B12	30	0	0	1	97
RNA-binding protein Raly (Fragment)	RALY	25	0	0	1	116
Condensin complex subunit 1	NCAPD2	157	0	0	1	112
60S ribosomal protein L22	RPL22	15	0	0	1	429
Ig gamma-1 chain C region secreted form OS=Mus musculus	**Ig Mouse**	36	0	0	1	-
Antileukoproteinase	SLPI	14	0	1	0	4
Band 4.1-like protein 2 (Fragment)	EPB41L2	92	0	0	1	191
L-lactate dehydrogenase		31	1	0	0	-
Protein PET100 homolog, mitochondrial	PET100	6	0	0	1	1
14-3-3 protein epsilon	YWHAE	11	0	0	1	447

Protein threshold: 95%, Peptide Threshold: 95%, 1 minimum peptide

A.4.3 Proteins identified via GFP-pulldown of TMEM131 with PreScission protease treatment

.

GFP-tagged fragments of TMEM131-GFP and TMEM131-3C-GFP were enriched by GFP pulldown. After washes, the beads were incubated with PreScission protease to separate tail fragments from the GFP tag. The supernatant was then incubated with glutathione-magnetic beads to remove some of the protease. The protease and heat-eluted bead fractions were run on a 4-12% polyacrylamide gel then stained with SYPRO ruby. Two gel pieces in the TMEM131-GFP fraction and three pieces in the TMEM131-3C-GFP cleared fraction were cut out and processed with an in-gel tryptic digest. The molecular weight of the centre of each piece was estimated from the protein ladder. Peptides were identified by LC-MS then analysed with Scaffold using a protein confidence threshold of 95% and a peptide confidence threshold of 95%. A minimum of one peptide was required for identification. The values presented here show the number of unique peptides of each protein found in that sample. The following lane shows the number of occurrences that protein appears in the Contaminant Repository for Affinity Purification (CRAPome) database out of 716 experiments that followed a Single Epitope tag AP-MS protocol.

Identified Protein	Alternate ID	MW (kDa)	3CG1 ~38kD	3CG2 ~30kD	3CG3 ~24kD	131-1 ~32kD	131-2 ~28kD	CRAPome Exp. Found/716 total
Keratin, type II cytoskeletal 1	KRT1	66	21	8	8	37	20	671
Keratin, type I cytoskeletal 10	KRT10	59	13	8	6	24	15	616
TMEM131-GFP	**(2095 aa)**	230	2	1	0	24	10	-
Keratin, type II cytoskeletal 2 epidermal	KRT2	65	12	4	6	30	13	628
Keratin, type I cytoskeletal 9	KRT9	62	8	3	5	13	9	577
Keratin, type II cytoskeletal 6A	KRT6A	60	11	0	5	21	20	515
Keratin, type II cytoskeletal 5	KRT5	62	7	1	2	26	13	508
Keratin, type I cytoskeletal 14	KRT14	52	8	4	5	17	12	523
Keratin, type I cytoskeletal 16	KRT16	51	7	3	5	17	8	516
Serum albumin	ALB	69	16	0	0	4	4	239
Caspase-14	CASP14	28	3	2	2	10	7	62
Keratin, type I cytoskeletal 17	KRT17	48	6	0	0	7	6	465
Desmoplakin	DSP	332	1	0	0	12	3	328
Lipocalin-1	LCN1	19	6	1	1	3	3	36
Desmoglein-1	DSG1	114	1	1	0	6	5	116
Protein S100-A9	S100A9	13	1	0	0	6	2	116
Lactotransferrin (Fragment)	LTF	77	2	0	0	7	5	61
Serpin B3	SERPINB3	45	0	0	0	8	1	45
Actin, cytoplasmic 1	ACTB	42	4	0	0	2	3	667
Protein S100-A8	S100A8	11	1	1	1	4	1	79
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	36	1	0	0	7	3	458
Cathepsin D	CTSD	44	2	1	0	3	3	85
Ketosamine-3-kinase	FN3KRP	34	0	0	0	6	0	87
Cystatin-A	CSTA	11	1	1	1	5	4	57
Annexin A2	ANXA2	39	1	0	0	8	1	398
Phosphate carrier protein, mitochondrial	SLC25A3	40	0	0	0	1	4	362
Serpin B4 (Fragment)	SERPINB4	43	0	0	0	6	0	34
Gasdermin-A	GSDMA	49	2	0	0	5	1	25
Isoform 2 of F-actin-capping protein subunit beta	CAPZB	31	0	0	0	7	0	366
Apolipoprotein D (Fragment)	APOD	24	0	0	0	3	1	16
Isoform 12 of cAMP-specific 3',5'-cyclic phosphodiesterase 4D	PDE4D	24	0	1	0	0	0	18
Junction plakoglobin	JUP	82	0	0	0	6	2	244
Tubulin beta chain	TUBB	50	0	0	0	6	3	685
F-actin-capping protein subunit alpha-1	CAPZA1	33	0	0	0	6	0	318
Polyubiquitin-B	UBB	17	1	1	1	2	2	457
PreScission Protease		46	3	0	0	5	1	-
Lysozyme C	LYZ	17	2	0	0	3	2	74
CCR4-NOT transcription complex subunit 9	CNOT9	34	0	1	0	2	5	31
Tubulin beta-4B chain	TUBB4B	50	0	0	0	5	3	678
Keratin, type II cytoskeletal 1b	KRT77	62	0	0	0	4	0	592
Elongation factor 1-alpha 1	EEF1A1	50	1	0	0	4	3	653
14-3-3 protein epsilon	YWHAE	29	0	1	0	1	6	447
Gamma-glutamylcyclotransferase	GGCT	21	0	0	0	5	2	96
Peroxiredoxin-2	PRDX2	22	0	0	0	4	2	482
ADP/ATP translocase 2	SLC25A5	33	0	0	0	0	4	525
Histone H4	H4C1	11	2	0	1	3	2	-
Desmocollin-1	DSC1	100	1	0	0	6	1	71
Keratinocyte proline-rich protein	KPRP	64	0	0	0	5	3	81
Peroxiredoxin-1 (Fragment)	PRDX1	19	0	0	0	3	3	549
Crk-like protein	CRKL	34	1	0	0	5	0	218
60S acidic ribosomal protein P0	RPLP0	30	0	0	0	5	0	436
Histone H2A	hCG_2039566	18	1	1	0	1	0	-
Immunoglobulin heavy constant alpha 1 (Fragment)	IGHA1	43	1	0	0	3	1	51
Protein S100-A7	S100A7	11	0	0	0	3	0	61
Zinc-alpha-2-glycoprotein	AZGP1	34	2	0	0	3	1	58
Hornerin	HRNR	282	0	0	0	5	0	223
Serine/threonine-protein phosphatase PGAM5, mitochondrial	PGAM5	32	0	0	0	3	1	223
Heterogeneous nuclear ribonucleoproteins C1/C2	HNRNPC	34	1	0	0	5	0	444
TGc domain-containing protein		79	0	0	0	4	0	-
Tropomyosin alpha-3 chain	TPM3	26	0	0	0	0	4	354
60S ribosomal protein L7	RPL7	29	0	0	0	0	3	335
F-actin-capping protein subunit alpha-2	CAPZA2	33	0	0	0	5	0	232
BTB/POZ domain-containing protein KCTD12	KCTD12	36	0	0	0	4	0	74
Tubulin alpha chain	TUBA1C	58	0	0	0	1	2	694
40S ribosomal protein S3a (Fragment)	RPS3A	24	0	0	0	3	3	434
Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1	37	1	0	0	4	0	536
Proteasome subunit alpha type-1	PSMA1	30	0	0	0	1	1	180
Keratin, type II cytoskeletal 78	KRT78	57	0	0	0	3	0	191
Prohibitin-2	PHB2	30	0	0	0	5	0	243
AH receptor-interacting protein	AIP	38	0	0	0	2	0	39
Histone H2B type 1-K	HIST1H2BK	14	0	1	0	0	0	515
Dermcidin	DCD	11	2	0	0	3	2	279
Prolactin-inducible protein	PIP	17	1	0	0	2	0	59
Suprabasin	SBSN	61	0	0	0	3	2	68
Filaggrin-2	FLG2	248	0	0	0	2	1	153
Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase	HACD3	40	0	0	0	3	1	176
Prohibitin (Fragment)	PHB	27	0	0	0	1	5	269
40S ribosomal protein S3	RPS3	25	0	0	0	1	2	491
Heterogeneous nuclear ribonucleoprotein A1	HNRNPA1	33	0	0	0	4	0	552
Zinc finger CCCH-type antiviral protein 1-like	ZC3HAV1L	33	0	0	0	5	0	21
Immunoglobulin lambda constant 2	IGLC2	11	1	0	0	1	2	33
Galectin-7	LGALS7	15	1	0	0	2	1	43
Protein/nucleic acid deglycase DJ-1	PARK7	18	0	0	1	0	0	236
Proteasome subunit beta type-6	PSMB6	25	0	0	0	2	2	189
Arginase-1	ARG1	35	1	0	0	3	0	57
60 kDa heat shock protein, mitochondrial	HSPD1	61	0	0	0	1	1	520
Protein PBDC1	PBDC1	26	0	0	0	4	0	30
40S ribosomal protein S4, X isoform	RPS4X	30	0	0	0	0	1	451
Rho guanine nucleotide exchange factor 17	ARHGEF17	222	0	0	0	1	0	15
14-3-3 protein zeta/delta (Fragment)	YWHAZ	28	0	0	0	1	2	410

Emerin	EMD	29	0	0	0	2	2	270
Isochorismatase domain-containing protein 1	ISOC1	32	0	0	0	2	2	40
Triosephosphate isomerase	TPI1	31	0	0	0	2	3	177
Fatty acid-binding protein 5	FABP5	15	0	0	0	3	1	106
F-box only protein 50	NCCRP1	31	0	0	0	2	1	21
Serpin B12	SERPINB12	46	0	0	0	1	0	49
Heat shock cognate 71 kDa protein	HSPA8	69	0	0	0	2	0	703
40S ribosomal protein SA	RPSA	33	0	0	0	2	0	335
Short-chain dehydrogenase/reductase family 9C member 7	SDR9C7	35	0	0	0	3	0	4
Tropomyosin alpha-4 chain	TPM4	29	0	0	0	0	3	311
Keratin, type II cytoskeletal 80	KRT80	51	0	0	0	1	0	281
Poly(rC)-binding protein 2	PCBP2	34	0	0	0	2	0	475
Cofilin-1 (Fragment)	CFL1	18	0	0	1	1	2	496
Immunoglobulin J chain	JCHAIN	18	1	0	0	2	1	16
Mammaglobin-B	SCGB2A1	11	1	0	0	0	0	2
Thioredoxin	TXN	12	0	0	0	2	2	432
Proteasome activator complex subunit 3 (Fragment)	PSME3	21	0	0	0	2	2	198
Endoplasmic reticulum chaperone BiP	HSPA5	72	0	0	0	2	1	661
14-3-3 protein sigma	SFN	28	0	0	0	1	3	314
Proteasome subunit alpha type (Fragment)	PSMA4	26	0	0	0	0	1	188
Trypsin-3 (Fragment)	PRSS3	19	0	0	0	1	1	91
CDSN	CDSN	52	0	0	0	3	0	50
Cornifin-B	SPRR1B	10	0	0	0	2	0	19
Cathepsin B	CTSB	38	0	0	0	2	0	47
Hemoglobin subunit alpha (Fragment)	HBA2	14	0	0	0	1	0	90
Four and a half LIM domains protein 3	FHL3	31	0	0	0	0	1	39
Alpha-enolase	ENO1	47	0	0	0	0	1	478
Lysophospholipase D GDPD3	GDPD3	37	0	0	0	2	0	2
Glyoxalase domain-containing protein 4	GLOD4	55	0	0	0	3	0	81
E3 ubiquitin-protein ligase SIAH1	**SIAH1**	31	0	0	0	0	3	5
Purine nucleoside phosphorylase	MTAP	27	0	0	0	0	2	50
Voltage-dependent anion-selective channel protein 2 (Fragment)	VDAC2	30	0	0	0	2	0	256
Secretoglobin family 1D member 1	SCGB1D1	10	2	0	0	0	0	1
Guanine nucleotide-binding protein subunit beta-like protein 1	GNB1L	36	0	0	0	1	0	35
Fe-S cluster assembly protein DRE2	CIAPIN1	26	0	0	0	1	0	138
CCR4-NOT transcription complex subunit 2 (Fragment)	CNOT2	30	0	0	0	1	0	31
G-patch domain and KOW motifs-containing protein	GPKOW	52	0	1	0	0	0	107
NACHT, LRR and PYD domains-containing protein 7	NLRP7	107	0	0	0	1	0	0
Cystatin-M	CST6	17	0	0	0	1	1	13
Protein S100-A14	S100A14	12	0	0	0	1	1	16
Cathepsin L2	CTSV	37	0	0	0	1	1	6
Skin-specific protein 32	XP32	26	0	0	0	1	0	54
Calmodulin-like protein 5	CALML5	16	0	0	0	0	1	150
E3 ubiquitin-protein ligase SIAH2	**SIAH2**	35	0	0	0	1	0	1
40S ribosomal protein S2	RPS2	25	0	0	0	1	0	447
Pro-cathepsin H	CTSH	36	0	0	0	1	0	1
Proteasome subunit alpha type-7	PSMA7	28	0	0	0	3	0	194
Carbonyl reductase [NADPH] 1	CBR1	30	0	0	0	0	3	143
Nucleophosmin	NPM1	33	0	0	0	3	0	554
Proteasome activator complex subunit 2	PSME2	29	0	0	0	0	2	82
Proteasome activator complex subunit 1	PSME1	29	0	0	0	0	2	69
Phosphotriesterase-related protein	PTER	39	0	0	0	2	0	12
Pyridoxal phosphate homeostasis protein	PLPBP	30	0	0	0	0	3	84
Leucine-rich repeat-containing protein 59	LRRC59	35	0	0	0	2	0	175
Serine/arginine-rich-splicing factor 1 (Fragment)	SRSF1	16	0	0	0	0	2	280
L-lactate dehydrogenase B chain	LDHB	37	0	0	0	2	0	340
Methionine adenosyltransferase 2 subunit beta	MAT2B	38	0	0	0	1	0	45
Replication factor C subunit 4 (Fragment)	RFC4	24	0	0	0	1	0	162
Protein SETSIP	SETSIP	34	0	0	0	1	0	372
Receptor of-activated protein C kinase 1	RACK1	26	0	0	0	1	0	352
60S ribosomal protein L7a	RPL7A	30	0	0	0	1	0	398
Mitochondrial glutamate carrier 1 (Fragment)	SLC25A22	21	0	0	0	1	0	7
60S ribosomal protein L8 (Fragment)	RPL8	26	0	0	0	0	1	390
26S proteasome non-ATPase regulatory subunit 14	PSMD14	35	0	0	0	2	0	168
Ribosomal protein L15 (Fragment)	RPL15	16	0	0	0	1	1	365
Desmocollin-3	DSC3	100	0	0	0	1	1	15
Inositol monophosphatase 2	IMPA2	13	0	0	0	1	1	2
60S ribosomal protein L13 (Fragment)	RPL13	15	0	0	0	1	1	474
BTB/POZ domain-containing protein KCTD7	KCTD7	28	0	0	0	0	1	0
Interleukin-37	IL37	24	0	0	0	1	1	0
Mitochondrial 2-oxoglutarate/malate carrier protein (Fragment)	SLC25A11	32	0	0	0	1	0	290
Proteasome subunit alpha type	PSMA6	23	0	0	0	0	1	193
Glutamine synthetase	GLUL	42	0	0	0	1	0	58
60S ribosomal protein L24	RPL24	18	0	0	0	1	0	447
Kallikrein-7	KLK7	28	0	0	0	1	0	4
Macrophage migration inhibitory factor	MIF	12	0	0	0	0	1	211
Interleukin-36 gamma	IL36G	19	0	0	0	0	1	1
Toll interacting protein, isoform CRA_b	TOLLIP	27	0	0	0	0	1	13
Hemoglobin subunit beta	HBB	12	0	0	0	1	0	113
Sacsin	SACS	521	0	0	0	1	0	6
Nascent polypeptide-associated complex subunit alpha, muscle-specific form	NACA	205	0	0	0	2	0	302
Phenylalanine--tRNA ligase alpha subunit	FARSA	26	0	0	0	2	0	174
Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	37	0	0	0	2	0	239
CCR4-NOT transcription complex subunit 7	CNOT7	33	0	0	0	0	2	12
Cytosolic Fe-S cluster assembly factor NUBP1	NUBP1	35	0	0	0	2	0	18
Peflin	PEF1	30	0	0	0	0	2	126
Tricarboxylate transport protein, mitochondrial	SLC25A1	34	0	0	0	0	2	175
DDB1- and CUL4-associated factor 7	DCAF7	39	0	0	0	2	0	135
Peroxisomal 2,4-dienoyl-CoA reductase	DECR2	29	0	0	0	0	1	4
ELAV-like protein 1	ELAVL1	36	0	0	0	2	0	322
Malectin	MLEC	32	0	0	0	0	2	7
Heterogeneous nuclear ribonucleoprotein K	HNRNPK	51	0	0	0	1	0	581

