Investigating tumour antigens in mismatch repair proficient gastrointestinal cancers through patient-derived organoids

Alice Newey

The Institute of Cancer Research, University of London

Supervisor: Professor Marco Gerlinger

A thesis submitted for the degree of
Doctor of Philosophy,
University of London

March 2023

Declaration

This thesis was completed under the supervision of Professor Marco Gerlinger, and the work was carried out at the Translation Oncogenomics Lab, Division of Molecular Pathology, The Institute of Cancer Research, 237 Fulham Road, SW3 6JB.

I, Alice Diana Anne Newey, confirm that the work presented in this thesis is my own. Where others have made contributions or information has been derived from other sources, this has been clearly referenced and acknowledged.

Signed: Alice Newey

March 2023

Acknowledgements

I would first like to thank my supervisor, Professor Marco Gerlinger, for the incredible opportunity to work in his group and benefit from his expertise. I am ever grateful to Marco for generously taking a chance on an unconfident undergraduate, as working as a part of his group has enabled me to achieve things I never imagined I would be capable of.

I would also like to thank all the members of my group, both past and present. I am immensely grateful for the expertise of Dr Louise Barber, Dr Marta Buzzetti, and Dr Maria Semiannikova, which was kindly shared throughout our scientific discussions. Their mentorship, advice, and wit has made my time in the group a pleasure, and I am a much more complete scientist for having spent time with them.

There are a number of groups at the ICR who have shown me unending generosity in sharing equipment and laboratory space. I credit these individuals hugely for their help in making this thesis possible. It has been a sheer privilege to work at the Institute of Cancer Research, and I would particularly like to thank those in the Administration and Academic Dean's Teams who have supported me through a COVID-19 extension and lab site moves.

I would like to thank my parents for being such smart, kind, and generous people. They have always shown me the value of hard work, and their compassion created an environment where I felt totally supported to work hard and achieve my goals. It is a privilege to be your daughter. Finally, to my stoic and thoughtful boyfriend Jack, your optimism and silliness have made these 4 years a complete joy. It is no exaggeration to say I could not have done this without your support, I'm so grateful for you.

Abstract

Immunotherapy of gastrointestinal (GI) cancers with immune-checkpoint inhibitors (ICIs) has shown clinical progress in mismatch repair deficient (MMRd) GI cancers, but not in MMRp GI cancers. It is understood that the difference in mutation load and associated neoantigen presentation is behind this difference in immunotherapy sensitivity. Therefore, this thesis focuses on advancing the understanding of peptide and neoantigen presentation in MMRp GI cancers through a multi-omics approach of transcriptomics, proteomics, and peptidomics, alongside patient-derived organoid (PDO) immune cell co-cultures.

Immunopeptidomics of 5 MMRp colorectal cancer (CRC) PDOs was able to reveal a sparse neoantigen landscape, and demonstrate that perturbation with IFNγ or trametinib treatment could not promote the presentation of novel neoantigens. Combining proteomics data with the immunopeptidomics data allowed me to resolve features that influenced peptide abundance change under IFNγ exposure. The data showed peptides generated by chymotryptic activity or presented on HLA-B were more likely to increase in intensity. Analysis also highlighted proline within the core of a peptide (positions 4-6) could promote its downregulation under IFNγ exposure. This data could be used to inform peptide selection for peptide or mRNA cancer vaccines, as it would allow for selection of specific peptides which are presented both in the presence or absence of IFNγ, and do not get downregulated. Furthermore, optimisation of an endoplasmic reticulum (ER) isolation protocol has illustrated that it is possible to detect peptides and neoantigens from organelle preparations. This should support the future study of neoantigen loss in the ER.

Generation of PBMCs and PDOs from gastro-oesophageal adenocarcinoma (GOA) patients enabled the testing of tumour-specific T-cell expansion from MMRp patient PBMCs. Whilst for the patient tested, no tumour-specific T-cell reactivity was seen, the study showed the expansion protocol generates T-cells which can then be interrogated for reactivity, and sensitivity to different immunotherapy combinations.

COVID-19 Impact Statement

During the extended period of COVID-19 disruption much of my work was impacted and I had to adjust my approach. Due to senior staff furlough during lockdown I had to increase my time contribution to clinical trial sample processing, as our clinical trials were still active. In addition to the COVID-19 disruption, in March 2022 my research group moved to a different institute across London. This meant a lot of time was spent preparing and helping with the move, and I spent most of the year working between sites, so a lot of time was lost on travel.

The first way I was affected was institution shutdown. At the time of the institute shutdown I had 4 ongoing pieces of work which had to be ceased: 1) I had optimised a concentration of histone deacetylase inhibitor (HDACi) treatment intended for PDO immunopeptidomics. Upon return to work I did not have sufficient time to expand the patient-derived organoids (PDOs) and submit them for immunopeptidomics analysis. 2) I had just started my first endoplasmic reticulum (ER) isolation protocol, with the view to using the isolates for peptidomics. Following this, I intended to grow up cells to perform more iterations of the ER isolation experiment to improve it. 3) Furthermore, before institute shutdown I had a long-term allogeneic bispecific antibody resistance co-culture running. It was 4 weeks in to a 6-week protocol and it had to be discarded. Had this experiment been able to continue it would have provided a large amount of data: cell growth tracking via Celigo, flow cytometry to assess T-cell response and exhaustion, and RNA sequencing data. 4) I had begun the initial readout development for autologous PBMC-PDO co-culture. This was intended to be followed by establishment of the full protocol, and then utilising it on our trial patient's samples.

Although the time spent in lockdown was used as productively as possible, the time lost to COVID-19 did negatively impact my thesis. Because I did not have the time to expand and HDACi-treat PDOs for immunopeptidomics, I was not able to discern the impact of my optimised HDACi dose on the immunopeptidome. Later, a press release outlining unpublished clinical data suggested the HDACi domatinostat did not provide clinical benefit to patients on immune checkpoint inhibitor treatment, so this research may have not provided clinically useful information. However, the assessment of 5 PDO lines could have shown different

responses, and the understanding the mechanism behind this could have been clinically useful.

Regarding the ER isolation procedure, not having enough time to grow PDO cells meant that I could not directly interrogate my hypothesis of the bottleneck in neoantigen presentation, and instead I had to use HCT116 cells for which I did not have access to the full panel of data. Furthermore, the slow return to work precluded me from re-starting the ER isolation sooner, meaning my optimisation progress was delayed. This meant I did not have time to generate replicates and critically analyse the variability between runs, and the reproducibility of the peptidomics analysis. Additionally, had I had enough time to perform mass spectrometry on an ER isolation run which exhibited more similarity to the prior runs, I could have performed a larger portion of my analysis on a specific subset of data. This could have helped me be able to more confidently assert findings on the ER peptidome. I ultimately believe the time lost to COVID-19 disruption and the lab move have prevented me from making biologically meaningful findings from a promising protocol.

Because I had to discard the long-term allogeneic co-culture, and the return to work was graduated, with risk of further lockdown, this entire project had to be discarded due to feasibility issues. I believe had this project been able to continue it could have uncovered validatable mechanisms behind bispecific antibody resistance, including CEA target downregulation. Finally, because of the delays, it was not possible to thaw precious early-passage PDO cells for the autologous co-culture for many months. I was only able to test one PDO, but if I had been able to start sooner I would have been able to test more PDO cells. Therefore, whilst I have been able to establish the basic function of this protocol, I have not been able to use it to characterise clinical samples. I believe with less COVID-19 disruption I could have generated more novel findings. However, I am immensely grateful to the ICR for the support they provided in this time.

Alice Newey

Professor Marco Gerlinger

Publications

Some of the work presented in this thesis has been published:

Newey A, Griffiths B, Michaux J, Pak HS, Stevenson BJ, Woolston A, et al. Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment. J Immunother Cancer. 2019;7(1):309.

Newey A, Yu L, Barber LJ, Choudhary JS, Bassani-Sternberg M, Gerlinger M. Multifactorial remodeling of the cancer immunopeptidome by interferon gamma. bioRxiv. 2022 Mar 26;2022.03.23.485466.
→ Accepted for publication in Cancer Research Communications in September 2023.

List of Tables

Table 5.2. Percentage of MS-detected gradient-derived and HLA-I-eluted peptides predicted to be strong or weak HLA-I-binders according to NetMHACpan4.1 BA rank.	**228**

List of Figures

Abbreviations

Abbreviation	Definition
Aa	Amino acids
ABC	ATP-binding cassette
APC	Antigen presenting cell
APM	Antigen processing machinery
ATCC	American Type Culture Collection
ATP	Adenosine tri-phosphate
BSA	Bovine serum antigen
CNX	Calnexin
CRT	Calreticulin
CRC	Colorectal cancer
CRISPR	Clustered regularly interspaced palindromic repeats
CTB	CellTitre Blue
CTL	Cytotoxic T-cells
CytC	Cytochrome C
DC	Dendritic cell
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DRiPs	Defective Ribosomal Products
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ERAP	Endoplasmic reticulum associated-aminopeptidases
FA	Formic acid
FBS	Foetal bovine serum
FC	Fold change
FDR	False discovery rate
FITC	Fluorescein isothiocyanate
GFP	Green fluorescent protein
GI	Gastro-intestinal
GOA	Gastro-oesophageal adenocarcinoma
GSEA	Gene Set Enrichment Analysis
HDAC	Histone deacetylase

HEPES	N-2-hydroxyethylpiperazine-N'-2-ethansulfonic acid
HLA	Human leukocyte antigen
HLA-I	HLA class I
HLA-II	HLA class II
HRP	Horseradish peroxidase
IAA	Iodoacetamide
ICI	Immune checkpoint inhibition
IFNγ	Interferon gamma
ISB	Isotonic sucrose buffer
IP	Immunopurification
LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry
LRS	Lymphocyte reduction cone
mAb	Monoclonal antibody
MHC	Major histocompatibility complex
MS	Mass spectrometry
MMR	Mismatch repair
MMRd	Mismatch repair deficient
MMRp	Mismatch repair proficient
NMD	Nonsense mediated decay
PBCP	Packed blood cell pellets
PBMC	Peripherical blood mononuclear cells
PBS	Phosphate buffered saline
PDOs	Patient-derived organoids
PLC	Peptide loading complex
PMSF	Phenylmethylsulfonyl fluoride
RNA	Ribonucleic acid
SD	Standard deviation
SDC	Sodium deoxycholate
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
TA	Tris-acetate
TAA	Tumour-associated antigens
TAP	Transporter associated with antigen processing
TAPBPR	TAP binding protein

TCB	T-cell bispecific
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride solution
TCR	T-cell receptor
TEAB	Triethylammonium bicarbonate
TME	Tumour microenvironment
TMT-MS	Tandem-mass-tag mass spectrometry
TSA	Tumour-specific antigen
Tsn	Tapasin
UHPLC	Ultra-high-performance liquid chromatography
WB	Western blot

Table 1. Abbreviations

Chapter 1 Introduction

1.1 Tumorigenesis

1.1.1 Process of tumorigenesis

Tumorigenesis is a multistep process; transformation of a healthy normal cell to a malignant cell begins when there has been an event of DNA damage or mistake during replication and a failure of accurate DNA repair, allowing a mutation to sustain in the DNA. Due to exposure to mutagens in the environment, over time, point mutations can accumulate in the DNA of healthy tissues (Martincorena *et al.*, 2018; Moore *et al.*, 2020). Most of the mutations gained are neutral passenger mutations, but in some cases this initial transformation event provides a selective advantage in the survival and replication of the cell (Nowell, 1976; Gerstung *et al.*, 2020). For example, some known preferentially early driver mutations seen in multiple cancers are *KRAS*, *TERT*, *CDKN2A*, and *TP53* loss with 17p deletion, which may even be present in pre-cancer evolution (Folkins *et al.*, 2008; Gerstung *et al.*, 2020). The initial changes then combine with an accumulation of subsequent advantageous genetic alterations (such as mutations, amplifications, deletions, inversions, translocations, chromothripsis (Li *et al.*, 2020)) and epigenetic alterations.

These genetic and epigenetic alterations will result in gain-of-function of oncogenes, and/or loss-of-function of tumour suppressor genes which give the transformed cells competitive growth advantage, enabling further clonal and subclonal expansions. This is part of the now widely-accepted Darwinian evolution of cancer (Nowell, 1976; Gerlinger *et al.*, 2012, 2014). Furthermore, the nature of the tumour is shaped by the fitness landscape in which the cancer cells evolve. Nutrient availability, oxygen availability (Lyssiotis and Kimmelman, 2017; Jing *et al.*, 2019; Grimes *et al.*, 2020), space for growth (Chkhaidze *et al.*, 2019), and selection pressure by the immune system (Grasso *et al.*, 2018), are just some of the microenvironmental factors that can influence the future characteristics of the evolving tumour. It is the heterogeneity of a tumour, with subclones exhibiting different advantageous driver mutations, that allows adaptability of the tumour to its niche, and bestows an advantage to select subclones enabling them to

become the dominant clone (Gerlinger *et al.*, 2012, 2014; Gerstung *et al.*, 2020; von Loga *et al.*, 2020). Development of cancer is further enhanced when these changes are supported by the tumour microenvironment (Whiteside, 2008; Wang *et al.*, 2017).

1.1.2 Hallmarks of cancer

Through the acquisition of genome instability and mutations, the resulting malignant cells evolve to develop the 10 hallmarks of cancer which enable the cancer cells to survive, grow, and metastasise (Hanahan and Weinberg, 2000, 2011). These hallmarks are:

(i) genome instability; genome instability can be acquired through failure of at least one component of the DNA repair machinery. The gradual accumulation of mutations through compromised DNA repair permits cells to develop the other cancer cell properties (Salk, Fox and Loeb, 2010).

(ii) an ability to exert tumour-promoting inflammation; the inflammatory response executed by immune cells targeting the cancer cells can paradoxically create a tumour-promoting environment. Inflammation can promote the secretion of growth and proangiogenic factors, and extracellular matrix-modifying enzymes which can aid metastasis (Kaplan *et al.*, 2005; Kitamura *et al.*, 2007; Umar *et al.*, 2009).

(iii) sustained proliferative signalling; cancer cells are able to promote their own division through upregulation or constitutive activation of growth factor receptors, autocrine signalling, and promotion of paracrine signalling through stimulation of the tumour stroma (Witsch, Sela and Yarden, 2010).

(iv) replicative immortality; cancer cells are able to divide without limit due to increased expression of telomerase, which adds protective telomeric repeats back to the eroded telomeres to prevent apoptosis and cell senescence (Blasco, 2005).

(v) resistance to cell death; cancer cells can evade apoptosis through the loss of the tumour suppressor genes like *TP53*, loss or downregulation of proapoptotic genes (e.g. *BAD*, *BAX*, *BAK1*), or by increasing expression of antiapoptotic or survival genes (e.g. *BCL2*, *BCL2L1*) (Adams and Cory, 2007).

(vi) deregulated cellular energetics; cancer cells can adapt their glucose metabolism to instead employ aerobic glycolytic fuelling, whereby glucose is

converted to lactate to provide energy, even in the presence of oxygen (Warburg, Wind and Negelein, 1927; Heiden, Cantley and Thompson, 2009). The result of glycolytic fuelling is the increased biosynthesis of precursors to support cell growth and division, such as amino acid precursors for the synthesis of new proteins, and ribose to make nucleotides for the synthesis of DNA.

(vii) an ability to evade growth suppressors; cancer cells can experience loss-of-function mutations to key tumour suppressor genes such as *RB* and *TP53* (Levine, Momand and Finlay, 1991; Burkhart and Sage, 2008). When functional, these genes could prevent the continuation of the cell division cycle when stress or aberration is detected, and redirect the cell to senescence or apoptosis. Many cancer cells are also resistant to contact inhibition to allow them to grow in an uncontrolled manner. Resistance to contact inhibition is often due to loss of tumour suppressor genes like *NF2*, which functions by stabilising cell-cell junctions and sequestering growth factor receptors (Curto *et al.*, 2007).

(viii) an ability to elude immune destruction; there are multiple mechanisms by which cancer cells can evolve to evade immune destruction, including 'hiding' from the immune system by mutation or downregulation of antigen presentation and human leukocyte antigen (HLA) genes to prevent presentation of tumour antigens (Grasso *et al.*, 2018), and secreting immunosuppressive factors such as TGF-β to supress the infiltrated immune cells (Tauriello *et al.*, 2018).

(ix) an ability to stimulate angiogenesis; cancer cells are able to support the growth of nascent blood vessels through the expression of proangiogenic genes such as *VEGF-A* to support the oxygen and nutrient needs of the growing mass (Carmeliet, 2005). The dysregulated nature of cancer angiogenesis can result in distorted and leaky vasculature which can lead to hypoxia, or enable metastases (Weis *et al.*, 2004).

(x) an ability to invade tissues and metastasise; cancer cells can spread to other areas of the body through the invasion-metastasis cascade. This involves the intravasation of cancer cells in to the blood stream, extravasation in to a distant tissue site, establishment of micrometastases, and finally formation of the full metastatic tumour (Fidler, 2003).

1.2 Peptide presentation

1.2.1 Structure of the HLA super-locus

There are three distinct regions within the HLA super-locus. First, the class I region, which contains classical and non-classical class I genes (HLA-A,B,C and HLA-E,F,G respectively). Second, the class II region which encodes the α and β chains of HLA-DP,DQ,DR molecules alongside genes associated with antigen degradation, processing, and transportation (e.g. peptide transporting proteins transporter associated with antigen processing 1 and 2 (TAP1 and TAP2), chaperone protein Tapasin, and interferon gamma (IFNγ)-inducible immunoproteasome subunits PSMB8 and PSMB9). Third, the class III region which contains genes for innate and adaptive immune responses such as tumour necrosis factor (TNF) and complement proteins C2 and C4A-B (Pagliuca *et al.*, 2022). In total, the HLA super-locus contains 132 genes relating to the function of the immune system, alongside additional genes involved in other cellular processes (Shiina *et al.*, 2009).

1.2.2 Human leukocyte antigen class I and II

Major histocompatibility complex (MHC) class I and II (MHC-I and MHC-II) are the proteins that are expressed on the cell surface to display peptides to the immune system. Human leukocyte antigen class I and II (HLA-I and HLA-II) specifically refers to human MHC. This is the term I will use going forwards as all the work and analyses was performed on human cell lines and organoids.

All nucleated cells express HLA-I, which is a mechanism by which the internal health of the cell is communicated to cytotoxic CD8+ T-cells and NK cells via the presentation of self and non-self-peptides. Only professional antigen presenting cells (APCs) (dendritic cells (DCs), macrophages, and B-cells) constitutively express HLA-II, and present peptides to CD4+ T-cells.

HLA-I and HLA-II are encoded as part of the HLA locus chromosome region 6p21 (chr6p21) (Shiina *et al.*, 2009). HLA-I and HLA-II are also polygenic, meaning that

human DNA contains multiple genes for each HLA class, endowing multiple peptide-binding specificities to an individual.

1.2.3 HLA polymorphisms

Furthermore, HLA-I and HLA-II are highly polymorphic genes. HLA-A, B, and C are the HLA-I alleles, and HLA-DP, DQ, and DR are the HLA-II alleles, and the different subtypes of those genes are termed HLA allotypes. There are currently 24,639 HLA-A, B, and C allotypes, and 10,327 HLA-DP, DQ, and DR allotypes recorded on the ImMunoGeneTics project (IMGT/HLA) HLA database from across the global population (as of 21st March 2023), with the database still growing. Between 1998 and 2014, the IMGT/HLA database grew and average of 29% per year (Robinson *et al.*, 2015).

HLA allotypes can differ by up to 20 amino acids, with the most polymorphic amino acids lining pockets in the peptide binding groove, defining the peptides the HLA is able to present (Charles A Janeway *et al.*, 2001).

In HLA-I the peptide binding groove is closed at both ends by tyrosine residues which restricts the binding capacity predominantly to peptides 8-10 amino acids (aa) in length, with some class I allotypes permitting 11aa peptides (Rammensee, Friede and Stevanoviíc, 1995; Wieczorek *et al.*, 2017). The binding groove of HLA-II is unrestricted at the ends, therefore it is able to bind peptides 13-25aa long, as the N-terminus of the peptide is able to extend out the HLA binding groove.

In HLA-I the binding specificities are defined by the ‘anchor’ residues required at positions P2, P9, and possibly also secondary anchor residues at P3/P5/P6 (Nguyen, Szeto and Gras, 2021). The HLA-II binding specificities are defined by the anchor residues required at P1, P4, P6, and P9 (Wieczorek *et al.*, 2017).

Certain HLA allotypes have even been implicated in disease susceptibility and treatment sensitivity which is often linked to the ability of the HLA to present peptides to the immune system. For example, HLA-DRB1 has been linked with susceptibility to rheumatoid arthritis, with evidence showing that some of the

allotypes of *HLA-DRB1*, such as 04:01, 04:04, 04:05, shared an increased ability to present citrullinated peptides to T-cells, provoking an immune response (Howell, 2014). Different HLA allotypes have also been implicated in cancer susceptibility. For example, HLA alleles *B*4402* and *DQB1*0301* have been shown to be risk factors for the development of cervical cancer, and the effect was enhanced when both HLA alleles occurred together (Madeleine *et al.*, 2008). Some studies have further linked specific HLA allotypes with patient sensitivity to immunotherapy. For example, although the mechanism is currently undefined, *HLA-A*03* has been linked with poor sensitivity to immune checkpoint inhibition (ICI) therapy (Naranbhai *et al.*, 2022).

1.2.4 Summary of peptide presentation on HLA-I

The immunopeptidome is defined as the collection of peptides presented on HLA molecules on the surface of cells. The first step in peptide presentation is translation of the source protein by the ribosomes. Peptides can be generated by the proteasome, a multicatalytic cleavage complex, either rapidly during translation due to premature translation termination, nonsense mediated decay (NMD) or defective protein folding (these peptides are known as Defective Ribosomal Products - DRiPs); or produced from the digestion of retired full-length proteins (known as retiree peptides – the source protein half-life defined by its protein stability) (Yewdell, Antón and Bennink, 1996; Reits *et al.*, 2000; Apcher *et al.*, 2011; Bourdetsky, Schmelzer and Admon, 2014). A 2022 publication has shown that DNA damage can initiate the pioneer round of translation, resulting in increased production of peptides and a corresponding quantifiable increase in cell surface HLA-I abundance (Uchihara *et al.*, 2022). This supports the functional relevance of the pioneer round of translation as a source of antigens.

In normal conditions, ubiquitinated proteins are actively unfolded by the 19S proteasome cap and fed in to the hollow catalytic 20S core particle of the proteasome, where different catalytic enzymes processively exert their caspase-like, trypsin-like, and chymotrypsin-like cleavage activities (PSMB6, PSMB7, PSM5 respectively) to break down the proteins in to peptides 8-25aa long. These peptides can then be further trimmed down or totally destroyed by cytosolic peptidases such as THOP1 (Saric *et al.*, 2001; Kim *et al.*, 2003) and Tripeptidyl

peptidase II (TPPII) (Reits *et al.*, 2004). The surviving peptides are actively transported in to the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP) complex (Ghanem *et al.*, 2010; Lehnert and Tampé, 2017). The peptide may then be further cleaved at the N-terminus by ERAP1 or ERAP2, before being loaded on the HLA-I peptide loading complex (PLC) (HLA-I complexed with β2M and chaperone proteins calreticulin (CRT), ERp57 and tapasin (Tsn)). The binding cleft of HLA-I is most tolerant of peptides 8-11aa long, with binding motifs specific to different HLA-I allotypes according to their highly-conserved biochemical structure (Rammensee, Friede and Stevanoviíc, 1995). Tsn has been shown to stabilise HLA-I in its peptide-receptive state until a peptide with high affinity to the HLA-I allotype binds (Purcell, 2000; Purcell *et al.*, 2001; Praveen *et al.*, 2010). Once a high affinity peptide has stably docked in the HLA-I, the final glucose of the N-linked glycan on asparagine 86 of HLA-I is removed by glucosidase II (Radcliffe *et al.*, 2002). This allows the peptide-HLA complex to dissociate from calreticulin and be released from the rest of the PLC for presentation on the cell surface (Fleischmann *et al.*, 2015; Thomas and Tampé, 2017). It is the specificity of these proteins, alongside the binding specificities of the HLA allotypes, that help to shape the characteristics of the immunopeptidome.

1.2.5 Shaping of the immunopeptidome by protein specificities

1.2.5.1 Proteasome

Experiments using proteasome inhibitors have implicated the proteasome in the proteolysis of the large majority of expressed intracellular proteins, and in producing a large proportion of HLA-I peptides (Rock *et al.*, 1994). The proteasome can generate peptides 4-24aa long, with only 30% of the proteasomally-derived peptides being ≥8aa long (Kisselev *et al.*, 1999). The constitutive proteasome has chymotryptic (PSMB5 cleaves after A, F, I, L, M, V, or Y), tryptic (PSMB7 cleaves after K or R), and caspase-like (PSMB6 cleaves after D or E) activity (Harris *et al.*, 2001; Huber *et al.*, 2012). The immunoproteasome catalytic site is composed of two inducible subunits which have trypsin-like activity (PSMB9 and PSMB10), and a subunit with chymotryptic-like activity (PSMB8), but with a larger binding pocket to facilitate the binding of a hydrophobic P1 residue, which is more compatible for HLA-I binding.

Specifically, mice deficient in PSMB8, but not PSMB9 and PSMB10, have shown a 50% lower expression of MHC-I (Fehling *et al.*, 1994; Huber *et al.*, 2012). The specificity of the proteasome and the balance of constitutive and immunoproteasome subunits begins the shaping of the immunopeptidome.

A 2018 publication utilised reversible cross-linking technology of peptides in close proximity to PSMA1. This was followed by PSMA1 immunoprecipitation (IP) to isolate proteasome-associated peptides to help advance the scientific understanding of the degradome (Wolf-Levy *et al.*, 2018). This technique was important as they showed they were able to analyse actively degrading proteins, particularly being able to capture high-turnover and short-lived proteins. The study was able to resolve ~ 5,000 peptides and successfully identified peptide C-termini generated by caspase and tryptic-like activity. The method was able to detect peptide abundance changes in IFNγ or TNF-α-treated HEK293 cells. This enabled resolution of changes in proteasome dynamics and revealed different proteasome substrate specificities in peripheral blood mononuclear cells (PBMCs) of healthy donors and patients with systemic lupus erythematosus. However, although this technique could identify proteasomal-adjacent peptides, it could not identify specifically post-proteasomal peptides, or track their further processing. This does not help to address my question of which peptides survive peptidase processing and get transported in to the ER.

1.2.5.2 Cytosolic peptidases

As will be mentioned in Chapter 2, studies have shown that up to 99% of all peptides produced by the proteasome are vulnerable to degradation by peptidases (Reits *et al.*, 2003). In addition, multiple peptidases are thought to help form the HLA-presented peptidome by cleaving the N-terminus of proteasomal peptides (Rock, Farfán-Arribas and Shen, 2010). For example, the peptidases LAP3 (Beninga, Rock and Goldberg, 1998); THOP1 (Saric *et al.*, 2001); BLMH (Stoltze *et al.*, 2000); NPEPPS (Towne *et al.*, 2008); NRDC (Kessler *et al.*, 2011; Van Endert, 2011); and (although controversial) TPPII (Seifert *et al.*, 2003; van Endert, 2008; Van Endert, 2011) have been associated with the production of HLA-compatible peptides. Leucine aminopeptidase 3 (LAP3) is a peptidase that cleaves leucine residues from the N-terminus of

peptides. Whilst it has been associated with peptide presentation (Beninga, Rock and Goldberg, 1998), research has shown it is not essential for class 1 peptide presentation (Towne *et al.*, 2005). Thimet oligopeptidase 1 (THOP1) cleaves any peptide between 5-22aa long, with a preference for 9-12aa long, with no amino acid sequence specificity (Ferro, Gewehr and Navon, 2020). THOP1 has been implicated in both the destruction and production of HLA-peptides (Kessler *et al.*, 2011). Both bleomycin hydrolase (BLMH) and aminopeptidase puromycin sensitive (NPEPPS) have been connected with the processing of class I peptides (Stoltze *et al.*, 2000; Kim, Kwak and Ahn, 2009), although their substrate septicity is not well characterised. TPPII is a cytosolic peptidase that cleaves triplets of amino acids from the N-terminus of peptides. Studies have shown that whilst TPPII can convert peptides from longer to shorter forms, it is not required for peptide presentation and it may only contribute to peptide presentation by the cleavage of very long precursor peptides (Kawahara *et al.*, 2009).

Whilst the contribution of each cytosolic peptidase may be contested, it is known that peptidases can help to shape the immunopeptidome through their cleavage specificities. Furthermore, the cleavage of peptides to an optimal length facilitates their transport in to the ER for further processing and HLA-loading.

1.2.5.3 TAP1/2

The TAP transport complex is a heterodimer of ATP-binding cassette (ABC) family proteins TAP1 and TAP2 which transports peptides from the cytosol in to the ER (Lehnert and Tampé, 2017). The TAP complex transports peptides with a length of 8-40aa, with a preference for those 9-12aa long (Uebel and Tampé, 1999; Lehnert and Tampé, 2017). Further to its length preference, TAP also has positional amino acid preferences. TAP has a preference for positively charged residues in P1-2 of the peptide, aromatic at P3, and hydrophobic and basic residues at the peptide C-terminus. TAP also has a bias against proline in P2, alongside a bias against acidic residues in P3 (Lehnert and Tampé, 2017). Proline is disfavoured in P2 because it causes distortion of the peptide backbone, preventing P1-3 binding to TAP. HLA-peptides with proline in P2 and acidic residues in P3 have been detected on HLA-I molecules, but likely due to the aminopeptidase activity of ERAPs cleaving longer precursor peptides. There is

little to no selectivity for specific amino acids within peptide positions 4-8. In the finally processed and HLA-I loaded peptide, these are important positions for T-cell receptor (TCR) recognition. The selection preferences of TAP1/2 mirror some of the preferences of HLA-I, such as preferences against proline in P2 and for hydrophobic or basic amino acids at the C-terminus. This essentially acts as a pre-selection step, presumably to transport possible HLA-binding candidates in to the ER for HLA-loading. Further to this, studies of TAP transport into proteoliposomes showed that TAP activity is controlled by a secondary low-affinity binding site in the ER luminal surface. This limited ER luminal peptide concentration to 16μM (Grossmann *et al.*, 2014). Therefore, when peptide concentration in the lumen gets too high, peptides saturate the low-affinity binding site and cause trans-inactivation of the TAP transporter. These binding specificities and activity limitations provide further restrictions on peptides' entry to the ER. However, due to the similarity in amino acid specificities to those of HLA-I, TAP selectivity cannot fully explain the absence of NetMHC-predicted HLA-I binders from expressed proteins in our PDOs. It is possible that analysis of ER peptidomes under different conditions may help us analyse peptide abundances under different conditions.

1.2.5.4 Tapasin and TAP-binding protein-related (TAPBPR)

Tapasin (Tsn - gene name TAPBP) is a multifunctional chaperone protein that connects HLA-I to the TAP1/2 heterodimer, stabilising TAP1/2 and bringing HLA-I close to the supply of new HLA-I-candidate peptides (Thomas and Tampé, 2017). Furthermore, Tsn facilitates the process of peptide editing – the process of optimising the peptide repertoire by selection of high affinity peptides. Tsn achieves this by stabilising the high-energy open conformation of the HLA-binding groove, enabling peptide exchange until a high affinity peptide binds the HLA-binding groove (Thomas and Tampé, 2017). When a high affinity peptide binds to the HLA peptide cleft, a conformational change occurs in the HLA, which enables its release from Tsn and the rest of the PLC. The result is that only peptides with the correct length and high affinity to HLA will enable release from the PLC and transport to the cell surface. This process is always in flux, controlled by peptide supply and Tsn abundance: when Tsn abundance is low relative to peptide supply, lower affinity peptides are able to bind HLA-I and be released to

the cell surface, however, as Tsn abundance increases, the potential for immunopeptidome editing increases (Boulanger *et al.*, 2022). It has also been shown that Tsn can help to promote the presentation of low-copy but high affinity peptides (Boulanger *et al.*, 2022). Some HLA allotypes have self-editing capability, meaning they are able to exert peptide selection activity through changes in conformation; it is the HLA allotypes with the largest conformational flexibility that have the highest capacity for self-editing (Williams *et al.*, 2002; Bailey *et al.*, 2014, 2015). Therefore, tapasin supply and its relationship to the HLA allotypes could be another important factor shaping the immunopeptidome.

Whilst tapasin performs the majority of the peptide loading and editing, TAPBR, a tapasin homologue, also contributes to shaping of the immunopeptidome. TAPBR is present in the more peptide-scarce secretory pathway and operates as a fine-tuning step in peptide selection (Boyle *et al.*, 2013; Hermann *et al.*, 2015; Morozov *et al.*, 2016). TAPBR is implicated as a peptide exchange catalyst via multiple mechanisms, including acting as a cap over the peptide-bound HLA binding cleft to prevent premature peptide release during selection (McShan *et al.*, 2018). Furthermore, TAPBPR has been shown to associate with UGT1, an ERGIC-resident enzyme that reglucosylates N-linked glycan moieties on incompletely folded glycoproteins (Hammond, Braakman and Helenius, 1994; Caramelo *et al.*, 2004). UGT1 can therefore reglucosylate suboptimally loaded or empty HLA-I, enabling recognition by calreticulin in the PLC, facilitating peptide exchange (Neerincx *et al.*, 2017). Therefore, the interaction of TAPBR with UGT1 provides another layer of peptide optimisation from TAPBR.