Alpha-ketoglutarate-dependent dioxygenase alkB homolog 4	ALKBH4	34	0	0	0	1	0	10
Sideroflexin-4	SFXN4	38	0	0	0	1	0	7
L-lactate dehydrogenase A chain	LDHA	37	0	0	0	2	0	304
Immunoglobulin heavy constant gamma 3 (Fragment)	IGHG3	49	0	0	0	1	0	54
Heat shock protein HSP 90-alpha	HSP90AA1	85	0	0	0	1	0	565
Carboxypeptidase A4	CPA4	36	0	0	0	1	0	13
Cyclin-dependent kinase 2	CDK2	27	0	0	0	2	0	234
Isoform 2 of Heterogeneous nuclear ribonucleoprotein D-like	HNRNPDL	34	0	0	0	1	0	369
Cell division cycle 2, G1 to S and G2 to M, isoform CRA_a	CDK1	34	0	0	0	2	0	282
Solute carrier family 35 member E1	SLC35E1	45	0	0	0	1	0	27
ATPase family AAA domain-containing protein 3B	ATAD3B	73	0	0	0	1	0	172
HCLS1-associated protein X-1	HAX1	32	0	0	0	1	0	63
Extracellular glycoprotein lacritin	LACRT	13	1	0	0	0	0	11
Antileukoproteinase	SLPI	14	1	0	0	0	0	4
Serine/threonine-protein phosphatase CPPED1	CPPED1	36	0	0	0	1	0	19
Glutamate--cysteine ligase regulatory subunit	GCLM	31	0	0	0	0	1	29
Malate dehydrogenase, cytoplasmic	MDH1	19	0	0	0	1	0	100
Glutathione S-transferase LANCL1 (Fragment)	LANCL1	22	1	0	0	0	0	94
Microtubule-associated protein RP/EB family member 1	MAPRE1	30	0	0	0	0	1	118
Heme oxygenase 2 (Fragment)	HMOX2	15	0	0	0	1	0	20
26S proteasome non-ATPase regulatory subunit 8	PSMD8	20	0	0	0	0	1	134
Putative transferase CAF17, mitochondrial	IBA57	38	0	0	0	1	0	46
5'-AMP-activated protein kinase subunit gamma-1	PRKAG1	32	0	0	0	1	0	43
Caspase-3	CASP3	21	0	0	0	0	1	20
10 kDa heat shock protein, mitochondrial	HSPE1	11	0	0	0	0	1	167
Copper chaperone for superoxide dismutase	CCS	27	0	0	0	0	1	9
Small proline-rich protein 2E	SPRR2E	8	0	0	0	0	1	28
Exosome complex component RRP40	EXOSC3	30	0	0	0	0	1	137
Proteasome subunit alpha type	PSMA2	26	0	0	0	1	0	193
Heterogeneous nuclear ribonucleoprotein M	HNRNPM	78	0	0	0	1	0	520
Abl interactor 1	ABI1	52	0	0	0	1	0	83
Synaptophysin-like protein 1 (Fragment)	SYPL1	15	0	0	0	1	0	20
Protein SEC13 homolog	SEC13	32	0	0	0	1	0	194
Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 (Fragment)	GNB1	12	0	0	0	1	0	140
StAR-related lipid transfer protein 7, mitochondrial (Fragment)	STARD7	18	0	0	0	1	0	1
Eukaryotic initiation factor 4A-II	EIF4A2	41	0	0	0	1	0	461
Methylosome subunit pICln	CLNS1A	20	0	0	0	1	0	205
Proteasome subunit alpha type-3	PSMA3	4	0	0	0	1	0	173
Mitochondrial import inner membrane translocase subunit TIM50 (Fragment)	TIMM50	12	0	0	0	1	0	354
Catalase	CAT	60	0	0	0	1	0	57
Calmodulin-like protein 3	CALML3	17	0	0	0	1	0	81
40S ribosomal protein S8	RPS8	24	0	0	0	1	0	498
Neutrophil gelatinase-associated lipocalin	LCN2	23	0	0	0	1	0	5
Translation initiation factor eIF-2B subunit alpha	EIF2B1	34	0	0	0	1	0	30
Isoamyl acetate-hydrolyzing esterase 1 homolog (Fragment)	IAH1	8	0	1	0	0	0	23
Histidine ammonia-lyase (Fragment)	HAL	18	0	0	0	1	0	11
UPF0600 protein C5orf51	C5orf51	17	0	0	0	1	0	12
Thioredoxin-related transmembrane protein 1	TMX1	12	0	0	0	0	1	53
Spermidine synthase	SRM	11	0	0	0	1	0	85
Acidic leucine-rich nuclear phosphoprotein 32 family member A	ANP32A	20	0	0	0	0	1	116
Translocon-associated protein subunit alpha	SSR1	32	0	0	0	1	0	57
Aminoacyl tRNA synthase complex-interacting multifunctional protein 2	AIMP2	35	0	0	0	1	0	172
L-aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase	AASDHPPT	36	0	0	0	1	0	51

A.4.4 Proteins identified by TMEM131 GFP-pulldown from an in-gel tryptic digest of gel pieces

Six pieces of the coomassie gel shown in figure 4.29B were cut up in the TMEM131 lanes and 5 in the TMEM131dC lanes then processed with an in-gel tryptic digest. The molecular weight of the centre of each piece was estimated from the protein ladder. Peptides were identified by LC-MS then analysed with Scaffold using a protein confidence threshold of 95% and a peptide confidence threshold of 95%. A minimum of two peptides was required for identification. The values show the number of unique peptides of each protein found in that sample.

The following lanes show:

- The number of occurrences that protein appears in the Contaminant Repository for Affinity Purification (CRAPome) database out of 716 experiments that followed a Single Epitope tag AP-MS protocol.
- Whether the protein was also detected in the later BioID experiment (table A.4.5)
- Whether the protein was in the 30 top upregulated or downregulated genes after TMEM131 knockdown listed in the K562 Genome-wide Perturb-Seq library by Replogle *et al.*[357]
- Proteins with the Biological Process Gene Ontology terms for "ERAD pathway" or "Protein Folding" from the Gene Ontology database using the PANTHER enrichment tool.

Some of the proteins discussed in the text are shown in bold. These may not represent all interesting hits.

Identified Protein	Alternate ID	MW kDa	DC1 ~200kD	DC2 ~165kD	DC3 ~100kD	DC4 ~65kD	DC5 ~42kD	FL1 ~250kD	FL2 ~140kD	FL3 ~95kD	FL4 ~65kD	FL5 ~42kD	FL6 ~35kD	CRAPome Exp. Found/716 total	Found in BioID	Found in PeterbSeq	GO terms ERAD pathway	GO terms Protein Folding
HA-TMEM131-GFP	(2095 aa)	230	191	91	49	14	54	219	93	144	56	37	104	-				
Keratin, type II cytoskeletal 1	KRT1	66	100	54	90	86	51	94	113	87	108	83	51	671	Y			
Keratin, type I cytoskeletal 10	KRT10	59	66	36	68	44	36	86	77	64	64	58	60	616	Y			
Keratin, type II cytoskeletal 2 epidermal	KRT2	65	80	25	58	85	31	82	93	54	99	53	33	628	Y			
Keratin, type I cytoskeletal 9	KRT9	62	62	27	59	50	35	55	66	51	52	51	30	577	Y			
Heat shock 70 kDa protein 1B	HSPA1B	70	4	2	10	118	6	9	12	14	139	10	3	698	Y			Y
Keratin, type I cytoskeletal 14	KRT14	52	38	21	39	21	38	36	36	35	32	48	19	523				
Keratin, type I cytoskeletal 16	KRT16	51	31	14	34	17	40	30	35	24	43	48	12	516				
Keratin, type II cytoskeletal 5	KRT5	62	38	21	31	25	17	35	40	39	41	37	13	508				
Keratin, type II cytoskeletal 6A	KRT6A	60	33	19	31	21	20	34	41	31	49	33	16	515				
Keratin, type II cytoskeletal 6B	KRT6B	60	0	0	29	0	0	31	0	0	46	35	0	570				
CCR4-NOT transcription complex subunit 1	CNOT1	267	0	0	0	0	0	218	54	0	0	0	0	130				
Isoform 2 of CCR4-NOT transcription complex subunit 1	CNOT1	266	0	0	0	0	0	210	52	0	0	0	0	130				
Isoform 2 of Neutral alpha-glucosidase AB	GANAB	109	0	0	125	4	2	0	0	28	0	0	0	216				
Neutral alpha-glucosidase AB	GANAB	107	0	0	123	0	0	0	0	29	0	0	0	216	Y			
Fatty acid synthase	FASN	273	0	0	0	0	0	181	34	0	0	0	0	436	Y			
Desmoplakin	DSP	332	21	2	9	2	3	45	18	3	9	12	0	328				
Keratin, type I cytoskeletal 17	KRT17	48	14	6	15	0	28	15	8	10	11	28	0	465				
Actin, cytoplasmic 2	ACTG1	42	0	2	5	0	50	0	4	6	6	44	0	667				
Histone deacetylase 6	HDAC6	131	0	20	4	0	0	0	77	23	4	0	0	27				
Heat shock cognate 71 kDa protein	HSPA8	71	0	0	0	57	0	0	3	0	67	0	0	703	Y			Y
Elongation factor 2	EEF2	95	0	0	51	0	0	0	0	106	0	0	0	488	Y			
Actin, cytoplasmic 1	ACTB	42	0	2	5	0	49	0	3	5	5	44	0	667	Y			
Tetratricopeptide repeat protein 13	TTC13	97	0	0	24	0	0	0	0	106	0	0	0	0				
Serum albumin	ALB	69	7	5	12	19	4	19	15	4	27	0	6	239	Y			
Tetratricopeptide repeat protein 17	**TTC17**	130	14	54	12	0	0	0	40	0	0	0	0	0	Y	Y		
Heat shock 70 kDa protein 6	HSPA6	71	0	0	0	26	0	0	0	0	30	0	0	687				Y
Isoform 2 of Tetratricopeptide repeat protein 17	**TTC17**	109	13	51	12	0	0	0	39	0	0	0	0	0		Y		
Lactotransferrin (Fragment)	LTF	77	4	2	17	0	2	34	24	16	7	4	4	61				
Isoform 5 of Golgin subfamily A member 4	GOLGA4	262	0	0	0	0	0	114	0	0	0	0	0	14				
Golgin subfamily A member 4	GOLGA4	261	0	0	0	0	0	114	0	0	0	0	0	14				
Desmoglein-1	DSG1	114	11	4	9	6	4	6	9	9	13	9	4	116				
Nucleolin	NCL	77	0	0	36	0	0	0	0	66	2	0	0	527				
CTP synthase 1	CTPS1	67	0	0	0	28	0	0	0	0	62	0	0	312	Y			
Endoplasmin	**HSP90B1**	92	0	0	41	0	0	0	0	47	0	0	0	460	Y	Y	Y	Y
Transitional endoplasmic reticulum ATPase	**VCP**	89	0	0	29	0	0	0	0	59	0	0	0	314			Y	
26S proteasome non-ATPase regulatory subunit 2	**PSMD2**	100	0	0	39	0	0	0	0	43	0	0	0	271				
Actin, alpha cardiac muscle 1	ACTC1	42	0	0	3	0	28	0	0	0	0	26	0	656				
Stress-70 protein, mitochondrial	HSPA9	74	0	0	0	26	0	0	0	0	40	0	0	552	Y			Y
Isoform 4 of Dynamin-like 120 kDa protein, mitochondrial	OPA1	118	0	0	31	0	0	0	0	44	0	0	0	41	Y			
Lamin-B1	LMNB1	66	0	0	0	42	0	0	0	0	36	0	0	284				
Heat shock cognate 71 kDa protein (Fragment)	HSPA8	20	0	0	0	25	0	0	0	0	28	0	0	703				
Polyubiquitin-C (Fragment)	UBC	14	8	6	7	3	5	8	8	9	8	4	2	457				
CAD protein	CAD	243	0	0	0	0	0	66	5	0	0	0	0	387				
CCR4-NOT transcription complex subunit 3	CNOT3	82	0	0	0	0	0	0	0	72	0	0	0	23				
Keratin, type II cytoskeletal 1b	KRT77	62	6	3	6	4	3	5	4	2	3	3	2	592				
Tubulin alpha-1B chain	TUBA1B	50	0	0	3	3	7	0	4	4	3	9	8	694				
Exportin-2	CSE1L	110	0	0	34	0	0	0	0	23	0	0	0	318				
DNA replication licensing factor MCM3	MCM3	91	0	0	28	0	0	0	0	29	0	0	0	322				
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	**RPN1**	69	0	0	0	21	0	0	0	0	27	0	0	246	Y	Y		
Heat shock 70 kDa protein 4	HSPA4	94	0	0	53	0	0	0	0	0	0	0	0	304				Y
AP-2 complex subunit alpha-1	AP2A1	108	0	0	0	0	0	0	0	55	0	0	0	80				
Creatine kinase B-type	CKB	43	0	0	0	0	28	0	0	0	0	21	0	308				
X-ray repair cross-complementing protein 6	XRCC6	70	0	0	0	18	0	0	0	0	33	0	0	388				
C-1-tetrahydrofolate synthase, cytoplasmic	MTHFD1	102	0	0	59	0	0	0	0	0	0	0	0	311	Y			
Cytoskeleton-associated protein 4	CKAP4	66	0	0	0	17	0	0	0	0	34	0	0	168	Y			
Isoleucine--tRNA ligase, cytoplasmic	IARS	145	0	0	0	0	0	0	50	0	0	0	0	257				
U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	245	0	0	0	0	0	42	0	0	0	0	0	333				
Zinc-alpha-2-glycoprotein	AZGP1	34	4	4	3	3	7	2	4	0	2	9	3	58				
CCR4-NOT transcription complex subunit 1 (Fragment)	CNOT1	50	0	0	0	0	0	42	9	0	0	0	0	130				
Glucosidase 2 subunit beta	**PRKCSH**	60	0	0	0	28	4	0	0	0	9	0	6	174	Y	Y		
Fructose-bisphosphate aldolase A	ALDOA	39	0	0	0	0	28	0	0	0	0	30	0	301				
Isoform 2 of Glucosidase 2 subunit beta	**PRKCSH**	59	0	0	0	29	0	0	0	0	9	0	6	174		Y		
Heat shock protein 105 kDa	HSPH1	97	0	0	37	0	0	0	0	7	0	0	0	347				Y
Phenylalanine--tRNA ligase beta subunit	FARSB	66	0	0	0	16	0	0	0	0	25	0	0	139				
Insulin-like growth factor 2 mRNA-binding protein 1	IGF2BP1	63	0	0	0	16	0	0	0	0	30	0	0	343				