These factors therefore influence the immunopeptidome and may contribute to the bottleneck in peptide presentation.

1.2.5.5 ERAPs

Endoplasmic reticulum associated-aminopeptidases (ERAPs) are aminopeptidases which are able to cleave HLA-precursor peptides to make them compatible with HLA-binding. ERAP1 and ERAP2 have a preference for peptides 9-16aa long, with very low activity against peptides smaller than 8mers (Saric *et al.*, 2002; Chang *et al.*, 2005). ERAP binds peptides with a hydrophobic C-

terminus in its hydrophobic pocket, which is spatially separated from the catalytic site (Chang *et al.*, 2005). This spatial separation acts as a molecular ruler mechanism where ERAPs can sequentially cleave amino acids from the N-terminus of the peptide until it reaches 8-9aa long. There has long since been a discussion as to whether ERAPs function by cleaving free peptides or peptides bound to MHC-I (Chen *et al.*, 2016), with one study suggesting this was made possible by partial dissociation of peptide N-termini from MHC-I (Papakyriakou *et al.*, 2018). However, a 2021 study utilising peptides linked to MHC-I by a disulphide bond, has showed that ERAP1 cannot cleave 12-mer peptides, which can be cleaved in solution, bound to MHC-I (Mavridis *et al.*, 2021). Further work should be performed to definitively determine the mechanism of ERAP1 cleavage of peptides, as it will indicate the first driver of selection in peptide presentation in the ER.

ERAP1 can cleave away all amino acids from the N-terminus (aside from proline) but has a strong preference for leucine, and to a lesser extent methionine, alanine, and phenylalanine (Schomburg *et al.*, 2000; Evnouchidou *et al.*, 2008; Nguyen *et al.*, 2011). Furthermore, data has shown the internal sequence of the peptide can determine the rate of aminopeptidase activity. ERAP1 has a preference for hydrophobic and positively charged residues most significantly at positions 2, 5, 7, with some preference also seen at positions 4, 8, and 9 (Evnouchidou *et al.*, 2008).

ERAP2 is another ERAP, but with much more limited expression, being absent in 25% of individuals. ERAP2 also has a much more restricted specificity, cleaving peptides 9aa in length, or less, with N-terminal arginine and lysine (Saveanu *et al.*, 2005; López de Castro, 2018). ERAP1 and ERAP2 primarily exist as monomers, but have been shown to be able to form heterodimeric complexes. Data suggests ERAP1 and ERAP2 are able to work cooperatively for efficient processing of HLA-I precursors with N-extensions containing both hydrophobic and basic residues (Saveanu *et al.*, 2005; Chen *et al.*, 2016).

ERAP1 and ERAP2 are polymorphic genes. 99% of ERAP1 in the global population is represented by 10 allotypes, and the polymorphisms can affect enzyme activity, specificity, or expression level (Reeves *et al.*, 2013;

Stamogiannos *et al.*, 2015; Hanson *et al.*, 2018; Hutchinson *et al.*, 2021). In ERAP1 most of the polymorphisms occur near the catalytic or peptide binding site (Ombrello, Kastner and Remmers, 2015). In the case of ERAP2, the effects of most polymorphisms are very minor. However, there are a few cases of polymorphisms causing changes in ERAP2 activity (K392N) (Evnouchidou *et al.*, 2012) or protein expression (rrs1004354) (Alvarez-Navarro *et al.*, 2015). Some ERAP polymorphisms have been linked to certain diseases, often in association with particular HLA-I allotypes. ERAP variants have been linked to ankylosing spondylitis (Burton *et al.*, 2007; Evans *et al.*, 2011; Reeves *et al.*, 2014), Behçet's disease (Kirino *et al.*, 2013; Guasp *et al.*, 2016), and psoriasis (Strange *et al.*, 2010; Lysell *et al.*, 2013; Wiśniewski *et al.*, 2018). This illustrates the large impact that ERAP activity can have on the immunopeptidome.

This information highlights that ERAP activity provides yet another layer of differential selection to peptide presentation. However, the peptide length preferences and C-terminus preferences mirror that of HLA-I, so it may not entirely explain the absence of NetMHC-predicted HLA-I binders.

1.3 Immunogenicity in cancer

1.3.1 Cancer antigens

An antigen is defined as a molecular structure that can be bound to by an antibody or TCR, whereas, an epitope is defined as the part of that antigen to which the antibody or TCR binds. This work focuses on HLA-presented antigens, which are the entire 8-25aa peptide sequences bound to HLA-I and HLA-II (Wieczorek *et al.*, 2017), and not the specific epitope to which TCRs bind.

Presentation of peptides on HLA class I (HLA-I) is a central component of the adaptive immune system. Presenting self-peptides on HLA-I signals that a cell is healthy, whereas presentation of non-self or mutated/overexpressed peptides signal that a cell is infected or malignant, marking it for destruction. There are multiple subclasses of cancer antigens. Non-mutated cancer-associated antigens, also known as tumour-associated antigens (TAAs), are antigens that can be derived from: aberrantly expressed proteins that normally are phase-restricted markers of differentiation – differentiation antigens (e.g. gp100);

proteins which are normally restricted to being expressed in the testes, foetal ovaries, or trophoblasts – cancer/testis antigens (e.g. MAGE-1); or proteins which are deleteriously overexpressed, demonstrating malignancy – overexpressed antigens (e.g. ERBB2) (Zamora, Crawford and Thomas, 2018). Neoantigens are a subclass of cancer antigen. Neoantigens are peptide antigens which contain mutations encoded from their source mutated gene. Neoantigens are a tumour-specific antigen (TSA) because they are only presented on the tumour cells which contain the encoding mutation.

Presentation of cancer antigens on HLA enables the recognition of the malignant cell by the immune system, leading to their destruction by immune cells, including cytotoxic CD8 T-cells. Non-mutated cancer antigens are often tumour type-specific, but not patient-specific, making these markers suitable for more broadly-applicable types of therapy. Neoantigens are often more patient-specific, so are more suitable for personalised immunotherapy.

1.3.2 The cancer immunity cycle

Understanding the full cancer immunity cycle is important to provide context to the discipline of immunopeptidomics. The term ‘Cancer-Immunity Cycle’ was first described in 2013 (Chen and Mellman, 2013), and it outlines the multi-step cycle that must be completed in order to successfully eliminate immunogenic tumour cells. In the first step, ongoing mutagenesis in a tumour generates mutation-containing proteins that can be broken down in to peptides, some of them being mutation-containing neoantigens (described above). As well as presentation on HLA, neoantigens can be released when cells die an immunogenic or necrotic cell death, and then be captured by DCs. If this occurs in the presence of secondary immune signalling such as proinflammatory cytokines or damage-associated molecular patterns released from dying tumour cells, then the DCs will mature, and travel to the lymph nodes where they will present the antigen on HLA-I or HLA-II to T-cells. This interaction will prime and activate the T-cells against the ‘foreign’ antigen. The nature of the T-cell response is controlled by the balance of stimulatory and inhibitory factors expressed on the DC and in the environment. Once activated, the T-cells travel through the vasculature to the tumour niche, and their antigen-specific TCR will recognise the cancer antigen

bound to HLA, stimulating them to exert their cytotoxic effector functions. The killing of the cancer cells results in the release of further TAAs. In some cases, novel neoantigens/TAAs may be released in a phenomenon known as epitope spreading.

1.3.3 The influence of the immune system on cancer evolution: the 3 E's of immunoediting (Dunn, Old and Schreiber, 2004)

Immunoediting is the theory that through the selection pressure that the immune system imparts on the tumour, classical Darwinian evolution allows the tumour to lose any features that may target it for destruction by the immune system, and also make it insensitive to immune attack. Data from non-small cell lung cancer (NSCLC), bladder cancer, and melanoma has shown that neoadjuvant immune checkpoint inhibition therapy induced a stronger T-cell response in patients with earlier stage disease, when compared against the patients with stage IV disease. It is possible that the depletion of neoantigens through immunoediting could be responsible for the poorer response in the advanced cancer, alongside the development of an immunosuppressive and immune-occlusive tumour microenvironment (Chalabi *et al.*, 2020). Immunoediting can be divided in to 3 main phases.

1.3.3.1 Stage 1: Immunosurveillance (elimination)

As described above, development of resistance to killing by the immune system is a major hallmark of cancer. The first stage of the immunoediting is the 'Elimination' of cancer cells bearing TAAs through immunosurveillance (Dunn *et al.*, 2002). Immunosurveillance theory states that the immune system is able to identify and eliminate malignant cells, first formally proposed in 1957 (Burnet, 1957; Dunn *et al.*, 2002). The importance of immunosurveillance for cancer elimination is made evident through the increased rates of cancers seen in immunocompromised patients (Vajdic and Van Leeuwen, 2009). Furthermore, a longitudinal study of a pre-invasive lung cancer precursor, lung carcinoma in situ, showed that half the patient's cancers regressed or did not progress, illustrating the role the immune system can have in cancer control and killing (Teixeira *et al.*, 2019).

Some of the first evidence used to support this theory was the illustration that immunising inbred mice strains against implanted syngeneic tumours could help to facilitate tumour rejection, exemplifying the existence of TAAs (Old and Boyse, 1964; Klein, 1966). Further to this, patient co-culture experiments of irradiated melanoma cells and autologous cytotoxic T-cells (CTLs) were able to show specific anti-tumour activity (Hérin *et al.*, 1987). When these CTLs were then used in immunoselection experiments, they were able to detect six different melanoma antigens, one of which was later shown to have derived from MAGEA1 (Van Der Bruggen *et al.*, 1991). Many more TAAs have since been identified.

1.3.3.2 Stage 2: Equilibrium

During the equilibrium phase, any surviving tumour cell variants and the cells of the immune system enter a dynamic equilibrium; the lymphocytes enacting cytotoxic function and secreting IFNγ exert significant enough selection pressure to contain the initial tumour mass. During this Darwinian evolution immunostimulatory tumour variants are destroyed, but the surviving variants continue to divide and mutate, developing features that allow them to subvert the immune response. This is likely to be the longest phase in tumour evolution, and it is where the majority of the driver and passenger mutations are acquired.

1.3.3.3 Stage 3: Immune escape

Once the surviving tumour variants have acquired resistance to immunological attack, the cells are able to divide uninhibited, resulting in tumour outgrowth (Matsushita *et al.*, 2012). This is often the stage where tumour cells are also able to leave their niche and establish secondary tumours in other tissues (often defined by classic tissue tropisms of particular tumour types).

1.3.3.4 The effect of treatments of antigenicity

One of the first examples of immunoediting was described in mice (Urban *et al.*, 1982). The study showed immune-competent mice selected against tumour-specific antigen-expressing tumour cells leading to antigen loss, but nude mice did not impart such selection. This biological process is extremely clinically

relevant as the immunoediting effect has now been observed in many clinical trials/settings. For example, reduced predicted neoantigen number was observed in the metastasis of a patient with a primary pancreatic cancer (Balachandran *et al.*, 2017). Additionally, loss or reduction of neoantigens in patients that have received immunotherapy has been seen across multiple tumour types, including melanoma (Verdegaal *et al.*, 2016), glioblastoma (Zhao *et al.*, 2019), and NSCLC (Anagnostou *et al.*, 2017; McGranahan *et al.*, 2017; Rosenthal *et al.*, 2019). It is for this reason that it is most clinically beneficial to target multiple antigens, as it reduces the chance of total antigen loss.

In contrast however, is the theory of 'epitope spreading' which is defined as the 'enhancement and diversification' of the endogenous T-cell response against antigens that are distinct from the initially targeted antigen, and were previously inaccessible to the immune system (Brossart, 2020). This can either be from the same protein (intramolecular spreading) (Inderberg-Suso *et al.*, 2012), or a different source protein (intermolecular spreading). In cancer immunotherapy, it has been shown that immune-induced tumour cell lysis can release novel antigens for the engulfment by APCs, continuing on the cancer immunity cycle. Epitope spreading has been seen across multiple types of clinical intervention, such as from treatment with a single-target chimeric antigen receptor (CAR) T-cells (Heckler and Dougan, 2018); immunization with peptide vaccines (Disis *et al.*, 2004; Ott *et al.*, 2020); chemotherapy and mAb combination therapies (Knutson *et al.*, 2016). Studies have shown that development of polyclonal and polyfunctional immune responses through epitope spreading can profoundly enhance the clinical success of a treatment (Ribas *et al.*, 2003; Corbière *et al.*, 2011). Furthermore, although HLA-II antigens are less therapeutically targeted, it has been shown that HLA-II antigens can promote anti-tumour immune responses (Kreiter *et al.*, 2015). HLA-II is not often expressed highly in the absence of IFNγ, therefore, immunotherapy-induced IFNγ signalling in the tumour microenvironment (TME) may promote HLA-II expression, revealing new neoantigens to the immune system.

1.4 Peptide binding prediction

To aid in the prediction of HLA restriction of peptides, peptide prediction tools have been developed and published online. In the early development, algorithms were informed by binding affinity data of known peptide and HLA sequences (Reche, Glutting and Reinherz, 2002; Govindarajan *et al*., 2003; Nielsen *et al*., 2007). Peptide binding affinity data can be collected in a number of ways, including competition assays with radiolabelled peptides can be used to determine the relative peptide affinities (IC50) (Sidney *et al.*, 2013; Gfeller and Bassani-Sternberg, 2018). Newer peptide prediction algorithms are now trained on a combination of binding affinity data and mass spectrometry (MS) data of HLA-eluted peptides (Gfeller and Bassani-Sternberg, 2018; Reynisson et al., 2020; Sarkizova *et al*., 2020). Eluted peptide data from cells expressing only a single HLA allele has enabled resolution of HLA-I binding motifs, used to advance the development of prediction algorithms (Abelin *et al*., 2017). One possible flaw with using eluted ligands as a training dataset is the bias towards high affinity peptides; this may cause a loss of resolution of predictions for lower affinity, but functionally relevant, HLA-binding peptides. Multiple tools have been designed, but the tool now used as the gold standard is NetMHCpan; the most recent version is NetMHCpan4.1b (Reynisson *et al*., 2020).

The development of *in silico* peptide binding prediction tools has enabled the prediction of neoantigen binding to HLA. Neoantigens are identified first by classifying the non-synonymous mutations from next generation sequencing data of tumour and non-malignant material (often PBMCs). This is followed by HLA-binding prediction of mutation-containing sequences. The vaccine-stimulated immune cells are then directed towards the neoantigen-presenting tumour cells for killing. Studies have shown neoantigen vaccines are able to generate neoantigen-specific T-cells and prevent re-occurrence in melanoma (Ott *et al.*, 2017), and induce tumour-specific CD4+ and CD8+ T-cell responses in glioblastoma (Keskin *et al.*, 2019). Therefore, accuracy of peptide binding predictions is important, as it supports analysis of neoantigen-generating mutations (Turajlic *et al.*, 2017), and facilitates the inclusion of true strong-binding HLA-presented peptides in the peptide vaccine.

Some research has even gone further in efforts to correctly identify strong-binding peptides, by filtering NetMHCpan4.0-generated candidate peptides through a 3D structure-based predictor (Aranha *et al.*, 2020). The publication described an increase in the positive predictive value of classification of peptides as strong binders (k_D <100nm) from 40 to 52%. Efforts have also gone towards genomically mapping MS-detected antigens from 26 different cancer datasets to try improve predictive value of neoantigens (Bedran *et al.*, 2022). This found that patient mutations in regions which overlapped with MS-detected peptides were a better prediction metric for immunotherapy response than neoantigen prediction alone. This also allowed for identification of pan-cancer mutations which generated MS-detectable neoantigens, which could provide data for the formulation of cancer vaccines which are less individualised.

1.5 The discipline of immunopeptidomics

Immunopeptidomics is the direct identification and quantification of eluted HLA-binding peptides by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Immunopeptidomics therefore identifies the immunopeptidome, which is compartment of peptides presented on HLA-I. The initial development of these research approaches (Hunt *et al.*, 1992; Caron *et al.*, 2011; Bassani-Sternberg *et al.*, 2015; Bassani-Sternberg and Coukos, 2016; Abelin *et al.*, 2017; Purcell, Ramarathinam and Ternette, 2019), has allowed for their use to interrogate biological questions (Liepe *et al.*, 2016; Pearson *et al.*, 2016; Chong *et al.*, 2017; Komov *et al.*, 2018, 2021), and discover targetable neoantigens for the development of customised patient mRNA or peptide vaccines, or TCR-based therapies (Ott *et al.*, 2017; Sahin *et al.*, 2017; Hilf *et al.*, 2018)

Immunopeptidomics can be performed to interrogate peptide presentation on patient resection material, patient-derived organoids (PDOs), or cell lines. Whilst groups may take slightly different immunopeptidomics approaches, the majority of established methodologies are centred around the immunoaffinity purification (IP) of HLA complexes from detergent-solubilised material. For patient material, endogenous antibodies are first depleted from the lysate. Following this, samples are subject to sequential IP of HLA-I followed by HLA-II using anti-pan-HLA-I and anti-pan-HLA-II antibodies. Peptides are then released from the peptide HLA and

prepared for LC-MS/MS. Peptides are separated by liquid chromatography, then directly injected in to the mass spectrometer for fragmentation and generation of spectra which can be interrogated. Peptides are identified by the comparison of the spectra with the *in silico*-generated spectra of peptides from protein databases with MS search tools such as MaxQuant and Proteome Discoverer (Gfeller and Bassani-Sternberg, 2018). For human peptide identification, the reference database for comparison is the updated annotated human proteome, and for neoantigen identification custom reference databases including patient mutation sequences are used.

Whilst peptide identification by MS has helped to improve the accuracy of peptide binding prediction, it itself is subject to certain MS biases. For example, highly hydrophobic or hydrophilic peptides may be lost during the liquid chromatography stage, as the solid phase of the column retains peptides based on their hydrophobicity. Furthermore, highly hydrophobic or hydrophilic peptides also do not ionise well, precluding their identification. In addition, MS is biased against the identification of cysteine, as it is readily oxidised, and because oxidised cysteine is often not included in reference databases (Gfeller and Bassani-Sternberg, 2018). This impedes the identification of cysteine-containing peptides.

With the knowledge of these biases, constant adjustments are being made to MS protocols to try and detect the most representative repertoire of peptides (Chong, Coukos and Bassani-Sternberg, 2022). For example, some protocols have included modified cysteines within their search databases (Abelin *et al.*, 2017). Recent studies have also incorporated double reduction steps in their peptide preparation protocols to stabilise the ‘reactive’ amino acids in peptides, to aid their identification by MS (Wang *et al.*, 2019). Studies have also examined the impact of different detergents and peptide purification strategies on HLA peptide enrichment (Nicastri *et al.*, 2020). Their work found the detergent CHAPs (3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate) enabled the isolation of the highest number of peptides, compared against Igepal, Triton X-100, and sodium deoxycholate (SDC).

Furthermore, tandem mass tag (TMT) MS has been used to enhance the sensitivity of MS, allowing the identification of a larger number of peptides, and

lower abundance peptides (Pfammatter *et al.*, 2020). In this study, from a preparation of $2x10^6$ B-lymphoblastoid cells, an average of 4,984 peptides per identified by TMT-MS, compared against 3,235 peptides identified by the standard approach. A recent study using TMT-MS was able to resolve 688 HLA peptides from approximately 1,000 cultured cells, or 1,257 HLA peptides from approximately 1mg of tumour tissue (Ramarathinam *et al.*, 2020).

Immunopeptidomics is therefore a fast-moving field with constant improvements being made to progress the sensitivity of peptide detection and identification, to most accurately represent the true HLA peptidome. As this work continues, the immunopeptidomics discipline is likely to be applied more and more in cancer research to help answer a diverse range of biological questions.

1.6 Colorectal cancer

1.6.1 CRC statistics and classification

Colorectal cancer (CRC) is the most common gastrointestinal (GI) cancer, and is the third most common cause of cancer-related mortality (Siegel *et al.*, 2021), despite incidence dropping between 2000-2018 due to the tripling of colonoscopy uptake (Siegel *et al.*, 2020). Furthermore, there is increasing incidence in younger adults, dropping the median ages of diagnosis from 72 years old in 2002 to 66 years old in 2010 (Siegel *et al.*, 2020).

Although advances in treatments have been made in localised and advance-stage CRCs through new chemotherapeutic agents and development of targeted therapies (Siegel *et al.*, 2020), median overall survival of metastatic CRC remains at 30 months (Cremolini *et al.*, 2015; Tsuji *et al.*, 2021).

1.6.2 CRC development

CRC develops through repeated accumulation of mutations over a period of years (Fearon and Vogelstein, 1990; Vogelstein *et al.*, 2013). The initial “gatekeeping” mutation provides a selective growth advantage to a normal epithelial cell, allowing it to outgrow surround cells. In CRC the most common gatekeeping mutation affects the APC gene. This results in the formation of a

small slow-growing adenoma. If a cell then acquires driver mutations, such as KRAS, then the growth rate accelerates and a large adenoma forms. Clonal expansion continues to occur, and often additional driver mutations in PIK3CA, SMAD4, and TP53 are acquired, allowing the generation of a malignant tumour which can invade tissue and metastasize to lymph nodes and organs. In CRC the most common site of metastasis is the liver.

1.6.3 CRC classification

Over 85% of early stage CRCs are mismatch repair proficient (MMRp), and approximately 95% of the metastatic CRCs are MMRp. Around 15% of early stage CRCs are mismatch repair deficient (MMRd). MMRd tumours are so-called as they have inactivation of the genes in DNA MMR pathway (such as MLH1, MSH6, MSH2 (Kim *et al.*, 2019)) due to promoter hypermethylation or inactivating mutations. This inability to accurately repair DNA mismatches generates a hypermutator phenotype, where cells rapidly accumulate mutations. MMRp CRCs have a lower mutation burden than the MMRd CRCs, at under 10 mutations per megabase of DNA (Kim *et al.*, 2019).

1.6.4 Standard of care for CRC

The current standard chemotherapy regimen for first-line treatment for metastatic CRC is 5-fluorouracil (5FU) combined with either oxaliplatin (FOLFOX) or irinotecan (FOLFIRI) (Xie, Chen and Fang, 2020). This chemotherapy regimen is also commonly combined with targeted monoclonal antibody (mAb) treatments, such as bevacizumab, a mAb against vascular endothelial growth receptor, or cetuximab and panitumumab, mAbs against epidermal growth factor receptor. A combination of chemotherapy with any of the targeted agents is approved for use as first-line treatment (Holch, Stintzing and Heinemann, 2016; Cervantes *et al.*, 2022).

1.6.5 Immunotherapy in CRC

Early work in immunotherapy of MMRp CRC had shown immunotherapy to be ineffective. For example, in MMRp CRC patients single-agent treatment with anti-

PD1 mAb pembrolizumab saw a 0% objective response rate (ORR), but a 40% ORR in MMRd patients (Le *et al.*, 2015).

However, more recent studies have showed some promise in combination immunotherapies. For example, the combined treatment of anti-PD-1 antibody balstilimab, with novel regulatory T-cell-depleting anti-CTLA-4 antibody botensilimab in heavily pre-treated MMRp CRCs showed an ORR of 24%, up to 42% in those without liver metastases (Bullock *et al.*, 2022; Lote *et al.*, 2022).

The final clinical analysis NICHE study recently showed promising results in MMRd and MMRp CRC. The trial investigated efficacy of neoadjuvant immunotherapy with the combination treatment of nivolumab, ipilimumab, and celecoxib, (Verschoor *et al.*, 2022). Pathological response was seen in 100% of MMRd patients, but also 30% of MMRp patients. None of the responders had disease recurrence within the median 25-month follow up period.

Clinical studies have also trialled combination therapy of anti-PD-L1 mAb pembrolizumab with novel anti-LAG3 mAb favezelimab (Garralda *et al.*, 2021). This treatment was trialled in MMRp metastatic CRC patients who had failed a least 2 lines of treatment. The median objective response rate was 6.3%, and patients with higher PD-L1 expression at screening saw greater benefit.

This illustrates the potential of immunotherapies in MMRp CRCs, and highlights the severe need to further understand their potential for immunogenicity.

1.7 Gastro-Oesophageal Adenocarcinoma

1.7.1 GOA statistics and classification

Gastro-Oesophageal Adenocarcinoma (GOA) is the 4th commonest cancer type worldwide. It is a remarkably aggressive disease that is one of the biggest causes of cancer mortality worldwide (Ferro *et al.*, 2014; Ferlay *et al.*, 2015). GOA has a significant prevalence in Asia, encompassing nearly 75% of all gastric cancer cases (Carcas, 2014). GOAs can be subclassified based on their anatomic location in to: gastric adenocarcinomas (GAs), gastro-oesophageal junction (GEJ cancers), or oesophageal cancers. Depending on the country, approximately 5.6-

33.3% of GOA cancers are of the MMRd subtype (Polom *et al.*, 2018), which is caused by the genetic activation of, MSH2, MSH6, and PMS2, and MSH1 (which can also be inactivated by methylation) (von Loga *et al.*, 2020). The genetic instability of MMRd GOAs allows them to exhibit extreme intratumour heterogeneity. MMRd GOA patients also have better survival rates than MMRp GOAs (Polom *et al.*, 2018), hypothesised to be due to their large number of mutations from which neoantigens can be sourced. This illustrates the clinical challenge which is associated with treating MMRp GOAs, which represent the larger proportion of GOAs. New testing platforms and treatment modalities need to be evaluated to rapidly advance the clinical benefit of immunotherapy to MMRp GOAs.

1.7.2 Treatment options for GOAs

Surgery is currently the only curative treatment choice in GOA. Surgical choice for GOA is defined by the stage of disease. For stage II and below, there is a focus on surgical resection, with total or subtotal gastrectomy depending on the location of the tumour and the depth of invasion (Sexton *et al.*, 2020). It has also been shown in stage II and III patients that neoadjuvant or adjuvant perioperative chemotherapy can provide an additional survival benefit (Sasako *et al.*, 2011). For those with locally advanced disease, it is recommended to give perioperative chemotherapy or preoperative chemoradiotherapy. Currently, FLOT (5-FU, leucovorin, oxaliplatin, and docetaxel) is the standard of perioperative chemotherapy. In the metastatic setting, first line treatment includes a combination therapy incorporating 5-FU and platinum-based agents such as FOLFOX or CAPOX. HER2 overexpression is found in approximately 4-54% of gastric cancers; in those overexpressing HER2, HER2-targeting mAb treatment Trastuzumab can also be added to the treatment regime.

1.8 Utilising organoids in cancer research

Only a small number of typically more aggressive cancer cell clones are able to be grown as cancer cell lines in the 2D setting (Yu *et al.*, 2019). This adds a bias to any research performed on cell lines which may limit the translational relevance of any findings.

Organoids are tissue-derived cells which are grown embedded in to a 3D matrix and supplemented with media rich in factors which recapitulate the in vivo stem cell niche (e.g. R-spondin, WNT, etc) (Drost and Clevers, 2018). The 3D matrix is commonly composed of laminin, collagen IV, heparin sulphate proteoglycans, amongst other factors, to help mimic the extracellular membrane (Hughes, Postovit and Lajoie, 2010). These growth conditions support the growth of organoids into organised organotypic structures, much like the epithelium from which they were derived. One of the first cancer organoid cultures was developed from CRCs, following the development of protocols of culturing healthy GI organoids (Sato *et al.*, 2009, 2011). Cancer organoids also exhibit the same structures as their source tissue, and due to the genetic dysregulation of the cancer, these structures are less organised than the healthy organoids from the same organ.

In contrast to cell lines, 3D organoid culture enhances the survival and long-term maintenance of a larger number of unique cancer cells and supports the maintenance of a cancer phenotype such as tissue organisation, morphology, and genetic stability. In addition, the survival of more tumour cells in culture facilitates the maintenance of the intratumour heterogeneity which allows for more accurate recapitulation of the tumour genetics, chemosensitivity, and susceptibility to immune cell attack. In a comparative genetics study of 14 CRC patients, it was shown that 90% of mutations were shared between the tumour tissue and tumour organoids, and their copy number profiles were highly correlated (0.9) This further illustrates how tumour organoids can better represent the tumour tissue (Weeber *et al.*, 2015). These beneficial features have made organoids a highly desirable model for experimentation, which has resulted in the development of many organoid biobanks (Van De Wetering *et al.*, 2015; Vlachogiannis *et al.*, 2018; Jacob *et al.*, 2020).

1.9 Outline of PhD aims

The goal of my PhD thesis is to advance the understanding of the immunopeptidome and potential immunogenicity of MMRp gastrointestinal cancers. The first aim is to define the immunopeptidome landscape in MMRp

colorectal cancer (CRC) through the use of transcriptomics, proteomics, and peptidomics on MMRp CRC patient-derived organoids (PDOs). A second aim is to develop an autologous PBMC-PDO co-culture platform to help understand immunogenicity in MMRp gastro-oesophageal adenocarcinomas (GOA). The first results chapter (Chapter 3) outlines the analysis of HLA-peptides on 5 MMRp CRC, including the analysis of neoantigen loads. Chapter 3 further scrutinises the effect of IFNγ treatment and MEK inhibition on canonical peptide and neoantigen presentation, and critiques the viability of histone deacetylase inhibition for an immunopeptidomics analysis. Chapter 4 analyses the impact of IFNγ treatment on remodelling of the immunopeptidome, focusing on the peptide-specific features that facilitate the IFNγ-influenced upregulation of peptides. Chapter 5 focuses on the development of a novel MS peptidomics-compatible ER isolation protocol for the development of the understanding of the bottleneck in antigen presentation. Finally, Chapter 6 illustrates the use of allogeneic T-cell-PDO co-cultures for the assessment of immunotherapy resistance in MMRp CRC, and autologous PBMC-PDO co-cultures for the interrogation of tumour immunogenicity in MMRp GOAs. Chapter 6 further demonstrates the possible potential of the autologous PBMC-PDO co-culture platform for the individualised assessment of potential immunotherapy combinations for MMRp GOAs.

Chapter 2 Materials and methods

2.1 Materials

2.1.1 Drugs

Drug	Manufacturer	Solvent	Product code
Doxycycline hyclate	Sigma	Ultrafiltered water	D9891
Interferon gamma (IFNγ)	R&D Systems	Ultrafiltered water	285-IF/CF
Puromycin dihydrochloride	Sigma	Ultrafiltered water	P8833
Trametinib	Cambridge Bioscience	DMSO	
Domatinostat	Selleckchem	DMSO	S7555
Atezolizumab	Selleckchem	PBS	A2004
Nivolumab	Selleckchem	PBS	A2002
BMS-986205	Selleckchem	DMSO	S8629

Table 2.1 Drugs.

2.1.2 Reagents

Reagent	Manufacturer	Product code
DMSO	Sigma	
Sodium deoxycholate (SDC)	Sigma	30970
DTT	Sigma	
BSA fraction V	Sigma	05482
DMEM/F-12	Thermofisher	11320033
ECL prime	Cytiva	GERPN2236
0.5M EDTA	In-house	05482
EDTA-free protease inhibitor	Roche	05892791001
Foetal bovine serum (FBS)	Thermo Fisher Scientific	10270106

Full-range rainbow marker	GE healthcare	GERPN800E
Glutamax	Thermo Fisher Scientific	35050061
Histodenz	Sigma	D2158
Igepal CA-630	Sigma	I8896
100% reduced growth factor Matrigel	Corning	356231
McCoy's 5A medium	Thermo Fisher Scientific	36600021
Methanol	VWR	20847.307
MOPS running buffer	Invitrogen	A53226
5M NaCl	In-house	
NuPAGE LDS sample buffer	Invitrogen	NP0007
Phenylmethylsulphonyl fluoride (PMSF)	Sigma	10837091001
Pierce Protein A/G magnetic beads	Thermo Fisher Scientific	88802
Phosphatase inhibitor Cocktail II	Sigma	P5726
Proteinase K from Tritirachium album	Sigma	P6556
PVDF membrane	Thermo Fisher Scientific	88518
RPMI 1640 Medium	Thermo Fisher Scientific	31870025
Sodium dodecyl sulphate (SDS)	Sigma	
Sodium azide	Sigma	
Sucrose	Sigma	S0389
Transfer buffer	In-house	
1M Tris pH 7.6	In-house	
Trypan blue	Invitrogen	T10282

TrypLE Express	Thermo Fisher Scientific	12605010
Tween 20	Sigma	P1379
Ultra-filtered (UF) water	In-house	WHA3030917
Whatman cellulose chromatography papers	Sigma	WHA3030917

Table 2.2 Reagents.

2.1.3 Buffer recipes

Buffer	Recipe
HEPES buffer	20mM HEPES, 1.2 mM $MgCl_2$, 1mM EDTA, 1µM PMSF, pH 7.4.
Hypotonic extraction buffer	10mM HEPES (pH 7.8), 25mM potassium chloride, and 1mM EDTA, 1µM PMSF pH 7.4.
Isotonic sucrose buffer	10mM HEPES, 250mM sucrose, 25mM potassium chloride, 1mM EDTA, 1µM PMSF, pH 7.4.
Sucrose solutions	Sucrose made up to 0.3M, 1.3M, 1.5M, and 2.0M with HEPES buffer as the diluent.
Histodenz solutions	Histodenz with HEPES buffer as the diluent.
NP40	250mM NaCl, 50mM Tris, 5mM EDTA, 1% Igepal CA-630, 1X One COMPLETE ULTRA EDTA-free Protease inhibitor tablet, 1X phosphatase inhibitor cocktail II, and 10mM NaF

RIPA buffer	150mM NaCl, 50mM Tris, 0.5% DOC, 0.1% SDS, 1% Igepal CA-630, 1X One COMPLETE ULTRA EDTA-free Protease inhibitor tablet, 1X phosphatase inhibitor cocktail II, and 10mM NaF.
Western blot sample buffer	4X NuPAGE LDS sample buffer with DTT.
Flow cytometry buffer	0.5% BSA, 5mM EDTA, PBS.

Table 2.3 Buffer recipes.

2.1.4 Commercial assays and kits

Reagent	Manufacturer	Product code
ER Enrichment Kit	Invent Biotechnologies	ER-036
Pierce Coomassie Plus protein assay	Thermo Fisher Scientific	23238
Pierce Rapid Gold BCA protein assay	Thermo Fisher Scientific	A53226
CellTiter blue	Promega	G8080
HDAC activity kit	BioVision	K331-100
QIFIKIT	Agilent	K0078

Table 2.4 Commercial assays and kits.

2.1.5 Western blot antibodies

Antibody	Manufacturer	Product code	Dilution
ABCB1	Abcam	ab129450	1:5000
ATP1A1	Abcam	ab76020	1:20,000
Anti-mouse HRP	Cytiva	NA931VS	1:3000
Anti-rabbit HRP	Cytiva	NA934VS	1:3000
β-tubulin HRP	Abcam	ab21058	1:5000
Calnexin	Cell Signalling Technologies	2679T	1:1000

Calreticulin	Cell Signalling Technologies	28918	1:800
Cas9	Cell Signalling Technologies	14697	1:1000
COPA	Atlas Antibodies	HPA028024	4μg/mL
Cytochrome C	Abcam	ab76107	1:500
ERK	Cell Signalling Technologies	9102	1:1000
p-ERK	Cell Signalling Technologies	9101	1:1000
GAPDH HRP	Abcam	ab9482	1:5000
GM130	Cell Signalling Technologies	12480T	1:1000
GP78/BiP	Abcam	ab108613	1:375
HLA-I-HC	Abcam	ab70328	1:1000
Histone H3	Novus Biologicals	NB500-171	1:1000
LAMP1	Cell Signalling Technologies	9091T	1:1000
SEC16A	Atlas Antibodies	HPA005684	4μg/mL
SEC61B	Cell Signalling Technologies	14648	1:1000
TAP1	Cell Signalling Technologies	12341	1:600

Table 2.5 Western blot antibodies.

2.1.6 Flow cytometry antibodies

Antibody	Manufacturer	Product code	RRID	Dilution
Anti-HLA-A,B,C	Biolegend	311402	AB_314871	1:125
Anti-HLA-DR,DP,DQ	Biolegend	361702	AB_2563139	1:200
IgG2a, κ isotype control	Biolegend	400202	-	1:125/ 1:200
AF700 anti-HLA-A,B,C	Biolegend	311438	AB_2566306	1:125

BV421 anti-CD274	Biolegend	329714	AB_2563852	1:20
Anti-CEA (CH1A1A)	Roche	-	-	1:20
PE F(ab')2 anti-human IgG	Jackson Immuno-Research	109-116-097	AB_2337677	1:250
Zombie near-infrared	Biolegend	423106	-	1:250
Zombie red	Biolegend	423110	-	1:250
BV785 anti-CD3	Biolegend	300472	AB_2687178	1:20
PerCP anti-CD4	Biolegend	300528	AB_893321	1:20
APC-Cy7 anti-CD4	Biolegend	300518	AB_314086	1:20
AF488 anti-CD8	Biolegend	344716	AB_10549301	1:20
PE anti-CD107a	BD Biosciences	555801	AB_396135	1:120
AF647 anti-CD107a	Biolegend	328612	AB_396135	1:120
APC anti-CD137	Biolegend	309810	AB_830672	1:20
BV421 anti-Granzyme B	Biolegend	396414	AB_2810603	1:20
PE anti-Granzyme B	Biolegend	396414	AB_2810603	1:20
BV650 anti-IFN-γ	Biolegend	502538	AB_2563608	1:30

Table 2.6 Flow cytometry antibodies.

2.2 Cell culture

2.2.1 Adherent cell line culture

Cell lines were cultured under the conditions listed in Table 2.7.

Cell line	Source	Culture conditions
NCI-N87	ATCC	RPMI 1640, 10% foetal bovine serum (FBS), 1% GlutaMAX, 1% penicillin/streptomycin in 5% CO2, at 37°C.
HCT116	Gift from J. Choudhary	In-house DMEM (supplemented with glutamine and penicillin/streptomycin), 10% FBS in 5% CO2, at 37°C.

Table 2.7. Details of cell line culture.

To split cells, media was aspirated, cells were washed with PBS, and then incubated with 1X Trypsin or TrypLE express until cells detached (3-5 minutes). Media was then added to neutralise the dissociation agent, then cells were spun down at 300g for 5 minutes. For re-seeding cells would then be resuspended, counted using the Invitrogen Countess, and diluted in to the appropriate volume of media for the selected culture vessel. For long-term storage, the cell pellet would be resuspended in freezing media (FBS with 10% DMSO), 1mL aliquoted in to freezing vials, and frozen at -80°C overnight in Mr Frosty slow freezing containers. The following day, cells were then transferred to liquid nitrogen for long-term storage.

2.2.2 Establishing patient-derived organoids (PDOs) from needle core biopsies

Samples were established from 2-4mm needle core biopsies derived from the primary tumour site or liver metastases. First, the tissue was washed by sequential submersion through 3 wells of 3-5mL Tumour Tissue Wash Buffer (Phenol free RPMI 1640 + 4% Penicillin/streptomycin+ 1% L-Glutamine + 50 µg/ml Gentamicin), before manual scalpel chopping whilst covered in 0.5mL of tumour dissociation buffer (TDB) (Phenol free RPMI 1640 + 1%

Penicillin/streptomycin+ 1% L-Glutamine + 50 µg/ml Gentamicin) to prevent the sample from drying out.

The chopped sample and covering TDB were added to a Miltenyi Biotec C-tubes containing 4.5mL of TDB supplemented with ROCK inhibitor and pre-prepared enzymes from the Miltenyi Biotec human tumour dissociation kit. The dissociation C-tube was then attached to the gentleMACS dissociator, and the gentle dissociation program 37C_BTDK_1 (22 minute 48 seconds) was run. At the end of the run, the mixture was transferred in to a fresh tube and the dissociation tube rinsed out with another 5mL of basis media with 10% FBS (Advanced DMEM/F12, 1x GlutaMAX, 100µM HEPES, 1% penicillin/streptomycin). The tube was then spun down at 300g for 5 minutes. If there was significant red blood cell contamination the pellet was resuspended in 2mL of ACK lysis buffer and incubated for 5 minutes at room temperature, followed by dilution with 5mL of basis media and centrifugation for 10 minutes at 300g. The supernatant was then poured off, and the tube spun again at 300g for 1 minute, and any residual supernatant was removed. The pellet was then resuspended in an appropriate volume of pre-thawed matrigel (routinely 60-80µL per biopsy), and 3-4 15-20µL domes were carefully placed in to the centre of a pre-warmed 24-well plate. The 24-well plate was then returned to the 37°C, 5% CO_2 incubator, inverted, to set for 20 minutes. Finally, the plate was righted and 0.5-1mL of the appropriate 3D media supplemented with 10µM ROCK inhibitor was added gently, and the plate returned to the incubator. Plates were checked visually daily, and photographed regularly on the Zeiss Observer microscope.

2.2.3 Thawing early-passage PDO lines

Vials were thawed rapidly in a 37°C water bath, leaving a small amount of ice crystals remaining. The thawed suspension was transferred to a 15mL Falcon tube and 10mL of cold 1% BSA in DMEM:F12 was added dropwise under gentle shaking. The cells were then spun down at 300g and 4°C for 5 minutes, the supernatant removed, then the tube was further briefly spun at 300g for 1 minute and any remaining supernatant removed. The cell pellet was then resuspended in an appropriate volume of matrigel, left to set, inverted, in a 37°C 5% CO_2

incubator for 20 minutes, before the righting and addition of 3D gastric/colorectal media supplemented with 10μM ROCK inhibitor.

2.2.4 3D culture of PDO lines

3D PDO cultures were sustained in a 100% matrigel dome, and covered in 0.5-1mL of the appropriate 3D media. Early cultures were managed with 3-4 15-20μL matrigel dots in a single well of a 24-well plate, and as the culture grew, it would be expanded to multiple wells, then into a 12-well, or 6-well plate according to the replication rate of the cells. As cultures grew in a heterogeneous way, cultures were split before the largest organoids became denser and darker under the microscope, but after the smallest organoids had a chance to grow in size. Split ratios were 1/2-1/4 depending on the growth rate of the PDO.

Cancer type	Culture conditions
Gastric	Advanced DMEM/F12, 1x GlutaMAX, 100μM HEPES, 1% penicillin/streptomycin, 1x B27, 1x N-2, 10mM Nicotinamide, 1mM NAC, 100ng/mL EGF, 10μM SB202190, 500nM A-83-01, 1μM PGE2, 10nM Gastrin-1, 1μg/mL R-Spondin, 100ng/mL Noggin, 100ng/mL FGF10, 100ng/mL Wnt3a, 10ng/mL FGF2, 50ng/mL HGF, 1μg/mL IGF1, 100μg/mL Primocin. Supplemented with 10μM Y27632 at the point of seeding/splitting only, and cultured at 5% CO2, and 37°C.
Colorectal	Advanced DMEM/F12, 1x GlutaMAX, 100μM HEPES, 1% penicillin/streptomycin, 1x B27, 1x N-2, 10mM Nicotinamide, 1mM NAC, 50ng/mL EGF, 10μM SB202190, 500nM A-83-01, 10nM PGE2, 10nM Gastrin-1, 1μg/mL R-Spondin, 100ng/mL Noggin, 100ng/mL FGF10, 100ng/mL Wnt3a, 10ng/mL FGF2, 100μg/mL Primocin. Supplemented with 10μM Y27632 at the point of seeding/splitting only, and cultured at 5% CO2, and 37°C.

Table 2.8. Details of PDO culture.

To passage the cells, the media was discarded, and the matrigel domes would be gently scratched off the plate surface with a pipette tip and be resuspended an appropriate volume of PBS. The matrigel-PBS mixture would then be transferred to a 15mL falcon spun down, and the supernatant removed. To remove the gel and part-dissociate the organoids, the pellet would be resuspended in pre-warmed TrypLE express an incubated at 37°C for an amount of time specific to each donor. This was due to different sensitivities to TrypLE express (often 5-15 minutes). Once the cells were suitably dissociated upon visual examination, the cells were diluted in 5mL of basis media with 10% FBS and spun down at 300g for 5 minutes. The supernatant was removed, the tube spun again for 1 minute, then any remaining supernatant was removed. The cell pellet was then resuspended in an appropriate volume of matrigel, left to set, inverted, in a 37°C 5% CO_2 incubator for 20 minutes, before the turning the plate back and addition of 3D gastric/colorectal media supplemented with 10μM ROCK inhibitor.

2.2.5 2D culture of PDO lines

Cells were maintained in 2D advanced media supplemented with 2% matrigel (Advanced DMEM/F12, 1x Glutamax, 1% penicillin/streptomycin, 10% FBS), at 37°C and 5% CO2.

To split cells, media was aspirated, cells were washed with PBS, and then incubated with 1X TrypLE express until cells detached (5-15 minutes). Media was then added to neutralise the TrypLE express, then cells were spun down at 300g for 5 minutes. For re-seeding, cells would then be resuspended and counted using the Invitrogen Countess. The correct volume of cells would then be spun down, resuspended in media supplemented with freshly thawed 2% matrigel, and added to the selected culture vessel. For long-term storage, the cell pellet would be resuspended in freezing media (FBS with 10% DMSO), 1mL aliquoted in to freezing vials, and frozen at -80°C overnight in Mr Frosty slow freezing containers. The following day, cells were then transferred to liquid nitrogen for long-term storage.

2.2.6 Isolation of peripherical blood mononuclear cells (PBMCs) from leukocyte reduction cone system (LRSs) and packed blood cell pellets (PBCPs)

The PBCPs were generated at the hospital by the tissue collection team spinning the blood samples for 10 minutes at 1,600g and collecting the plasma, leaving behind the PBCP. For PBMC isolation, PBCPs from the vials would then be combined and diluted 1:1 in PBS, often 1 gradient would be required for 3-5 EDTA tubes.

LRSs were purchased from the blood bank, and were generated by filtration of blood products through the cone filters. The concentrated leukocytes were released from the LRS by snipping the sealed tubes and allowing it to drip out in to a 50mL Falcon, yielding 8-10mL of blood which is very thick and dense (as LRSs contain cells from 1L of whole blood). 1 LRS was diluted to a final volume 70mL of PBS. For each LRS, 2 Ficoll gradients were prepared.

For each Ficoll gradient, 15mL of Ficoll reagent was gently added to the base of the 50mL Falcon tube, and 35mL of the diluted LRS blood was layered on top by pipetting slowly at a shallow angle. Each gradient was then centrifuged at 800g for 25 minutes with no brake. After centrifugation, there was a plasma layer, a dense PBMC layer, and a red blood cell layer. Excess plasma was taken off, and then the PBMC layer was aspirated in to a 10/25mL stripette tip, before transferring in to a fresh tube. The PBMCs yielded were diluted to a total volume of 50mL and spun at 120g for 10 minutes with no acceleration or break. After the spin, the platelet-rich supernatant was discarded, and the PBMC pellet was resuspended in 8mL of ACK lysis buffer for removal of red blood cells, and incubated at room temperature for 5 minutes. After the incubation, the suspension was diluted by addition of PBS to a final volume of 50mL, then spun down at 300g for 8 minutes, with break and acceleration set to 5. As many PBS-washes of the PBMC pellet would continue until the supernatant was visually clear of platelet contamination. Cells were then diluted to an appropriate volume for counting, and counted on the Invitrogen countess. The cell suspension was then spun down, supernatant removed, and the pellet resuspended in freezing media (FBS + 10% DMSO) at $5x10^6$-$5x10^7$ cells per 1mL in a freezing vial. Cells were frozen at -80°C overnight, before transferring to liquid nitrogen the next day for long-term storage.

2.2.7 Isolation of CD8+ T-cells from PBMCs

PBMCs were thawed quickly in a 37°C water bath, then 10mL of pre-warmed T-cell media was added to the cells dropwise, before spinning down at 300g for 5 minutes. CD8+ T-cells were isolated according to the instructions in the FlowComp Human CD8 isolation kit. In summary: Up to $5x10^7$ PBMCs were diluted in to 500μL of isolation buffer, and 25μL of the kit CD8 antibody was added and mixed well. After a 10-minute 4°C incubation, cells were washed by addition of 2mL of isolation buffer and spun down. During the spin, the Dynabeads were washed and resuspended in isolation buffer. After the spin, the supernatant was removed and the PBMC pellet was resuspended in 75μL of Dynabead solution before a 15-minute incubation of rolling and tilting at room temperature. Next, the now bead-bound CD8+ T-cells were retained in the tube by the DynaMagnet as the non-CD8+ T-cells were washed off. Once the cells had been washed twice,

the bead-bound CD8+ T-cells were incubated with bead release buffer under rolling and tilting at room temperature for 10 minutes. The suspension was then incubated on the DynaMagnet which sequestered the beads, leaving the CD8+ T-cells in supernatant. The supernatant was then washed in 2mL of isolation buffer, spun down, and the CD8+ T-cell pellet was resuspended in to 1mL of T-cell media supplemented with IL-2 (for a final concentration of 30IU/mL). The T-cell suspension was then added to a 12-well plate and incubated at 37°C and 5% CO_2.

2.2.8 Activation of CD8+ T-cells

CD8+ T-cells were counted on the Invitrogen Countess, then they were resuspended at a density of $1x10^6$ cells/mL. For every $1x10^6$ T-cells cells, 25μL of CD3/CD28 activation Dynabeads were washed and resuspended in T-cell media for combination with T-cells. The cells were then returned to the incubator incubated at 37°C and 5% CO_2. The cells were checked on daily, media changed every 1-3 days with 30IU/mL IL-2 T-cell media, and split 1:2 every time the cells formed a dense lawn on the base of the 12-well plate. Once the cells had been incubated with the CD3/CD28 activation Dynabeads for 10 days, the beads were removed on the DynaMagnet. The bead-free activated cells were either used immediately, or resuspended at $1\text{-}5x10^6$ cells/mL in freezing media (FBS + 10% DMSO) and frozen at -80°C overnight, before transferring to liquid nitrogen the next day for long-term storage.

2.3 DNA analysis

2.3.1 Exome sequencing

Sequencing libraries were prepared by Dr Louise Barber from >=500ng PDO or tumour DNA and matched blood using the Agilent SureSelectXT Human All Exon v5 kit, according to the manufacturer's recommendations. Paired-end sequencing was performed on an Illumina HiSeq2500 with a depth of 100x.

2.3.2 Somatic mutation and copy number aberration analysis

Mutation and copy number analysis were performed by Dr Andrew Woolston and Dr Louise Barber as described in (Woolston *et al.*, 2019). Mutations with a cancer cell fraction >0.7 were considered clonal.

2.4 RNA analysis

2.4.1 RNA sequencing

Two different RNA sequencing experiments were carried out for this thesis: first, untreated PDOs and IFNγ-treated or trametinib-treated PDOs; second, untreated PDOs and domatinostat-treated PDOs.

PDOs were harvested with TrypLE express as normal and pelleted down by centrifugation. RNA extraction was performed in accordance with the Qiagen RNEasy plus kit protocol. RNA was eluted in to 30μL of RNase-free water before quantification with the Qubit RNA broad range assay. 250ng of each sample was then processed for RNA sequencing, following the Lexogen Quantseq 3' mRNA-Seq library for Illumina preparation protocol. In summary 250ng of RNA in 5μL of RNase-free water was used for input, and optimal PCR cycle number was determined by qPCR before PCR amplification of the library. Resultant libraries were quantified using Qubit and Bioanalyzer DNA High sensitivity kits before pooling in an equimolar ratio. The pools were sequenced by the Tumour Profiling Unit at the ICR on an Illumina HiSeq500 in rapid 100 base pair single-end mode with dual indexing.

2.4.2 Analysis of RNAseq data

The Quantseq 2.3.6 FWD pipeline on the BlueBee cloud platform was used to normalise the raw readcount data, and differential expression analysis was performed with the Quantseq DE 1.4.0 pipeline. Gene annotations were exported from Ensembl with BioMart in R. Protein coding genes were selected by filtering for the 'gene_biotype' label.

2.5 Protein analysis by western blot (WB)

2.5.1 Extraction of protein from cells

All cell pellets for WB were lysed with a 30-minute incubation with NP-40 buffer on ice, with vortexing every 5-10 minutes. The lysates were then spun down at 13,000g for 10 minutes at 4°C, and the supernatant transferred to a fresh, pre-labelled and pre-chilled Eppendorf, and either used immediately or stored at -80°C.

2.5.2 Extraction of protein/nuclear proteins from cells/organelle pellets

All the pellets were lysed with a 30-minute incubation with RIPA buffer on ice, with aggressive vortexing every 5-10 minutes. The lysates were then spun down at 13,000g for 10 minutes at 4°C, and the supernatant transferred to a fresh, pre-labelled and pre-chilled Eppendorf, and either used immediately or stored at -80°C.

2.5.3 Protein quantification with Bradford protein assay

The protein concentration of lysates in NP-40 buffer were quantified using the BioRad Quick Start Bradford Protein Assay.

Commercial BSA standards (of concentrations 0.25, 0.5, 0.75, 1, 1.5, 2 mg/mL) were used to generate a standard curve. Where necessary lysates were always diluted with NP-40 buffer. 5μL of the standard, sample, or lysis buffer alone or lysis buffer alone were added to a 96-well plate in triplicate, and 250μL of the Bradford Reagent was added. The plate was then incubated at room temperature for 5 minutes, after which the absorbance was read on the Cytation3 at 595nm. A standard curve was created by plotting the average absorbance values of the standards against their known concentrations in μg/mL and used to determine the protein concentration of the prepared lysates.

2.5.4 Protein quantification with Pierce Rapid Gold BCA protein assay

Protein concentration of lysates in RIPA buffer were quantified using the Pierce Rapid Gold BCA protein assay according to the manufacturer's instructions. The Pierce Rapid Gold BCA protein assay was selected for quantification where RIPA buffer was used as it contains levels of SDS and sodium deoxycholate (SDC) which are incompatible with Bradford Reagent.

The working reagent was generated by adding 50 parts of reagent A to 1 part reagent B, and vortexing. The working reagent was always used within half an hour. Commercial BSA standards (of concentrations 0.25, 0.5, 0.75, 1, 1.5, 2 mg/mL) were used to generate a standard curve. Lysates were always diluted with RIPA buffer. 10μL of the standard, sample, or lysis buffer alone were added to a 96-well plate in triplicate (or duplicate for scarce samples), and 200μL of the working reagent was added, and the plate was mixed on the plate shaker for 30 seconds. The plate was then incubated at room temperature for 5 minutes, after which the absorbance was read on the Cytation3 at 480nm. The average values for the replicates were calculated, and the standard and lysate sample absorbances were blank-corrected by subtracting the absorbance values of the blank (working reagent and lysis buffer) – necessary as RIPA buffer has a high background signal. A standard curve was created by plotting the blank-corrected absorbance values of the standards against their known concentrations in μg/mL and used to determine the protein concentration of the prepared lysates.

2.5.5 Sample preparation and SDS PAGE separation

Samples were prepared by combining 12-30μg of the protein lysates with 1X western blot sample buffer and UF water to reach a final volume. Samples were heated to 70°C for 10 minutes, before being pulsed down in a microcentrifuge and loaded in to the wells of the gel. 6μL of the rainbow full-range molecular marker was added directly to the gel.

Samples were separated on a 4-12% NuPAGE Bis-Tris gel, run in 1X MOPS running buffer (diluted by UF water) or a 3-8% Tris-Acetate gel in 1X Tris-Acetate buffer, to best resolve the required range of molecular weights. Gels were run in the Invitrogen XCell SureLock mini-cell at 120V for 90-120 minutes until the proteins had run to the base of the gel.

2.5.6 Western blotting

PVDF membrane (polyvinylidene fluoridine) was the membrane type selected. Before the transfer, the PVDF was activated with a 2-minute incubation with

100% methanol, and briefly washed in transfer buffer (192mM glycine, 25mM Tris and 200mM methanol in UF water). The gel and membrane were sandwiched between 3mm chromatography filter paper and transfer sponges inside the XCell II Blot transfer cassette. Cold transfer buffer was then added inside the transfer cassette and the tank. The protein from the gel was transferred to the PVDF membrane at room temperature and 30V for 90 minutes, or 30V overnight at 4°C.

The membranes were then incubated in Ponceau stain solution for 60 seconds, before being washed off 2-3 times in TBS-T and imaged. Ponceau stain was used to illustrate the evenness of loading, as when blotting for subcellular compartments there were no possible loading controls.

2.5.7 Protein detection and visualisation

Membranes were first blocked in 5% BSA TBS-T (0.1% Tween-20 in TBS) for 1 hour at room temperature. Primary antibodies were diluted by specified dilution factor (Section 2.1.5) in 5% BSA TBS-T and incubated overnight, rolling at 4°C. Membranes were then washed 3 times for 6 minutes each in a high volume of TBS-T, then incubated in HRP-linked secondary antibody for an hour at room temperature (1:5,000 for anti-rabbit, and 1:3,000 for anti-mouse, diluted in 5% BSA TBS-T). The protein bands were then visualised using ECL Prime reagent. Equal volumes of detection reagent A and B were combined (1mL of each for a full 8X9cm blot) and mixed by vortexing. The reagent was applied to the membrane and left to incubate for 1 minute, before blotting off the excess and placing between two sheets of transparency film. The sandwiched membranes were then imaged on the Azure Biosystems c300 Gel Imaging System. Later in my project I used the ChemiDoc Touch Imaging system.

2.5.8 Blot stripping and re-probing

Because often the yield of certain steps from the ER isolation was quite low, it was required to re-use some blots to probe for additional markers to interrogate the success for the ER enrichment and depletion of other organelles. The developing reagent was washed off for 6 minutes in TBS-T, then the membranes were incubated in ReBlot stripping solution for 15 minutes. Next, the membranes

were washed three times for 6 minutes each in TBS-T, then blocked for 1 hr in 5% BSA TBS-T. Following this, the membrane was incubated overnight in primary antibody, and the rest of the protein detection and visualisation process proceeded as normal.

2.6 Protein analysis by mass spectrometry (MS)

2.6.1 Protein analysis by tandem-mass-tag mass spectrometry (TMT-MS)

Proteomics was performed by our collaborator Professor Jyoti Choudhary. Cell pellets were lysed with SDC lysis buffer (consisting of 1% SDC, 100 mM triethylammonium bicarbonate (TEAB), 10% glycerol, 50mM NaCl) containing Halt Protease and phosphatase inhibitor cocktail. Samples were then completely homogenized with probe sonication: 15 seconds at 40% power with 1 second on / 1 second off, heated at 90^{o}C for 5 minutes and then sonicated again. Proteins were quantified using Quick Start™ Bradford Protein Assay.

100µg protein was taken from each sample and samples were made to the same volume with lysis buffer. Proteins were reduced with 10mM tris(2-carboxyethyl)phosphine hydrochloride solution (TCEP) at room temperature (RT) for 10 minutes, then alkylated with 5mM iodoacetamide (IAA) for 30 minutes at RT. Protein was precipitated with 20% trichloroacetic acid precipitation. The pellet was resuspended in 100mM TEAB buffer, and digested by 3.3µg trypsin (Pierce) at a ratio of 1:30 (trypsin:protein by weight) at 37°C for 18 hours.

40µg of the protein digest was labelled with 0.5mg TMTpro^{TM} 16-plex reagents according to the manufacturer's instruction. The digests were incubated for 1 hour at RT, then quenched for 15 minutes by incubation with by 4µl of 5% hydroxylamine. Following labelling, the samples were combined. SDC was precipitated out by addition of formic acid (FA). After centrifugation, the supernatant was collected and dried by speed-vacuum.

The prepared samples were resuspended in 0.1% ammonium hydroxide, and fractionated on an XBridge BEH C18 column. Further details of the fractionation were described in (Newey *et al.*, 2022). Fractions were collected and dried in SpeedVac. Samples were then resuspended in 0.5% FA for LC-MS/MS analysis.

2.6.2 LC-MS/MS Analysis and mass Spectral Data Processing

The LC-MS/MS analysis was performed on the Orbitrap Fusion Lumos mass spectrometer coupled with U3000 RSLCnano UHPLC system. Details of the LC-MS/MS analysis were described in (Newey *et al.*, 2022).

All raw files were processed in Proteome Discoverer 2.4 using the Sequest HT search engine to search against reviewed Uniprot database of Homo Sapiens (Version February 2020) and contaminate database. Protein FDR was set at 1%. Reporter ion intensities were normalized by total peptide intensity to correct the variation in different protein loading in each channel. Further details of the data processing were described in (Newey *et al.*, 2022).

2.7 Protein analysis by enzyme activity

2.7.1 Histone deacetylase (HDAC) activity assay

Cells were seeded in to a 6-well plate, left to attach overnight then visually assessed in the morning. Cells were then media changed with media supplemented with DMSO or domatinostat (always utilising an identical volume of DMSO as the vehicle and vehicle control). Cells were treated with 0.625-2.5μM domatinostat for 24H, 3 days, or 5 days.

Following treatment, cells were harvested with TrypLE express as normal, pelleted, and washed twice with ice-cold PBS. The supernatant was removed, and then the cells were lysed with RIPA buffer. Cells were incubated with RIPA buffer for 30 minutes, with harsh vortexing every 5 minutes before being spun down at 13,000g for 10 minutes. The supernatant was then taken aside in to a clean, pre-chilled tube and quantified by Pierce Gold BCA Assay (as in Section 2.5.4).

The protocol was followed according to kit instructions. To summarise: an equal amount of whole cell lysate (50-200μg) was diluted to a final volume of 85μL with RNase-free water. 10X HDAC assay buffer and the colorimetric substrate were

added to the lysate mixture to reach a final volume of 100uL, before being incubated at 37°C for 1 hour. The reaction was then stopped by addition of lysine developer, followed by a further incubation 37°C for 30 minutes before reading on the plate reader at 405nm. HeLa nuclear extract used as the positive control, and 2μL of 1mM HDAC inhibitor (Trichostatin A) used in the negative control wells. The negative control was used as a blank, so that value was subtracted from all the test readings. Then for each line all the test values were normalised to their respective DMSO control, so the DMSO value was at 1.0 for each PDO.

2.8 Cellular assays

2.8.1 CellTitre Blue assay

Cells were plated in 50μL of appropriate media in to 96-well Optilux black plates, at a density relevant to the cell line/PDO growth rate and the duration of the experiment. Cell lines were seeded in media alone, and PDOs were seeded in media supplemented with 2% matrigel. A reference plate and experimental plates were always seeded at the same time, with every condition having 6 duplicate wells per plate. The outer wells were filled with 250μL of UF water to prevent 'edge effects' caused by excessive evaporation. Cells were left to incubate overnight, then treated with 50μL of 2X-concentrated drug-supplemented media. Immediately after, the reference plate was tested for the metabolic rate of the cells with application of CellTitre Blue (CTB). CTB was pre-warmed to 37°C in the water bath, then 20μL was added to every well, and the plate(s) were incubated for 1-2 hours, before reading at 590nM on the Cytation3 (BioTek Instruments Inc.). After the designated culture period, the test plates were also examined with CTB application and fluorescence reading. The mean value of the repeats were taken, then the mean value was normalised to the vehicle control (DMSO/UF water), where the value of 1 indicated full, un-inhibited growth. GraphPad Prism versions 8-9 were used to plot the response curves using the simple linear regression function.

2.9 Flow cytometry

2.9.1 HLA quantification by flow cytometry

Cells were grown, treated, harvested identically as in Section 2.2. HLA quantification was performed with the QIFIKIT quantitative flow cytometry assay (Agilent) according to the manufacturer's instructions. Antibodies used were pan-HLA-A/B/C (BioLegend, W6/32), pan-HLA-DR/DP/DQ (BioLegend, TU39), and IgG2aκ isotype control (BioLegend, MOPC-173). In summary: cells were filtered, counted on the Invitrogen Countess, and the appropriate volume spun down. Cells were then washed once in PBS, and aliquoted so $5x10^5$ cells per well were placed in a 96-well U-bottom plate. Cells were then resuspended in 50μL of diluted HLA or isotype control antibody (diluted as in Section 2.1.5 – with isotype control antibody matching the dilution of the pan-HLA-A/B/C or pan-HLA-DR/DP/DQ), and left to incubate for 30 minutes at 4°C in the dark. After the incubation, 100μL of setup beads, and 100μL of calibration bead solution were then added to separate additional wells of the plate. The primary-incubated cells and QIFIKIT beads were then washed twice with flow cytometry buffer. Next, cells and beads were incubated with 25μL of 1/50 anti-mouse secondary, and left to incubate for 45 minutes at 4°C in the dark. Cells and beads were then washed twice with flow cytometry buffer, resuspended in 200μL of flow cytometry buffer, and 2μL of DRAQ7 was added to the cells and incubated at 37°C for 10 minutes, before the samples were run on the Sony SH800. First the setup beads were run to adjust the photomultiplier tubes (PMT) voltage to get both the negative and positive populations on the scale, then all samples were run under those identical parameters. The data for the calibration beads and samples was then recorded. First, the MFI values for each HLA-stained sample were subtracted by the values of the corresponding IgG2aκ isotype control-stained sample to determine the specific antibody binding capacity of the calibration beads and the PDO samples. The calibration beads consisted of 5 groups of beads with different densities of antibody-binding molecules; these groups of beads had different MFI values, and this data was then used to generate a standard curve, to which the PDO samples were compared; this enabled approximation of the surface density of HLA-A, B, and C or HLA-DR, DP, and DQ.

2.10 Methods specific for Chapter 3 and 4

2.10.1 IFNγ treatment of cells

Cells for RNA sequencing, immunopeptidomics, and flow cytometry were treated with 600ng/mL IFNγ (R&D Systems) and incubated for 48 hours.

2.10.2 Trametinib treatment of cells

Cells for RNA sequencing, WB, and immunopeptidomics were treated with 30nM/mL trametinib (Cayman Chemical) and incubated for 48 hours.

2.10.3 Domatinostat treatment of cells

After optimisation, cells for RNA sequencing were treated with 2.5μM domatinostat (Selleckchem) and incubated for 72 hours.

2.10.4 HLA typing and HLA mutation calling

4-digit HLA typing of the PDOs was performed by our collaborator Professor Michal Bassani-Sternberg. The TruSight HLA v2 panel for sequencing was used on the Illumina MiniSeq system. For calling of possible HLA mutations, the sequencing data was then entered into the shell_call_hla_mutations_from_type script in the POLYSOLVER suite (Shukla *et al.*, 2015).