Arginine--tRNA ligase, cytoplasmic	RARS	75	0	0	0	6	0	0	0	0	33	0	0	333				
Nucleophosmin	NPM1	33	0	0	0	0	8	0	0	0	0	12	19	554				
Junction plakoglobin	JUP	82	3	0	2	2	0	0	4	2	11	4	0	244				
Transketolase	TKT	68	0	0	0	15	0	0	0	0	26	0	0	168				
Isoform 2 of ATPase family AAA domain-containing protein 3A	ATAD3A	66	0	0	0	27	0	0	0	0	8	0	0	229	Y			
Nuclear pore complex protein Nup155	NUP155	155	0	0	0	0	0	0	40	0	0	0	0	175	Y			
Lipocalin-1	LCN1	19	4	0	3	4	3	5	6	0	4	5	4	36				
Microtubule-associated protein	MAP4	245	3	10	0	0	0	0	21	0	0	0	0	312	Y			
Ketosamine-3-kinase	FN3KRP	34	0	0	0	0	0	0	0	0	0	0	28	87	Y			
Myosin-9	MYH9	227	0	0	0	0	0	37	0	0	0	0	0	448				
AP-2 complex subunit beta	AP2B1	105	0	0	7	0	0	0	0	33	0	0	0	183				
Ubiquitin carboxyl-terminal hydrolase 5	USP5	96	0	0	17	0	0	0	0	19	0	0	0	148				
26S proteasome regulatory subunit 10B	**PSMC6**	46	0	0	0	0	25	0	0	0	0	18	0	179			Y	
Aminopeptidase	NPEPPS	103	0	0	21	0	0	0	0	17	0	0	0	123				
Alpha-enolase	ENO1	47	2	2	3	2	2	0	3	2	4	2	0	478				
Transcription intermediary factor 1-beta	TRIM28	89	0	0	22	0	0	0	0	20	0	0	0	466	Y			
Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	ATP2A2	115	0	0	17	0	0	0	0	26	0	0	0	302				
Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	SDHA	73	0	0	0	14	0	0	0	0	26	0	0	163	Y			
Immunoglobulin heavy constant alpha 1 (Fragment)	IGHA1	43	2	2	3	2	0	4	5	3	4	2	2	51				
Pre-mRNA-processing-splicing factor 8	PRPF8	274	0	0	0	0	0	35	0	0	0	0	0	339				
Polyamine deacetylase HDAC10	HDAC10	66	0	0	0	11	0	0	0	0	22	0	0	3				
Sodium/potassium-transporting ATPase subunit alpha-1	ATP1A1	113	0	0	17	0	0	0	0	23	0	0	0	335				
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	36	0	0	5	0	5	0	2	4	2	7	2	458	Y			
60 kDa heat shock protein, mitochondrial	HSPD1	61	0	0	0	6	0	0	3	0	6	6	0	520	Y			Y
Plastin-3	PLS3	71	0	0	0	15	0	0	0	0	25	0	0	208				
Schlafen family member 11	SLFN11	103	0	0	14	0	0	0	0	19	0	0	0	44				
AP-2 complex subunit alpha-2	AP2A2	104	0	0	0	0	0	0	0	33	0	0	0	80				
Tubulin beta chain	TUBB	50	0	0	0	0	5	0	0	0	0	6	11	685	Y			
Heat shock protein HSP 90-alpha	HSP90AA1	85	0	0	7	0	0	0	2	18	0	0	0	565				Y
Ubiquitin carboxyl-terminal hydrolase 24	USP24	294	0	0	0	0	0	30	0	0	0	0	0	48				
Leucine--tRNA ligase, cytoplasmic	LARS	134	0	0	0	0	0	0	31	0	0	0	0	348				
Filamin-A	FLNA	281	0	0	0	0	0	32	0	0	0	0	0	508				
Heat shock 70 kDa protein 4L	HSPA4L	95	0	0	19	0	0	0	0	13	0	0	0	181				
eIF-2-alpha kinase activator GCN1	GCN1	293	0	0	0	0	0	35	0	0	0	0	0	246	Y			
Elongation factor 1-alpha 1	EEF1A1	50	0	0	4	0	4	2	5	2	3	9	5	653	Y			
Phosphoribosylformylglycinamidine synthase	PFAS	145	0	0	0	0	0	0	23	0	0	0	0	166				
26S proteasome non-ATPase regulatory subunit 1	**PSMD1**	106	0	0	34	0	0	0	0	0	0	0	0	217				
Mini-chromosome maintenance complex-binding protein	MCMBP	73	0	0	0	11	0	0	0	0	24	0	0	92				
Structural maintenance of chromosomes protein 1A	SMC1A	143	0	8	0	0	0	0	22	0	0	0	0	248				
Heat shock protein HSP 90-beta	HSP90AB1	83	0	0	8	0	0	0	0	14	2	0	0	573				Y
Coatomer subunit beta'	COPB2	102	0	0	16	0	0	0	0	13	0	0	0	164				
Clathrin heavy chain	CLTC	192	33	0	0	0	0	0	0	0	0	0	0	406				
ATP-dependent RNA helicase A	DHX9	141	0	0	0	0	0	0	36	0	0	0	0	470				
Polyadenylate-binding protein 1	PABPC1	71	0	0	0	3	0	0	0	0	22	0	0	408				
Coatomer subunit beta	COPB1	107	0	0	21	0	0	0	0	11	0	0	0	278				
Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC	158	0	0	0	0	0	0	30	0	0	0	0	249				
Methionine--tRNA ligase, cytoplasmic	MARS	101	0	0	11	0	0	0	0	19	0	0	0	295				
Lysine--tRNA ligase	KARS1	68	0	0	0	10	0	0	0	0	21	0	0	-				
ATPase family AAA domain-containing protein 3B	ATAD3B	73	0	0	0	16	0	0	0	0	7	0	0	172	Y			
Peptide-N(4)-(N-acetyl-beta-glucosaminyl)asparagine amidase	NGLY1	74	0	0	0	22	0	0	0	0	5	0	0	9	Y		Y	Y
Programmed cell death 6-interacting protein	PDCD6IP	96	0	0	11	0	0	0	0	13	0	0	0	227				
Structural maintenance of chromosomes protein 2	SMC2	136	0	0	0	0	0	0	29	0	0	0	0	194				
Poly(rC)-binding protein 2	PCBP2	39	0	0	0	0	13	0	0	0	0	16	0	475				
Nuclear migration protein nudC	NUDC	38	0	0	0	0	13	0	0	0	0	18	0	296				Y
Importin subunit beta-1	KPNB1	97	0	0	4	0	0	0	0	27	0	0	0	418				
Tyrosine-protein phosphatase non-receptor type 11	PTPN11	68	0	0	0	11	0	0	0	0	14	0	0	158				
Annexin A2	ANXA2	39	2	2	0	0	2	3	2	2	2	2	8	398				
L-lactate dehydrogenase A chain	LDHA	37	0	0	0	0	0	0	0	0	0	0	24	304				
Immunoglobulin heavy constant alpha 2 (Fragment)	IGHA2	37	0	0	3	0	0	0	0	0	0	0	0	49				
Adenosine deaminase	ADA	41	0	0	0	0	16	0	0	0	0	14	0	35				
Acetyltransferase component of pyruvate dehydrogenase complex (Fragment)	DLAT	51	0	0	0	3	0	0	0	0	17	0	0	137				
CTP synthase 2	CTPS2	66	0	0	0	7	0	0	0	0	18	0	0	136				
Splicing factor, proline- and glutamine-rich	SFPQ	76	0	0	6	0	0	0	0	28	0	0	0	484				
60S ribosomal protein L5	RPL5	34	0	0	0	0	0	0	0	0	0	0	28	395				
Tubulin beta-4B chain	TUBB4B	50	0	0	0	0	0	0	0	0	0	5	6	678	Y			
Insulin-like growth factor 2 mRNA-binding protein 3	IGF2BP3	64	0	0	0	8	0	0	0	0	15	0	0	288				
26S proteasome non-ATPase regulatory subunit 6	**PSMD6**	46	0	0	0	0	17	0	0	0	0	13	0	187				
Probable ATP-dependent RNA helicase DDX5	DDX5	69	0	0	0	2	0	0	0	0	20	0	0	528				

Dermcidin	DCD	11	4	4	3	0	6	0	4	0	0	4	3	279				
High density lipoprotein binding protein (Vigilin), isoform CRA_a	HDLBP	141	0	5	0	0	0	0	12	0	0	0	0	218				
Endoplasmic reticulum chaperone BiP	HSPA5	72	0	0	0	4	0	0	0	0	19	0	0	661	Y		Y	Y
Leucine-rich repeat-containing protein 40	LRRC40	68	0	0	0	3	0	0	0	0	18	0	0	88				
Neurofilament light polypeptide	NEFL	62	0	0	0	5	0	0	0	0	18	0	0	187				
Bifunctional glutamate/proline--tRNA ligase	EPRS	171	17	0	0	0	0	0	0	0	0	0	0	393				
Coatomer subunit alpha	COPA	138	0	0	0	0	0	0	26	0	0	0	0	202				
Keratin, type II cytoskeletal 78	KRT78	57	3	0	0	0	0	0	3	2	6	0	0	191				
L-lactate dehydrogenase B chain	LDHB	37	0	0	0	0	0	0	0	0	0	2	26	340				
Structural maintenance of chromosomes protein 3	SMC3	142	0	2	0	0	0	0	21	0	0	0	0	194				
Lamin-B2	LMNB2	70	0	0	0	11	0	0	0	0	12	0	0	156				
AMP deaminase	AMPD2	99	0	0	0	0	0	0	0	19	0	0	0	11				
Splicing factor 3B subunit 1	SF3B1	146	0	0	0	0	0	0	26	0	0	0	0	361				
Aminopeptidase B	RNPEP	68	0	0	0	7	0	0	0	0	17	0	0	50				
Desmocollin-1	DSC1	100	5	0	0	2	5	0	4	0	0	5	0	71				
Heterogeneous nuclear ribonucleoprotein A1	HNRNPA1	39	0	0	0	0	0	0	0	0	0	4	13	552				
Opioid growth factor receptor	OGFR	73	0	0	5	0	0	0	0	20	0	0	0	31	Y			
Histone H2A type 1-B/E	HIST1H2AB	14	0	0	2	0	3	3	3	0	3	2	3	550				
Arginase-1	ARG1	35	2	0	0	0	0	0	0	0	0	3	3	57				
Protein pelota homolog	PELO	43	0	0	0	0	13	0	0	0	0	12	0	63				
Isoform 4 of Acetyl-CoA carboxylase 1	ACACA	270	0	0	0	0	0	22	0	0	0	0	0	296	Y			
Myosin-10	MYH10	229	0	0	0	0	0	23	0	0	0	0	0	379				
Vimentin	VIM	50	0	0	0	0	3	0	0	0	0	3	0	543				
Interleukin enhancer-binding factor 2	ILF2	39	0	0	0	0	6	0	0	0	0	14	0	313				
Ferrochelatase, mitochondrial	FECH	48	0	0	0	0	7	0	0	0	0	13	0	15				
Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	45	0	0	0	0	11	0	0	0	0	12	0	161				
MTHFD1L protein	MTHFD1L	106	0	0	19	0	0	0	0	0	0	0	0	150				
Lysozyme C	LYZ	17	0	0	0	0	2	2	2	0	0	2	5	74				
1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1	PLCG1	149	0	0	0	0	0	0	16	0	0	0	0	15				
SNW domain-containing protein 1	SNW1	61	0	0	0	5	0	0	0	0	12	0	0	191	Y			
Erlin-2	ERLIN2	38	0	0	0	0	12	0	0	0	0	3	0	113			Y	
Transmembrane protein 131 (Fragment)	**TMEM131**	14	0	0	0	0	0	7	0	0	0	0	0	0		Y		
DNA mismatch repair protein Msh2	MSH2	105	0	0	18	0	0	0	0	0	0	0	0	202				
Isoform 2 of F-actin-capping protein subunit beta	CAPZB	31	0	0	0	0	0	0	0	0	0	0	24	366	Y			
Probable ubiquitin carboxyl-terminal hydrolase FAF-X	USP9X	292	0	0	0	0	0	19	0	0	0	0	0	177				
FACT complex subunit SPT16	SUPT16H	120	0	0	0	0	0	0	24	0	0	0	0	175				
Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial	PDHA1	43	0	0	0	0	9	0	0	0	0	10	0	126				
Cytoplasmic FMR1-interacting protein 1	CYFIP1	145	0	0	0	0	0	0	17	0	0	0	0	217	Y			
Caspase-14	CASP14	28	5	0	0	0	0	0	0	0	0	3	2	62				
Poly(rC)-binding protein 1	PCBP1	37	0	0	0	0	8	0	0	0	0	15	0	472	Y			
Calmodulin-like protein 5	CALML5	16	6	0	7	0	0	0	0	0	0	0	0	150				
Melanoma-associated antigen D2	MAGED2	65	0	0	0	3	0	0	0	0	9	0	0	133	Y			
Stress-induced-phosphoprotein 1	STIP1	63	0	0	0	7	0	0	0	0	7	0	0	237				Y
AP-1 complex subunit beta-1	AP1B1	105	0	0	8	0	0	0	0	10	0	0	0	147				
2',5'-phosphodiesterase 12	PDE12	67	0	0	0	7	0	0	0	0	13	0	0	71				
Cell division cycle 5-like protein	CDC5L	92	0	0	0	0	0	0	0	17	0	0	0	209				
Glycogen phosphorylase, liver form	PYGL	97	0	0	5	0	0	0	0	11	0	0	0	74				
Malate dehydrogenase, mitochondrial	MDH2	36	0	0	0	0	0	0	0	0	0	0	15	198				
60S acidic ribosomal protein P0	RPLP0	34	0	0	0	0	0	0	0	0	0	0	18	436				
4F2 cell-surface antigen heavy chain	**SLC3A2**	65	0	0	5	5	0	0	0	2	0	0	0	189	Y	Y		
Catenin delta-1	CTNND1	105	0	0	10	0	0	0	0	6	0	0	0	99				
GDP-mannose 4,6 dehydratase	GMDS	42	0	0	0	0	7	0	0	0	0	9	0	33				
Probable ATP-dependent RNA helicase DDX17	DDX17	80	0	0	0	3	0	0	0	0	14	0	0	523	Y			
Calcium/calmodulin-dependent serine protein kinase (MAGUK family), isoform CRA_c	CASK	104	0	0	8	0	0	0	0	2	0	0	0	21				
Heat shock protein 75 kDa, mitochondrial	TRAP1	80	0	0	0	4	0	0	0	0	8	0	0	478				Y
E3 ubiquitin/ISG15 ligase TRIM25	**TRIM25**	71	0	0	0	5	0	0	0	0	14	0	0	136			Y	
Septin-2	SEPTIN2	41	0	0	0	0	9	0	0	0	0	7	0	-				
DnaJ homolog subfamily B member 11	**DNAJB11**	41	0	0	0	0	8	0	0	0	0	7	0	117		Y		Y
TRPM8 channel-associated factor 1	TCAF1	102	0	0	6	0	0	0	0	4	0	0	0	14				
Ubiquitin carboxyl-terminal hydrolase 7	USP7	128	0	0	0	0	0	0	19	0	0	0	0	201				
Isoform 2 of DNA mismatch repair protein Msh2	MSH2	97	0	0	14	0	0	0	0	0	0	0	0	202				
Isoform 2 of 4F2 cell-surface antigen heavy chain	**SLC3A2**	58	0	0	4	6	0	0	0	0	0	0	0	189		Y		
Insulin receptor substrate 4	IRS4	134	0	0	0	0	0	0	14	3	0	0	0	378				
Exportin-T	XPOT	110	0	0	9	0	0	0	0	2	0	0	0	101				
Endophilin-B1	SH3GLB1	44	0	0	0	0	9	0	0	0	0	9	0	30				
Protein phosphatase 1G	PPM1G	59	0	0	0	6	0	0	0	0	13	0	0	155				
Isoform 2 of Heterogeneous nuclear ribonucleoprotein A3	HNRNPA3	37	0	0	0	0	9	0	0	0	0	6	0	447				
Heterogeneous nuclear ribonucleoprotein A3	HNRNPA3	40	0	0	0	0	9	0	0	0	0	6	0	447				