2.10.5 Purification of HLA-I and HLA-II peptides for LC-MS/MS analysis

Purification and LC-MS/MS of HLA-I and HLA-II peptides was performed by our collaborator Professor Michal Bassani-Sternberg. Each PDO cell pellet (biological replicate, $3.85x10^7$–$1x10^8$ cells/pellet) was split into two technical replicates that were processed. To summarise the process of HLA-peptide isolation and LC-MS/MS: the prepared cells were gently lysed, the lysate run through HLA-I complex capture with antibody-coated beads, and the flow-through then went on to HLA-II capture. The HLA-peptide complexes were then loaded on to a C18 plate for peptide recovery, released from the antibody coating through the addition of 1% trifluoracetic acid (TFA), before elution of peptides from the

HLA with the addition of acetonitrile (ACN) (28% for HLA-I and 32% for HLA-II) and 0.1% TFA. The recovered peptides would then be dried by vacuum centrifugation and stored at -20°C. Peptides were then resuspended in 0.1% FA and placed in to the UHPLC autosampler for analysis.

2.10.6 Analysis of MS immunopeptidomics data

For the MS immunopeptidomics MaxQuant analysis, I utilised the 'match between runs' (MBR) option for all replicates and treatment conditions. MBR was selected as it facilitates increased depth of detection and is most applicable where samples are similar. Therefore, MBR was performed separately per PDO line, and separately between HLA-I and HLA-II datasets. For peptide identification, the immunopeptidomics list was searched in MaxQuant against a fasta file containing the human proteome (Homo_sapiens_UP000005640_9606), with a 1% false discovery rate (FDR). In addition, a fasta file containing a custom reference database containing patient somatic mutations was used to define the neoantigen search space, set with an FDR of 5% to enhance the chance of peptide identification. Methionine oxidation and N-terminal acetylation were considered as modifications. Protocol described in detail in (Chong *et al.*, 2017; Newey *et al.*, 2019)

2.10.7 Prediction of NetMHC percentile ranks by NetMHCpan4.0

All HLA-I MS-detected peptides were entered into NetMHCpan4.0. The HLA allotypes determined for each PDO line were used for the prediction. Binding percentile rank predictions were used; the HLA restriction of the peptide was defined by the HLA for which the peptide had the lowest rank.

2.10.8 Prediction of NetMHC percentile ranks by NetMHCpan4.1

All HLA-I MS-detected peptides were entered into NetMHCpan4.1. The HLA allotypes determined for each PDO line were used for the prediction. Binding affinity (BA) rank predictions were used; the HLA restriction of the peptide was defined by the HLA for which the peptide had the lowest rank.

2.10.9 Computational prediction of neoantigens

Neoantigen prediction was performed by collaborator Dr Marta Punta for publication (Newey *et al.*, 2019). Neoantigen sequences were predicted from somatic mutations (including non-silent substitutions and indels, but not splice-site mutations or stop-gains), and binding percentile rank scores were generated with NetMHCpan4.0 by running all neoantigens for each PDO against corresponding HLA-I allotypes. Strong binders were defined as having predicted percentile rank $\leq 0.5\%$.

2.10.10 HLA-II motif deconvolution

HLA-II motif deconvolution was performed by collaborator Professor Michal Bassani-Sternberg for publication (Newey *et al.*, 2019). HLA-II peptidomics samples that contained ≥ 500 peptides with a length of ≥ 12 amino acids were submitted to motif deconvolution with MoDec (Racle *et al.*, 2019). Peptides that do not fit a motif inferred by MoDec are assigned to a flat motif, and represent possible contaminants or derivation from HLA-II molecules which provide few peptides to the sample.

2.11 Methods specific to Chapter 5

2.11.1 Cell culture and IFNγ treatment

Cells were cultured in 4-5 T175 flasks (or a single 5-layered T175 flask) per condition, depending on the growth characteristics of the cell lines/PDOs. Flasks were incubated at 37°C and 5% CO2, checked regularly, and when the cells reached 80% confluence, IFNγ was added. Fewer cells were required from the PDOs to yield a similar amount of product.

2.11.1.1 PDOs

Organoids were maintained in Advanced DMEM:F12 media supplemented with 10% FCS and 2% matrigel. PDOs were treated with 200IU IFNγ (10ng) per ml, for 48 hours before harvesting. This treatment dose and timeframe was selected as we had access to proteomics data for PDOs treated by IFNγ in this regime.

2.11.1.2 Cell lines

The different cell lines were maintained as in Section 2.2.1. Cells were treated with 250IU IFNγ (12.5ng) per ml for 24 hours before harvesting. This treatment dose and timeframe was selected as we had access to immunopeptidomics data for HCT116 treated by IFNγ in this regime.

2.11.1.3 Harvesting cells

To harvest the cells, the media was removed and flasks were washed with cold PBS. Cells were then incubated with pre-warmed TrypLE express for 5-15 minutes to enable the cells to detach. The TrypLE express was then neutralised with a 3-times volume of ice-cold PBS, and the cells were spun down at 300g and 4°C for 5 mins. The pellet was then resuspended once with ice-cold PBS, and pelleted down at 300g and 4°C for 5 mins. Pellets were then processed differently according to different runs.

TrypLE express was used instead of trypsin, as the reagent protects surface protein epitopes from degradation (*Thermo Fisher Scientific*, 2017), and it was previously used successfully for our immunopeptidomics study (Chapter 3 and (Newey *et al.*, 2019)), which was able to resolve the HLA-I and HLA-II immunopeptidome.

2.11.2 Use of the ER minute kit

In this protocol, and all protocols following, all the lysis reagents and all the buffers contained protease inhibitors (cOmplete™, EDTA-free protease inhibitor cocktail, Roche) and 1μM PMSF.

The PDO cells were maintained and IFNγ-treated as in Section 2.11.1.1. The highest end of the manufacturer-recommended cell number was used ($35x10^6$ cells). Cells were pelleted and snap frozen on dry ice with 100% ethanol, and stored in the -80°C freezer overnight. The protocol was executed according to the manufacturer's instructions, and all centrifugations took place at 4°C.

Cell pellets were thawed inverted, resuspended in buffer A, vortexed vigorously, applied to a spin column, and centrifuged at 16,000g for 30 seconds. The filter was discarded, any pellet was resuspended in the supernatant, and the tube was centrifuged at 2,000g for 5 minutes to eliminate nuclei, cell debris, and un-lysed cells (pellet 1). The supernatant was then moved to a fresh tube, and the pellet was snap frozen for further analysis. The supernatant was then centrifuged at 8,000g for 10 minutes to eliminate cell debris, mitochondria, and plasma membranes (pellet 2). This pellet was also snap frozen for further analysis.

To obtain the 'enriched ER' 400μL of the supernatant was transferred to a fresh tube and mixed well with 40μL of buffer B, incubated at room temperature for 30 minutes, then centrifuged at 8,000g for 10 minutes. Following this, the supernatant was removed and the pellet was resuspended in 400μL of cold buffer A by pipetting repetitively 40-50 times, followed by vortexing for 20 seconds. Then 40μL of buffer C was added and left to incubate at room temperature for 15 minutes, vortexing every 5 minutes. The mixture was then centrifuged at 800g for 5 minutes, and 400μL of that supernatant was combined with 400μL of buffer D, before being incubated at 4°C for 20 minutes.

Following the 4°C incubation the tube was centrifuged at 10,000g for 10 minutes, the supernatant discarded, and the tube spun briefly to allow the removal of any residual supernatant. The 'ER enriched' pellet (pellet 3) was then snap frozen.

2.11.3 Preparation and MS of Minute™ ER isolation kit product

Proteomics and peptidomics was performed by our collaborator Professor Jyoti Choudhary. The ER pellet was permeabilised by probe sonication (10 rounds of 1 second on, 1 second off, at 40% amplitude) in a solution of 30% methanol, 0.2% SDC, and 100mM TEAB in UF water. The solution was then centrifuged at 14,000g at 4°C for 10 minutes, and the pellet was solubilized in 0.5% SDS in UF water, vortexed, and centrifuged. SDS was removed from the supernatant with the HiPPR Detergent Removal kit, then it was applied through an Amikon 3K size exclusion filter. The filter allowed short peptides through (≈ 27 amino acids and below), and proteins were left inside the filter. The proteins were then recovered, reduction-alkylated, and trypsin-digested.

MS spectra for the proteome and the peptidome were searched against UniProt annotated reference proteomes of *Homo Sapiens* in Proteome Discoverer v2.4. For the peptidome search, to reduce the search space and increase the power, only peptides 7-36 amino acids long were analysed.

2.11.4 Sucrose gradient centrifugation approach 1

2.11.4.1 Cell lysis

The NCI-H747 cells were maintained and IFNγ-treated as in Section 2.11.1.2. One dense 5-layer T175 flask ($100x10^6$ NCI-H747 cells) was harvested as in Section 2.11.1.3, snap frozen on dry ice with 100% ethanol, and stored in the -80°C freezer overnight. The pellet was thawed inverted on ice, before resuspension in 5mL isotonic sucrose buffer (ISB) and homogenisation with 40 strokes of a Dounce homogeniser using the tight pestle.

2.11.4.2 Nuclei exclusion

The first stage was differential centrifugation: un-lysed cells and nuclei were pelleted down by a 10-minute centrifugation at 800g and 4°C. At this stage, the supernatant was removed and the pellet was snap frozen on dry ice with 100% ethanol.

2.11.4.3 Mitochondrial exclusion

The supernatant from the nuclei exclusion step was diluted to 6mL and aliquoted equally in to 1.5mL Eppendorf tubes and the mitochondria were pelleted down with a 20-minute centrifugation at 12,000g and 4°C. The supernatant was taken off, and the pellet was snap frozen on dry ice with 100% ethanol.

2.11.4.4 ER enrichment by sucrose gradient

The main components of the discontinuous sucrose gradient were assembled at room temperature ahead of time, before pre-chilling for an hour. The gradient was made by pipetting lightest to heaviest, adding the heavier layers underneath the lightest with a syringe (sucrose layers 1.3M (3.2mL), 1.5M (3.2mL), 2M (2.2mL)).

The supernatant from the mitochondrial exclusion step was diluted with cold HEPES buffer to a final volume of 26.3mL, then pelleted down at 150,000g and 4°C for 1 hour using the Ti70 rotor. This step was done to pellet down the organelles to enable their resuspension in a smaller volume of the ISB (1.8mL), for layering over the gradient. This was topped off with 1.6mL of HEPES buffer,

then spun for 90 mins at 180,000g and 4°C in a SW41 swinging bucket rotor, with acceleration and deceleration set to 1. The visible layers from each interface were then extracted with a fine needle and syringe, diluted with cold HEPES buffer, and pelleted down at 126,000g and 4°C in the SW41 rotor. The pellets were left to dry inverted, then diluted in appropriate volumes of RIPA buffer, before processed for WB as described in Section 2.5.

2.11.5 Sucrose gradient centrifugation approach 2

2.11.5.1 Cell lysis

The protocol utilised a frozen pellet of 1 dense 5-layer T175 flask of IFNγ-treated NCI-H747 cells. The pellet was thawed inverted on ice, before resuspension in ice-cold hypotonic extraction buffer (3x the volume of the packed cell pellet, according to (Sigma-Aldrich, 2006)), and incubated at 4°C for 20 minutes. Following this, cells were spun down at 600g for 5 minutes at 4°C, resuspended in ISB, then homogenised with 40 strokes of a Dounce homogeniser.

2.11.5.2 Nuclei exclusion

This step was kept the same as sucrose gradient method 1 (Section 2.11.4.2).

2.11.5.3 Mitochondrial exclusion

Following on from the results from sucrose gradient method 1, the speed and duration of the mitochondrial exclusion step was reduced to 10,000g for 10 minutes.

2.11.5.4 ER enrichment by Sucrose gradient

This step similar as in run 1 (Section 2.11.4.4), however, the speed of the gradient spin was reduced from 180,000g to 152,000g, and the duration was reduced from 90 minutes to 70 minutes. These settings were selected based on (Williamson *et al.*, 2015).

2.11.6 Sucrose gradient centrifugation approach 3

The steps of cell culture, cell lysis, nuclei exclusion, and ER enrichment gradient steps proceeded as in sucrose gradient method 2 (Section 2.11.5).

2.11.6.1 Mitochondrial exclusion

The speed of the mitochondrial exclusion step was reduced to 5,000g for 10 minutes.

2.11.7 Sucrose gradient centrifugation approach 4

The HCT116 and NCI-H747 cells were cultured as in the Section 2.2.1, and IFNγ-treated as in Section 2.11.1.2. The steps of cell lysis and nuclei exclusion proceeded as sucrose gradient method 2 (Section 2.11.5).

2.11.7.1 Mitochondrial exclusion

Th speed of the mitochondrial exclusion centrifugation was reduced to 3,000g for 10 minutes.

2.11.7.2 ER enrichment by Sucrose gradient

For version 1 (V1) of the protocol, the steps proceeded as in as in sucrose gradient method 3 (Section 2.11.6). For version 2 (V2) of the protocol, the post mitochondrial supernatant was spun at 5,400g and 4°C for 15 minutes. The pellet from this centrifugation was resuspended in 2mL of ISB and kept on ice whilst the supernatant was prepared. The post-5,400g supernatant from was spun at 150,000g and 4°C for 1 hour in a Ti70 rotor, and resuspended in 2mL with ISB. 1.8mL of each suspension were layered over separate pre-chilled gradients, then topped off with 1.6mL of HEPES buffer. The gradients were then processed alongside the gradient from V1, treated as in Section 2.11.5.4. In this iteration, approximately 80% of the top layer was discarded prior to needle collection of the gradient layers.

2.11.8 Histodenz gradient centrifugation approach 1

The cell lysis, nuclei exclusion, and mitochondrial exclusion were kept the same as in sucrose gradient method 4 (Section 2.11.7).

2.11.8.1 ER enrichment by Histodenz gradient

During the mitochondrial exclusion centrifugation, the Histodenz discontinuous gradient was formed at room temperature, layering the lightest layers first, slowly syringing the heavier layers underneath (5.5% HD (4.4mL), 17% HD (1.7mL), 35% HD (1.7mL)). The gradient was then refrigerated at 4°C.

The supernatant from the mitochondrial exclusion step was subjected to an additional 5,400g spin for 15 minutes at 4°C. The pellet and supernatant were treated as in Section 2.11.7.2, and layered over a gradient (each in 4.2mL of ISB), but in this case, it was the Histodenz gradient.

The gradient was centrifuged at 50,500g and 4°C for 40 minutes in a SW41 swinging bucket rotor, with the acceleration and deceleration set to 1. The tubes were carefully removed from the rotor, and the visible layers at the interfaces were fractionated with a needle and syringe, taking up 1-2mL per layer, according to the visible density of the layers. Only layers 2-4 were visible and collected.

The fractions were then diluted with cold HEPES buffer to fill to the top of the 13.2mL centrifuge tubes, and spun down at 126,000g and 4°C for 45 minutes. The pellets were then resuspended in RIPA lysis buffer, aliquoted in to Eppendorf tubes, and immediately snap frozen. The pellets were then all processed for western blotting as described in the Section 2.5. The WB membrane was cut and probed once, then stripped with ReBlot Solution, blocked, and subsequently re-probed with a new selection of antibodies.

2.11.9 Histodenz gradient centrifugation approach 2

The IFNγ-treated cell pellets were prepared as mentioned in Section 2.11.1.2, but for this experiment the pellets were not frozen, they were instead lysed fresh. For this experiment I used 1 dense 5-layer T175 flask of HCT116 cells.

2.11.9.1 Cell lysis

The fresh cell pellet was resuspended in 5mL of ISB. The cavitation chamber was pre-chilled to 4°C, the cell suspension was loaded in to the core of the chamber, and the chamber was screwed tight. The chamber was then attached to a nitrogen tank and the nitrogen was slowly released in to the chamber; the pressure in the cavitation chamber was set to 450 psi. The pressurised chamber was incubated on ice for 8 minutes, then the suspension was released in to a 50mL falcon.

2.11.9.2 Nuclei exclusion

The lysate was then diluted with ISB to a 10mL final volume and centrifuged at 800g and 4°C for 10 minutes. The supernatant was collected, being careful not to disturb the pellet. Any excess supernatant was removed, and the pellet was snap frozen on dry ice with 100% ethanol.

2.11.9.3 Mitochondrial exclusion

After the nuclei exclusion step, the post-nuclei supernatant was topped up to 10mL with more ISB, and aliquoted identically in to 5 x 2mL Eppendorf tubes. This was then spun at 3,000g and 4°C for 10 minutes.

2.11.9.4 ER enrichment by Histodenz gradient

During the mitochondrial exclusion centrifugation, the Histodenz discontinuous gradient was formed at room temperature, layering the lightest layers first, slowly syringing the heavier layers underneath (5.5% HD (4.4mL), 17% HD (1.7mL), 35% HD (1.7mL)). The gradient was then refrigerated at 4°C.

The post-mitochondrial supernatant was diluted to a final volume of 4.2mL with ISB and layered over the Histodenz gradient. The mitochondrial pellet was snap frozen on dry ice for later analysis. The gradient was centrifuged at 50,500g and 4°C for 40 minutes in a SW41 swinging bucket rotor, with the acceleration and deceleration set to 1. The tubes were removed from the rotor, and the visible layers were fractionated with a needle and syringe, taking up 1-2mL per layer, according to the visible density of the layers.

After ultracentrifugation samples were resuspended in 50μL of PBS to remove the pellet from the base of the ultracentrifugation tube before snap freezing on dry ice with ethanol. For WB processing, the layers were thawed in ice before pelleting the organelles at 16,100g for 15 minutes at 4°C. WB processing then was executed as described in the Section 2.5. The first blot was run with all the lysates, and the second blot was run with just the lysates from the discard or final pellets.

2.11.10 MS preparation of Histodenz gradient-derived fractions

The fractions were solubilized using probe sonication in lysis buffer (100mM TEAB, 1% SDC, 10% isopropanol, 50mM NaCl) supplemented with Halt Protease and Phosphatase inhibitor cocktail (Thermo Scientific), boiled for 5 min at 90°C, and sonicated once more. Proteins were precipitated by adjusting acetonitrile concentration to 70% and incubating at -20°C for 15 min, followed by 15 min centrifugation at 10,000g. The supernatants containing endogenous peptides were collected and speed-vacuum dried, re-suspended in 0.1% FA and centrifuged for 5 min at 10,000g to precipitate SDC, then speed-vacuum dried again. The protein pellets were re-suspended in 100mM TEAB, reduced with TCEP for 1 hour, followed by alkylation with 10mM IAA for 30 min, then digested

by adding trypsin (Pierce) at final concentration 75ng/µl to each sample and incubating the samples for 18 hours at RT.

The endogenous and trypsin-digested peptides were re-suspended in 100mM TEAB and then labelled with TMTpro (Thermo Scientific) for 1h according to manufacturer recommendations. Both endogenous and trypsin digested peptide samples were combined into their respective TMT sets, speed-vacuum dried and fractionated using Pierce High pH Reversed-phase peptide fractionation Kit (Thermo Scientific, #84868) according to manufacturer recommendations. Following this, the fractions were speed-vacuum dried and were resuspended in 0.1% FA.

LC-MS analysis was performed Dionex UltiMate UHPLC 3000 system coupled with the Orbitrap Lumos Mass Spectrometer (Thermo Scientific). Peptides were loaded to a C18 column and subjected to a gradient elution before analysis by MS.

2.11.11 MS analysis of gradient-derived fractions

Database searching was performed by Zuza Kozik for our collaborator Professor Jyoti Choudhary. MS spectra were searched against UniProt annotated reference proteomes of *Homo Sapiens* (January 2023) in Proteome Discoverer v2.4 (Thermo Scientific) with the SEQUEST-HT engine using following parameters: 20 ppm precursor mass tolerance and 0.5 Da fragment ion mass tolerance, the static modifications were specified as carbamidomethylation of Cys residues and TMT modification of peptide N-term and Lys residues, and dynamic modifications were specified as oxidation of Met and deamidation of Asp and Glu residues. For enzymatically-digested peptides, the search was restricted to fully tryptic peptides with maximum of two missed-cleavages, while for endogenous peptides No-Enzyme search was performed. Based on the *q* value and decoy database search confidence of peptide discovery was estimated at 1% FDR with percolator node. Protein groups with only unique peptides were used for quantification.

2.11.12 Purification and analysis of HLA-I peptides from HCT116 for LC-MS/MS analysis

HLA-I peptide purification and MS analysis were performed by Zuza Kozik for our collaborator Professor Jyoti Choudhary. This protocol is currently unpublished so I cannot provide exact details of the methodology. However, I am able to share that it runs similarly to that in 2.10.5 - 2.10.6, but instead, using a C18 column in place of the C18 plate, and Proteome Discoverer was used for database searching instead of MaxQuant.

2.12 Methods specific to Chapter 6

2.12.1 Allogeneic T-cell and 2D PDO co-culture

PDOs maintained as in Section 2.2.5, and PBMCs were harvested, and CD8+ T-cells isolated and activated as in Sections 2.2.6-2.2.8. T-cells were frozen in FBS with 10% DMSO at day 10 post-activation, then thawed 24 hours before treatment in to T-cell media supplemented with 30IU/mL IL-2. Before freezing, purity of CD8 cells was assessed by antibody staining and flow cytometry with APC anti-human CD8 (BioLegend, SK1).

2.12.1.1 GFP-tagging of PDO cells

PDO cells were previously labelled with nuclear eGFP as described in (Gonzalez-Exposito *et al.*, 2019).

2.12.1.2 Testing the effect of culture conditions on CEA expression

To test effect of cell density on CEA expression cells were seeded at a density of $1x10^5$, $2x10^5$, $4x10^5$, $8x10^5$, $1x10^6$ cells in DMEM/F12 media with 20% foetal bovine serum (FBS), Glutamax, 100 units/ml penicillin/streptomycin and 2% matrigel, and grown in a 6-well plate for 5 days. To test effect of media conditions on CEA expression cells were seeded in DMEM/F12 media with 20% FBS, Glutamax, 100 units/ml penicillin/streptomycin (+/- 2% matrigel), or RPMI 1640 media with 10% FBS, Glutamax, 100 units/ml penicillin/streptomycin (+/- 2% matrigel), and were grown in culture for 5 days.

2.12.1.3 Flow cytometry for CEA expression

DRAQ7 was used for dead cell exclusion. CEA was stained for with anti-human CEA CH1A1A (Roche) primary antibody and an anti-human PE-conjugated secondary antibody (Jackson ImmunoResearch). Cells were analysed on a Sony SH800 flow cytometer. Data analysis performed on FlowJo 10.5.3.

2.12.1.4 E:T titration for treatment with CEA-TCB

PDOs were seeded in T-cell media (RPMI 1640 media with 10% FBS, Glutamax, 100 units/ml penicillin/streptomycin) at a cell number to give a 5% confluence in a 96-well plate on day 0, before treatment of the PDOs. On day 1 PDOs were treated with 20nm of CEA-TCB (Roche) or 20nm of untargeted control-TCB (DP47) (Roche). Effector cells used were pre-activated allogeneic CD8 T-cells. To test treatment ratios cells were treated with a 2:1, 1:1, or 1:2 E:T ratio. Growth of the GFP-tagged organoid cells was tracked on the confluence application on the Celigo Imaging Cytometer (Nexcelom Bioscience).

2.12.1.5 CEA-TCB Resistance Assay

PDO cells were seeded as above, but in to a 24-well plate on day 0. On day 1 allogeneic CD8 T-cells were added (at E:T ratio 1:1) alongside drug treatment of 20nm CEA-TCB or 20nm control-TCB, with or without 2µg/mL Atezolizumab (Selleckchem). Growth of the GFP-tagged PDOs was tracked on the Celigo Cytometer (Nexcelom Bioscience). Cells were re-treated with T-cells and drugs when they had grown at least 1.25-fold compared to the last treatment timepoint. At re-treatment, the E:T was increased to 1.5:1 and then to 3:1. Cells were harvested at day 20 for flow cytometry and further culture.

2.12.1.6 Flow cytometry for CEA and HLA-I expression

PDOs were stained with Zombie Red (BioLegend) for dead cell exclusion, PE-conjugated anti-CEA antibody (Sino Biological, IgG2a Clone #02), Brilliant Violet anti-PD-L1 (BioLegend, 29E.2A3), and analysed on a BD LSR II flow cytometer.

For IFNγ-sensitivity testing cells were treated with 10ng/mL IFNγ for 48H then stained with Zombie Red and AF700 anti-human-HLA-A,B,C (BioLegend, W6/32).

2.12.2 Autologous T-cell and PDO co-culture

Where LRS cones have been used, they were prepared as follows: PBMCs were thawed 24H before treatment. Cells were thawed quickly in a 37°C water bath, then 10mL of pre-warmed T-cell media was added to the cells dropwise, before spinning down at 300g for 5 minutes. PBMCs were incubated overnight at 37°C and 5% CO_2 in T-cell media supplemented with 30IU/mL IL-2.

2.12.2.1 Extracellular readout for autologous PBMC-expanded T-cell and PDO co-culture test

On the day of the experiment, PBMCs were spun down and resuspended in T-cell media at $1x10^6$ cells per mL, and 50μL was added to a U-bottom plate. 50μL of PE anti-CD107a (diluted in T-cell media) was added to the plate, to a final dilution of 1/120. For the negative control wells 100μL of T-cell media alone was added. For the positive control wells 100μL of T-cell media supplemented with PMA and ionomycin, to a final concentration of 50ng/mL and 1μg/mL, was added. The plate was then centrifuged at 200g for 1 minute to loosely pellet cells, and the plate was left to incubate at 37°C and 5% CO_2 for 24 hours.

Following the incubation, the cells were spun down at 330g for 5 minutes. Some cells were also stained as single stains and fluorescence minus one controls (FMOs) for compensation and gating. Cells were washed once with flow buffer, then incubated with 20μL of viability dye zombie near-infrared (Z-NIR) (1/250 in PBS), for 20 minutes at room temperature, protected from light. After the incubation, cells were washed once with flow buffer, then resuspended in 20μL antibody single stains or master mixes as appropriate (BV421 anti-CD3, AF488 anti-CD8, and APC anti-CD137 were prepared at 1/20 in flow buffer). Cells were incubated at 4°C for 30 minutes, protected from light. The cells were then washed once with 200μL flow buffer. Cells were then fixed in 80μL BD

Fix/Permeabilization buffer at 4°C for 20 minutes, protected from light. Following this, cells were then washed twice with 200μL Perm/Wash buffer, and resuspended in 200μL flow buffer for storage at 4°C overnight before running on the BDSymphony flow cytometer the following day.

2.12.2.2 Intracellular readout for autologous PBMC-expanded T-cell and PDO co-culture test

On the day of the experiment, PBMCs were spun down and resuspended at 10^6 cells per mL, and 50μL was added to a U-bottom plate. 50μL of PE anti-CD107a (diluted in T-cell media) was added to the plate, to a final dilution of 1/120. For the negative control wells 100μL of T-cell media alone was added. For the positive control wells 100μL of T-cell media supplemented with PMA and ionomycin, to a final concentration of 50ng/mL and 1μg/mL, was added. The plate was then centrifuged at 200g for 1 minute to loosely pellet cells, and the plate was left to incubate at 37°C and 5% CO_2 for 1 hour. Following this, GolgiSTOP/GolgiPLUG solution was prepared in T-cell media, and added to reach a final concentration of 1/1,500 and 1/1,000. Cells were then incubated for another 5 hours at 37°C and 5% CO_2.

Following the incubation, the cells were spun down at 330g for 5 minutes. Some cells were also stained as single stains and fluorescence minus one controls (FMOs) for compensation and gating. Cells were washed once with flow buffer, then incubated with 20μL of viability dye Z-NIR (1/250 in PBS), for 20 minutes at room temperature, protected from light. After the incubation, cells were washed once with flow buffer, then resuspended in 20μL of surface antibody single stains or master mixes as appropriate (AF700 anti-CD3, BV650 anti-CD4, AF488 anti-CD8 were prepared at 1/20 in flow buffer). Cells were incubated at 4°C for 30 minutes, protected from light. The cells were then washed once with 200μL flow buffer, and fixed in 80μL BD Fix/Permeabilization buffer at 4°C for 20 minutes, protected from light. Following this, cells were then washed twice with 200μL Perm/Wash buffer, resuspended in 20μL of intracellular antibody stain APC anti-IFNγ (1/40 in Perm/Wash buffer), and incubated at 4°C for 30 minutes, protected from light. Following this, cells were then washed twice with 200μL Perm/Wash

buffer, and resuspended in 200μL flow buffer for storage at 4°C overnight before running on the BDSymphony flow cytometer.

2.12.2.3 Optimisation of extracellular and intracellular readout for autologous PBMC-expanded T-cell and PDO co-culture test

On the day of the experiment, PBMCs were spun down and resuspended at 10^6 cells per mL, and 50μL was added to a U-bottom plate. 50μL of PE anti-CD107a (diluted in T-cell media) was added to the plate, to a final dilution of 1/120. For the negative control wells 100μL of T-cell media alone was added. For the positive control wells 100μL of T-cell media supplemented with PMA and ionomycin, to a final concentration of 50ng/mL and 1μg/mL, was added. The plate was then centrifuged at 200g for 1 minute to loosely pellet cells, and the plate was left to incubate at 37°C and 5% CO_2 for 16 hours.

Following this, GolgiSTOP/GolgiPLUG solution was prepared in T-cell media, and added to reach a final concentration of 1/1,500 and 1/1,000. Cells were then incubated for another 5 hours at 37°C and 5% CO_2. Following the incubation, the cells were spun down at 330g for 5 minutes. Some cells were also stained as single stains and fluorescence minus one controls (FMOs) for compensation and gating. Cells were washed once with flow buffer, then incubated with 20μL of viability dye Z-NIR (1/250 in PBS), for 20 minutes at room temperature, protected from light. After the incubation, cells were washed once with flow buffer, then resuspended in 20μL of surface antibody single stains or master mixes as appropriate (BV785 anti-CD3, PerCP anti-CD4, AF488 anti-CD8, and APC anti-CD137 were prepared at 1/20 in flow buffer). Cells were incubated at 4°C for 30 minutes, protected from light. The cells were then washed once with 200μL flow buffer, and fixed in 80μL BD Fix/Permeabilization buffer at 4°C for 20 minutes, protected from light. Following this, cells were then washed twice with 200μL Perm/Wash buffer, resuspended in 20μL of intracellular antibody single stains or master mixes as appropriate (BV650 anti-IFNγ 1/40, and BV421 anti-Granzyme B 1/20, in Perm/Wash buffer), and incubated at 4°C for 30 minutes, protected from light. Following this, cells were then washed twice with 200μL Perm/Wash

buffer, and resuspended in 200μL flow buffer for storage at 4°C overnight before running on the BDSymphony flow cytometer.

2.12.2.4 Autologous PBMC and PDO co-culture for expansion of tumour-reactive T-cells

PDOs were establish and maintained as in Section 2.2.2 - 2.2.4, and PBMCs were harvested as in Section 2.2.6.

The protocol used for this co-culture was based around the work performed in (Dijkstra *et al.*, 2018; Cattaneo *et al.*, 2020; Chalabi *et al.*, 2020).

To begin the co-culture, a PDO that was growing well was selected. On D-2 fully formed PDOs were removed from matrigel but kept intact by incubation with 2mg/mL dispase (in PBS) for 15 minutes at 37°C. Then the reaction was ceased by the addition of 100μL of 0.5M EDTA, and the PDOs washed with PBS. The PDOs were pelleted by spinning at 300g for 5 minutes, and were re-plated in 3D complete media supplemented with 10μM ROCK inhibitor. On D-1 the PDOs were then treated with 200ng/mL IFNγ overnight. The PBMCs were gently thawed a 37°C, diluted with dropwise with 14mL pre-heated cell culture medium, and pelleted at 200g for 12 minutes with intermediate deceleration. The autologous PBMCs were then resuspend at $2x10^6$/mL in T-cell culture medium supplemented with 150IU/mL of IL-2 and incubated overnight at 37°C. A U-bottom 96-well plate was coated with 5μg/mL of anti-CD28 (in PBS), covered with parafilm, and incubated overnight at 4°C.

On D0 of the assay, the IFNγ-treated organoids were harvested and dissociated to single cells and small clusters by a 5-15-minute incubation with pre-warmed TrypLE Express. The PBMCs were then resuspended in T-cell media with 150IU/mL IL-2 and 20μg/ml anti-PDL1 antibody Atezolizumab, and combined with the PDOs at a ratio of 20:1. The pre-coated anti-CD28 plate was washed twice, and then 200μL of the PDO/PBMC suspension was plated per well. The plate was incubated at 37°C for a week, with the medium refreshed every 2-3

days. After 1 week in culture, the PBMCs were harvested and incubated again with a fresh supply of IFNγ-treated organoids for a further week.

2.12.2.5 Autologous PBMC-expanded T-cell and PDO co-culture

After a total of 2 weeks in culture, the PBMCs were harvested, some cryopreserved, and some assessed for tumour reactivity. By the end of the 2 week co-culture 90% of the expanded PBMCs were expected to be T-cells, and therefore I will refer to them as T-cells for this assay (Cattaneo *et al.*, 2020).

For the tumour reactivity co-culture, PDOs were prepared as in the T-cell expansion protocol (Section 2.12.2.4). However, for this assay, the T-cells were combined with the PDOs at a ratio of 2:1 in another pre-coated anti-CD28 plate. For this assay, different combinations for the co-culture were tested:

	PDO	T-cells
T-cell and PDO alone	Standard PDOs	Untreated T-cells
T-cell and PDO with anti-PDL1 (Atezolizumab)	Standard PDOs	T-cells with final concentration 20μg/mL anti-PDL1
T-cell and PDO with anti-PD1 (Nivolumab)	Standard PDOs	T-cells with final concentration 20μg/mL anti-PD1
Positive control	Standard PDOs	T-cells with final concentration 50ng/mL PMA and 1μg/mL ionomycin
Negative control	HLA-I-blocked PDOs	T-cells with final concentration 20μg/mL anti-PDL1

Table 2.9. Details of PDO PBMC co-culture variables.

As the assay was plated, AF647 anti-CD107a was also added to the culture, to a final dilution of 1/120.

Once all the wells had been plated, the plate was then centrifuged at 200g for 1 minute to loosely pellet cells, and left to incubate at 37°C and 5% CO_2 for 16 hours. Following this, GolgiSTOP/GolgiPLUG solution was prepared in T-cell media, and added to reach a final concentration of 1/1,500 and 1/1,000. Cells were then incubated for another 5 hours at 37°C and 5% CO_2. Cells were stained according to the protocol in Section 2.12.2.3, but with viability stain Zombie-red, and surface antibodies: BV785 anti-CD3, APC-Cy7 anti-CD4, AF488 anti-CD8 (all 1/20), and intracellular antibodies: BV650 anti-IFNγ (1/40) and PE anti-Granzyme B (1/20).

Chapter 3 Mass spectrometry analysis of the HLA-presented immunopeptidome in mismatch repair proficient patient-derived colorectal cancer organoids

3.1 Introduction

Mismatch repair proficient (MMRp) colorectal cancers (CRCs) represent 95% of the metastatic CRC population. Although survival rates are improving in CRCs overall, MMRp CRC is lagging behind, with median overall survival remaining at 30 months (Cremolini *et al.*, 2015; Tsuji *et al.*, 2021). Immunotherapy with immune-checkpoint inhibitors (ICIs) has been successful in mismatch repair deficient (MMRd) CRCs, but remains ineffective in MMRp CRC (Lote *et al.*, 2022).

As MMRd tumours consist of cancer cells that are deficient in DNA mismatch repair mechanisms, the tumours possess a higher somatic mutation load, averaging a mean of 1,158 non-silent mutations in MMRd CRC. Conversely, MMRp CRC are MMR-proficient, and as a result have a lower mutation count, averaging 123 mutations (Grasso *et al.*, 2018). Neoantigens are human leukocyte antigen (HLA)-binding peptides that are derived from proteins encoded by a non-silent somatic mutation. Neoantigens are considered a key subset of tumour-specific antigens, that are able to identify the tumour cell as foreign to T-cells, marking the cell for destruction by the immune system. Therefore, the comparatively lower mutation load seen in MMRp CRC could explain the poor sensitivity to ICIs. However, computational algorithms that predict the binding of potential neoantigens to autologous HLA class I (HLA-I) molecules have predicted that many MMRp CRCs harbour over 100 potential HLA-I-binding neoantigens, which is at odds with the poor ICI sensitivity seen in MMRp CRCs (Woolston *et al.*, 2019; Glaire *et al.*, 2022). Therefore, it was important to assess the disparity between the number of predicted neoantigens and ICI sensitivity in MMRp CRCs, to better understand their true HLA-presented neoantigen load.

Mass spectrometry (MS) immunopeptidomics is an alternative method to neoantigen prediction that directly evaluates the repertoire of the immunopeptidome, which encompasses all HLA-presented peptides and

neoantigens. Previously this has been applied to cell lines which may not accurately represent patient tumours. This is because typically, only aggressive cancer cell clones can persist in 2D culture, and after many years of culture they may not represent the tumours from which they were derived. MS immunopeptidomics has also been performed on large amounts of material from surgical resections (>1g) (Bassani-Sternberg *et al.*, 2016; Chong *et al.*, 2017; Müller *et al.*, 2017) which far exceeds the mass of tissue that can be recovered from biopsies. This is an important distinction, because biopsies can be taken prior to treatment and be more easily incorporated in to the treatment regime of clinical trials. Furthermore, CRC tumours can have high stromal content, including fibroblasts and immune cells. As HLA-I is expressed on all nucleated cells, direct analysis of clinical tissue samples which contains stromal components can impede the analysis of the cancer-specific immunopeptidome.

Patient-derived organoids (PDOs) are a model system which can be established from CRC specimens, including even small (3-5mm) biopsies, with success rates of up to 90% reported (Sato *et al.*, 2011; Van De Wetering *et al.*, 2015). PDOs can be grown from patient tumours that match the stage and pre-treatment history of CRC subtypes in which ICIs have been tested in clinical trials (Le *et al.*, 2015). Therefore, combining PDO culture with the MS immunopeptidomics technique has the potential to provide clinically-useful information about the immunopeptidome of patient tumours and how they are affected by different treatments.

The aim of this initial study was to use MMRp CRC PDOs to directly measure their HLA-presented neoantigen and cancer/testis antigen (C/T) landscape by high-depth immunopeptidomic analysis. Culture techniques that facilitated growth of CRC PDOs to large numbers for MS analysis was important, as PDOs are usually cultured in a 3D matrigel matrix which is prohibitively expensive and time-intensive. Prior to my joining the Gerlinger group, a method of culture whereby media is supplemented with 2% matrigel was developed by Beatrice Griffiths. This addition of 2% matrigel functioned to support attachment of the PDOs to the plastic culture substrate. The matrigel supplement is then easily removed with standard dissociation reagents. This study showed that this culture method enables large-scale expansion of PDOs to hundreds of millions of cells,

which are sufficient for high-depth immunopeptidomic analysis. Data on the breadth of neoantigen presentation in MMRp CRCs from this study could deliver additional information to help more accurately predict success of ICI immunotherapy, and provide more robust targets for peptide vaccines. Another advantage of the PDO culture model, was the ability to test the effect of different cellular perturbations on the immunopeptidome, which may help to inform treatment choices in the clinic. For example, understanding how a treatment may increase the number of HLA-presented neoantigens may provide a good candidate for combination treatment with immune checkpoint inhibitor treatment in MMRp cancer types. Therefore, further to understanding the CRC MMRp immunopeptidome, I endeavoured to understand if some perturbations could increase the number of neoantigens and C/T antigens presented on HLA.

It is known that IFNγ exposure can increase expression of HLA-I and HLA-II, alongside immunoproteasome catalytic components PSMB8-10 which may enhance processing and presentation of neoantigens (Ferrington and Gregerson, 2012). Furthermore, genetic inactivation of IFNγ signalling in cancer cells has been correlated with failed tumour rejection by the immune system, and more recently with ICI immunotherapy resistance (Dighe *et al.*, 1994; Gao *et al.*, 2016), supporting the clinical relevance of IFNγ in tumour rejection. Based on this knowledge, IFNγ was selected as one of the treatments to test for neoantigen-boosting ability.

Trametinib is an inhibitor of the mitogen-activated protein kinase (MAPK) pathway that acts downstream of RAF kinases by inhibiting MEK 1/2. The MAPK pathway is often constitutively activated through mutations or amplifications of MAPK pathway components such as *KRAS* , *NRAS*, or BRAF (Armaghany *et al.*, 2012), as seen in the majority of CRCs. Alternatively, autocrine section of EGF receptor tyrosine kinase ligands can constitutively activate the downstream MAPK pathway. In vitro data has shown that MEK inhibitor treatment can increase HLA expression, and combination treatment with immune checkpoint inhibiting antibodies could significantly improve tumour growth inhibition when compared against the single agent treatments (Liu *et al.*, 2015). Therefore, perturbation with trametinib treatment was included in this study to better

understand how of MEK inhibition may impact the presentation of neoantigens and C/T antigens in MMRp CRC.

Therefore, I applied MS immunopeptidomics to five MMRp CRC PDOs to identify the scale of the immunopeptidome, and the nature and size of the neoantigen and C/T repertoire. Further to this, comparison of the five PDOs facilitated interrogation of the reasons behind differences between disparities in immunopeptidome depth. I then subsequently investigated the effects of IFNγ and MEK-inhibitor treatment on the immunopeptidome and neoantigen landscape.

3.2 Results

3.2.1 The basal HLA class-I immunopeptidome in 5 CRC PDOs

For this analysis I utilised previously-propagated patient derived organoids (PDOs) derived from MMRp CRC patient biopsies; one from a chemotherapy-naïve patient (CRC-08), and the other four from chemotherapy-resistant metastatic CRC (CRC-01, -03, -04, -05).

A non-synonymous mutation in DNA produces a protein with an altered amino acid sequence. When the mutated protein is degraded by the proteasome, a peptide containing that mutation can be generated and presented on HLA-I, targeting the mutation-harbouring cell for destruction by T-cells; therefore, understanding the genetic background of these PDOs was important. Furthermore, understanding how many and which specific genes were mutated could help to explain the size, diversity and characteristics of the immunopeptidomes. For this reason, exome sequencing and clonality assessment had been performed by Louise Barber and Andrew Woolston of the Gerlinger group.

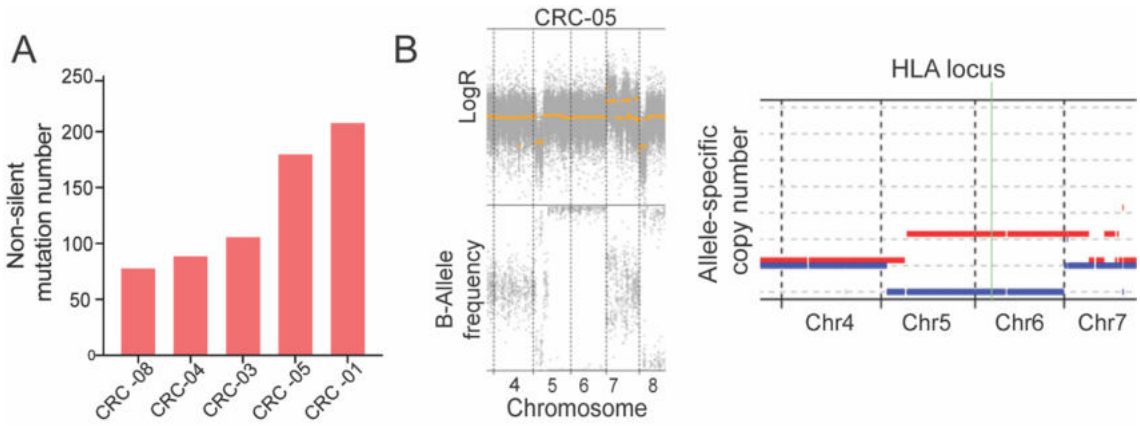


Figure 1.1: Genetic context across our five PDOs. A: Number of non-silent mutations detected by exome sequencing, per PDO. **B:** DNA copy number data of PDO CRC-05, generated from exome sequencing.

Exome sequencing of the PDOs, at a 100X target depth, detected 78-209 non-silent mutations per PDO (Figure 1.1A), including typical MMRp CRC driver mutations (e.g. *APC*, *TP53*, and *KRAS* mutations – Table 3.1).

	CRC-01	CRC-03	CRC-04	CRC-05	CRC-08
Stage	IV	IV	IV	IV	IV
Prior chemo-therapy	Yes	Yes	Yes	Yes	No
Age at biopsy	60	61	59	52	51
Non-silent mutation load	208	106	89	180	78
APC	p.Y935X, p.S1411fs	p.R876X	p.S1356X	p.Q1367X	p.Y1075fs
TP53		p.G245S	p.T284fs	p.R210X	p.R205C
KRAS	p.G12C	Amplification	p.A18D	p.G12D	p.G12D
TCF7L2		p.F105fs			
SMAD4			p.G365R		

Table 3.1. Clinical characteristics of source tumour and mutation load in the 5 PDOs.

The mutations on the chromosomes with loss of heterozygosity (LOH) (shown Figure 1.1B), each exhibited variant allele frequency of 99-100%, which demonstrated that the organoid cultures consisted of highly pure cancer cells without substantial stromal cell contribution. This means I could be confident that the peptides detected were derived specifically from cancer cells without contamination from other cell types. In addition, the exome sequencing showed 93% of all the mutations were clonal, illustrating that there was not a substantial amount of heterogeneity within the organoid populations.

To enable the PDOs to be sent for immunopeptidomic analysis of HLA-I and HLA-II peptides, they were expanded over 8-16 weeks to $1x10^8$ cells per replicate in 20% FCS media DMEM:F12 media, supplemented with 2% matrigel. The matrigel was used to support attachment of cells to the plastic substrate. Cells were carefully harvested by incubation with TrypLE express to enable detachment from matrigel and the plastic substrate whilst protecting cell surface

epitopes (*Thermo Fisher Scientific*, 2017). Cells were pelleted and snap frozen before being sent for immunopeptidomics. All cell pellet preparation was performed by Beatrice Griffiths before my arrival at the institute, and the MS immunopeptidomics was performed by our collaborator Dr Bassani-Sternberg at the Ludwig Institute in Lausanne. All further analysis and experiments were performed by myself.

	CRC-01	CRC-03	CRC-04	CRC-05	CRC-08
Untreated - Biological replicates	2	3	2	2	3
- Median viability	85%	74%	88%	92%	96%
IFNγ - Biological replicates	5	-	4	3	6
- Median viability	82%		84%	87%	92%
Trametinib - Biological replicates	4	-	4	4	4
- Median viability	96%		82%	88%	86%

Table 3.2. Total cell numbers used and number of repeats for PDO immunopeptidomics.

For purification of HLA-I and HLA-II peptides, each biological replicate (numbers of biological replicates detailed in Table 3.2) was split in to two technical replicates to enable critical analysis of similarity.

To briefly summarise the process of HLA-peptide isolation and LC-MS/MS: the prepared cells were gently lysed. The lysate was run through HLA-I complex capture with antibody-coated beads, and the flow-through then goes on to HLA-II capture. The HLA-peptide complexes were then loaded on to a C18 plate for peptide recovery, released from the antibody coating through the addition of 1% trifluoracetic acid (TFA) which denatures the antibodies, before elution of peptides from the HLA with the addition of acetonitrile (ACN) (28% for HLA-I and 32% for HLA-II) and 0.1% TFA. The recovered peptides would then be dried by vacuum centrifugation and stored at -20C. Peptides were then resuspended in 0.1% formic acid (FA) and placed in to the UHPLC autosampler for analysis.

For the MS immunopeptidomics MaxQuant analysis, I utilised the 'match between runs' (MBR) option for all replicates and treatment conditions. MBR was selected as it facilitates increased depth of detection, allowing the identities of peptides found in other conditions to be assigned to a missing value in another technical replicate/condition based on the m/z ratio, charge state, and retention time (Lim, Paulo and Gygi, 2019; Yu, Haynes and Nesvizhskii, 2021). MBR is most applicable where samples are similar, therefore it was performed separately per PDO line, and separately between HLA-I and HLA-II datasets. For peptide identification, the immunopeptidomics list was searched against a fasta file containing the human proteome (Homo_sapiens_UP000005640_9606), with a 1% false discovery rate (FDR). In addition, a fasta file containing a custom reference database containing patient somatic mutations was used to define the neoantigen search space, set with an FDR of 5% to enhance the chance of peptide identification.

To determine the number of unique peptide sequences, peptides were simply classified by a binary criterion of detected or not detected; a peptide was only defined as detected if it was detected in both technical replicates, of at least a one of the biological replicates. Furthermore, peptides were only considered for analysis if they could map to a single source protein. In addition, peptides 8-25aa in length were considered when counting HLA-I peptides, whereas only peptides 12-25aa were considered for HLA-II peptide counts.

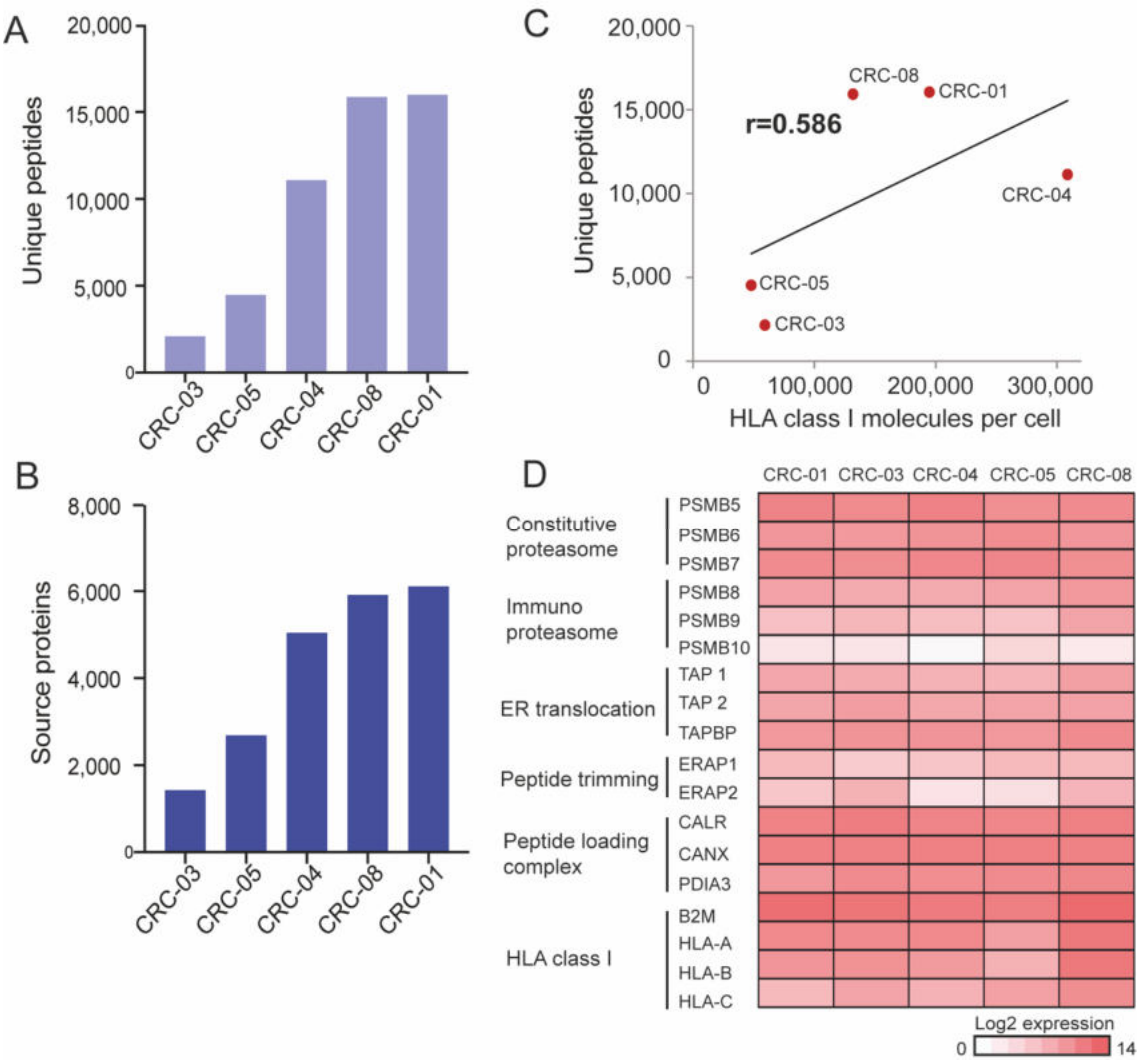


Figure 1.2: HLA-I immunopeptidome across our five PDOs. A: Number of unique peptides per PDO detected by MS immunopeptidomics across all biological repeats. **B:** Number of source proteins to which peptides could be uniquely mapped. **C:** Correlation of HLA-I molecules per cell (assessed by flow cytometry with quantification beads), against the number of unique peptides. The Pearson correlation coefficient between the 5 CRC PDOs is highlighted. **D:** Heatmap of RNA expression of genes central in the antigen and HLA-I peptide presentation process.

I first analysed the immunopeptidome of unperturbed cells to interrogate the total number of peptides and of neoantigens presented on HLA-I and on HLA-II. To give the best depth of detection I chose to count the number of unique peptides (UPs) that appeared in one or more biological replicates (Figure 1.2A). The number of peptides recovered from HLA-I ranged between 2,124 and 16,030 per

PDO, with a mean of 9,936 peptides. CRC-01 presented 16,030 UPs, the highest number of UPs on HLA-I, mapping to 6,124 unique source proteins (Figure 1.2B). In contrast CRC-05 and CRC-03 presented a dramatically lower number of UPs (4,504 and 2,124 UPs mapping to 2,700 and 1,439 proteins respectively). Despite CRC-05 and CRC-03 only presenting 28.1% and 13.3% of the number of UPs of CRC-01, they retained 44.1% and 23.5% of the source protein diversity. For CRC-01, each protein sampled on HLA-I was represented by a mean of 2.6 peptides, compared to a mean of 1.67 and 1.48 peptides per presented protein for CRC-05 and CRC-03 respectively. The ability of PDOs, CRC-03 and CRC-05, with a smaller immunopeptidome to present a higher proportion of the encoded proteome in their smaller immunopeptidome, illustrated that representation of a diverse range of proteins on HLA-I must be biologically important.

To try to understand the molecular determinants of the large variation in the number of HLA-I-presented UPs between the PDOs, I first assessed the median number of HLA-I molecules presented on the cell surface of each PDO line (Figure 1.2C). I was able to determine the number of HLA-I molecules on each cell by performing flow cytometry using a pan-HLA-I antibody and a bead set (QIFIKIT) which had a fixed number of antibody binding sites per bead. This allowed for the generation of a standard curve, from which to compare the median fluorescence intensity values from each PDO. 48,202-308,847 HLA-I molecules (mean of 148,789) were detected across the 5 PDOs, which showed modest correlation with the number of UPs presented on each PDO (with a Pearson's correlation coefficient of 0.586). This data suggested that the abundance of surface HLA-I molecules can influence the diversity of the proteasome.

Pathway	CRC-01	CRC-03	CRC-04	CRC-05	CRC-08
Kegg antigen processing and presentation	***CIITA*** (p.E854G)	-	-	-	***HSPA6*** (p.G226R)
Reactome Class I MHC-mediated antigen8 processing and presentation	-	***KLHL41*** (p.E60D - within BTB domain), ***RNF115*** (p.S61T)	-	***FGB*** (p.E275K), ***UBR4*** (exon87: c.12901+1G>T), ***HERC5*** (p.E137fs), ***FBXL14*** (p.S282A)	***UBE2Z*** (p.S341X)

Table 3.3. Table of mutated antigen processing and presentation-related genes across the 5 PDOs.

Next, I assessed the exome sequencing data for mutations or copy number alterations in antigen processing machinery (APM) genes, and the HLA-I genes themselves (HLA gene mutations evaluated through use of the POLYSOLVER suite).

No inactivating mutations were found in the shortlist of genes central to the HLA-I peptide processing and presentation process (depicted in Figure 1.2D). However, a small number of the mutated genes matched the genes from the 'KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION' and 'REACTOME_CLASS_I_MHC_MEDIATED_ANTIGEN_PROCESSING_PRESENTATION' genesets (Table 3.3). CRC-01 harboured a p.E854G mutation in *CIITA*. *CIITA* is only involved in HLA-II presentation, and data from UniProt showed this mutation site did not lie within a functional domain and the amino acid at that position was not evolutionarily conserved. The Catalogue of Somatic Mutations in Cancer (COSMIC) database showed only one other mutation at this site was a p.E854K mutation which was predicted to be of neutral consequence (0.26 on FATHMM prediction score). Therefore, this mutation was unlikely to have a significant functional impact on the CIITA protein, and the antigen presentation process as whole. The only 3 inactivating mutations found were *UBR4* and *HERC5* in CRC-

05, together with *UBE2Z* in CRC-08. UBE2Z is an E2 ubiquitin conjugating enzyme that is involved in the ubiquitination of proteins, commonly targeting them for destruction. For *UBE2Z*, the mutation was located at the very end of the protein within the disordered region. Therefore, it is likely to not have had a dramatic effect on protein function, and COSMIC recorded 0 cases of copy number loss. CRC-05 had a splicing mutation in *UBR4*, an E3 ubiquitin-protein ligase enzyme that can target N-terminal degrons, and is so termed an N-recognin. COSMIC showed 5 cases of copy number loss across different cancer types. It is highly likely that this mutation will have impacted the function of the enzyme, however, there is functional redundancy in the class of N-recognins so *UBR1*, *UBR2*, and *UBR5* can carry out the same functions (Leboeuf *et al.*, 2020; Zhang *et al.*, 2023). Therefore, the impact of this mutation on peptide generation is likely to have been small. HERC5 is major E3 protein ligase enzyme for the ISG15 conjugation of proteins, which is not well understood. Some data suggests that ISG15 conjugation enhances MHC-I antigen presentation (Burks, Reed and Desai, 2015), but newer data suggests ISG15 does not direct its substrates for proteasomal destruction and may instead direct peptides in to the co-translational processing pathway (Held *et al.*, 2021). In CRC-05 the gene *HERC5* had a frameshift after position 137 in the protein, rendering it inactive, and COSMIC showed 20 cases of copy number loss across different cancer types. Although HERC5 was rendered inactive by the mutation, it is unlikely to have expressly affected the amount of antigen presented, as *ISG15* knockout cells do not show significantly reduced antigen presentation (Held *et al.*, 2021). The genetic aberration with the highest chance of impacting antigen presentation was the loss of heterozygosity (LOH) on chromosome 6 of CRC-05, which resulted in the complete loss of one allele across the entire HLA gene locus (Figure 1.1B). The presence of only a single allele for HLA-A, B, C was confirmed by the molecular subtyping (Table 3.4 – performed by our collaborator Professor Michal Bassani-Sternberg). The antigen presentation potential of each HLA allotype is constrained by the specificity of the anchor residues within the binding groove of each HLA allotype, therefore the reduced number of HLA-I allotypes will limit the diversity of peptide characteristics that are able to be presented. This therefore explained, at least in part, the lower number of UPs presented by CRC-05.

	CRC-01	CRC-03	CRC-04	CRC-05	CRC-08
HLA-A	31:01	02:01	03:01	32:01	03:01
	32:01	29:02	24:02	-	29:01
HLA-B	14:01	15:01	18:01	40:01	35:03
	27:05	44:03	35:08	-	44:02
HLA-C	02:02	03:04	04:01	03:04	04:01
	08:02	16:01	05:01	-	05:01

Table 3.4. Table depicting the HLA typing of PDOs.

After assessing the intactness of APM genes, I went on to compare their expression levels, to understand if this could provide the explanation for varying immunopeptidome diversity (Figure 1.2D). It should be noted that whilst the cells used for transcriptomics were harvested from the same cell stock as used for the immunopeptidomics analyses, and so best represent their transcriptome, a second batch of transcriptomics was performed in Section 3.2.6 and the untreated cells had a slightly different result for APM and PLC gene expression. The most highly expressed gene across the 5 PDOs was *B2M*, the non-membrane integral component of HLA-I. Conversely, the most weakly expressed gene was *PSMB10*, an immunoproteasome subunit with tryptic-like cleavage specificity. Across the PDO lines, the log2 expression of APM genes was similar, showing no clear differential expression to explain the difference in immunopeptidome complexity.

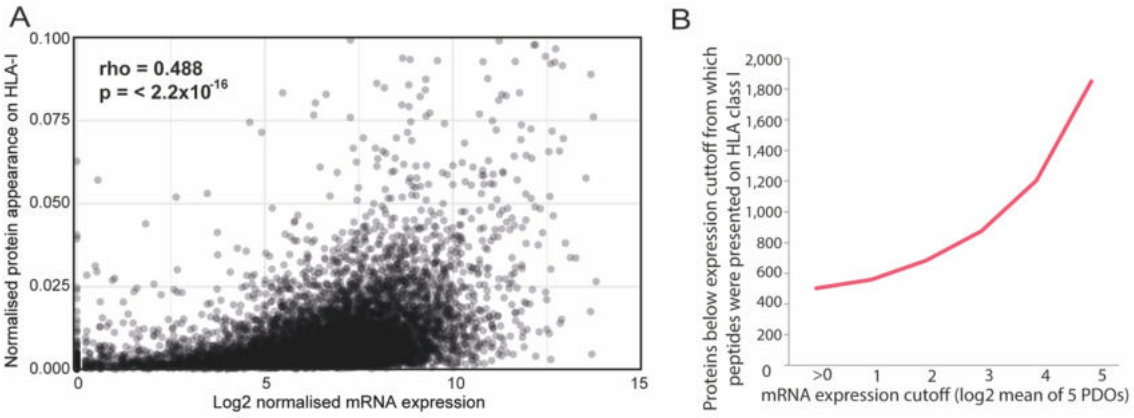


Figure 1.3: The relationship between mRNA expression and surface HLA-I peptide presentation. A: Analysis of normalised peptide appearance on HLA-I

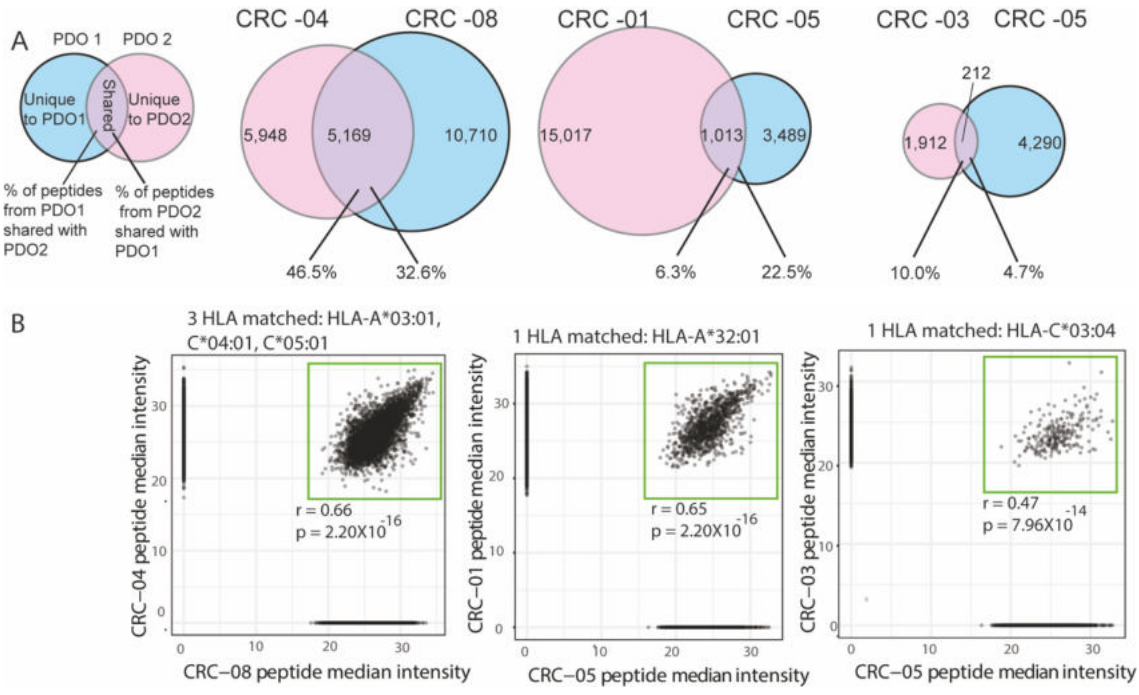
compared against log2 normalised gene expression. **B:** Number of source proteins that generated MS-detected peptides expressed at, or below the indicated log2 mRNA expression value.

In previous studies, gene expression levels have been statistically associated with levels of HLA-I peptide presentation (Leone *et al.*, 2013; Abelin *et al.*, 2017). Plotting the normalised peptide appearance of each gene (calculated as a number of unique MS-detectable peptides detected from each protein across 5 PDOs, divided by source protein length) against mean mRNA expression showed a similar, but small, link in our PDOs (Figure 1.3A). This suggested that the level of mRNA expression is a poor predictor of the extent of representation of a source protein as a peptide on HLA-I. For this reason, I then went on to investigate whether there is a minimum mRNA expression cut-off, below which peptides from a protein are not detected on surface HLA-I. Out of the 13,761 genes that were detectably expressed across the 5 PDOs, 61.5% of the genes (8,464 genes) were able to generate a MS-detectable HLA-I peptide. However, 502 genes which were not detectably expressed were also represented in the HLA-I immunopeptidome. This could be due to RNA expression levels below the detection limit of the RNA sequencing assay, or due to erroneously identified peptide sequences (possible due to the allowed FDR of 1%) which could lead to assignment of a peptide to the wrong protein. Increasing the log2 mRNA expression cut-off, drastically increases the number of genes expressed below this cut-off, from which peptides were presented (Figure 1.3B). This suggests a simple binary classification of genes into those that are, or are not detectably expressed at the mRNA level as the most useful metric to predict which proteins can be represented on HLA-I on the cell surface.

I next investigated the effect of the associated HLA-I allotype on peptide presentation by looking at PDOs which shared HLA alleles. 4-digit HLA typing of the PDOs was performed with the TruSight HLA v2 panel on the Illumina MiniSeq. These encoded HLA allotypes and the MS-detected peptides were input in to the NetMHCpan4.0 prediction algorithm, which was used to assign peptides to the HLA allotype for which they were predicted to have the lowest % rank. CRC-04 and CRC-08 were both found to express *HLA-A*03:01*, *HLA-C*04:01*, and *HLA-*

*C*05:01*. 23.7% of all the MS-detected peptides were identical between the two PDOs, constituting 46.5% of all peptides displayed on HLA-I of CRC-04, matching CRC-08 (Figure 1.4A). CRC-01 and CRC-05 only shared a single allele, *HLA-A*32:01*. 22.5% all the MS-detected peptides detected from CRC-05 (the PDO with only a single HLA-A, B, and C allele) were identical to the peptides found in CRC-01. In contrast, CRC-03 and CRC-05 shared a single HLA-C allele (*HLA-C*03:04*) and only a maximum of 10% of all the peptides from a single PDO were shared between the PDOs. This data indicated that the HLA specificity to particular anchor residues strongly constrains the immunopeptidome of cells, and the more HLA allotypes different cells have in common, the more highly similar their immunopeptidomes will be.