RNA cytosine C(5)-methyltransferase NSUN2	NSUN2	86	0	0	0	0	0	0	0	12	0	0	0	254				
Serpin B3	SERPINB3	45	0	0	0	0	3	0	0	0	0	3	2	45				
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	DDOST	49	0	0	0	0	9	0	0	0	0	4	0	148				
Dual specificity protein kinase TTK	TTK	97	0	0	3	0	0	0	0	8	0	0	0	87				
Coatomer subunit gamma-1	COPG1	98	0	0	5	0	0	0	0	10	0	0	0	228				
Isoform III of Ubiquitin-protein ligase E3A	**UBE3A**	100	0	0	7	0	0	0	0	10	0	0	0	14				
Erlin-1	ERLIN1	39	0	0	0	0	14	0	0	0	0	4	0	81	Y		Y	
Insulin receptor substrate 2	IRS2	137	0	3	0	0	0	0	10	0	0	0	0	21				
UDP-glucose:glycoprotein glucosyltransferase 1	**UGGT1**	177	10	0	0	0	0	0	0	0	0	0	0	69	Y		Y	Y
Filamin-B	FLNB	278	0	0	0	0	0	13	0	0	0	0	0	417				
Valine--tRNA ligase	VARS	140	0	0	0	0	0	0	14	0	0	0	0	265				
Receptor of activated protein C kinase 1	RACK1	35	0	0	0	0	0	0	0	0	0	0	12	352				
Far upstream element-binding protein 2	KHSRP	73	0	0	0	0	0	0	0	15	0	0	0	403	Y			Y
Src substrate cortactin	CTTN	62	0	0	0	2	0	0	0	0	7	0	0	318	Y			
Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 2	INPPL1	139	0	0	0	0	0	0	13	0	0	0	0	22				
Alpha-actinin-4	ACTN4	105	0	0	11	0	0	0	0	0	0	0	0	335				
Elongator complex protein 1	ELP1	150	0	2	0	0	0	0	12	0	0	0	0	63				
Leucine-rich repeat-containing protein 47	LRRC47	63	0	0	0	5	0	0	0	0	4	0	0	82				
Immunoglobulin heavy constant gamma 1 (Fragment)	IGHG1	32	0	0	0	0	3	0	0	2	0	2	2	62				
Interleukin enhancer-binding factor 3	ILF3	95	0	0	7	0	0	0	0	0	0	0	0	379				
Crk-like protein	CRKL	34	0	0	0	0	3	0	0	0	0	8	0	218	Y			
Probable ATP-dependent RNA helicase DDX20	DDX20	92	0	0	0	0	0	0	0	8	0	0	0	108				
Aspartate aminotransferase, mitochondrial	GOT2	48	0	0	0	0	8	0	0	0	0	7	0	78				
Leukocyte elastase inhibitor	SERPINB1	43	0	0	0	0	8	0	0	0	0	9	0	17				
V-type proton ATPase catalytic subunit A	ATP6V1A	68	0	0	0	3	0	0	0	0	7	0	0	153				
Spectrin beta chain, non-erythrocytic 1	SPTBN1	275	0	0	0	0	0	12	0	0	0	0	0	250				
Cytoplasmic FMR1-interacting protein	CYFIP2	143	0	0	0	0	0	0	13	0	0	0	0	216				
Putative transferase CAF17, mitochondrial	IBA57	38	0	0	0	0	0	0	0	0	0	0	17	46	Y			
Leucine-rich repeat-containing protein 59	LRRC59	35	0	0	0	0	0	0	0	0	0	0	15	175				
Protein S100-A9	S100A9	13	3	0	2	0	0	0	2	0	0	0	0	116				
Keratin, type II cytoskeletal 4	KRT4	57	0	3	2	0	0	2	0	0	2	0	0	391				
Structural maintenance of chromosomes protein	SMC4	144	0	4	0	0	0	0	4	0	0	0	0	253				
DNA replication licensing factor MCM6	MCM6	93	0	0	4	0	0	0	0	8	0	0	0	282				
Cytoskeleton-associated protein 5	CKAP5	226	0	0	0	0	0	14	0	0	0	0	0	196				
Thioredoxin domain-containing protein 11	TXNDC11	111	0	0	16	0	0	0	0	0	0	0	0	1				
RNA-binding motif protein, X chromosome	RBMX	42	0	0	0	0	0	0	0	0	0	10	0	377				
Inorganic pyrophosphatase	**PPA1**	33	0	0	0	0	0	0	0	0	0	0	9	200		Y		
Zinc finger CCCH-type antiviral protein 1-like	ZC3HAV1L	33	0	0	0	0	0	0	0	0	0	0	17	21	Y			
Proline-rich protein 4	PRR4	17	0	0	0	0	2	0	0	0	0	0	2	11				
ATP-dependent DNA helicase Q1	RECQL	73	0	0	0	2	0	0	0	0	7	0	0	69				
Replication protein A 70 kDa DNA-binding subunit	RPA1	68	0	0	0	3	0	0	0	0	5	0	0	242				
Zinc-hook domain-containing protein		143	0	0	0	0	0	0	12	0	0	0	0	-				
COP9 constitutive photomorphogenic-like protein subunit 4 isoform 2	COPS4	48	0	0	0	0	7	0	0	0	0	5	0	102				
Reticulocalbin-2	RCN2	37	0	0	0	0	6	0	0	0	0	6	0	208	Y			
F-box only protein 22	FBXO22	45	0	0	0	0	5	0	0	0	0	10	0	45				
Golgi resident protein GCP60	ACBD3	61	0	0	0	3	0	0	0	0	8	0	0	107				
Polyadenylate-binding protein	PABPC4	68	0	0	0	0	0	0	0	0	11	0	0	376				
Angiomotin	AMOT	118	0	0	0	0	0	0	15	0	0	0	0	44				
Cyclin-dependent kinase 2	CDK2	39	0	0	0	0	0	0	0	0	0	0	11	234				
Filaggrin-2	FLG2	248	4	0	2	0	0	0	0	2	2	3	0	153				
Keratinocyte proline-rich protein	KPRP	64	0	0	0	0	0	2	2	0	0	0	0	81				
Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1	37	0	0	0	0	6	0	0	0	0	9	0	536				
TBC1 domain family member 4	TBC1D4	147	0	2	0	0	0	0	8	0	0	0	0	113	Y			
Deoxynucleoside triphosphate triphosphohydrolase SAMHD1	SAMHD1	72	0	0	0	4	0	0	0	0	9	0	0	120				
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	88	0	0	7	0	0	0	0	0	0	0	0	578	Y			
Programmed cell death protein 2-like	PDCD2L	39	0	0	0	0	5	0	0	0	0	7	0	107	Y			
Isoform C of Prelamin-A/C	LMNA	65	0	0	0	7	0	0	0	0	5	0	0	240				
ATP-dependent RNA helicase DDX3X	DDX3X	81	0	0	0	2	0	0	0	0	6	0	0	498	Y			
Proline-, glutamic acid- and leucine-rich protein 1	PELP1	120	0	0	0	0	0	0	14	0	0	0	0	87				
Isoform 2 of Glyoxalase domain-containing protein 4	GLOD4	33	0	0	0	0	0	0	0	0	0	0	10	81				
CDSN	CDSN	52	0	0	0	0	2	0	0	0	0	2	3	50				
Kinectin	KTN1	156	0	2	0	0	0	0	5	0	0	0	0	132				
Serpin H1	**SERPINH1**	46	0	0	0	0	5	0	0	0	0	6	0	281	Y			
Cyclin-G-associated kinase	GAK	143	0	0	0	0	0	0	6	0	0	0	0	84				Y
Protein arginine N-methyltransferase 1	PRMT1	38	0	0	0	0	5	0	0	0	0	8	0	262				
Activator of 90 kDa heat shock protein ATPase homolog 1	AHSA1	38	0	0	0	0	7	0	0	0	0	5	0	230				Y
Mitogen-activated protein kinase 3	MAPK3	43	0	0	0	0	7	0	0	0	0	10	0	20				
Galactokinase	GALK1	42	0	0	0	0	6	0	0	0	0	4	0	51				

Protein arginine N-methyltransferase 5	PRMT5	73	0	0	0	2	0	0	0	0	9	0	0	285				
Heterogeneous nuclear ribonucleoprotein M	HNRNPM	78	0	0	0	2	0	0	0	0	4	0	0	520	Y			
26S proteasome non-ATPase regulatory subunit 14	**PSMD14**	35	0	0	0	0	0	0	0	0	0	0	12	168				
Malate dehydrogenase, cytoplasmic	MDH1	36	0	0	0	0	0	0	0	0	0	0	12	100				
F-actin-capping protein subunit alpha-1	CAPZA1	33	0	0	0	0	0	0	0	0	0	0	15	318	Y			
Spectrin alpha chain, non-erythrocytic 1	SPTAN1	285	0	0	0	0	0	11	0	0	0	0	0	324				
Mucosal addressin cell adhesion molecule 1	MADCAM1	41	0	0	0	0	3	0	0	0	0	2	0	0				
Vascular cell adhesion protein 1	VCAM1	81	0	0	0	0	2	0	0	0	0	0	0	0				
Histone H1.4	HIST1H1E	22	0	0	0	0	3	0	0	0	0	0	7	605				
Protein sel-1 homolog 1	SEL1L	89	0	0	7	0	0	0	0	7	0	0	0	8			Y	
C-Jun-amino-terminal kinase-interacting protein 4	SPAG9	145	0	4	0	0	0	0	6	0	0	0	0	151				
S-adenosylmethionine synthase isoform type-2	MAT2A	44	0	0	0	0	3	0	0	0	0	10	0	228				
Probable ATP-dependent RNA helicase DDX59	DDX59	69	0	0	0	2	0	0	0	0	6	0	0	0				
Aspartate--tRNA ligase, mitochondrial	DARS2	74	0	0	0	2	0	0	0	0	4	0	0	78				
Clustered mitochondria protein homolog	CLUH	150	0	3	0	0	0	0	4	0	0	0	0	54				
Heterogeneous nuclear ribonucleoprotein D0 (Fragment)	HNRNPD	30	0	0	0	0	0	0	0	0	0	9	0	489				
Elongation factor 1-delta (Fragment)	EEF1D	29	0	0	0	0	0	0	0	0	0	0	12	353				
Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	PDHB	39	0	0	0	0	0	0	0	0	0	0	12	141				
Sulfotransferase 1A1	SULT1A1	34	0	0	0	0	0	0	0	0	0	0	9	78				
Lipoprotein lipase	LPL	53	2	0	0	0	0	2	2	0	2	0	0	2				
Phosphoglycerate kinase 1	PGK1	45	0	0	0	0	0	0	0	0	0	7	0	200				
F-actin-uncapping protein LRRC16A	CARMIL1	152	0	0	0	0	0	0	7	0	0	0	0	42				
Nucleolar transcription factor 1	UBTF	87	0	0	4	0	0	0	0	5	0	0	0	68				
Enhancer of mRNA-decapping protein 4	EDC4	152	0	2	0	0	0	0	5	0	0	0	0	163				
Isoform 2 of Protein unc-45 homolog A	UNC45A	102	0	0	3	0	0	0	0	4	0	0	0	136				
Glutamate--cysteine ligase catalytic subunit	GCLC	73	0	0	0	5	0	0	0	0	8	0	0	2				
Ubiquitin carboxyl-terminal hydrolase 10	USP10	87	0	0	3	0	0	0	0	4	0	0	0	114				
RNA-binding protein 14	RBM14	69	0	0	0	2	0	0	0	0	5	0	0	293				
Signal recognition particle subunit SRP68	SRP68	71	0	0	0	4	0	0	0	0	6	0	0	192				
RNA-binding protein FUS	FUS	53	0	0	0	0	0	0	0	0	7	0	0	428				
SWI/SNF complex subunit SMARCC2	SMARCC2	133	0	0	0	0	0	0	9	0	0	0	0	164				
Clustered mitochondria protein homolog	CLUH	147	0	3	0	0	0	0	4	0	0	0	0	54				
Eukaryotic translation initiation factor 3 subunit B	EIF3B	92	0	0	11	0	0	0	0	0	0	0	0	282				
Prohibitin-2	PHB2	30	0	0	0	0	0	0	0	0	0	0	9	243				
Cell division cycle 2, G1 to S and G2 to M, isoform CRA_a	CDK1	34	0	0	0	0	0	0	0	0	0	0	10	282				
Alpha-actinin-1	ACTN1	103	0	0	7	0	0	0	0	2	0	0	0	314				
Alpha-centractin	ACTR1A	43	0	0	0	0	5	0	0	0	0	4	0	208				
Casein kinase II subunit alpha	CSNK2A1	44	0	0	0	0	3	0	0	0	0	8	0	150				Y
Transmembrane protein 39B	**TMEM39B**	56	0	0	0	0	3	0	0	0	0	5	0	0				
Transportin-1	TNPO1	102	0	0	0	0	0	0	0	8	0	0	0	130				
Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3	HACD3	43	0	0	0	0	0	0	0	0	0	0	8	176				
Polymeric immunoglobulin receptor	PIGR	83	0	0	0	0	0	2	2	3	0	0	0	12				
Protein S100-A8	S100A8	11	0	0	0	0	0	3	2	0	0	0	3	79				
Nucleolar RNA helicase 2	DDX21	87	0	0	0	0	0	0	0	5	0	0	0	438				
Actin-related protein 2	ACTR2	45	0	0	0	0	5	0	0	0	0	4	0	140				
Protein diaphanous homolog 1	DIAPH1	137	0	0	0	0	0	0	5	0	0	0	0	127				
Splicing factor 1	SF1	68	0	0	0	2	0	0	0	0	6	0	0	256				
Uncharacterized protein	LOC102724159	102	0	0	5	0	0	0	0	4	0	0	0	49				
Glutamine synthetase	GLUL	42	0	0	0	0	7	0	0	0	0	5	0	58	Y			
Zinc finger protein 618	ZNF618	105	0	0	4	0	0	0	0	3	0	0	0	3				
Lysophosphatidic acid phosphatase type 6	ACP6	49	0	0	0	0	5	0	0	0	0	4	0	5				
Glycerol-3-phosphate dehydrogenase, mitochondrial	GPD2	81	0	0	0	0	0	0	0	0	3	0	0	47				
Signal recognition particle receptor subunit alpha	SRPRA	70	0	0	0	2	0	0	0	0	6	0	0	115				
Insulin-like growth factor 2 mRNA-binding protein 2	IGF2BP2	67	0	0	0	2	0	0	0	0	9	0	0	257				
Gephyrin	GPHN	80	0	0	0	0	0	0	0	9	0	0	0	73				
DNA-directed RNA polymerase subunit beta	POLR2B	121	0	0	0	0	0	0	8	0	0	0	0	130				
60S ribosomal protein L3	RPL3	46	0	0	0	0	0	0	0	0	0	8	0	402	Y			
ATP-dependent RNA helicase DHX30	DHX30	131	0	0	0	0	0	0	10	0	0	0	0	188				
Splicing factor 3A subunit 1	SF3A1	89	0	0	9	0	0	0	0	0	0	0	0	313				
Staphylococcal nuclease domain-containing protein 1	SND1	102	0	0	8	0	0	0	0	0	0	0	0	265				
Serine/arginine-rich-splicing factor 1	SRSF1	28	0	0	0	0	0	0	0	0	0	0	8	280				
Alanine--tRNA ligase, cytoplasmic	AARS	107	0	0	12	0	0	0	0	0	0	0	0	122				
Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	PPP2CA	36	0	0	0	0	0	0	0	0	0	0	8	193				
Melanoma-associated antigen D1	MAGED1	86	0	0	0	0	0	0	0	10	0	0	0	90				
DNA-directed RNA polymerase subunit beta	POLR2B	133	0	0	0	0	0	0	9	0	0	0	0	130				
Immunoglobulin kappa constant	IGKC	12	0	0	0	0	0	2	0	0	0	0	0	36				
Condensin complex subunit 1	NCAPD2	157	0	0	0	0	0	0	6	0	0	0	0	112				
Coiled-coil domain-containing protein 22	CCDC22	71	0	0	0	2	0	0	0	0	2	0	0	10				

Isoform 2 of Drebrin	DBN1	72	0	0	0	0	0	0	0	6	0	0	0	204				
Developmentally-regulated GTP-binding protein 1	DRG1	41	0	0	0	0	3	0	0	0	0	7	0	172				
Wiskott-Aldrich syndrome protein family member 2	WASF2	54	0	0	0	2	0	0	0	0	5	0	0	103				
Leukotriene A-4 hydrolase	LTA4H	69	0	0	0	3	0	0	0	0	0	0	0	66				
Sorting nexin-2	SNX2	58	0	0	0	2	0	0	0	0	6	0	0	141				
Calcium-binding mitochondrial carrier protein Aralar2	SLC25A13	74	0	0	0	6	0	0	0	0	0	0	0	160				
Sister chromatid cohesion protein PDS5 homolog A	PDS5A	151	0	0	0	0	0	0	5	0	0	0	0	136				
Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 1	PAPSS1	71	0	0	0	0	0	0	0	0	3	0	0	19				
Talin-1	TLN1	270	0	0	0	0	0	10	0	0	0	0	0	217				
Nitric oxide synthase-interacting protein	NOSIP	33	0	0	0	0	0	0	0	0	0	0	11	69	Y			
Nardilysin	NRDC	132	0	0	0	0	0	0	8	0	0	0	0	29				
E3 ubiquitin-protein ligase CHIP	**STUB1**	35	0	0	0	0	0	0	0	0	0	0	9	125			Y	
Host cell factor 1	HCFC1	209	0	0	0	0	0	0	2	2	0	0	0	278	Y			
Isoform 2 of SUMO-activating enzyme subunit 1	SAE1	30	0	0	0	0	5	0	0	0	0	4	0	102				
Zinc finger CCCH-type antiviral protein 1	ZC3HAV1	114	0	0	0	0	0	0	0	4	0	0	0	210				
Ancient ubiquitous protein 1	AUP1	46	0	0	0	0	4	0	0	0	0	3	0	9			Y	
HCG2044799	HNRNPUL2-BSCL	85	0	0	0	0	0	0	0	2	0	0	0	135				
DNA helicase	MCM4	101	0	0	0	0	0	0	0	2	0	0	0	293				
DNA fragmentation factor subunit alpha	DFFA	37	0	0	0	0	3	0	0	0	0	3	0	44				Y
Aspartate aminotransferase, cytoplasmic	GOT1	46	0	0	0	0	4	0	0	0	0	4	0	39				
Histone deacetylase	HDAC8	46	0	0	0	0	4	0	0	0	0	6	0	1				
MAP7 domain-containing protein 2	MAP7D2	82	0	0	3	0	0	0	0	5	0	0	0	51				
Sorting nexin-1	SNX1	59	0	0	0	2	0	0	0	0	6	0	0	140				
2',3'-cyclic-nucleotide 3'-phosphodiesterase	CNP	48	0	0	0	0	2	0	0	0	0	7	0	70				
DnaJ homolog subfamily A member 3, mitochondrial	DNAJA3	52	0	0	0	0	0	0	0	0	0	4	0	116				Y
Transcription elongation regulator 1	TCERG1	124	0	0	0	0	0	0	4	0	0	0	0	183				
Nuclear pore complex protein Nup205	NUP205	228	0	0	0	0	0	8	0	0	0	0	0	158				
Scaffold attachment factor B1	SAFB	103	0	0	0	0	0	0	4	0	0	0	0	206				
ELAV-like protein 1	ELAVL1	36	0	0	0	0	0	0	0	0	0	0	5	322				
Anillin	ANLN	124	0	0	0	0	0	0	5	0	0	0	0	72				
Isoform 3 of Cold shock domain-containing protein E1	CSDE1	91	0	0	0	0	0	0	0	8	0	0	0	191				
Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	37	0	0	0	0	0	0	0	0	0	0	9	239				
Hypoxia up-regulated protein 1	HYOU1	105	0	0	0	0	0	0	6	0	0	0	0	129	Y			Y
Annexin A5	ANXA5	36	0	0	0	0	0	0	0	0	0	0	8	80				
Voltage-dependent anion-selective channel protein 1	VDAC1	31	0	0	0	0	0	0	0	0	0	0	10	213				
F-actin-capping protein subunit alpha-2	CAPZA2	33	0	0	0	0	0	0	0	0	0	0	8	232				
RWD domain-containing protein 2B	RWDD2B	36	0	0	0	0	0	0	0	0	0	0	9	0				
Cystatin-A	CSTA	11	0	0	0	0	0	0	0	0	0	2	0	57				
Peroxiredoxin-1 (Fragment)	PRDX1	19	0	0	0	0	0	0	0	0	0	0	2	549	Y			
Probable ATP-dependent RNA helicase DDX41	DDX41	70	0	0	0	0	0	0	0	0	3	0	0	171				
Eukaryotic translation initiation factor 4 gamma 1	EIF4G1	172	0	0	0	0	0	5	0	0	0	0	0	293				
Zinc finger CCCH-type antiviral protein 1	ZC3HAV1	101	0	0	0	0	0	0	0	4	0	0	0	210				
HLA class I histocompatibility antigen, Cw-6 alpha chain	HLA-C	41	0	0	0	0	6	0	0	0	0	2	0	118				
General transcription factor 3C polypeptide 3	GTF3C3	101	0	0	4	0	0	0	0	6	0	0	0	113				
N-alpha-acetyltransferase 15, NatA auxiliary subunit	NAA15	101	0	0	3	0	0	0	0	5	0	0	0	68				
Voltage-dependent anion-selective channel protein 2 (Fragment)	VDAC2	30	0	0	0	0	0	0	0	0	0	0	6	256				
SWI/SNF complex subunit SMARCC1	SMARCC1	123	0	0	0	0	0	0	5	0	0	0	0	188				
Tripeptidyl-peptidase 2	TPP2	130	0	0	0	0	0	0	6	0	0	0	0	59				
Protein PBDC1	PBDC1	26	0	0	0	0	0	0	0	0	0	0	11	30				
Protein S100-A7	S100A7	11	0	0	0	0	0	2	0	0	0	0	0	61				
TGc domain-containing protein		79	0	0	2	0	0	0	0	0	0	0	0	-				
14-3-3 protein epsilon	YWHAE	29	0	0	0	0	0	0	0	0	0	0	3	447				
RNA 3'-terminal phosphate cyclase	RTCA	39	0	0	0	0	3	0	0	0	0	2	0	31				
Liprin-alpha-1 (Fragment)	PPFIA1	139	0	0	0	0	0	0	2	0	0	0	0	14				
COP9 signalosome complex subunit 3	COPS3	48	0	0	0	0	4	0	0	0	0	2	0	109				
Prolactin regulatory element-binding protein	PREB	45	0	0	0	0	0	0	0	0	0	2	0	7				
E3 ubiquitin-protein ligase TRIM32	**TRIM32**	72	0	0	0	0	0	0	0	0	4	0	0	1				
Diphosphomevalonate decarboxylase	MVD	43	0	0	0	0	4	0	0	0	0	3	0	0				
Protein SGT1 homolog	SUGT1	41	0	0	0	0	3	0	0	0	0	3	0	190				
Cystathionine gamma-lyase	CTH	45	0	0	0	0	3	0	0	0	0	3	0	4				
Nucleoside diphosphate kinase 7	NME7	38	0	0	0	0	2	0	0	0	0	2	0	2				
RNA-binding protein 4	RBM4	40	0	0	0	0	2	0	0	0	0	4	0	82				
2-amino-3-ketobutyrate coenzyme A ligase, mitochondrial	GCAT	45	0	0	0	0	2	0	0	0	0	3	0	6				
Dynactin subunit 1	DCTN1	139	0	0	0	0	0	0	5	0	0	0	0	232				
Cleavage and polyadenylation specificity factor subunit 2	CPSF2	88	0	0	0	0	0	0	0	6	0	0	0	166				
Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	70	0	0	0	0	0	0	0	0	5	0	0	459				
CCR4-NOT transcription complex subunit 2	CNOT2	57	0	0	0	0	0	0	0	0	0	5	0	31				
Ribose-phosphate pyrophosphokinase 1	PRPS1	31	0	0	0	0	0	0	0	0	0	0	5	225				