Following on from this, I tested to see if there was any similarity between the median intensities of peptides in PDOs with matched HLA allotypes (Figure 1.4B). Each plot shows a high level of correlation between the matched peptide intensities (a mean r value of 0.59). This suggested that the properties of the peptide (which may affect peptide processing or its quantification by MS); the HLA allotype to which the peptides bind; and therefore, the affinity of the peptide for the allotype, affects the peptide abundance on HLA-I on the cell surface.



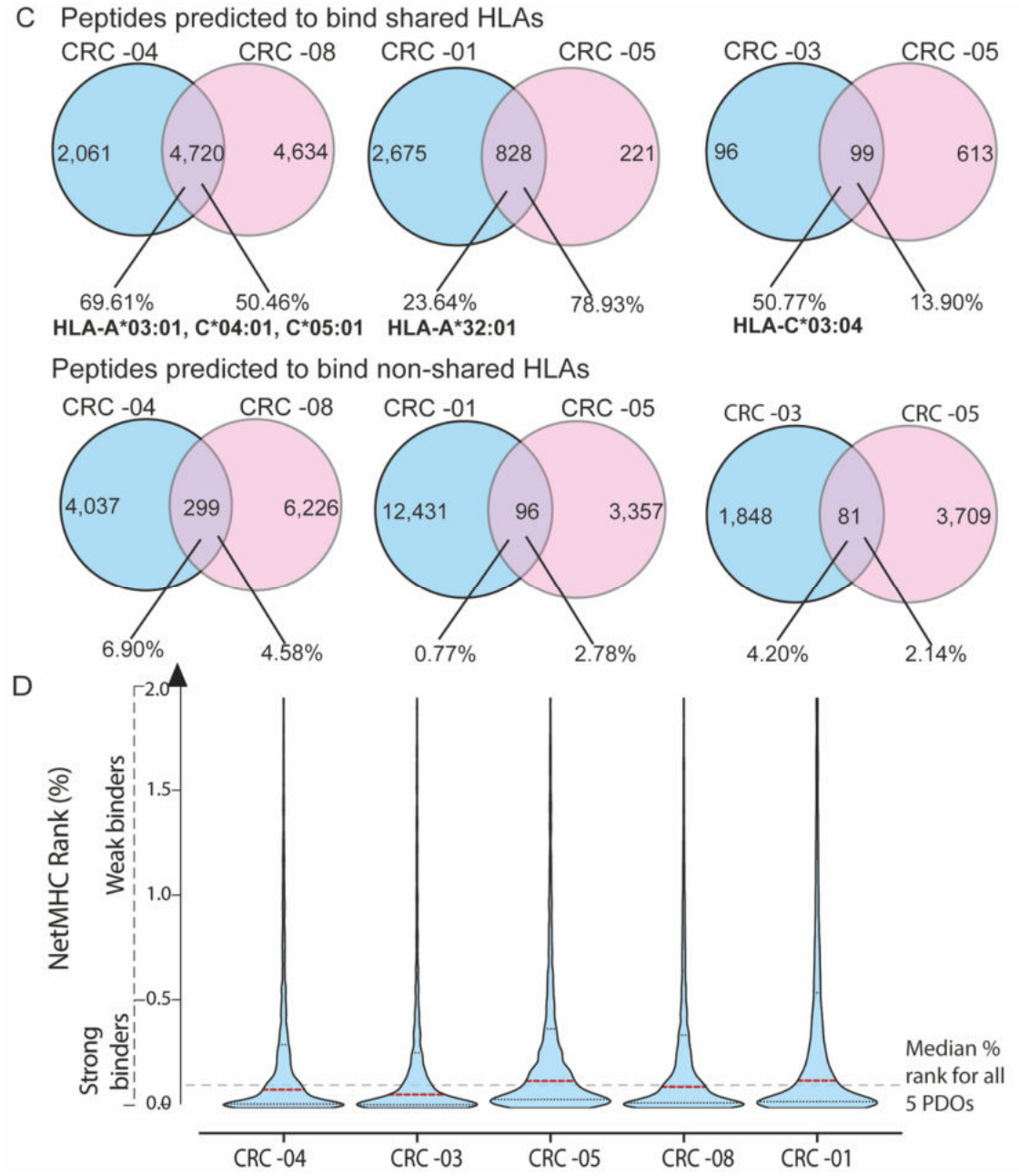


Figure 1.4: HLA-I-binding predictions by NetMHCpan4.0. **A:** Venn diagrams showing the numbers of matching and non-matching peptides between PDO pairs which share the indicated allele(s). **B:** Dot plot of the normalised intensities of peptides in PDOs that share the indicated HLA-I allele(s). **C:** Venn diagrams showing the concordance and discordance of peptides attributed to matching HLA between PDOs, as predicted by NetMHCpan4.0, and of peptides attributed to non-matching HLA between PDOs. **D:** Violin plot of lowest NetMHCpan4.0-predicted binding rank for their cognate HLA-I.

I next used NetMHCpan4.0 HLA binding predictions to subset peptides in to their predicted HLA-I allotypes, so I could expressly assess the overlap of peptides

predicted to bind shared and non-shared HLA-I allotypes (Figure 1.4C). In all cases, the percentage overlap of peptides increased when comparing the matching HLA-subset peptides with the total peptide population. 13.9-78.93% (mean: 47.89%) of the peptide population from the matching HLAs were shared between PDOs with matching HLA allotypes, whereas only 0.77-6.90% (mean: 3.56%) of the peptide population from the non-matching HLA were shared between PDOs. Major factors influencing the level of similarity of the immunopeptidomes are therefore the number of HLA allotypes in common, the expression level of that allotype (because it is known that HLA-A is expressed to a higher level than HLA-C for example) (Apps *et al.*, 2015), and the size of the immunopeptidome. This confirms that HLA allotypes are therefore a major determinate in peptide presentation in cells of the same cancer type.

I next analysed the NetMHCpan4.0-predicted binding ranks of all MS-identified peptides across the PDOs to establish whether the algorithm was able to identify them as HLA-binders (Figure 1.4D). Standard settings were used for this analysis, where peptides with a predicted rank of <0.5% were defined as strong binders, and peptides with a predicted rank 0.5-2% were defined as weak binders. 78.1% of the 49,682 peptides were predicted to strongly bind to at least one HLA, and 93% of all peptides had a rank <2%, meaning they were predicted to either strongly or weakly bind to at least one HLA. The median percentile rank across the 5 PDOs was 0.1115% (with the median for each PDO ranging from 0.06650 to 0.1372). This result showed the NetMHCpan4.0 algorithm is able to accurately classify the majority of MS-detected peptides and HLA binders which also reinforced the validity of the identification of these MS-detected peptides from PDOs with the indicated HLA allotypes.

3.2.2 The basal HLA class-II immunopeptidome in 5 CRC PDOs

Figure 1.5: The HLA-II immunopeptidome across 5 PDOs. A: The number of unique HLA-II peptides detected across all biological repeats, per PDO. **B:** The number of source proteins to which peptides from panel A can be uniquely mapped. **C:** Heatmap of RNA expression of genes central in the antigen and

HLA-II peptide presentation process. **D:** Motifs elucidated through the input of MoDec analysis on the HLA-II immunopeptidomics from CRC-08 (Racle *et al.*, 2019). The percentage above the motif specifies the proportion of the HLA-II peptidome that contributed to the illustrated motif. The peptides assigned to the flat motif are peptides that are displayed on lowly expressed HLA-II allotypes, or potential contaminants. Also displayed is the known matching motif, adapted from (Racle *et al.*, 2019).

I next went on to investigate the HLA-II immunopeptidome on our 5 PDO lines. As mentioned above, HLA-II peptides were isolated in a sequential manner, being isolated from the flow-through from the HLA-I complex capture. This is important because it means the HLA-II results are directly representative of the HLA-II immunopeptidome in the same cells from which the HLA-I results were obtained. For the analysis, only peptides >=12aa long were considered as HLA-II binders.

Although HLA-II are most commonly expressed on professional antigen presenting cells (APCs), this analysis was important because HLA-II expression has been shown on 23% of CRCs, and expression in those patients proved to be a marker of a favourable prognosis (Sconocchia *et al.*, 2015).

MS analysis only revealed 6-24 UPs in CRC-01, CRC-03, and CRC-05, whereas 392 UPs from 140 source proteins and 713 UPs from 247 source proteins were identified in CRC-04 and CRC-08 respectively (Figure 1.5A-B). This confirmed that the HLA-II peptidome of unperturbed cells is much smaller than that on HLA-I. Even in the PDO lines CRC-04 and CRC-08 with more peptides, HLA-II surface expression for all 5 PDOs was below the detection limit of flow cytometric QIFIKIT analysis. RNA expression analysis showed only one PDO to detectably express the class II transactivator protein (CIITA – the master regulator of HLA-II expression), which did not correlate with an elevated HLA-II peptide number, and RFX5 (another known activator of HLA-II transcription) did not appear to be differentially expressed between the PDOs. However, low level HLA-II expression was detected in both the PDOs with a higher level of HLA-II peptides, and a complete absence in the other PDOs. HLA-II expression in the absence of detectable CIITA expression may be explained by the limited sensitivity in the

chosen RNA sequencing assay, or poor transcriptional control in the CRC PDOs (Bradner, Hnisz and Young, 2017).

Next, I utilised the MoDec platform to perform motif deconvolution on HLA-II datasets with >= 500 peptides which were 12aa or longer. Analysing the binding motifs of the MS-detected peptides was executed to confirm their legitimacy as HLA-II binding peptides. Only CRC-08 presented enough peptides to formulate a motif. 85% of the detected peptides contributed to the resultant motif, which was highly similar to the known HLA-II motif of DRB1*04:04, supporting the identification of peptides as genuine HLA-II binders. This is also supported by the detectable RNA expression of HLA DRB1 in CRC-08. Expression of HLA-II molecules, and the corresponding peptide presentation were therefore found to be limited in our CRC PDOs, perhaps even lower than within the tumour microenvironment where cells may be exposed to IFNγ-producing immune cells (Cioni *et al.*, 2019; Leivonen *et al.*, 2019).

3.2.3 Assessing mutation load and neoantigen presentation in 5 CRC PDOs

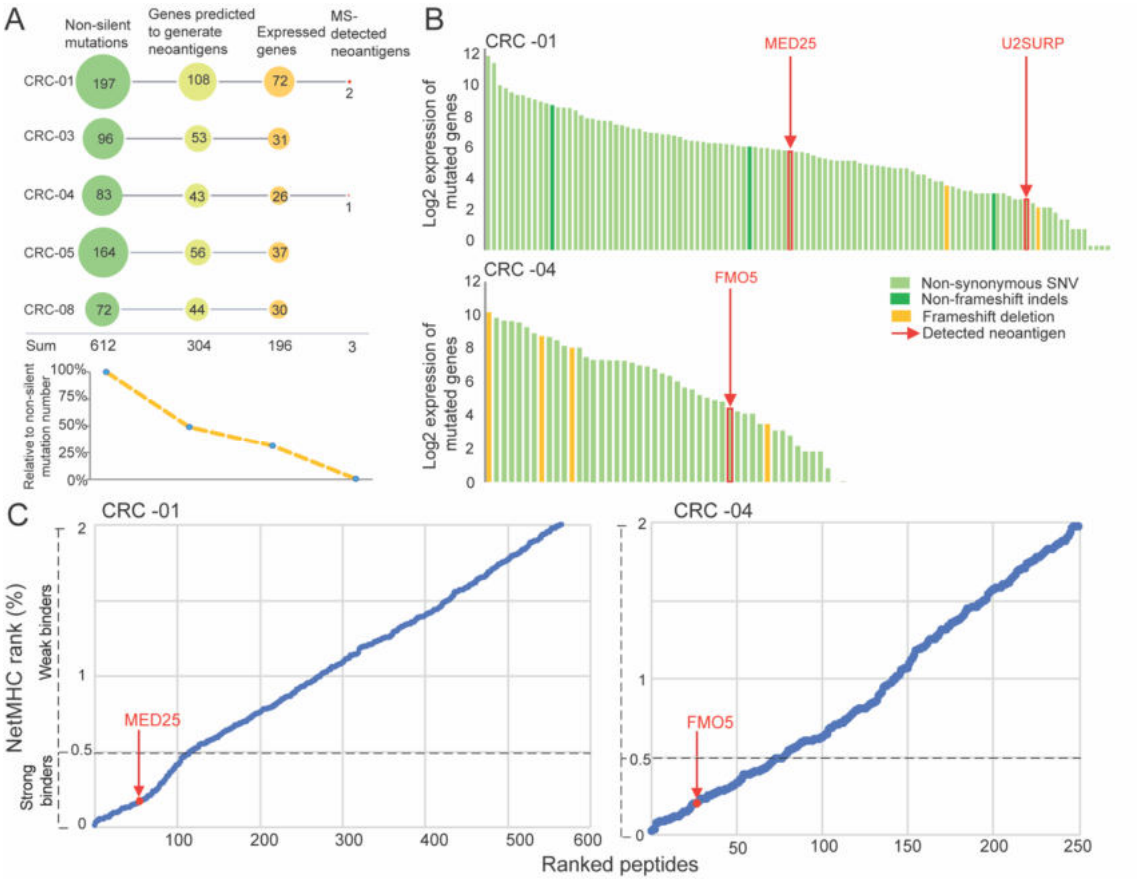


Figure 1.6: The disparity between MS-detected and predicted neoantigens in 5 PDOs. A: Number of mutations that encode for changes in amino acid composition (missense, frame-shift and stop-loss mutations), genes predicted to generate strong binders (as predicted by NetMHCpan4.0, defined as having a percentile rank < 0.5), and number of strong binder-candidate genes that are expressed, compared against the number of MS-detected neoantigens. **B:** Log2 gene expression of all non-silent gene mutations (depicting non-synonymous point mutations, non-frameshift indels, and frameshift deletions). The genes from which neoantigens were derived were outlined in red. **C:** NetMHCpan4.0 HLA percentile ranks of all strong and weak HLA-binding predicted neoantigens, ordered from the lowest to the highest rank, with the data point for the MS-detected neoantigens highlighted in red.

The analysis of the canonical, non-mutated immunopeptidome revealed the immunopeptidomics approach was successful in achieving good sensitivity and depth. I next assessed whether the MS immunopeptidomics approach could

elucidate neoantigens encoded by somatic mutations within our PDOs. For analysis of neoantigens, non-silent substitutions and indels were considered, of which there were 612 across the 5 PDOs. To increase the sensitivity of the identification of neoantigens from the mass spectra, a relaxed FDR of 5% was used, as described (Chong *et al.*, 2017). Despite the depth of the non-mutated immunopeptidome, only 3 neoantigens were detected across the 5 PDOs (Figure 1.6A). This represented 0.49% of all the non-synonymous mutations. All neoantigens were derived from missense mutations and none were derived from the 33 frameshift mutations. On CRC-01, the PDO with the highest mutation load, only 2 HLA-I-presented mutation-neoantigens were detected: one 8-mer derived from a K422T mutation in the *MED25* gene, and one 11-mer from a T224R mutation in *U2SURP*. The only other neoantigen detected was found from CRC-04: a 10-mer from a S423N mutation in *FMO5* (Table 3.5). MS analysis detected no neoantigens presented on HLA-II. Analysis of the mRNA expression values of all the mutated genes showed the neoantigen-source genes to be either moderately or lowly expressed (Figure 1.5B), compared against the other mutated genes that were not represented on HLA-I, illustrating that neoantigen presence cannot be explained by higher mRNA expression.

PDO	Source gene	Peptide length (amino acids)	Mutation	Neoantigen	Lowest NetMHC rank (%) WT	Lowest NetMHC rank (%) neo-antigen
CRC-01	*U2SURP*	11	T224R	IQEERDERHKR	75.8495	5.7765
CRC-01	*MED25*	8	K422T	SVDANTTL	0.5336	0.1586
CRC-04	*FMO5*	10	S423N	RYVENQRHTI	0.31911	0.2692

Table 3.5. Peptide properties of the MS-detected neoantigens.

No non-mutated peptides from the wildtype alleles of the mutated proteins were detected. Therefore, I went on to compare the NetMHCpan4.0-predicted percentile ranks of the MS-detected peptides with the percentile ranks of the NetMHCpan4.0-predicted neoantigens (Table 3.5). Only the *MED25* and *FMO5* neoantigens were predicted to be binders of the PDO's HLA-I allotypes, with

ranks of 0.16% and 0.27% respectively, positioning them in the top 1/3 of all predicted neoantigens (Figure 1.6C). For all 3 neoantigens, the mutation improved the predicted percentile rank, changing the weaker binding peptide to a stronger binding peptide - turning the *MED25* neoantigen from a weak binder to a strong binder, *FMO5* improved rank by 0.15-fold, and *U2SURP* improved by 13.13-fold. Some studies have shown that neoantigens that have increased affinity for HLA-I as opposed to their WT equivalent may be predicted to be more immunogenic (Balachandran *et al.*, 2017; Luksza *et al.*, 2017), especially if it leads to the presentation of a not before-encountered epitope (Duan *et al.*, 2014). However, I was not able to test this myself due to the lack of autologous PBMCs.

To understand why MS immunopeptidomics was only able to reveal 3 neoantigens, I wanted to investigate what limits the production of neoantigens. First, I analysed the NetMHCpan4.0 predictions in order to see how many of the non-synonymous mutations were predicted to generate neoantigens, predicted as described (Woolston *et al.*, 2019). 49.67% of the mutations were predicted to generate at least one strong binder of HLA-I (possessing a percentile rank <0.5%). CRC-05, which showed LOH of HLA locus, had the smallest percentage of predicted strong-binding neoantigens from its mutated genes, only 34.14% of the non-synonymous mutations were predicted to generate a strong HLA-binding neoantigens, compared amongst a mean of 55.74% in the other PDOs. This is likely due to the constraints on peptide binding imposed on the immunopeptidome of CRC-05 by its reduced diversity in HLA-I. Having only 3 alleles resulted in a reduced number of possible peptide binding motifs on the cell surface. Therefore, this data suggests that in cells with a full panel of 6 HLA-I, just under half of the mutated genes are not predicted to generate neoantigens, providing one source of limitation to HLA-I presentation.

Next, as the data in Figure 1.3B showed, undetectable RNA expression was a good predictor that a specific protein is unlikely to be presented on HLA-I. Owing to this, I discounted any genes without detectable RNA expression from the counts. This further reduced the number of possible mutated genes from which HLA-I-presentable neoantigens could be predicted to be derived to 196/612 (32.03%) of all mutated genes (Figure 1.6A). Therefore, only 2/196 (1.02%) of

the expressed genes computationally predicted to generate neoantigens were actually detected, alongside 1 neoantigen that was not predicted to be a good binder of HLA-I. Together, this shows that even with using additional selection criteria to reduce the number of viable neoantigen source genes, a very small proportion of these mutations are represented on HLA-I, even with the use of high-sensitivity MS. This highlights the low neoantigen load in MMRp CRC which limits the potential success of immune checkpoint inhibition treatment, and illustrates the need for further selection criteria to enable more accurate predictions of neoantigen presentation.

3.2.4 Assessing the effects of IFNγ exposure on the immunopeptidomes of 4 CRC PDOs

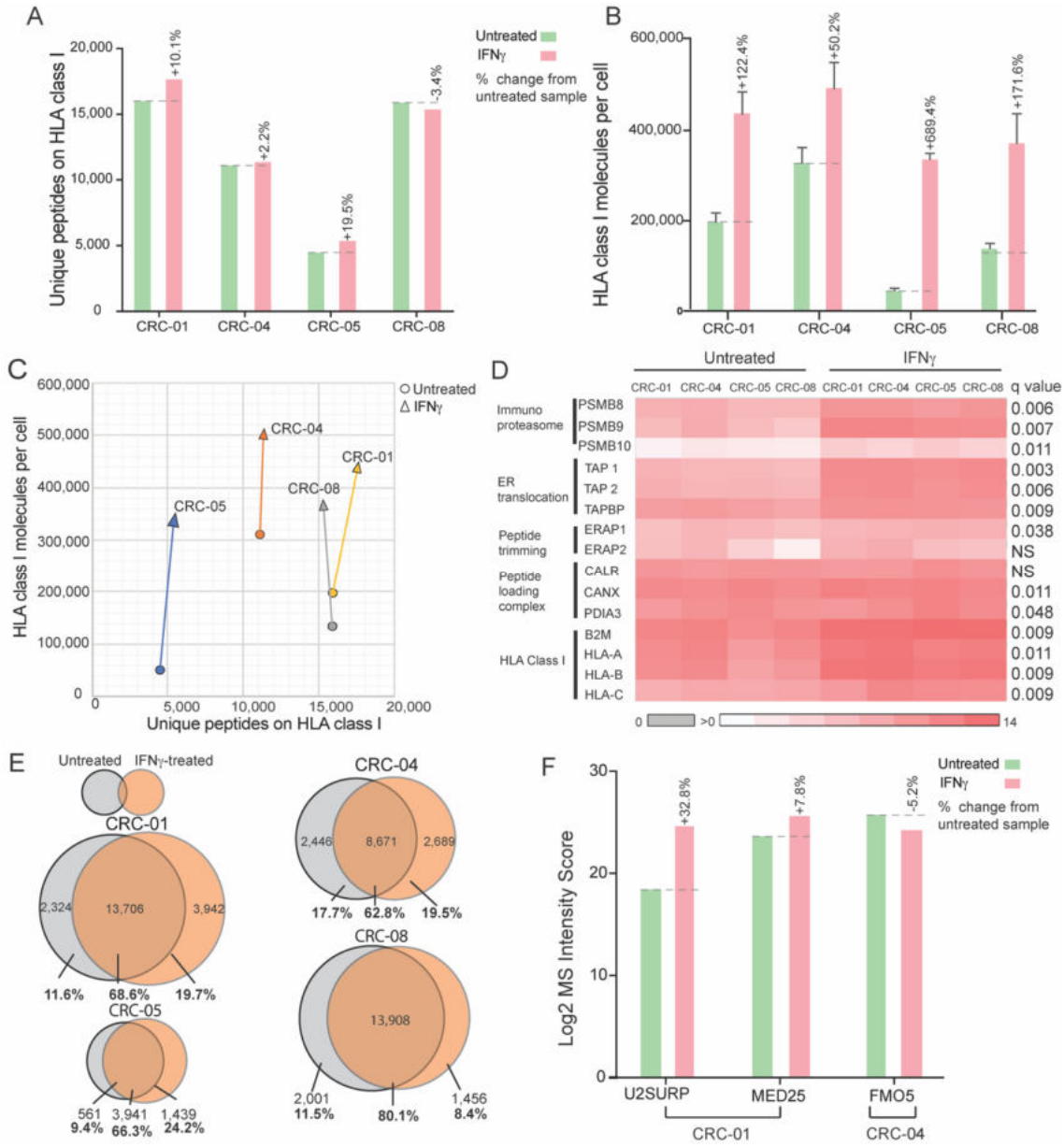


Figure 1.7: The effects of IFNγ treatment (600ng/mL) on HLA-I peptide presentation in 4 PDOs. A: Number of unique HLA-I peptides detected across all biological repeats, per PDO in the presence or absence of IFNγ. **B:** Flow cytometric quantification of HLA-I molecule number, in the presence or absence of IFNγ (bars display mean value of 4 biological replicates, the error bars display the SEM). **C:** Change in peptide diversity and HLA-I surface abundance in the presence or absence of IFNγ. **D:** Heatmap of mRNA expression of genes central in the antigen and HLA-I peptide presentation process, in the presence or absence of IFNγ Significant different analysis performed with paired t-test,

followed by FDR analysis set to 5%. **E:** Proportionally scaled Venn diagrams comparing the peptides detected in untreated and IFNγ-treated cells. **F:** Log2 MS intensities of the MS-detected neoantigens in the presence or absence of IFNγ.

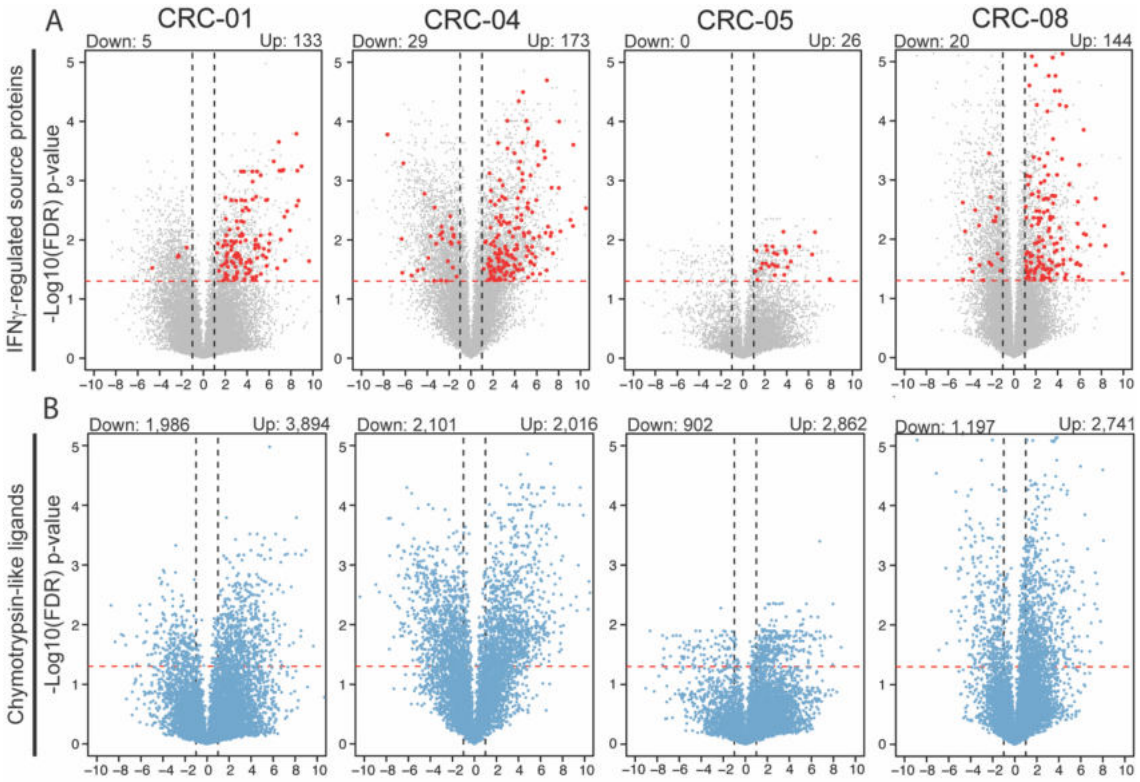
The immunopeptidomics analysis of the IFNγ-treated PDOs revealed only modest changes in unique peptide number (Figure 1.7A), with the largest increase of +19.5% seen in CRC-05, and, in contrast, a slight decrease in CRC-08 (-3.4%). These modest increases in peptide diversity were at odds with the larger changes in HLA-I surface abundance (Figures 1.7B-C), with the mean increase in abundance at 258.4%. The biggest increase was seen in the Chr6 LOH PDO, CRC-05, which increased 689.4%. The discrepancy between HLA-I surface abundance and unique peptide number was particularly surprising in CRC-08 which experienced a decreased peptide diversity in spite of a 171.6% increase in HLA-I abundance (Figure 1.7C). Furthermore, after IFNγ treatment, the number of HLA-I surface molecules rose to a more similar number (332,028 - 491,269 molecules), irrespective of the number of molecules present on the cell surface in untreated samples.

Transcriptomic analysis of the untreated and IFNγ-treated PDOs confirmed the IFNγ response of each PDO was intact, significantly increasing all selected APM genes and HLA-I genes ($q < 0.05$), aside from ERAP2 and calreticulin (Figure 1.7D). mRNA expression of immunoproteasome genes increased 2.77-5.08-fold, which again was at odds with the small changes in peptide diversity. This however did reassert that the modest increase in unique peptide number across the 4 PDOs was not due to a lack of IFNγ insensitivity.

However, differential peptide analysis revealed there were changes in which peptides were presented (Figure 1.7E). On average, only 69.45% of the peptides were shared between the untreated and IFNγ-treated samples, leaving certain peptides being exclusively present in only one of the conditions. MS intensities of the peptides were also up- or downregulated by IFNγ treatment; a mean of 1,371 peptides were upregulated at least 2-fold, and 1,169 peptides were downregulated at least 2-fold.

I next wanted to analyse the effects of IFNγ treatment on MS intensity of the detected neoantigens. One key finding was that all 3 neoantigens were detected in both untreated and IFNγ-treated conditions, and across the 4 PDOs IFNγ treatment did not reveal any additional neoantigens. This was despite the transcriptional increases in APM and HLA-I genes, and expected shifts in proteasome specificity.

To better assess whether there was an identifiable change in the dominant catalytic activity of the proteasome due to the increased mRNA expression of immunoproteasome components through IFNγ exposure, I next analysed the characteristics of all up- or down-regulated peptides (>=2-fold or <=0.5-fold change respectively) in each PDO line. The exchange of constitutive proteasome catalytic subunits for immunoproteasome subunits under IFNγ treatment, is supposed to increase in chymotryptic-like activity compared when compared against the constitutive proteasome (Mishto *et al.*, 2014; Chong *et al.*, 2017).



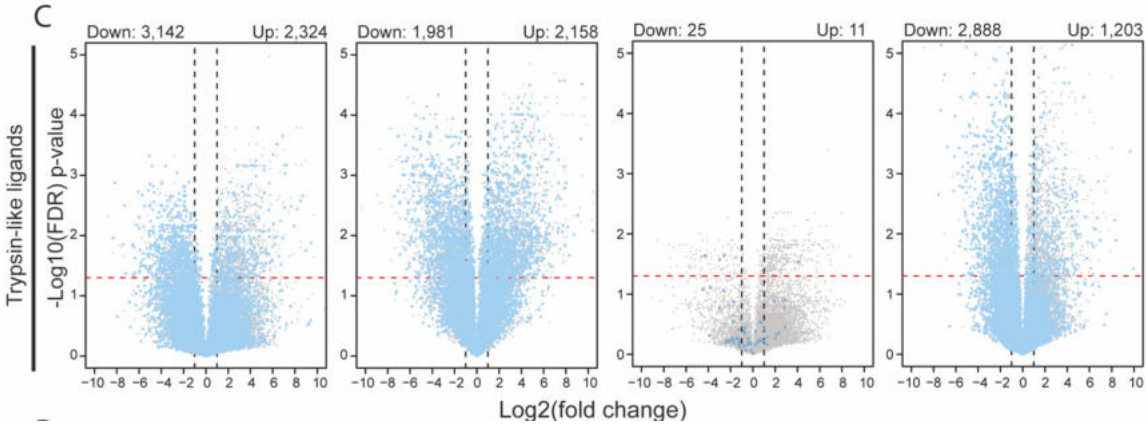


Figure 1.8: Changes in proteasome specificity induced by IFNγ treatment. **A:** Volcano plots showing the fold change of normalized peptide abundance with IFNγ treatment. Known IFNγ-inducible genes which showed a statistically significant (q<0.05) fold change, above +/-2 are highlighted in red. **B:** Volcano plots showing the fold change of normalized peptide abundance with IFNγ treatment. Chymotryptic-like ligands were highlighted in dark blue (defined as the C-terminal amino acid consisting of “A”, “F”, “I”, “L”, “M”, “V”, or “Y”). Numbers of peptides which show a statistically significant fold change above +/-2 are indicated above the volcano plots. **C:** Volcano plots showing the fold change of normalized peptide abundance with IFNγ treatment. Tryptic-like ligands were highlighted in light blue (defined as the C-terminal amino acid consisting of “K” or “R”). Numbers of peptides which show a statistically significant fold change above +/-2 are indicated above the volcano plots.

To see if any IFNγ-response genes were represented in the upregulated peptides I counted the number of significantly upregulated genes present in the HALLMARK_INTERFERON_GAMMA_RESPONSE geneset (Liberzon *et al.*, 2015) – a mean of 119 IFNγ-response genes was found upregulated, and 13 downregulated (Figure 1.8A). This provides further support that each PDO had an intact IFNγ response. Next, I analysed the characteristics of C-terminus for all up- or down-regulated peptides to see whether there was the expected increase in chymotryptic activity and corresponding decrease in tryptic activity of the proteasome. A mean of 2,878 peptides with a C-terminus generated by chymotryptic-like cleavage activity increased in intensity, whereas 1,547 peptides decreased in intensity (Figure 1.8B). This demonstrated that peptides with a C-terminus generated by chymotryptic-like activity are 1.86-times more likely to

increase in intensity in cells exposed to IFNγ. Conversely, a mean of 1,424 peptides with a C-terminus generated by tryptic-like cleavage activity increased in intensity, whereas 2,009 decreased in intensity (Figure 1.8C). This means peptides with a C-terminus generated by tryptic-like cleavage activity are more likely to decrease in abundance under IFNγ treatment. The exception being CRC-04, which displayed a small increase in the abundance of tryptic-like ligands, which could not be explained by the mRNA expression of proteasome components. Furthermore, CRC-05 had very few tryptic peptides at all, which also could not be explained by mRNA expression analysis. Taken together, this data suggested that the changes in proteasome activity caused by IFNγ treatment, only influenced small changes in the HLA-I peptidome.