Armadillo repeat-containing protein 5	ARMC5	108	0	0	0	0	0	0	0	8	0	0	0	0					
Thioredoxin-like protein 1	TXNL1	32	0	0	0	0	0	0	0	0	0	0	5	150					
DnaJ homolog subfamily C member 8	DNAJC8	30	0	0	0	0	0	0	0	0	0	0	5	167					
Electron transfer flavoprotein subunit alpha, mitochondrial	ETFA	35	0	0	0	0	0	0	0	0	0	0	6	148					
Ataxin-2-like protein	ATXN2L	113	0	0	0	0	0	0	6	0	0	0	0	305	Y				
DnaJ homolog subfamily C member 9	DNAJC9	30	0	0	0	0	0	0	0	0	0	0	8	86	Y				
Testis-expressed protein 10	TEX10	106	0	0	0	0	0	0	0	5	0	0	0	115					
Fatty acid-binding protein 5	FABP5	15	0	0	0	0	0	0	2	0	0	0	0	106					
Catalase	CAT	60	2	0	0	0	0	0	0	0	0	0	0	57					
Spermatogenesis-associated serine-rich protein 2	SPATS2	60	0	0	0	0	0	0	0	0	2	0	0	24					
Desmocollin-3	DSC3	100	0	0	0	0	0	0	0	0	0	2	0	15					
Serpin A12	SERPINA12	47	0	0	0	0	2	0	0	0	0	0	0	3					
Transcriptional repressor p66-beta	GATAD2B	63	0	0	0	0	0	0	0	0	2	0	0	178					
Ubiquitin thioesterase otulin	OTULIN	40	0	0	0	0	3	0	0	0	0	2	0	9					
Monocarboxylate transporter 1	SLC16A1	54	0	0	0	0	4	0	0	0	0	0	0	172	Y				
Protein phosphatase methylesterase 1	PPME1	42	0	0	0	0	0	0	0	0	0	2	0	35					
Creatine kinase U-type, mitochondrial	CKMT1A	47	0	0	0	0	2	0	0	0	0	2	0	71					
DNA-directed RNA polymerases I and III subunit RPAC1	POLR1C	39	0	0	0	0	4	0	0	0	0	4	0	84					
mRNA export factor	RAE1	41	0	0	0	0	3	0	0	0	0	4	0	112					
Heterogeneous nuclear ribonucleoprotein A/B	HNRNPAB	30	0	0	0	0	3	0	0	0	0	2	0	442					
Alkyldihydroxyacetonephosphate synthase, peroxisomal	AGPS	73	0	0	0	2	0	0	0	0	3	0	0	56					
Inositol-tetrakisphosphate 1-kinase	ITPK1	46	0	0	0	0	2	0	0	0	0	3	0	3					
Eukaryotic translation initiation factor 3 subunit G	EIF3G	36	0	0	0	0	3	0	0	0	0	2	0	289					
Interferon-induced, double-stranded RNA-activated protein kinase	EIF2AK2	62	0	0	0	2	0	0	0	0	3	0	0	76					
Isoform 4 of Cytosolic acyl coenzyme A thioester hydrolase	ACOT7	37	0	0	0	0	2	0	0	0	0	4	0	128					
Protein NEDD1	NEDD1	72	0	0	0	2	0	0	0	0	4	0	0	36					
Mitochondrial import receptor subunit TOM70	TOMM70	67	0	0	0	0	0	0	0	0	3	0	0	22					
Dual specificity mitogen-activated protein kinase kinase 4	MAP2K4	44	0	0	0	0	0	0	0	0	0	3	0	2					
Mucosa-associated lymphoid tissue lymphoma translocation protein 1	MALT1	92	0	0	0	0	0	0	0	5	0	0	0	10					
Exosome component 10	EXOSC10	101	0	0	0	0	0	0	0	8	0	0	0	210					
Serine/threonine-protein kinase mTOR	MTOR	289	0	0	0	0	0	4	0	0	0	0	0	11					
E3 ubiquitin-protein ligase AMFR	**AMFR**	73	0	0	0	3	0	0	0	0	0	0	0	0				Y	
Serine/arginine-rich-splicing factor 7	SRSF7	16	0	0	0	0	0	0	0	0	0	0	6	351					
Coatomer protein complex, subunit epsilon, isoform CRA_g	**COPE**	37	0	0	0	0	0	0	0	0	0	0	6	93		Y			
Cyclin-dependent kinase 4	CDK4	34	0	0	0	0	0	0	0	0	0	0	5	163					
Hexokinase-1	HK1	102	0	0	3	0	0	0	0	0	0	0	0	103					
Vesicular integral-membrane protein VIP36	LMAN2	40	0	0	0	0	0	0	0	0	0	0	5	22					
Guanine nucleotide-binding protein-like 3	GNL3	62	0	0	0	0	0	0	0	0	6	0	0	256	Y				
Alpha-ketoglutarate-dependent dioxygenase alkB homolog 4	ALKBH4	34	0	0	0	0	0	0	0	0	0	0	5	10					
Peroxiredoxin-2	PRDX2	22	0	0	0	0	0	0	0	0	0	0	2	482					
Dedicator of cytokinesis protein 7	DOCK7	243	0	0	0	0	0	3	0	0	0	0	0	103					
Choline-phosphate cytidylyltransferase A (Fragment)	PCYT1A	34	0	0	0	0	2	0	0	0	0	3	0	59					
Sorting nexin-9	SNX9	67	0	0	0	2	0	0	0	0	2	0	0	73					
F-box only protein 21 (Fragment)	FBXO21	59	0	0	0	0	0	0	0	0	2	0	0	3					
RCC1 domain-containing protein 1	RCCD1	40	0	0	0	0	0	0	0	0	0	4	0	3					
Gamma-tubulin complex component 2	TUBGCP2	103	0	0	0	0	0	0	0	3	0	0	0	41					
TATA-binding protein-associated factor 2N	TAF15	49	0	0	0	2	0	0	0	0	0	0	0	383					
Striatin	STRN	86	0	0	0	0	0	0	0	4	0	0	0	38					
Isoform 2 of Regulator of nonsense transcripts 1	UPF1	123	0	0	0	0	0	0	4	0	0	0	0	117					
Structural maintenance of chromosomes flexible hinge domain-containing protein 1	SMCHD1	226	0	0	0	0	0	3	0	0	0	0	0	155					
HLA class I histocompatibility antigen, A-3 alpha chain	HLA-A	37	0	0	0	0	6	0	0	0	0	0	0	114					
La-related protein 1	LARP1	124	0	0	0	0	0	0	5	0	0	0	0	251					
Transcription elongation factor SPT5	SUPT5H	121	0	0	0	0	0	0	5	0	0	0	0	158					
Ornithine aminotransferase, mitochondrial	OAT	49	0	0	0	0	0	0	0	0	0	6	0	202					
U1 small nuclear ribonucleoprotein A	SNRPA	31	0	0	0	0	0	0	0	0	0	0	6	194					
Ras GTPase-activating-like protein IQGAP1	IQGAP1	189	4	0	0	0	0	0	0	0	0	0	0	197					
Heterogeneous nuclear ribonucleoprotein A0	HNRNPA0	31	0	0	0	0	0	0	0	0	0	0	5	405					
Ribosome biogenesis protein BOP1	BOP1	84	0	0	0	0	0	0	0	3	0	0	0	96					
ER degradation-enhancing alpha-mannosidase-like protein 2	**EDEM2**	65	0	0	0	3	0	0	0	0	0	0	0	0				Y	
Inorganic pyrophosphatase 2, mitochondrial	PPA2	38	0	0	0	0	0	0	0	0	0	0	3	66					
Vacuolar protein sorting-associated protein 37B	VPS37B	31	0	0	0	0	0	0	0	0	0	0	5	2					
Immunoglobulin lambda-like polypeptide 5	IGLL5	23	0	0	0	0	0	0	0	0	0	0	2	20					
DNA mismatch repair protein	MSH6	143	0	0	0	0	0	0	2	0	0	0	0	222	Y				
NAD kinase 2, mitochondrial	NADK2	28	0	0	0	0	2	0	0	0	0	3	0	17					
tRNA methyltransferase 10 homolog C	TRMT10C	47	0	0	0	0	4	0	0	0	0	2	0	76					
Transmembrane protein 39A	**TMEM39A**	56	0	0	0	0	2	0	0	0	0	2	0	0					
Alanine--tRNA ligase, mitochondrial	AARS2	107	0	0	0	0	0	0	0	3	0	0	0	44					
Membralin	TMEM259	46	0	0	0	4	0	0	0	0	0	0	0	1					

Gamma-taxilin	TXLNG	61	0	0	0	2	0	0	0	0	3	0	0	51				
Fragile X mental retardation syndrome-related protein 1	FXR1	70	0	0	0	0	0	0	0	0	4	0	0	140				
Protein AAR2 homolog	AAR2	45	0	0	0	0	0	0	0	0	0	3	0	50				
Ribonucleoside-diphosphate reductase subunit M2	RRM2	43	0	0	0	0	0	0	0	0	0	2	0	64				
PDZ and LIM domain protein 5	PDLIM5	64	0	0	0	0	0	0	0	0	3	0	0	116				
Dynamin-2	DNM2	98	0	0	0	0	0	0	0	3	0	0	0	113				
E3 ubiquitin-protein ligase synoviolin	**SYVN1**	68	0	0	0	2	0	0	0	0	0	0	0	0			Y	
Gamma-tubulin complex component 3	TUBGCP3	104	0	0	0	0	0	0	0	2	0	0	0	46				
Cleavage and polyadenylation-specificity factor subunit 6	CPSF6	52	0	0	0	0	0	0	0	0	3	0	0	299				
Procollagen galactosyltransferase 1	COLGALT1	72	0	0	0	0	0	0	0	0	3	0	0	74				
Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 3	AGAP3	95	0	0	2	0	0	0	0	0	0	0	0	2				
3'-5' RNA helicase YTHDC2	YTHDC2	160	0	0	0	0	0	0	3	0	0	0	0	133				
Cleavage stimulation factor subunit 2 tau variant	CSTF2T	64	0	0	0	0	0	0	0	0	2	0	0	184				
HEAT repeat-containing protein 1	HEATR1	233	0	0	0	0	0	3	0	0	0	0	0	111				
mRNA-decapping enzyme 1A	DCP1A	63	0	0	0	0	0	0	0	0	3	0	0	79				
Ubiquitin carboxyl-terminal hydrolase	USP11	105	0	0	2	0	0	0	0	0	0	0	0	78				
Helicase SKI2W	SKIV2L	138	0	0	0	0	0	0	2	0	0	0	0	18				
PSME3-interacting protein	FAM192A	29	0	0	0	0	0	0	0	0	0	0	2	55				
BAG6	BAG6	119	0	0	0	0	0	0	3	0	0	0	0	76			Y	
Protein flightless-1 homolog	FLII	145	0	0	0	0	0	0	5	0	0	0	0	81				
Serine/threonine-protein phosphatase CPPED1	CPPED1	36	0	0	0	0	0	0	0	0	0	0	5	19	Y			
Mitochondrial glutamate carrier 1 (Fragment)	SLC25A22	33	0	0	0	0	0	0	0	0	0	0	5	7				
Interferon-inducible double-stranded RNA-dependent protein kinase activator A	PRKRA	34	0	0	0	0	0	0	0	0	0	0	4	58				
Spermidine synthase	SRM	34	0	0	0	0	0	0	0	0	0	0	4	85				
Caprin-1	CAPRIN1	78	0	0	0	0	0	0	4	0	0	0	0	282				
Nuclear pore membrane glycoprotein 210	NUP210	205	0	0	0	0	0	4	0	0	0	0	0	82				
Pyrroline-5-carboxylate reductase 2	PYCR2	34	0	0	0	0	0	0	0	0	0	0	2	98				
Inositol polyphosphate-4-phosphatase type I A	INPP4A	110	0	0	2	0	0	0	0	0	0	0	0	3				
Calponin-3	CNN3	36	0	0	0	0	0	0	0	0	0	2	0	213				
mRNA-capping enzyme	RNGTT	69	0	0	0	0	0	0	0	0	2	0	0	21				
Heterogeneous nuclear ribonucleoprotein U-like protein 1	HNRNPUL1	86	0	0	0	0	0	0	0	2	0	0	0	254				
Tropomodulin-3	TMOD3	40	0	0	0	0	2	0	0	0	0	2	0	169				
Serine-threonine kinase receptor-associated protein	STRAP	38	0	0	0	0	2	0	0	0	0	3	0	319				
Wiskott-Aldrich syndrome protein family member 1	**WASF1**	62	0	0	0	0	0	0	0	0	3	0	0	13		Y		
Haloacid dehalogenase-like hydrolase domain-containing 5	HDHD5	46	0	0	0	0	0	0	0	0	0	3	0	53				
Heterogeneous nuclear ribonucleoprotein R	HNRNPR	71	0	0	0	2	0	0	0	0	3	0	0	422				
Exosome complex component RRP43	EXOSC8	30	0	0	0	0	0	0	0	0	0	0	2	93				
Xaa-Pro aminopeptidase 1	XPNPEP1	70	0	0	0	0	0	0	0	0	2	0	0	14				
OTU domain-containing protein 5	OTUD5	61	0	0	0	0	0	0	0	0	3	0	0	4				
Band 4.1-like protein 2	EPB41L2	104	0	0	0	0	0	0	3	0	0	0	0	191				
Medium-chain-specific acyl-CoA dehydrogenase, mitochondrial	ACADM	42	0	0	0	0	0	0	0	0	0	2	0	83				
Paired amphipathic helix protein Sin3a	SIN3A	145	0	0	0	0	0	0	5	0	0	0	0	147				
Moesin	MSN	68	0	0	0	0	0	0	0	0	3	0	0	230				
E3 ubiquitin-protein ligase TRIM33 (Fragment)	**TRIM33**	99	0	0	0	0	0	0	3	0	0	0	0	69				
Serine/threonine-protein phosphatase PGAM5, mitochondrial	PGAM5	32	0	0	0	0	0	0	0	0	0	0	4	223				
Cullin-associated NEDD8-dissociated protein 1	CAND1	136	0	0	0	0	0	0	2	0	0	0	0	275				
Glycerol-3-phosphate phosphatase	PGP	34	0	0	0	0	0	0	0	0	0	0	4	11				
Condensin complex subunit 2 (Fragment)	NCAPH	68	0	0	0	0	0	0	0	4	0	0	0	212				
Isoform 4 of Myotubularin-related protein 5	SBF1	211	0	0	0	0	0	2	0	0	0	0	0	4				
Isochorismatase domain-containing protein 1	ISOC1	32	0	0	0	0	0	0	0	0	0	0	4	40				
Myeloid differentiation primary response protein MyD88	MYD88	33	0	0	0	0	0	0	0	0	0	0	2	0				
Integrin-linked kinase-associated serine/threonine phosphatase 2C	ILKAP	43	0	0	0	0	0	0	0	0	0	5	0	32				
Deoxyribose-phosphate aldolase	DERA	35	0	0	0	0	0	0	0	0	0	0	5	10				
GTPase Era, mitochondrial	ERAL1	48	0	0	0	0	2	0	0	0	0	3	0	40				
Catenin (Cadherin-associated protein), alpha 1, 102kDa, isoform CRA_a	CTNNA1	93	0	0	2	0	0	0	0	2	0	0	0	114				
Inositol polyphosphate 5-phosphatase OCRL	OCRL	104	0	0	0	0	0	0	0	2	0	0	0	21				
Mitogen-activated protein kinase kinase kinase 7	MAP3K7	67	0	0	0	0	0	0	0	0	2	0	0	23				
YTH domain-containing family protein 2	YTHDF2	62	0	0	0	2	0	0	0	0	0	0	0	250				
Extended synaptotagmin-2	ESYT2	102	0	0	0	0	0	0	0	2	0	0	0	77				
tRNA N(3)-methylcytidine methyltransferase METTL2B	METTL2B	43	0	0	0	0	0	0	0	0	0	4	0	47				
DNA-dependent protein kinase catalytic subunit	PRKDC	469	0	0	0	0	0	4	0	0	0	0	0	406				
Polyadenylate-binding protein 2	PABPN1	33	0	0	0	0	0	0	0	0	0	2	0	206				
MAP kinase-activated protein kinase 3 (Fragment)	MAPKAPK3	37	0	0	0	0	0	0	0	0	0	2	0	0				
Tyrosine-protein kinase receptor	IGF1R	155	4	0	0	0	0	0	0	0	0	0	0	2				
60S ribosomal protein L7a (Fragment)	RPL7A	22	0	0	0	0	0	0	0	0	0	0	4	398	Y			
Protein CIP2A	CIP2A	102	0	0	0	0	0	0	0	3	0	0	0	-				
Protein transport protein Sec24B	SEC24B	137	0	0	0	0	0	0	5	0	0	0	0	139				
Nascent polypeptide-associated complex subunit alpha, muscle-specific form	NACA	205	0	0	0	0	0	0	0	0	0	0	4	302	Y			

Fanconi anemia group I protein	FANCI	149	0	0	0	0	0	0	4	0	0	0	0	115				
UPF0600 protein C5orf51	C5orf51	17	0	0	0	0	0	0	0	0	0	0	4	12				
SLIT-ROBO Rho GTPase-activating protein 2	SRGAP2	121	0	0	0	0	0	0	4	0	0	0	0	69				
DNA topoisomerase 1	TOP1	91	0	0	0	0	0	0	0	4	0	0	0	285				
40S ribosomal protein S2	RPS2	31	0	0	0	0	0	0	0	0	0	0	2	447				
DNA-directed RNA polymerase II subunit RPB3	POLR2C	31	0	0	0	0	0	0	0	0	0	0	2	73				
Emerin	EMD	29	0	0	0	0	0	0	0	0	0	0	3	270				
Myb-binding protein 1A (Fragment)	MYBBP1A	140	0	0	0	0	0	0	2	0	0	0	0	301	Y			
Protein FAM49B	FAM49B	37	0	0	0	0	0	0	0	0	0	0	2	9				
Exosome complex component RRP42	EXOSC7	32	0	0	0	0	0	0	0	0	0	0	2	84				
ORC ubiquitin ligase 1	**OBI1**	81	0	0	0	0	0	0	0	2	0	0	0	-				
Nucleolar protein 56	NOP56	53	0	0	0	0	0	0	0	0	2	0	0	208				
Methionine aminopeptidase 1D, mitochondrial	METAP1D	37	0	0	0	0	0	0	0	0	0	0	3	1				
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	MAN1A1	73	0	0	0	2	0	0	0	0	0	0	0	0			Y	
Ubiquilin-4	UBQLN4	64	0	0	0	0	0	0	0	0	2	0	0	22				
Deoxyhypusine synthase	DHPS	31	0	0	0	0	0	0	0	0	0	2	0	32				
Proteasomal ubiquitin receptor ADRM1	**ADRM1**	42	0	0	0	0	0	0	0	0	0	2	0	196				
Nucleoporin Nup43	NUP43	42	0	0	0	0	0	0	0	0	0	2	0	37				
Nuclear protein localization protein 4 homolog	NPLOC4	68	0	0	0	3	0	0	0	0	0	0	0	97			Y	
BTB/POZ domain-containing protein KCTD12	KCTD12	36	0	0	0	0	0	0	0	0	0	2	0	74				
CDKN2A-interacting protein	CDKN2AIP	61	0	0	0	0	0	0	0	0	3	0	0	104				
Immunity-related GTPase family Q protein	IRGQ	63	0	0	0	0	0	0	0	0	3	0	0	8				
Isoform 4 of Tumor protein D54	TPD52L2	21	0	0	0	0	0	0	0	0	0	0	2	68				
tRNA wybutosine-synthesizing protein 4	LCMT2	76	0	0	0	0	0	0	0	0	2	0	0	24				
Peroxisomal multifunctional enzyme type 2	HSD17B4	77	0	0	0	0	0	0	0	0	0	0	2	139				
Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase 1	CMTR1	95	0	0	0	0	0	0	0	2	0	0	0	36				
Symplekin	SYMPK	141	0	0	0	0	0	0	2	0	0	0	0	111				
Geranylgeranyl transferase type-2 subunit beta	RABGGTB	37	0	0	0	0	0	0	0	0	0	0	3	13				
Eukaryotic initiation factor 4A-II	EIF4A2	46	0	0	0	0	0	0	0	0	0	3	0	461				
Double-stranded RNA-specific adenosine deaminase (Fragment)	ADAR	128	0	0	2	0	0	0	0	0	0	0	0	235	Y			
40S ribosomal protein S3a (Fragment)	RPS3A	22	0	0	0	0	0	0	0	0	0	0	3	434				
Heme oxygenase 2	HMOX2	36	0	0	0	0	0	0	0	0	0	0	3	20				
E3 ubiquitin-protein ligase HECTD3	**HECTD3**	97	0	0	0	0	0	0	0	4	0	0	0	3				
Phosphatidylinositol transfer protein beta isoform	PITPNB	32	0	0	0	0	0	0	0	0	0	0	2	61				
WD repeat-containing protein 61	WDR61	34	0	0	0	0	0	0	0	0	0	0	3	112				
RNA polymerase II-associated protein 1	RPAP1	123	0	0	0	0	0	0	3	0	0	0	0	29				
Gem-associated protein 5	GEMIN5	169	2	0	0	0	0	0	0	0	0	0	0	145				
26S proteasome non-ATPase regulatory subunit 13	**PSMD13**	43	0	0	0	0	0	0	0	0	0	2	0	160				
S-formylglutathione hydrolase	ESD	31	0	0	0	0	0	0	0	0	0	0	3	77				
Ankyrin repeat domain-containing protein 54 (Fragment)	ANKRD54	23	0	0	0	0	0	0	0	0	0	0	2	4				
Nucleoprotein TPR	TPR	267	0	0	0	0	0	2	0	0	0	0	0	256				
Eukaryotic translation initiation factor 5B	EIF5B	139	0	0	0	0	0	0	2	0	0	0	0	200				
cAMP-dependent protein kinase catalytic subunit alpha	PRKACA	41	0	0	0	0	0	0	0	0	0	2	0	43				
Terminal uridylyltransferase 4	TUT4	185	0	0	0	0	0	2	0	0	0	0	0	-				
Endoribonuclease Dicer	DICER1	219	0	0	0	0	0	3	0	0	0	0	0	70				
Glutamine-rich protein 1	QRICH1	86	0	0	0	0	0	0	0	2	0	0	0	12				
E3 ubiquitin-protein ligase SIAH2	**SIAH2**	35	0	0	0	0	0	0	0	0	0	0	2	1				
60S ribosomal protein L6	RPL6	33	0	0	0	0	0	0	0	0	0	3	0	432	Y			
Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	GNB1	37	0	0	0	0	0	0	0	0	0	0	2	140				
Serine/threonine-protein phosphatase 6 catalytic subunit	PPP6C	35	0	0	0	0	0	0	0	0	0	0	3	37				
Mitochondrial import receptor subunit TOM34	TOMM34	35	0	0	0	0	0	0	0	0	0	0	2	32				
E3 ubiquitin-protein ligase UHRF1	**UHRF1**	97	0	0	0	0	0	0	0	2	0	0	0	89				
Secretory carrier-associated membrane protein 2	SCAMP2	37	0	0	0	0	0	0	0	0	0	0	2	9				
Four and a half LIM domains protein 3	FHL3	31	0	0	0	0	0	0	0	0	0	0	2	39				
Pyridoxal kinase	PDXK	31	0	0	0	0	0	0	0	0	0	0	2	58				
Interferon regulatory factor 2-binding protein 1	IRF2BP1	62	0	0	0	0	0	0	0	0	2	0	0	13				
Serine/arginine-rich-splicing factor 2 (Fragment)	SRSF2	15	0	0	0	0	0	0	0	0	0	0	2	293				
3-mercaptopyruvate sulfurtransferase	MPST	33	0	0	0	0	0	0	0	0	0	0	2	34				
CAAX prenyl protease 1 homolog	ZMPSTE24	55	0	0	0	0	2	0	0	0	0	0	0	11				
Minor histocompatibility antigen H13	HM13	46	0	0	0	0	2	0	0	0	0	0	0	52			Y	
Microtubule-associated protein 1B	MAP1B	271	0	0	0	0	0	0	0	0	0	0	2	260				
Tyrosine-protein phosphatase non-receptor type 23	PTPN23	179	2	0	0	0	0	0	0	0	0	0	0	57				
Multivesicular body subunit 12B	MVB12B	36	0	0	0	0	0	0	0	0	0	0	2	0				
Phosphorylase b kinase regulatory subunit	PHKA1	133	0	0	0	0	0	0	2	0	0	0	0	32				
Large neutral amino acids transporter small subunit 1	SLC7A5	55	0	0	0	0	0	0	0	0	0	0	2	73				
DNA ligase 3	LIG3	113	0	0	2	0	0	0	0	0	0	0	0	101				
Double-strand-break repair protein rad21 homolog	RAD21	72	0	0	0	0	0	0	0	2	0	0	0	133				
Protein argonaute-2	AGO2	97	0	0	0	0	0	0	0	2	0	0	0	92				

A.4.5 Proteins identified in a BioID assay with TMEM131

T175 flasks were transfected with BioID-TMEM131-GFP, BioID-TMEM131ntd-GFP, TCRα BioID-TCRβ--GFP and TMEM131ntd-GFP. After waiting 24h for the constructs to be expressed, biotin was added to the media to a final concentration of 50µM. Cells were treated for 16h overnight. Cells were lysed and biotinylated proteins enriched with streptavidin-agarose beads. These beads were washed extensively then treated with the on-bead tryptic digest protocol.

Peptides were identified by LC-MS then analysed with Scaffold using a protein confidence threshold of 95% and a peptide confidence threshold of 95%. A minimum of one peptide was used for identification. The values show the number of unique peptides of each protein found in that sample. Using these thresholds no peptides were identified in the TCRα BioID-TCRβ--GFP or TMEM131ntd-GFP samples.

The following lanes show:

- The number of occurrences that protein appears in the Contaminant Repository for Affinity Purification (CRAPome) database out of 716 experiments that followed a Proximity Dependent Biotinylation protocol.
- Whether the protein was also detected in the previous GFP-pulldown experiment (section A.4.4)
- Whether the protein was in the 30 top upregulated or downregulated genes after TMEM131 knockdown listed in the K562 Genome-wide Perturb-Seq library by Replogle *et al.*[357]
- Proteins with the Biological Process Gene Ontology terms for "ERAD pathway" or "Protein Folding" from the Gene Ontology database using the PANTHER enrichment tool.

Some of the proteins discussed in the text are shown in bold. These may not represent all interesting hits.

Identified Proteins	Alternate ID	MW (kDa)	Sample		CRAPome Exp.	Found in GFP	Found in	GO terms	
			B131G	BntdG	Found/716 total	Pulldown	PeterbSeq	Protein Folding	ERAD PATHWAY
Acetyl-CoA carboxylase 1	ACACA	266	6	62	296	Y			
BioID-TMEM131ntd-GFP	**(1603 aa)**	257	0	75	-				
Isoform 3 of Acetyl-CoA carboxylase 1	ACACA	257	6	57	296	Y			
Propionyl-CoA carboxylase alpha chain, mitochondrial	PCCA	80	6	40	284				
Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial	MCCC1	80	4	42	280				
Pyruvate carboxylase, mitochondrial	PC	130	4	37	306				
Macrophage migration inhibitory factor	MIF	12	4	11	211				
Heat shock 70 kDa protein 1B	HSPA1B	70	1	24	698	Y		Y	
Endoplasmin	**HSP90B1**	92	0	36	460	Y	Y	Y	Y
Histone H2B type 1-D	HIST1H2BD	14	0	14	515				
Histone H2B type 2-F	HIST2H2BF	14	0	14	515				
Histone H2B type 1-L	HIST1H2BL	14	0	14	515				
Peptidyl-prolyl cis-trans isomerase A	PPIA	18	1	14	439			Y	
Histone H2B type 1-J	HIST1H2BJ	14	1	9	513				
Endoplasmic reticulum chaperone BiP	HSPA5	72	0	23	661	Y		Y	Y
Elongation factor 1-alpha 1	EEF1A1	50	0	13	653	Y			
UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1	177	0	20	69	Y		Y	Y
Protein RCC2	RCC2	56	0	12	336				
Cofilin-1	CFL1	19	2	6	496				
Tubulin beta chain	TUBB	50	1	13	685	Y			
Peroxiredoxin-1	PRDX1	22	1	8	549	Y			
BolA-like protein 2	BOLA2	10	3	6	250				
Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	MCCC2	61	0	11	251				
Profilin	PFN2	10	0	9	186	Y			
Tubulin beta-4B chain	TUBB4B	50	0	12	678	Y			
Cysteine and glycine-rich protein 2	CSRP2	27	1	8	217				
Profilin	PFN2	10	0	11	186	Y			
60S ribosomal protein L13	RPL13	24	0	11	474				
60 kDa heat shock protein, mitochondrial	HSPD1	61	0	7	520	Y		Y	
Isochorismatase domain-containing protein 2	ISOC2	22	0	7	90				
Isoform 2 of ATPase family AAA domain-containing protein 3A	ATAD3A	66	0	13	229	Y			
Peroxiredoxin-4	**PRDX4**	31	0	8	394		Y	Y	
Actin, cytoplasmic 1	ACTB	42	0	7	667	Y			
Crk-like protein	CRKL	34	0	7	218	Y			
Zinc finger CCCH-type antiviral protein 1-like	ZC3HAV1L	33	0	5	21	Y			
ATPase family AAA domain-containing protein 3B	ATAD3B	73	0	11	172	Y			
Stress-70 protein, mitochondrial	HSPA9	74	0	8	552	Y		Y	
Acetyl-CoA carboxylase 2	ACACB	277	0	8	265				
Heterogeneous nuclear ribonucleoprotein M	HNRNPM	78	0	4	520	Y			
ADP/ATP translocase 2	SLC25A5	33	0	8	525				
Alpha-2-macroglobulin receptor-associated protein	LRPAP1	41	0	6	31				
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	36	0	4	458	Y			
Propionyl-CoA carboxylase beta chain, mitochondrial	PCCB	61	0	3	263				
Ketosamine-3-kinase	FN3KRP	34	0	3	87	Y			
Transport and Golgi organization protein 1 homolog	**MIA3**	214	0	11	51				
Isoform 2 of Transcription factor BTF3	BTF3	18	0	4	117				
Keratin, type II cytoskeletal 1	KRT1	66	0	6	671	Y			
Far upstream element-binding protein 2	KHSRP	73	0	5	403	Y		Y	
60S ribosomal protein L24	RPL24	18	0	6	447				
60S ribosomal protein L6	RPL6	33	0	5	432	Y			
Neutral alpha-glucosidase AB	GANAB	96	0	6	216	Y			
C-1-tetrahydrofolate synthase, cytoplasmic	MTHFD1	102	1	4	311	Y			
Trifunctional purine biosynthetic protein adenosine-3	GART	108	0	5	336				
Elongation factor 2	EEF2	95	0	5	488	Y			
Serum albumin	ALB	69	0	4	239	Y			
CTP synthase 1	CTPS1	67	0	6	312	Y			
Protein disulfide-isomerase A6	**PDIA6**	48	0	7	357		Y	Y	
DnaJ homolog subfamily C member 10	DNAJC10	91	0	5	131			Y	Y
60S ribosomal protein L29	RPL29	19	0	5	399				
Polyubiquitin-B	UBB	17	0	4	457				
Fatty acid synthase	FASN	273	0	5	436	Y			
Transcription factor BTF3	BTF3L4	16	0	2	10				
Nascent polypeptide-associated complex subunit alpha, muscle-specific form	NACA	205	0	4	302	Y			
Isoform 2 of Tetratricopeptide repeat protein 17	**TTC17**	109	0	7	0	Y	Y		
Histone H2A	H2AFZ	13	0	3	491				
Heat shock cognate 71 kDa protein	HSPA8	71	0	6	703	Y		Y	
Protein disulfide-isomerase A3	**PDIA3**	57	0	5	281	Y	Y	Y	
Poly [ADP-ribose] polymerase 1	PARP1	113	0	2	438				
40S ribosomal protein S27	RPS27	9	0	4	411				
Microtubule-associated protein	MAP4	245	0	2	312	Y			
Thioredoxin	TXN	12	0	2	432				
60S ribosomal protein L38	RPL38	8	0	3	300				
60S ribosomal protein L8	RPL8	28	0	4	390				
Inactive C-alpha-formylglycine-generating enzyme 2	SUMF2	34	0	4	18				
Protein disulfide-isomerase	P4HB	57	0	5	247			Y	
Thyroid receptor-interacting protein 6	TRIP6	50	0	4	85				
Tyrosine-protein kinase CSK	CSK	51	0	2	53				
Opioid growth factor receptor	OGFR	73	0	1	31	Y			
Poly(rC)-binding protein 1	PCBP1	37	0	2	472	Y			
Cytoplasmic FMR1-interacting protein 1	CYFIP1	145	0	1	217	Y			
60S ribosomal protein L27a	RPL27A	17	0	3	430				
40S ribosomal protein S11	RPS11	18	0	4	340				
Ataxin-2-like protein	ATXN2L	113	0	1	305	Y			
Probable ATP-dependent RNA helicase DDX17	DDX17	80	0	4	523	Y			
Coiled-coil domain-containing protein 47	CCDC47	23	0	3	100			Y	Y
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	**RPN1**	69	0	4	246	Y	Y		
MICOS complex subunit MIC60	IMMT	79	0	4	93				
Transcription intermediary factor 1-beta	TRIM28	89	0	1	466	Y			
Protein disulfide-isomerase A3 (Fragment)	PDIA3	14	0	1	281	Y	Y		
Keratin, type I cytoskeletal 9	KRT9	62	0	2	577	Y			
14-3-3 protein theta (Fragment)	YWHAQ	17	0	2	410				
Endoplasmic reticulum lectin 1	ERLEC1	55	0	1	2				Y
Cystatin-B	CSTB	11	0	2	164				
Stathmin (Fragment)	STMN1	10	0	4	276				
Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	SDHA	64	0	1	163	Y			
Nitric oxide synthase-interacting protein	NOSIP	33	0	3	69	Y			
Melanoma-associated antigen D2	MAGED2	65	0	2	133	Y			
Guanine nucleotide-binding protein subunit beta-like protein 1	GNB1L	36	0	4	35				
Isoform 2 of Src substrate cortactin	CTTN	71	0	4	318	Y			
Isoform 2 of F-actin-capping protein subunit beta	CAPZB	31	0	1	366	Y			
Protein-L-isoaspartate(D-aspartate) O-methyltransferase	PCMT1	25	0	5	299				