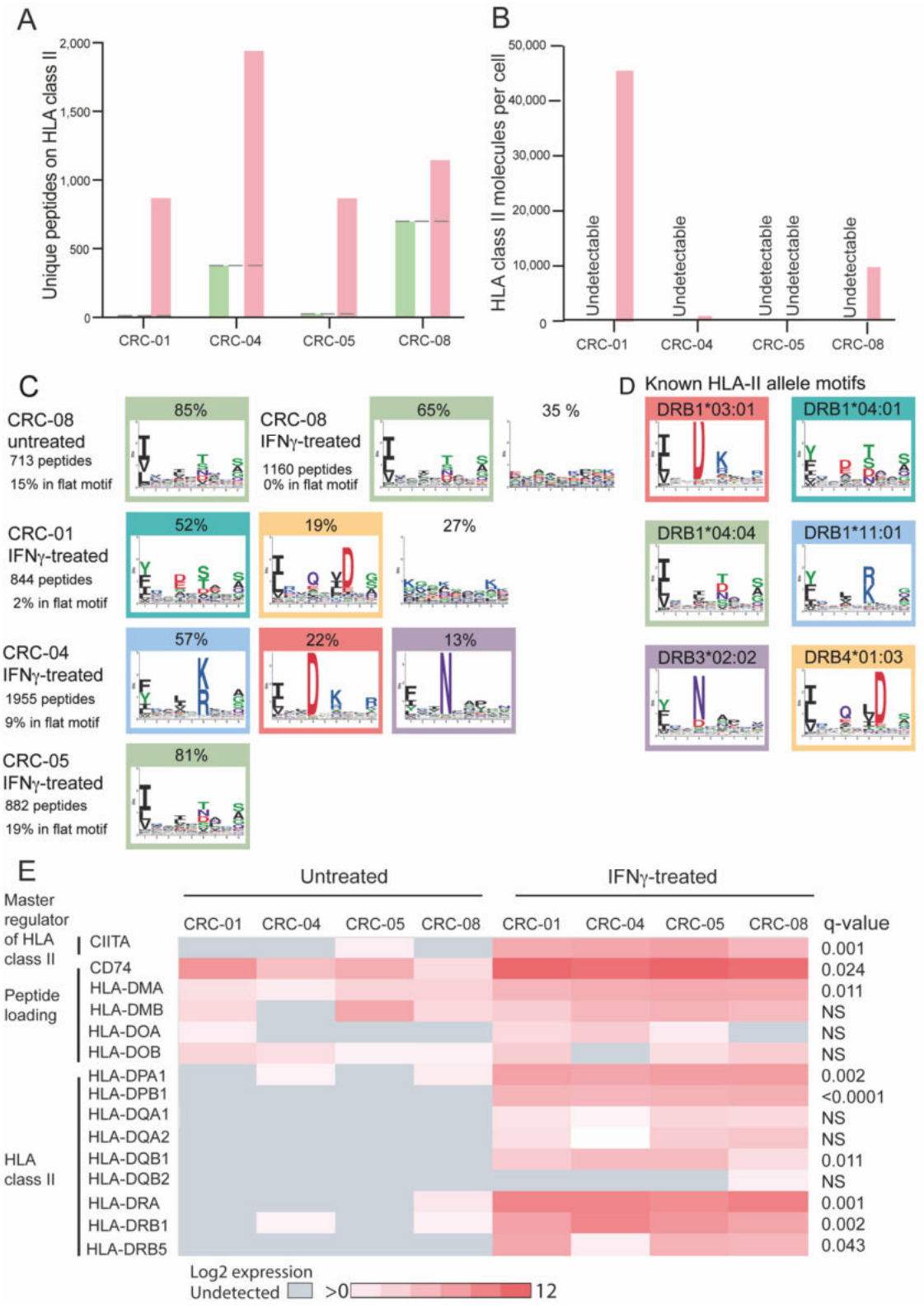


Figure 1.9: The effects of IFNγ treatment (600ng/mL) on HLA-II peptide presentation in 4 PDOs. A: Number of unique HLA-II peptides per PDO in the presence or absence of IFNγ detected across all biological repeats. **B:** Flow cytometric quantification of HLA-II molecule number, in the presence or absence of IFNγ (bars display mean value of 2 biological replicates). **C:** Motifs determined

through the input of all HLA-II peptidomes >500aa to MoDec analysis (Racle *et al.*, 2019). The percentage above the motif specifies the proportion of the HLA-II peptidome that contributed to the illustrated motif. The peptides assigned to the flat motif are peptides that are displayed on lowly expressed HLA-II allotypes, or potential contaminants. **D:** Known matching motifs, adapted from (Racle *et al.*, 2019). The coloured boxes indicate highly similar motifs between the MS-detected peptides and a known HLA-II motif. **E:** Heatmap of mRNA expression of genes central in the HLA-II peptide presentation process, in the presence or absence of IFNγ. Significant different analysis performed with paired t-test, followed by FDR analysis set to 5%.

I next analysed whether treatment with IFNγ could cause an increase in the size of the HLA-II peptidome and facilitate the presentation of novel neoantigens on HLA-II. For every PDO, there was a large increase in the number of unique MS-detected HLA-II peptides, reaching 882-1,955 peptides >= 12aa (Figure 1.9A). This was paired with a detectable increase in surface HLA-II abundance in 3/4 PDOs, however, the number of HLA-II still remained below the detection limit for CRC-05, despite an increase in MS-detected HLA-II peptides (Figure 1.9B). Furthermore, CRC-04 experienced a much smaller increase in surface HLA-II abundance, compared against a much larger increase in unique HLA-II peptides. Most of the peptides formed distinct motifs (Figure 1.9C), and could be attributed to known HLA-II allotypes (Figure 1.9D). These changes were accompanied with a significant increase in CIITA expression (Figure 1.9E, q = 0.001) and increases in HLA-II genes. 4/6 of the known HLA-II to which MS-detected peptides could be attributed were HLA-DRB1 allotypes, which significantly increased in expression (Figure 1.9E, q = 0.002), and the other 2 HLA-II were HLA-DRB3 and HLA-DRB4 which did not show detectable mRNA expression in untreated or IFNγ-treated conditions. This may again be explained by mRNA expression below the detection limit of the mRNA sequencing assay. Furthermore, despite the large increases in HLA-II peptidomes, no new neoantigens were discovered on HLA-II following IFNγ exposure.

3.2.5 Assessing the effects of MEK inhibition on the immunopeptidomes of 4 CRC PDOs

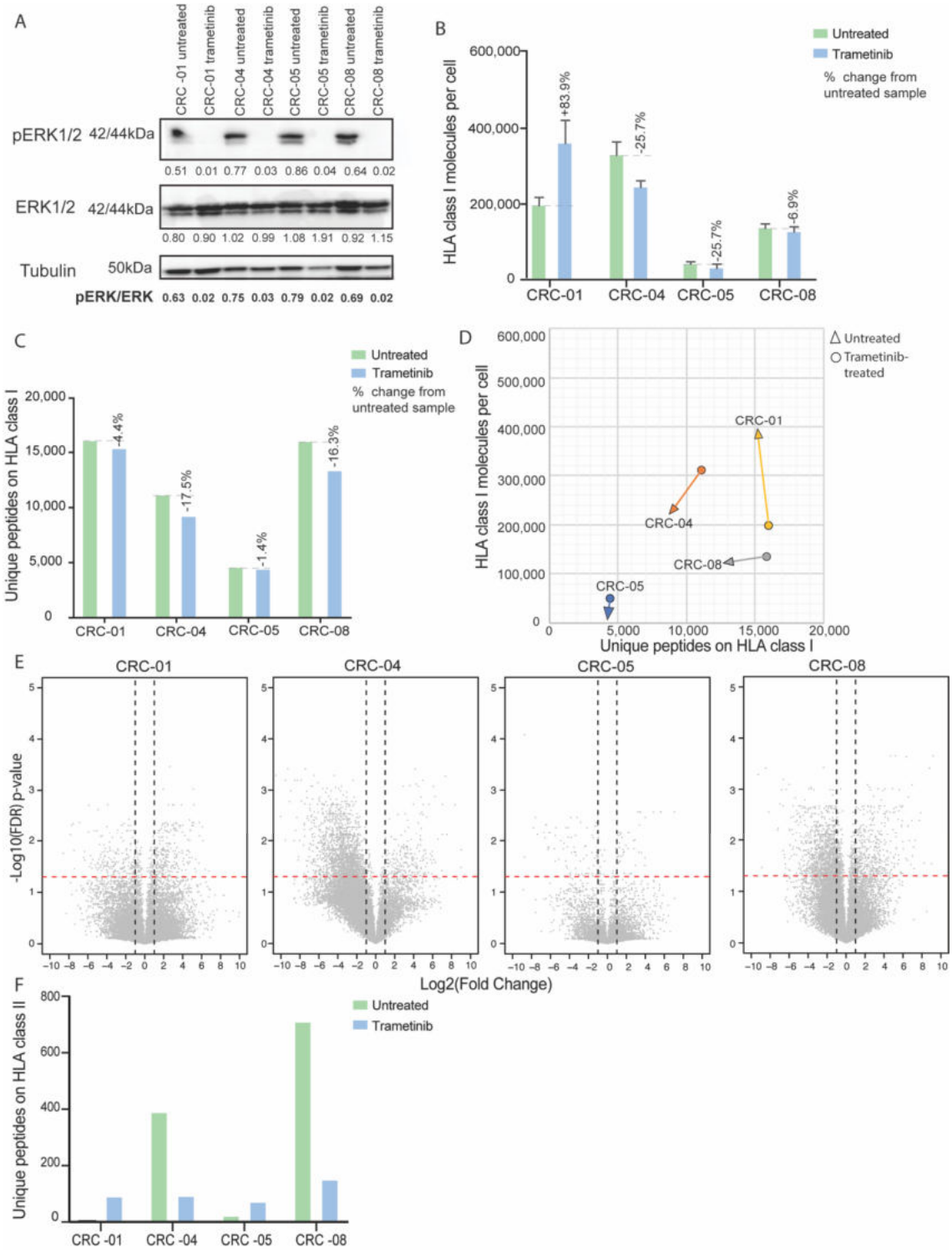


Figure 1.10: The effects of trametinib treatment (30nM) on HLA-I peptide presentation in 4 PDOs. A: Western blot showing inhibition of ERK phosphorylation (pERK) through trametinib, with tubulin as the illustrative loading control. Densitometry readings from ImageJ analysis, normalised by band intensity of tubulin loading control. pERK/ERK ratio shown below. **B:** Flow

cytometric quantification of HLA-I molecule number, in the presence or absence of IFNγ (bars display mean value of 4 biological replicates, the error bars display the SEM). **C:** Number of unique HLA-I peptides in the untreated or trametinib-treated PDOs, detected across all biological repeats. **D:** Change in peptide diversity and HLA-I surface abundance in the absence or presence of trametinib. **E:** Volcano plots showing the fold change of normalized peptide abundance with trametinib treatment. **F:** Number of unique HLA-II peptides per PDO in the presence or absence of IFNγ detected across all biological repeats.

Trametinib is an inhibitor of the mitogen-activated protein kinase (MAPK) pathway, which functions by inhibition of MEK1/2. Data has shown that MEK inhibitor treatment can increase HLA expression, and could significantly improve tumour growth inhibition when combined with immune checkpoint inhibiting antibodies (Liu *et al.*, 2015). Based off this data, I wanted to ascertain whether MEK inhibition with trametinib could increase immunopeptidome diversity and promote the presentation of novel neoantigens. The 4 PDO lines were treated with 30nM of trametinib for 48 hours. This treatment concentration and duration resulted in strong inhibition of phosphorylation of the downstream effector protein ERK in all 4 PDOs (Figure 1.10A), yet despite this there was no consistent effect on HLA-I surface abundance (Figure 1.10B). HLA-I surface abundance increased in CRC-01 (+83.9%), yet decreased in the other 3 PDOs (-6.9-25.7%). Furthermore, trametinib treatment decreased the number of unique peptides presented across all 4 PDOs (Figure 1.10C-D). CRC-04 showed the strongest decrease in peptide diversity (-17.5%) and the strongest downregulation of peptide abundance at the cell surface (Figure 1.10E). There was a small increase in peptide abundance in CRC-01 (Figure 1.10E), which could to explain how such increased HLA-I surface abundance was not able to increase peptide diversity, as some peptides increased more in relative abundance. No additional neoantigens were detected from trametinib treatment.

Trametinib treatment also had an inconsistent effect on the number of unique HLA-II peptides, slightly increasing the number of HLA-II peptides presented on the PDOs with a very low HLA-II peptide number, and decreasing the number on the PDOs with a higher number (Figure 1.10F). The number of HLA-II molecules

present on the cell surface remained below the detection limit of the flow cytometric quantification assay after trametinib treatment in all PDOs. Furthermore, no additional neoantigens were detected on HLA-I or HLA-II in trametinib-treated PDOs.

3.2.6 Optimising dose of HDAC inhibitor treatment on our 4 CRC PDOs

After learning that treatment with IFNγ and Trametinib did not succeed in increasing the number of unique neoantigens presented on the cell surface, I decided to pursue other drug treatments. A combination seen in clinic was the use of epigenetic modulating drugs histone deacetylase inhibitors (HDACi), and immune checkpoint inhibitors. Histone modifications are an important part in the dynamic remodelling of the chromatin structure. Acetylation of histones N-terminal tails on certain positions (e.g. H3K9) by histone acetyltransferases (HATs) puts the chromatin in a transcription-permissive state. This occurs as the addition of an acetyl group to a lysine residue neutralises the positive charge, which reduces the attraction to negatively charged DNA, loosening up the structure allowing access of transcription proteins. Histone deacetylases (HDACs) are lysine deacetylases that can act on any protein, but it is specifically class I HDACs that are located primarily in the nucleus and act upon on histone tails. Deacetylation of histone tails allows the chromatin structure to tighten back up in to a condensed, restricted state. Therefore, inhibition of HDACs with HDAC inhibitors allows DNA in to the transcription-permissive state and can lead to the expression of different genes.

Studies have shown that HLA-I and APM gene expression can be reduced by epigenetic silencing in cancer, and relieving this via HDAC inhibition can restore expression of APM genes and HLA-I surface expression (Ritter *et al.*, 2017; Souri *et al.*, 2020). A study of 3 uveal melanoma cell lines showed treatment with HDACi quisinostat increased HLA-A and HLA-B gene expression, and HLA-I surface expression (Souri *et al.*, 2020). Therefore, HDACi appeared as a desirable candidate for immunopeptidomics interrogation as the ability to unlock or amplify expression of certain genes, combined with the potential to increase HLA expression, could lead to increased surface expression of neoantigens.

The particular drug that was selected for this study was domatinostat, a small molecule class I HDAC inhibitor, which is active against HDAC1, HDAC2, HDAC3. Domatinostat was chosen due to its early use in the Phase Ib/II EMERGE study (NCT03812796) of oesophago-gastric carcinoma (OGA) and colorectal cancers (CRC) (Cartwright *et al.*, 2021). Furthermore, in an early study of 40 melanoma patients in combination with pembrolizumab, domatinostat treatment achieved 1 complete response, 2 partial responses, and 9 cases of stable disease. This was paired with a higher intratumoural expression of HLA and antigen processing machinery, and increased T-cell infiltration (Hassel *et al.*, 2021).

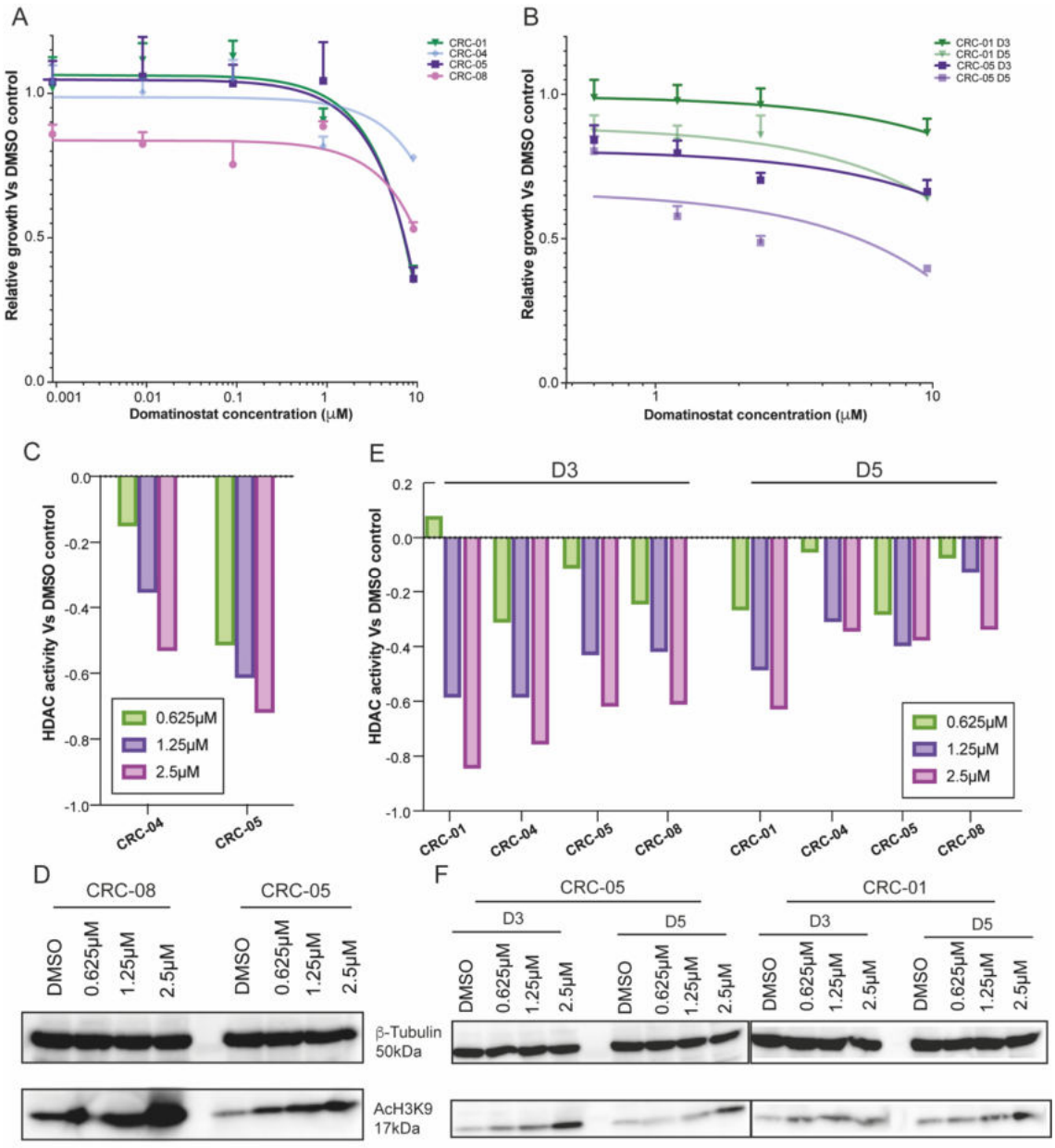


Figure 1.11. Assessing the sensitivity of CRC PDOs to HDACi domatinostat. **A:** 24H growth assay of 4 CRC PDOs treated with domatinostat, with cell viability assessed by CellTitre Blue (CTB) analysis. Data was normalised to the DMSO control, so that 1.0 represents the CTB reading of the DMSO control. Error bars illustrate the standard deviations (SD) of 6 technical replicates, and the graphs indicate a single illustrative repeat. The graphs were fit with a simple linear regression line by GraphPad Prism. **B:** 3- and 5-day growth assay of 2 CRC PDOs treated with domatinostat, with cell viability assessed by CTB analysis. Data was normalised to the DMSO control, so that 1.0 represents the CTB reading of the DMSO control. Error bars illustrate the SD of 6 technical replicates, and the graphs indicate a single illustrative repeat. The graphs were fit with a simple linear regression line by GraphPad Prism. **C:** HDAC activity assay of 2 CRC PDOs after 24H treatment with domatinostat, using the BioVision HDAC activity colorimetric assay kit. The value for the negative control was subtracted from the test values, then each condition was normalised to the DMSO control so that 1.0 represents the colorimetric values of the DMSO control. **D:** Western blot (WB) analysis of AcH3K9 in 2 CRC PDOs after 24H treatment with domatinostat. β-tubulin was additionally used as a loading control. **E:** HDAC activity assay of 4 CRC PDOs after 3- or 5-days of treatment with domatinostat, using the BioVision HDAC activity colorimetric assay kit. The value for the negative control was subtracted from the test values, then each condition was normalised to the DMSO control so that 1.0 represents the colorimetric values of the DMSO control. **F:** WB analysis of AcH3K9 in 4 CRC PDOs after 3- or 5-days of treatment with domatinostat. β-tubulin was used as a loading control.

I began by performing a drug titration to assess the sensitivity of the 4 PDOs to HDACi domatinostat over an incubation of 3 days (Figure 1.11A). The goal of this drug treatment was to diminish the HDAC functionality to enable modification of the transcriptome of the cells, but without causing a significant decrease in cell viability which could affect antigen processing and presentation. All 4 PDOs responded to domatinostat treatment in the 0.1-10μM range. To investigate a more discrete range of values for further use, I treated 2 PDO lines with 0.625μM-10μM for a period of 3 or 5 days (Figure 1.11B). The data showed even at 0.625μM domatinostat was able to slightly reduce cell growth in both PDO lines,

and the effect was undesirably strong at 10μM. Furthermore, the growth-inhibitory effect of domatinostat was far more pronounced over a 5-day incubation period, suggesting a shorter incubation period may be a beneficial mode of treatment for the immunopeptidomics study.

With the defined drug treatment range of 0.625μM-2.5μM, I tested the direct effect of domatinostat on HDAC activity after 24H of incubation (Figure 1.11C). 2.5μM of domatinostat had best efficacy, causing a -0.53 and -0.72-fold reduction in HDAC activity vs DMSO control in CRC-04 and CRC-05. This resulted in an accumulation of acetylated H3K9, strongest in the 2.5μM concentration (Figure 1.11D). The HDAC assay was then performed on all 4 PDOs after a 3-day or 5-day incubation to assess whether the inhibitory effect was sustained. In all 4 PDO lines HDAC inhibition was strongest at day 3 post drug application, not day 5, and once again strongest at 2.5μM (Figure 1.11E). This was paired with highest accumulation of acetyl-H3K9 in the 2.5μM sample at day 3 (Figure 1.11F).

All this data together asserted that the optimal domatinostat treatment was a 2.5μM dose incubated for 3 days, as this dose and duration resulted in strong HDAC inhibition, and a consequential accumulation of acetylated H3K9 which is likely to affect the transcriptome, without a strong detriment to cell viability. This was therefore the treatment applied to PDOs for RNAseq analysis. RNA library preparation was performed by myself, with help from Dr Louise Barber.

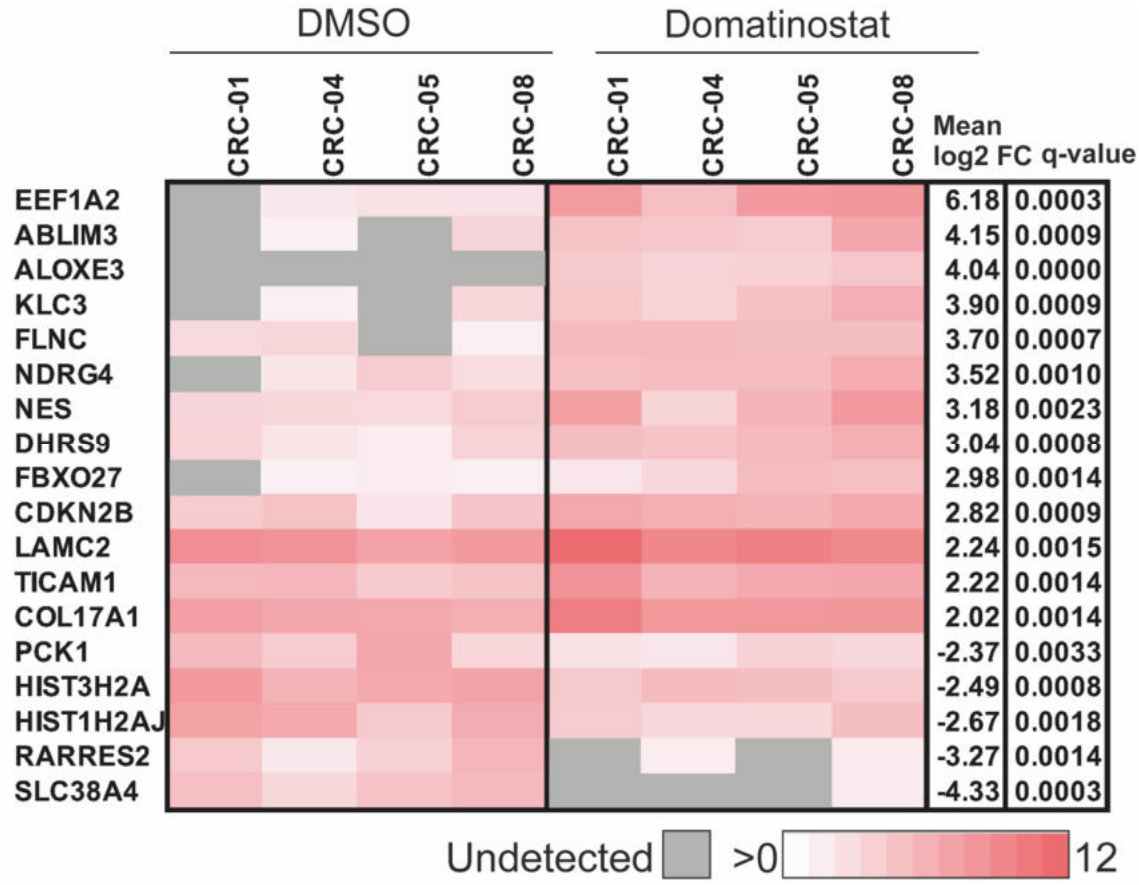


Figure 1.12. Transcriptome-wide differential expression analysis of CRC PDOs after treatment with domatinostat. Heatmap of log2 normalised 3' RNA expression read counts of the top 20 differentially expressed genes across the 4 CRC PDOs – analysis performed on the Lexogen QuantSeq BlueBee platform. Data illustrating untreated PDOs and PDOs treated with 2.5μM domatinostat for 3 days. Significant difference analysis performed with a paired t-test, followed by FDR analysis set to 5%.

To first understand the wider effects of domatinostat on the transcriptome, I analysed the RNA QuantSeq data of the 4 PDOs treated with 2.5μM domatinostat for 3 days. As only a single sample was analysed for each PDO due to budget constraints, DEseq analysis was performed across a single replicate from 4 PDOs. DEseq analysis highlighted 0.12% of the 19,634 protein coding genes were differentially expressed, with 18 genes upregulated and 5 downregulated in DMSO vs domatinostat. The most upregulated genes experienced a 2.02 to 6.18 log2 FC (log2 FC mean = +3.38), and the most downregulated genes experienced a -2.37 to -4.33 log2 FC (log2 FC mean = -4.18) (Figure 1.12). Gene set enrichment analysis (GSEA) through the molecular signatures database

found no significant gene set overlaps in the upregulated or downregulated genes.

Next, as previous studies had indicated domatinostat treatment resulted in increased HLA and APM component expression (Ritter *et al.*, 2017; Souri *et al.*, 2020), I wanted to observe if this also applied to our CRC PDOs. Therefore, I plotted a heatmap of the log2 normalized readcounts of proteasome, PLC, HLA-I, HLA-II genes to see if their expression was enhanced as expected (Figure 1.13A). All 4 lines saw only minimal changes in gene expression, none of the changes being statistically significant across the 4 treated PDOs. There was a mean increase of log2 1.16 in HLA-A, B, C, a smaller induction than seen under IFNγ treatment (a mean increase of log2 2.51), and not paired with strong increases in APM genes as observed under IFNγ. There was also an induction of some previously undetected HLA-II genes, mostly in CRC-01, however the level of expression still remained low (below log2 2.0). Overall, the induction of HLA and APM genes was minor and was unlikely to enable the presentation of additional neoantigens alone, as higher increases in the same genes under IFNγ exposure did not generate surface presentation of further neoantigens.

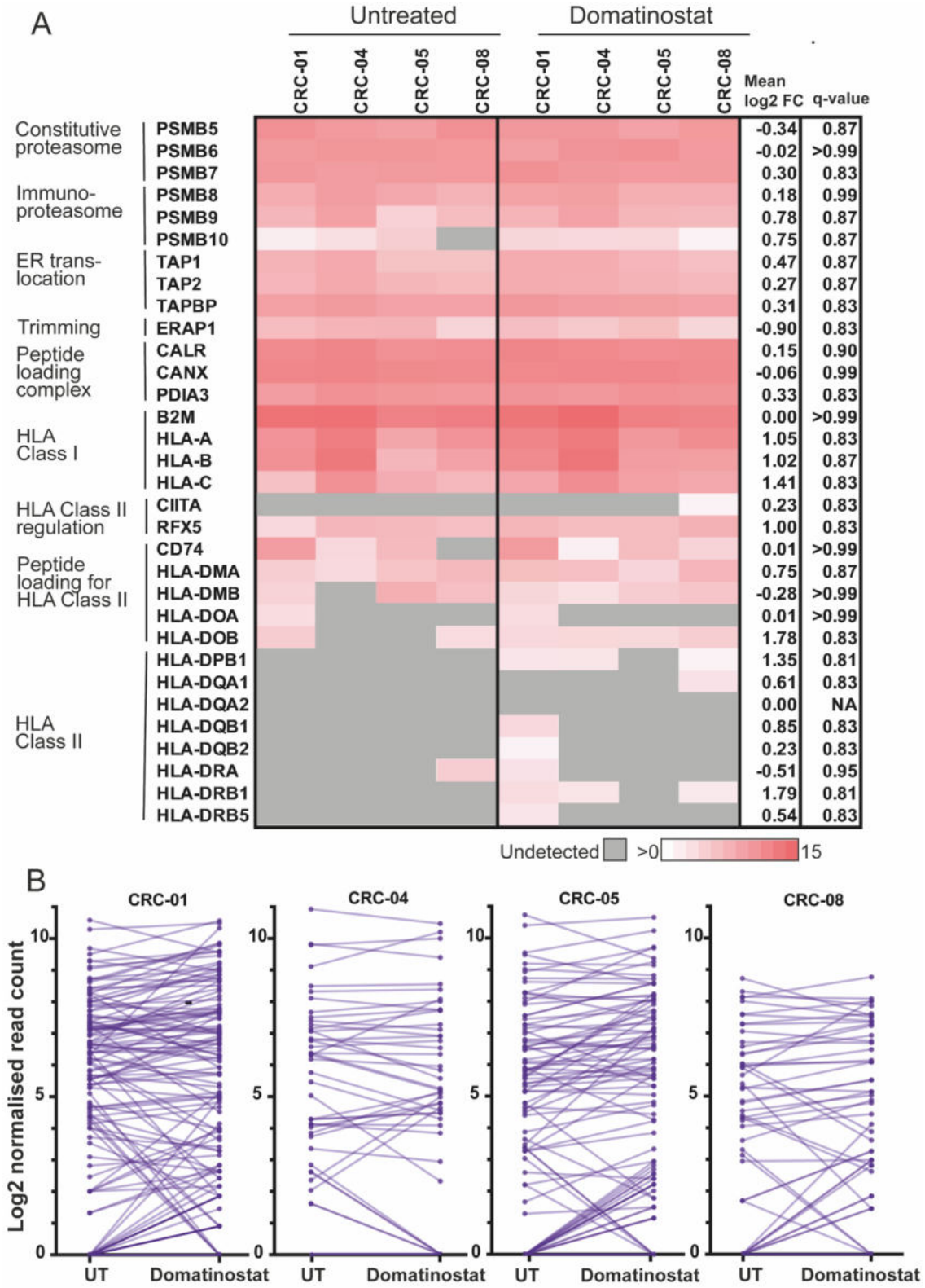


Figure 1.13. Differential gene expression analysis of mutated genes CRC PDOs after treatment with domatinostat. A: Heatmap of log2 normalised 3' RNA expression read counts of HLA and antigen processing machinery (APM) genes, of untreated and domatinostat-treated PDOs (3-day treatment with 2.5μM domatinostat). Significant difference analysis performed with a paired t-test,

followed by FDR analysis set to 5%. **B**: Before/after plots of log2 expression of genes containing nonsynonymous point mutations (specifically selected for mutated genes in each PDO) from untreated and domatinostat-treated PDOs (3-day treatment with 2.5μM domatinostat).