Ribosomal protein L19	RPL19	23	0	1	458				
60S ribosomal protein L28	RPL28	16	0	1	316				
60S ribosomal protein L17 (Fragment)	RPL17	15	0	3	424				
D-3-phosphoglycerate dehydrogenase	PHGDH	56	0	3	394				
60S ribosomal protein L7a	RPL7A	30	0	4	398	Y			
Protein O-mannosyl-transferase TMTC3	TMTC3	104	0	2	2				
Calnexin	CANX	68	0	2	310		Y	Y	Y
Keratin, type II cytoskeletal 2 epidermal	KRT2	65	0	3	628	Y			
Serine/threonine-protein kinase PLK1	PLK1	19	0	1	54				
Trypsin-2	PRSS2	26	0	1	57				
Putative transferase CAF17, mitochondrial	IBA57	38	0	1	46	Y			
60S ribosomal protein L32 (Fragment)	RPL32	16	0	2	184				
Heterogeneous nuclear ribonucleoprotein F	HNRNPF	46	0	1	528				
60S ribosomal protein L3	RPL3	46	0	2	402	Y			
Peptide-N(4)-(N-acetyl-beta-glucosaminyl)asparagine amidase (Fragment)	NGLY1	72	0	3	9	Y		Y	Y
Heterogeneous nuclear ribonucleoproteins C1/C2	HNRNPC	32	0	3	444				
Inner nuclear membrane protein Man1	LEMD3	100	0	3	82				
Phosphate carrier protein, mitochondrial	SLC25A3	40	0	3	362				
Serpin H1 (Fragment)	**SERPINH1**	17	0	3	281	Y			
Cytoskeleton-associated protein 4	CKAP4	66	0	2	168	Y			
MICOS complex subunit	CHCHD3	27	0	2	61				
Nucleobindin-2	NUCB2	50	0	2	34				
Myb-binding protein 1A (Fragment)	MYBBP1A	140	0	2	301	Y			
Glucosidase 2 subunit beta	**PRKCSH**	60	0	2	174	Y	Y		
Neudesin	NENF	19	0	1	27				
Trypsin-3 (Fragment)	PRSS3	19	0	1	91				
Tudor-interacting repair regulator protein	NUDT16L1	19	0	1	46				
Proteasome activator complex subunit 3	PSME3	20	0	1	198				
Mitochondrial 2-oxoglutarate/malate carrier protein (Fragment)	SLC25A11	32	0	1	290				
Heterogeneous nuclear ribonucleoprotein U (Fragment)	HNRNPU	20	0	2	578	Y			
Programmed cell death protein 2-like	PDCD2L	39	0	1	107	Y			
Destrin	DSTN	15	0	1	131				
Nucleoplasmin-3	NPM3	19	1	1	209				
Anaphase-promoting complex subunit 10	ANAPC10	14	0	1	2				
60S ribosomal protein L18 (Fragment)	RPL18	15	0	2	399				
SNW domain-containing protein 1	SNW1	61	0	1	191	Y			
Palladin	PALLD	151	0	2	85				
60S ribosomal protein L14	RPL14	23	0	2	333				
Matrin-3	MATR3	95	0	1	441				
F-actin-capping protein subunit alpha-1	CAPZA1	33	0	1	318	Y			
ATP-dependent RNA helicase DDX3X (Fragment)	DDX3X	20	0	1	498	Y			
Nucleobindin-1	NUCB1	54	0	1	43				
Amino acid transporter	SLC1A5	39	0	2	236				
Thioredoxin domain-containing protein 5	TXNDC5	48	0	3	149			Y	
60S ribosomal protein L11	RPL11	20	0	3	495				
Ubiquitin-like modifier-activating enzyme 1	UBA1	118	0	3	301				
Erlin-1	ERLIN1	39	0	2	81	Y			Y
Beta-sarcoglycan	SGCB	35	0	2	0				
Elongation factor Tu, mitochondrial	TUFM	50	0	2	408				
Haloacid dehalogenase-like hydrolase domain-containing protein 3	HDHD3	28	0	2	11				
60S ribosomal protein L23	RPL23	15	0	2	507				
60S ribosomal protein L10a	RPL10A	25	0	2	280				
60S ribosomal protein L34	RPL34	13	0	1	247				
Protein OS-9	OS9	46	0	2	1				Y
Group XIIA secretory phospholipase A2	PLA2G12A	21	0	1	0				
40S ribosomal protein S14 (Fragment)	RPS14	16	0	1	514				
KH domain-containing, RNA-binding, signal transduction-associated protein 1	KHDRBS1	48	0	1	348				
T-complex protein 1 subunit beta	CCT2	57	0	1	428			Y	
39S ribosomal protein L14, mitochondrial	MRPL14	16	0	1	74				
Protein transport protein SEC23 (Fragment)	SEC23B	6	0	1	95				
Pyridoxal phosphate homeostasis protein (Fragment)	PLPBP	17	0	1	84				
Nuclear pore complex protein Nup155	NUP155	148	0	1	175	Y			
Keratin, type I cytoskeletal 10	KRT10	59	0	2	616	Y			
Interferon regulatory factor 2-binding protein 2	IRF2BP2	61	0	1	51				
Threonine--tRNA ligase, mitochondrial	TARS2	50	0	1	18				
SHC-transforming protein 1 (Fragment)	SHC1	33	0	1	20				
HIG1 domain-containing protein		9	0	1	-				
Phosphotriesterase-related protein	PTER	39	0	1	12				
Mitochondrial-processing peptidase subunit beta	PMPCB	22	0	1	109				
TBC1 domain family member 4	TBC1D4	120	0	1	113	Y			
60S ribosomal protein L18a	RPL18A	18	0	1	300				
Isocitrate dehydrogenase [NAD] subunit beta, mitochondrial	IDH3B	42	0	1	69				
ADP-ribosylation factor GTPase-activating protein 3 (Fragment)	ARFGAP3	16	0	1	11				
Cancer-related nucleoside-triphosphatase	NTPCR	25	0	1	101				
Protein/nucleic acid deglycase DJ-1	PARK7	18	0	2	236				
40S ribosomal protein S16	RPS16	14	0	1	411				
40S ribosomal protein S6	RPS6	25	0	2	466				
Guanine nucleotide-binding protein-like 3	GNL3	62	0	1	256	Y			
3-hydroxyacyl-CoA dehydrogenase type-2	HSD17B10	17	0	2	290				
60S ribosomal protein L13a (Fragment)	RPL13A	24	0	2	294				
Very-long-chain enoyl-CoA reductase	TECR	12	0	2	192				
Chloride channel CLIC-like protein 1	CLCC1	62	0	2	103				
DnaJ homolog subfamily C member 3	DNAJC3	58	0	2	21			Y	
G2/mitotic-specific cyclin-B1	CCNB1	48	0	2	33				
Barrier-to-autointegration factor	BANF1	10	0	1	128				
Calcyclin-binding protein	CACYBP	26	0	1	187				
Dynamin-like 120 kDa protein, mitochondrial	OPA1	99	0	1	41	Y			
DnaJ homolog subfamily C member 9	DNAJC9	30	0	1	86	Y			
Protein disulfide-isomerase	PDIA4	73	0	1	168			Y	
Peptidyl-prolyl cis-trans isomerase FKBP7	FKBP7	26	0	1	0				
Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	SDHB	32	0	1	87				
DnaJ homolog subfamily B member 1	DNAJB1	38	0	1	143			Y	
Glutamate dehydrogenase 1, mitochondrial	GLUD1	61	0	2	217				
Protein canopy homolog 4	CNPY4	28	0	1	0				
Glycoprotein endo-alpha-1,2-mannosidase	MANEA	54	0	1	1				
Kinesin-like protein KIF2C	KIF2C	81	0	1	47				
Glutathione reductase, mitochondrial	GSR	56	0	1	109				
60S ribosomal protein L10	RPL10	20	0	1	369				
Mitochondrial import inner membrane translocase subunit TIM14	DNAJC19	12	0	1	22			Y	
Isoform 3 of Serine/threonine-protein phosphatase CPPED1	CPPED1	13	0	1	19	Y			
FLYWCH family member 2	FLYWCH2	16	0	1	82				
Histone H2A	H2AFY	22	0	1	148				
Hydroxyacylglutathione hydrolase, mitochondrial (Fragment)	HAGH	26	0	1	96				

Protein LSM14 homolog A	LSM14A	51	0	1	227				
Glutaryl-CoA dehydrogenase, mitochondrial	GCDH	48	0	1	24				
Glutamine synthetase	GLUL	57	0	1	58	Y			
26S proteasome non-ATPase regulatory subunit 10	PSMD10	20	0	1	38				
Pre-B-cell leukemia transcription factor-interacting protein 1	PBXIP1	81	0	1	2				
Monocarboxylate transporter 1	SLC16A1	54	0	1	172	Y			
Segment polarity protein dishevelled homolog DVL-2	DVL2	78	0	1	130				
Methyltransferase-like 26	METTL26	20	0	1	4				
Hypoxia up-regulated protein 1	HYOU1	111	0	1	129	Y		Y	
Double-stranded RNA-specific adenosine deaminase (Fragment)	ADAR	128	0	1	235	Y			
Signal peptidase complex subunit 2	SPCS2	25	0	1	17				
Glutamine--tRNA ligase	QARS	7	0	1	291				
4F2 cell-surface antigen heavy chain	**SLC3A2**	19	0	1	189	Y	Y		
Endoplasmic reticulum resident protein 29	ERP29	18	0	1	85			Y	
Golgi integral membrane protein 4	GOLIM4	79	0	1	10				
45 kDa calcium-binding protein	SDF4	28	0	1	45				
Enhancer of rudimentary homolog	ERH	8	0	1	454				
Regulating synaptic membrane exocytosis protein 1 (Fragment)	RIMS1	68	0	1	1				
Reticulocalbin-2 (Fragment)	RCN2	18	0	1	208	Y			
Protein unc-50 homolog	UNC50	32	0	1	0				
Thioredoxin domain-containing protein 12	TXNDC12	19	0	1	194				
Prohibitin	PHB	30	0	1	269				
60S ribosomal protein L22	RPL22	15	0	1	429				
T-complex protein 1 subunit theta	CCT8	60	0	1	451			Y	
Protein canopy homolog 3	CNPY3	31	0	1	1				
DNA mismatch repair protein Msh6	MSH6	153	0	1	222	Y			
Snurportin-1 (Fragment)	SNUPN	7	0	1	7				
eIF-2-alpha kinase activator GCN1	GCN1	293	0	1	246	Y			
Integrin beta-1 (Fragment)	ITGB1	11	0	1	38				
Thioredoxin-dependent peroxide reductase, mitochondrial	PRDX3	28	0	1	315				
m-AAA protease-interacting protein 1, mitochondrial	MAIP1	33	0	1	25				
Transgelin-2	TAGLN2	22	0	1	410				
Host cell factor 1	HCFC1	213	0	1	278	Y			
Stromal cell-derived factor 2-like protein 1	SDF2L1	24	0	1	29			Y	Y