Next, I interrogated whether HDAC inhibition could unlock expression of previously unexpressed genes containing nonsynonymous point mutations, which could provide a source of additional neoantigens. Therefore, I plotted the log2 normalised expression of these mutated genes, specific to each PDO, before and after domatinostat treatment (Figure 1.13B). The data showed substantial variation in the expression levels of the mutated genes: increases, decreases, gains, and losses. Therefore, I analysed the overall effect of domatinostat on the total number of mutated genes expressed. There was a small gain in total expression of mutated genes, with a mean log2 FC of 0.14% in expression of mutated genes (median 0.18%) – due to the fact that a mean of 13.75% of mutated genes increased more than 1 SD, whereas only a mean of 10.92% decreased <-1SD (Table 3.6). Most importantly, a mean of 8.68% (median 11.20%) of the total mutated gene population became expressed (went from zero to non-zero), leading to an overall gain of a mean of 6.06% (median 8.10%) of mutated genes as expressed. However, CRC-04 did not have any genes transition from zero expression to non-zero expression under trametinib treatment, despite being the second most sensitive PDO to HDACi over a 3-day period, in terms of both viability and HDAC activity. Furthermore, CRC-04 conversely lost expression of the largest proportion of its mutated genes.

Overall, the small increase in the total number of nonsynonymous point mutated genes expressed (a net gain of 15, 12, and 5 genes in CRC-01, -05, and -08 respectively, representing an 8.3%, 8.1% and 7.5% gain respectively), and the fact that only 3/4 PDOs experienced more nonsynonymous point mutated genes becoming detectably expressed, meant the pool of source proteins capable of generating additional neoantigens was only slightly larger. Therefore, given the minimal upregulation of HLA and APM genes, paired with the only small gain in expressed nonsynonymous point mutated genes, I decided that the

transcriptional impact of domatinostat was too modest to pursue any further in relation to immunopeptidomics.

	CRC-01	CRC-04	CRC-05	CRC-08	Mean	Median
Mean log2 FC	0.16	-0.13	0.35	0.19	**0.14**	**0.18**
% genes increased >1 SD	12.85	8.22	17.02	16.92	**13.75**	**14.89**
% genes decreased >1 SD	7.82	16.44	7.09	12.31	**10.92**	**10.07**
% of genes zero → non-zero expression	11.05	0	11.35	12.31	**8.68**	**11.20**
% of genes non-zero → zero expression	2.21	6.85	2.84	4.62	**4.13**	**3.73**
Overall gain in % of genes expressed	8.84	-6.85	8.51	7.69	**6.06**	**8.10**
% genes sustained as unexpressed	32.60	45.21	40.43	30.77	**37.25**	**36.52**

Table 3.6. Table depicting the expression changes in nonsynonymous point mutated genes in each of our 4 PDOs seen after treatment with 2.5μM domatinostat for 3 days. The number of mutated genes in each PDO was counted, then the number of genes experiencing described changes were counted and displayed as the percentage of the total number for each PDO. Both mean and median shown.

The evidence provided from the optimisation process suggested that domatinostat treatment of cells was unlikely to show a swathe of new and clinically-relevant neoantigens. Therefore, I decided to attenuate this component of the study, especially as growing cells to the hundreds of millions could take from weeks to multiple months, which was not compatible with my goals, due to the time lost to institute closure during the pandemic.

3.3 Discussion

This work has shown that it is feasible to grow CRC PDOs, established from small biopsies, to large cell numbers to enable MS immunopeptidomics analysis. Growing the cells to such large numbers facilitated the detection of a high number of unique peptides, higher than numbers found in other MS immunopeptidomics studies in cell lines (mean: 7,593/sample, range: 3,293-13,696 (Chong *et al.*, 2017)) , melanoma (mean: 3,144/sample, range: 121-23,971 (Bassani-Sternberg *et al.*, 2016)), ovarian (median: 1,381/sample, range: 183-4,289 (Schuster *et al.*, 2017)) or CRC tumour samples (mean: 1,171 peptides/cancer, range: 322-2,407 (Loffler *et al.*, 2018)). The sensitivity was also paired with a high level of cancer cell specificity due to the absence of stromal components. Up to 16,030 peptides per PDO were identified in the untreated condition, and a further 3,942 additional peptides after IFNγ exposure. With individual HLA alleles predicted to present between 1,000 and 10,000 peptides, it can be estimated that 6 different HLA-I molecules can present (Abelin *et al.*, 2017), on average, ~30,000 unique peptides, suggesting over 50% of the estimated peptide repertoire had been sampled in some of our PDOs.

Despite this sensitivity, only 3 mutated neoantigens were identified across the 5 PDOs, that together possessed 612 non-silent somatic mutations, and no additional neoantigens detected from IFNγ or MEK-inhibitor treatment. However, MS has several known limitations which does reduce the potential sensitivity of detection. MS has known biases for peptides with high affinity to the HLA allotypes, with good solubility in aqueous solution, and peptides which can be well ionized, alongside biases against peptides containing unstable amino acids methionine and cysteine. I therefore cannot exclude the possibility that additional HLA-I-presented neoantigens with less MS-compatible characteristics were present and not detected.

The low neoantigen load, or total absence in some PDOs, contrasts with published data showing that neoantigen-specific T-cells could be found amongst tumour-infiltrating lymphocytes in 5/5 CRCs (Tran *et al.*, 2015). However, this study was limited in that it only assessed the specificity of T-cells against

minigene-encoded predicted neoantigens presented on APCs, which does not expressly evaluate the recognition and reactivity of the autologous cancer cells. Furthermore, an MS immunopeptidomics study of hepatocellular carcinomas with moderate mutation loads (151 + 40 somatic variants) has similarly revealed a scarce neoantigen profile (Löffler *et al.*, 2019). Combining *in vitro* T-cell recognition assays of autologous cancer PDOs and T-cells with immunopeptidomics would be the next important step in to defining the CRC neoantigen landscape. It would help to reveal the features of the most immuno-stimulatory neoantigens, and define whether T-cells are able to recognise tumours without MS-detectable neoantigens which would be an additional validation of MS immunopeptidomics.

It should however be acknowledged that this study did lack an MMRd CRC PDO line as a positive control to assert the sensitivity of the MS immunopeptidomics technique in resolving to mutated neoantigens. Beyond validation of the technique, analysis of samples from MMRd CRC, which has been shown to be sensitive to immunotherapies, could allow for comparison of the quantity and quality of neoantigens which may explain the disparity of the ICI response. In addition, organoid establishment permitting, it could be possible to test the genetic and immunopeptidomic heterogeneity of a tumour by testing biopsies from different regions of the tumour, or to prospectively grow PDOs from patients undergoing treatment at different timepoints, to enable tracking of immunopeptidome changes and correlative analysis. Specifically, comparison between pre-metastatic samples and later-stage metastatic samples could show whether neoantigens are lost over time, which would indicate that tumour evolution by immunoediting is a major mechanism of neoantigen loss (Dunn, Old and Schreiber, 2004).

The repertoire of non-mutated cancer/testis (C/T) antigens also proved to be incredibly scant, with only 1 antigen identified each from 2 C/T proteins (FAM46D and SPANXN3). Furthermore, only antibody responses have been described against antigens from these proteins, so whether these can elicit a cytotoxic T-cell response is unknown (Van Rhee *et al.*, 2005; Bettoni *et al.*, 2009).

Comparing the number and characteristics of MS-detected neoantigens with NetMHCpan-predicted neoantigens, which is one of the current gold standards for neoantigen prediction, indicated the possible over-prediction of neoantigen load by computational analysis. 304/612 mutations were predicted to generate neoantigens that strongly bind autologous HLA-I allotypes, with 196 of those predicted in genes with detectable mRNA expression. This suggested that only 1.52% of non-silent mutations in expressed mutated genes predicted to generate neoantigens, actually produce MS-detectable neoantigens. This highlighted the need to generate more MS immunopeptidomics data to better train prediction tools, as any biological biases governing peptide generation and presentation would then be accounted for in the 'eluted ligand likelihood' component of the algorithm.

Although only a small number of MMRp CRC PDOs were analysed, the information gleaned from this study was clinically relevant, as 4/5 of the PDOs were established from metastatic tumours that exhibited resistance to palliative chemotherapy prior to biopsy sampling. Furthermore, the chemotherapy-naïve PDO had previously been exposed to immunotherapeutic treatment. The mutational load and characteristics of the PDOs were therefore representative of some of the clinical features observed in treatment-resistant CRCs, which have been a primary focus of novel immunotherapy trials. The sparse neoantigen and C/T antigen landscape observed across the 5 MMRp CRC PDOs suggested these advanced CRCs exhibited low endogenous immunogenicity, which could help to explain the low success rate of ICI immunotherapy in advanced MMRp CRCs (Le *et al.*, 2015). With the overwhelming data suggesting that the canonical HLA-I immunopeptidome only presents a small fraction of the HLA-compatible neoantigens, it is clear there is a pressing need to evaluate antigens from alternative sources (e.g. peptides associated with defects in antigen processing (Marijt *et al.*, 2018), fusion genes, de-repressed endogenous retroviruses, transposable elements, post-translationally modified peptides, and peptides from novel open reading frames (Laumont *et al.*, 2016)). Furthermore, alternative approaches could be used to find treatments that enhance the endogenous immunogenicity of patient tumours, or to uncover alternative target antigens that can be found in the absence of sufficient neoantigen load. Bispecific antibodies

or CAR-T-cells are examples of therapies that target overexpressed or aberrantly expressed cell surface molecules, such as CEA in CRC (Bacac *et al.*, 2016).

PDO growth from MMRp CRC patients provided a model on which to screen how drug or cytokine treatments can alter the peptidome. I found that IFNγ treatment increased the number of surface HLA-I molecules in all 4 PDOs, yet only increased the number of unique peptides in 2/4 PDOs, in which changes were only modest. This restricted increase suggested that although HLA-I expression is known to be a limiting factor in peptide diversity (Komov *et al.*, 2018), antigen processing and HLA allotype binding specificities provide more stringent constraints on the diversity of the immunopeptidome. The data did not illustrate that the switch from constitutive to immunoproteasome was able to generate many more HLA-I-compatible peptides, as would have been expected from the current understanding of proteasome function (Chong *et al.*, 2017). Furthermore, MEK inhibitor treatment did not exert any consistent effect on surface HLA abundance, or peptide diversity, across the 4 PDOs. This could explain the lack of clinical impact of MEK-inhibitor treatment, when used in combination with a PD-L1 in a recent clinical trial (Eng *et al.*, 2019), a combination which had initially been conceived due to the observation that MEK inhibition could increase the expression of HLA-I in a CRC mouse model (Ebert *et al.*, 2016).

HDACi treatment was another treatment trialled in our PDOs, due to early clinical success in combination with ICI immunotherapy (Cartwright *et al.*, 2021), and data showing it could increase HLA-I expression and APM genes (Souri *et al.*, 2020; Hassel *et al.*, 2021). Additionally, HDACi treatment had the potential to unlock/augment the expression silenced/low expressed genes which could encourage the presentation of novel neoantigens. The results showed HDACi treatment did not succeed in dramatically increasing the expression of HLA or APM genes as expected. Furthermore, although it did succeed in initiating the expression of some previously unexpressed nonsynonymous point mutated genes, this only facilitated a net 6.06% gain in the number of nonsynonymous point mutated genes expressed (mean across 4 PDOs). As in the initial study only 3 neoantigens were detected from 196 expressed mutated genes predicted to generate neoantigens, this small gain in previously unexpressed mutated

genes was unlikely to reveal many novel neoantigens. Given the modest findings from the optimisation experiments, the time limitation due to time lost during laboratory closures during lockdown, and the considerable cost of immunopeptidomics, this research was taken no further. Later, further clinical results (currently unpublished) found domatinostat, the chosen HDACi, was unlikely to provide clinical benefit (*Discontinuation Domatinostat Program - 4SC AG*, 2022), which suggested it could not sufficiently raise neoantigen presentation to improve immune recognition and destruction of cancer cells.

The recapitulation of poor neoantigen load in therapy-resistant advanced CRCs suggested that low neoantigen load could be predictive of response to ICI immunotherapies. Furthermore, the data thus far indicated that the treatments which fail to augment neoantigen presentation, correspondingly fail to provide clinical benefit in combination therapies. This suggested that immunopeptidomics analysis for treatment design may be best placed as a step in pre-clinical validation, before instigating clinical trials. Additionally, immunopeptidomics also does provide the best hope of identifying clinically-relevant peptides for cancer vaccine design, as neoantigen prediction does currently appear to overpredict the number of HLA-compatible neoantigens.

Chapter 4 Understanding the factors involved in interferon gamma-mediated immunopeptidome remodelling

4.1 Introduction

All nucleated cells express HLA class I (HLA-I), which is a mechanism by which the internal health of the cell is communicated to the immune system via the presentation of self and non-self-peptides. Presentation of non-mutated cancer-associated antigens or neoantigens on HLA-I is therefore is essential for the detection of malignant cells by the adaptive immune system, facilitating activation of CD8 T-cells to target and destroy the malignant cells. An important component of the cytokine milieu released in to the tumour microenvironment (TME) by activated CD8 T-cells and other immune cell subsets, is interferon gamma (IFNγ). Secreted IFNγ binds IFNγ receptors on the cell surface, which then activates the downstream JAK/STAT pathway, resulting in the expression of interferon response factor (IRF) transcription factors. IRFs such as IRF1, IRF8, and IRF9 stimulate the expression of a large number of IFNγ-regulated genes which causes large-scale changes throughout the transcriptome and proteome (Der *et al.*, 1998; Olsson *et al.*, 2021). One specific subset of genes that are strongly upregulated by IFNγ signalling, are those involved in the processing and presentation of peptides on HLA-I (Arellano-Garcia *et al.*, 2014; Megger *et al.*, 2017). The antigen processing machinery (APM) includes the catalytic immunoproteasome components PSMB8, PSMB9, and PSMB10 which facilitate an increase in total proteasomal output, alongside a specific increase in chymotryptic-like cleavage activity (Früh and Yang, 1999; Ferrington and Gregerson, 2012; Mishto *et al.*, 2014); and peptidases such as LAP3 (Beninga, Rock and Goldberg, 1998), THOP1 (Saric *et al.*, 2001), ERAP1 (Saric *et al.*, 2002), and ERAP2 (Saveanu *et al.*, 2005). Also upregulated are the peptide transporters TAP1 and TAP2 (Ma *et al.*, 1997; Seliger *et al.*, 1997), which transport peptides to the endoplasmic reticulum (ER) for loading on to HLA-I, and the chaperone proteins of the ER-resident peptide loading complex (PLC), TAPBP, CALR and CANX. Furthermore, it is well documented that IFNγ increases HLA-I and HLA-II expression (Steimle *et al.*, 1994; Früh and Yang,

1999; Zhou, 2009). The expected outcome of the increased proteasomal abundance and activity, peptide processing and transport, and HLA abundance, is a strong increase in HLA-I and HLA-II peptide presentation. Further to this, IFNγ also exerts anti-cancer functions, inhibiting the cell cycle, and triggering apoptosis through JAK-STAT1-caspase signalling (Burke *et al.*, 1997; Zaidi and Merlino, 2011; Jorgovanovic *et al.*, 2020).

Conversely, IFNγ can also exert pro-tumorigenic effects by promoting the expression of immunosuppressive genes in cancer cells. Key examples include PD-L1, whose ligation with PD-1 immune checkpoint molecule on T-cells results in decreased T-cell activation, proliferation, cytokine production, and survival (Freeman *et al.*, 2000; Latchman *et al.*, 2004; Butte *et al.*, 2007); and IDO1, whose expression on cancer and stromal cells in the TME suppresses cytotoxic T-cell effector function and induces Treg development (Baban *et al.*, 2009; Brody *et al.*, 2009; Munn and Mellor, 2016; Zhai *et al.*, 2018). Immunotherapy with PD1/PD-L1 inhibitors has been shown to be clinically effective in several cancer types (Eggermont *et al.*, 2018; Gandhi *et al.*, 2018; André *et al.*, 2020), with PD1/PD-L1 abundance often associated with response level (Robert *et al.*, 2015; Fehrenbacher *et al.*, 2016; Weber *et al.*, 2017). This highlights the central role the PD1/PDL-1 axis in restraining the immune system, consequently impeding tumour rejection. IFNγ is intrinsically linked with PD-L1, as IFNγ has shown to induce PD-L1 expression (Skoulidis *et al.*, 2015; Kadara *et al.*, 2017), and tumours with good response to PD1/PD-L1 inhibitors often show high IFNγ signal in the TME (Ayers *et al.*, 2017). IFNγ expression and response has been shown to be an important determinant of tumour rejection; critically, several studies have shown that defective IFNγ signalling in cancer cells can foster resistance to immunotherapy with immune checkpoint inhibitors (ICIs) (Duncan *et al.*, 2007; Manguso *et al.*, 2017; Patel, S., Sanjana, N., Kishton *et al.*, 2017). Although intact IFNγ signalling in cancer cells is important for response to ICIs, it is still unclear by which specific mechanisms IFNγ signalling achieves this. Gaining a deeper and more detailed understanding of how IFNγ is able to remodel the immunopeptidome may bring this to light.

Novel therapies have been designed to promote the recognition and destruction of cancer cells presenting specific peptide antigens on their surface HLA; antigen-targeting treatments such as cancer vaccines, and TCR-based therapies like ImmTAX (Table 4.1), have made significant advancements in the field of immunotherapy. These recent clinical trials have shown that it is feasible to generate tumour-specific CD4 and CD8 T-cell responses by TAA- and neoantigen-based vaccines, even in low tumour mutational burden (TMB) tumours like glioblastoma. Whilst these antigen-directed therapies provide great promise, they are limited by the uncertainty over the likelihood of HLA-presentation of individual cancer antigens and neoantigens. Computational HLA binding prediction tools have been advancing in development, the algorithms trained by machine learning to try and best predict peptides which are most likely to appear on HLA-I. Previously, algorithms had been built solely on peptide binding affinity data, but this did not encompass intrinsic biases in peptide processing, therefore now algorithms like NetMHCpan are also trained on MS-detected peptides eluted from HLA (Abelin *et al.*, 2017; Bulik-Sullivan *et al.*, 2018; Reynisson *et al.*, 2020). Despite these advancements, studies have shown that these techniques can overpredict the presentation of neoantigens (Bassani-Sternberg *et al.*, 2016; Teku and Vihinen, 2018; Newey *et al.*, 2019). Therefore, being able to distinguish between characteristics of peptides which are able to be presented in the presence or absence of IFNγ, and ones which are lost or downregulated under IFNγ exposure, could help fine-tune the target-antigen selection process to determine antigens most likely to be presented in both immune cold and immune hot environments.

Chapter 3 illustrated the IFNγ-influenced increase in HLA-I and HLA-II, and APM gene expression did not massively increase peptide diversity as may have been expected. Therefore, I wanted to try and resolve the changes in HLA-I-presented peptide characteristics under IFNγ exposure, to try to elucidate the molecular mechanisms through which IFNγ alters HLA-I peptide presentation. This was achieved through new global proteomics data, obtained by tandem-mass-tag (TMT) MS, which was combined with transcriptomic and MS immunopeptidomic data from untreated and 48-hour IFNγ-treated CRC PDOs. Whilst other multi-omics studies have begun to investigate how the immunopeptidome changes

under IFNγ exposure, no such study has been able to identify HLA-independent peptide features that can help to better predict peptide up- or downregulation.

Trial	Phase	Tumour type	Treatment type	Findings	Reference
NCT00683670	I	Melanoma	Dendritic cell vaccine	Induced T-cell responses against the high affinity patient-specific mutant peptides and additional subdominant neoantigens.	(Carreno *et al.*, 2015)
NCT020035956	I	Melanoma	mRNA: TAAs and neoantigens	mRNA vaccines produced peptides which induced CD4+ and CD8+ T-cell responses against multiple neoantigens. Enhanced efficacy with ICIs observed.	(Sahin *et al.*, 2017)
NCT02287428	I/IB	Glioblastoma	Peptide vaccine of 20 15-30mers	Even in 'immunologically cold' tumours with low TMB, it induced tumour-infiltrating CD4+ and CD8+ T-cells which effected anti-tumour responses.	(Keskin *et al.*, 2019)
NCT02149225	I	Glioblastoma	Peptide: TAAs and neoantigens	Neoantigen vaccine induced tumour-specific CD4+ responses, and TAA vaccine induced CD8+ T-cell responses in low TMB tumours.	(Hilf *et al.*, 2018)
NCT03070392	III	Uveal melanoma	Bispecific TCR – anti-CD3 fusion protein (ImmTAX)	Influenced statistically significant improved overall survival and progression free survival in a low TMB tumour. Now FDA approved.	(Nathan *et al.*, 2021)

Table 4.1. Examples of neoantigen-targeting immunotherapeutic treatments. Table inspired by (Blass and Ott, 2021).

4.2 Results

4.2.1 Analysing changes in the transcriptome and proteome induced by IFNγ

TMT-MS detected 8,477 proteins across the 3 PDOs. Plotting the IFNγ-induced fold-changes (FC) of mRNA transcripts against those of protein abundances showed a significant positive correlation (Spearman's rank r = 0.34-0.69, p $<2.2x10^{-16}$ for all 3 PDOs) (Figure 2.1A). IFNγ exposure increased the expression/abundance of a large number of transcripts/proteins, and downregulation was only notable in a smaller number of transcripts/proteins, and the downregulation was to a lesser degree.

I next analysed if known IFNγ-regulated transcripts/proteins, specifically those which have roles in antigen processing and presentation or immune evasion, undergo the expected protein abundance changes. Treatment with IFNγ increased the mRNA expression of most APM genes.

The heatmap illustrated that all the proteasome components increased in expression, including the constitutive proteasome catalytic subunits (PSMB5-7), and immunoproteasome catalytic subunits (PSMB8-10) (Figure 2.1B). This was linked to an increase in protein abundance, with the constitutive subunits being the exception. This disconnect between mRNA expression and protein abundance can be explained by the fact that immunoproteasome assembly is four times faster than that of the constitutive proteasome (Ferrington and Gregerson, 2012). This is due to the higher affinity of immunoproteasome subunit PSMB8 for the proteasome assembly chaperone POMP, enabling it to outcompete PSMB5, resulting in the rapid degradation of the unbound constitutive subunits (also known as orphan subunits) (Komov *et al.*, 2021). The switch from constitutive to immuno- proteasome activity alters the cleavage specificity towards an increased chymotryptic activity, as observed in in Chapter 3. Furthermore, the increased abundance of proteasome proteins would result in an increase in the total number of proteasomes, enabling increased output of peptide processing. Furthermore, the regulatory caps of the proteasome also change in abundance with IFNγ treatment. In unperturbed conditions the 26S

proteasome is created by the association of the 19S cap, to either or both ends of the 20S proteasome core particle (Besche, Peth and Goldberg, 2009; Matyskiela, Lander and Martin, 2013; Tomko and Hochstrasser, 2013). The 19S cap acts as a recognition platform for proteins marked for destruction by polyubiquitination (on K48 of ubiquitin) (Martinez-Fonts *et al.*, 2020) and dramatically increases the rate of peptide cleavage (McCarthy and Weinberg, 2015). The 19S cap captures these proteins, and in an ATP-dependent fashion, deubiquitylates them, unfolds them, and transports them in to the core of the 20S particle for cleavage (Rubin *et al.*, 1998; Smith *et al.*, 2005; Peth, Uchiki and Goldberg, 2010). There were only small changes in the levels of the 19S cap transcripts and proteins. In contrast, IFNγ strongly increased the expression of the 11S cap subunits. The 11S cap facilitates the ubiquitin-independent degradation of proteins and augments proteasomal throughput (Schwarz *et al.*, 2000; Förster, Whitby and Hill, 2003). IFNγ treatment induced a decrease in the abundance of the protein PSME4, another proteasome cap subunit, which has recently been revealed to have an inhibitory effect of the generation of HLA-I peptides (Javitt *et al.*, 2021). Therefore, it is possible that decreased PSME4 levels could further amplify HLA-I peptide production.

From the cytosolic peptidases, LAP3 (with a cleavage specificity towards hydrophobic N-terminal amino acids, primarily leucine) increased the most in transcript and protein abundance (Figure 2.1B). DNPEP (with a cleavage specificity towards acidic N-terminal amino acids, specifically aspartic acid) is the only other peptidase that increased in protein abundance, with the other proteins decreasing in abundance. Both ER aminopeptidases ERAP1 and ERAP2, which shapes the final immunopeptidome by trimming on HLA-I (Saric *et al.*, 2002; Saveanu *et al.*, 2005), increased in abundance.

PLC genes all increased in mRNA expression and protein abundance, including transporter-associated with antigen processing (TAP) proteins, which would help to facilitate accelerated peptide import in to the ER. In addition, NLCR5, the master transcription factor for HLA-I, and in concert HLA-A, -B and -C increased strongly which expands surface presentation capacity. Despite this, Chapter 3 showed that these gains in protein abundance did not influence an increased

peptide diversity as may have been expected, so I focused on pursuing an understanding of the characteristic changes that may instead occur under IFNγ.

I also investigated whether transcripts/proteins of immune evasion genes were upregulated by IFNγ. Most of the immune evasion genes increased in expression, including IDO1 and CD274/PD-L1 which increased most in expression and abundance. Thus, IFNγ did cause the expected changes in known IFNγ-regulated genes across the 3 PDO lines.

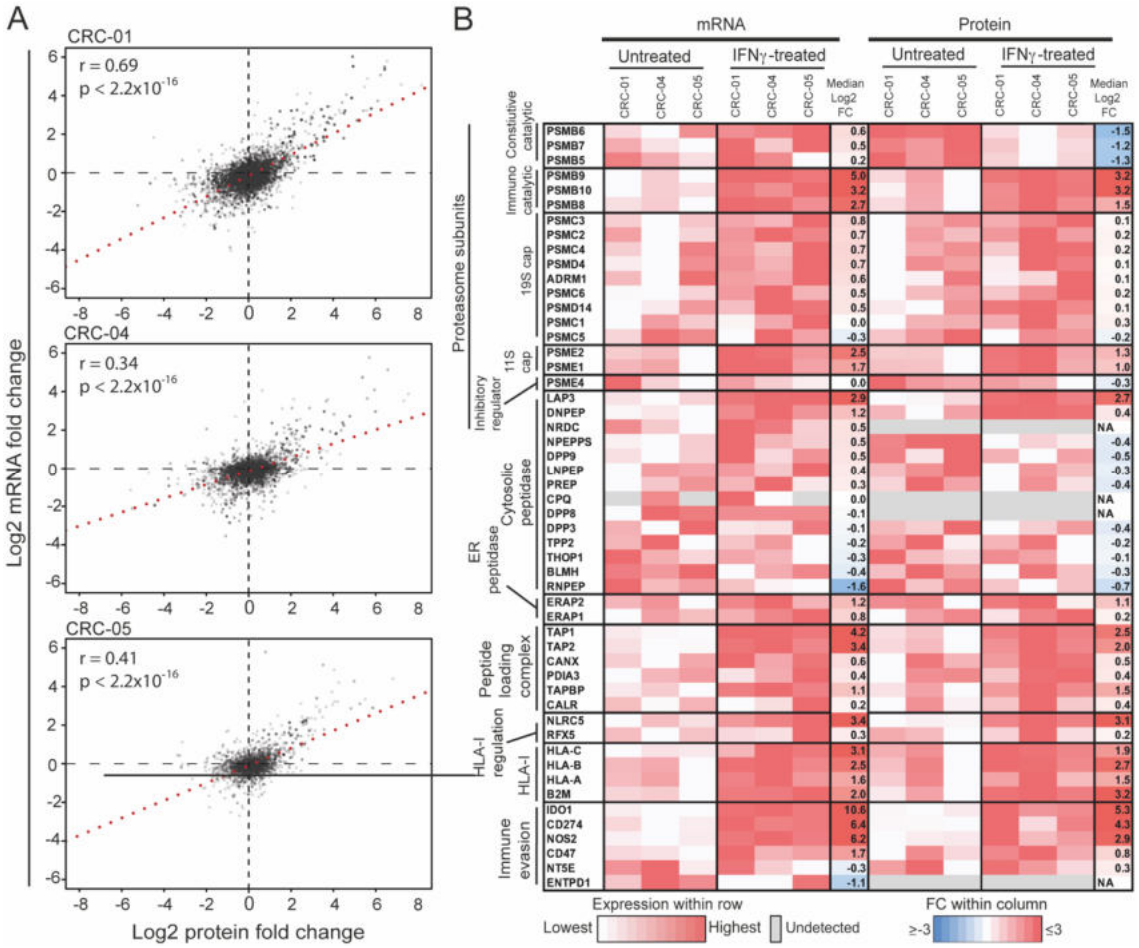


Figure 2.1: Transcriptomic and proteomic changes with IFNγ treatment. A: Correlation of the fold-change (FC) in normalised mRNA read numbers against the fold change in normalized protein intensity, between untreated and IFNγ conditions. Spearman's rank testing used for statistical analysis. **B:** mRNA expression and protein intensity of selected genes in untreated and IFNγ conditions.

4.2.2 Analysis of NetMHCpan4.0 Vs NetMHCpan4.1

For this study I wanted to analyse how the peptide characteristics of the non-mutated immunopeptidome changed under IFNγ exposure. Therefore, it was important to be able to understand the characteristics of the peptides, including their predicted affinities for their encoded HLA allotypes. Therefore, I chose to use NetMHCpan4.1 to analyse the predicted binding affinities, as it has had additional training and improvements over NetMHCpan4.0 (Reynisson *et al.*, 2020). As I wanted to analyse the HLA types for which peptides were predicted to have the highest affinity for, binding affinity (BA) was chosen instead of eluted ligand likelihood (ELL). This was because it was based on quantitative binding affinity data, and not eluted ligand MS data. BA prediction, for each HLA, compares the prediction score for the input peptide against a distribution of prediction scores of random MS-detected peptides for that same HLA; meaning a 2% rank illustrates that the input peptide had a BA score for the indicated HLA that was in the top 2% of those scores obtained from the set of random peptides. I also selected for BA rank, rather than score, as it normalised BA score across HLA allotypes, facilitating comparison across HLA allotypes. Peptides were assigned to the HLA allotype for which they had the lowest BA rank.

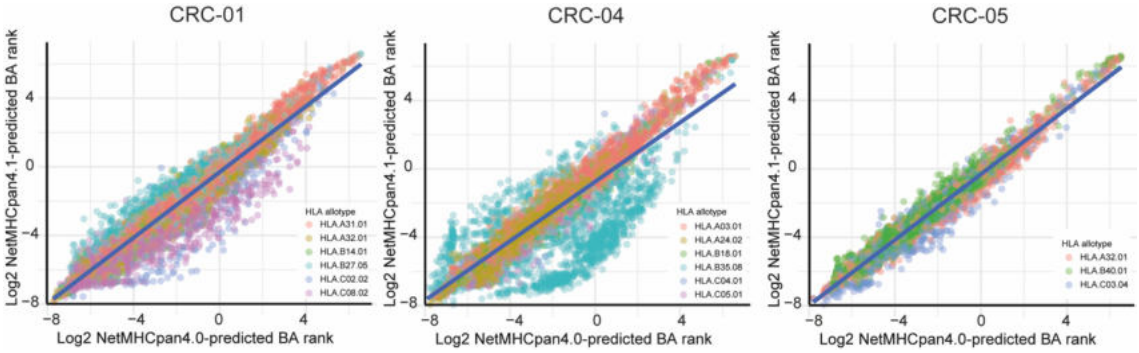


Figure 2.2. Correlation of log2 BA ranks for MS-detected peptides NetMHCpan4.0 Vs NetMHCpan4.1. Log2 NetMHCpan4.0-predicted BA rank plotted against log2 NetMHCpan4.1-predicted BA rank, where HLA allotypes are assigned by selecting the allotype for which the peptide has the lowest BA rank; only peptides for which both prediction algorithms predicted the peptide to bind the same HLA allotype were included.

To understand how the change in choice of NetMHCpan4.0 and NetMHCpan4.1 prediction algorithms affected the predicted BA ranks, I first assessed the

proportion of peptides that were predicted to be attributed to the same HLA allotype. 89.67%, 96.76%, and 98.86% of the total peptide repertoire for CRC-01, CRC-04, and CRC-05, respectively, were predicted to bind the same HLA allotype, indicating a good degree of similarity between the two prediction algorithms. To analyse whether there was uniformity by the way of BA rank, I then plotted the log2 BA rank for the peptides that were predicted to bind the same HLA allotype by both algorithms (Figure 2.2). CRC-05 had the highest degree of similarity of BA rank between NetMHCpan4.0 and NetMHCpan4.1, with a Spearman's rank correlation coefficient of 0.978, whereas both CRC-01 and CRC-04 had a weaker correlation, with a Spearman's rank of 0.886 and 0.962, respectively (p-value < 2.2×10^{-16} for each PDO). For CRC-01, peptides attributed to HLA.C02.02 and HLA.C08.02 appeared to have a slightly lower rank in NetMHCpan4.1, and HLA.B27.05 had a slightly higher rank in NetMHCpan4.1. For CRC-04, HLA.B35.08 exhibited the clearest divergence, exhibiting some peptides with higher ranks in NetMHCpan4.1, but most having higher ranks in NetMHCpan4.0. Using BA rankings to define peptides as strong or weak binders (SB or WB) is more sensitive to any changes in BA rank that may occur through using the updated algorithm, as they are defined by strict boundaries (typically BA rank ≤0.5 or ≤2% respectively). Because strong-binding and weak-binding was not interrogated in this analysis, this was not a concern. Overall, the changes between HLA allotype allocation and BA ranks between the two prediction algorithms were mostly modest, suggesting it was appropriate to move forwards with the updated NetMHCpan4.1.

4.2.3 Analysis of peptide and protein abundance changes under IFNγ

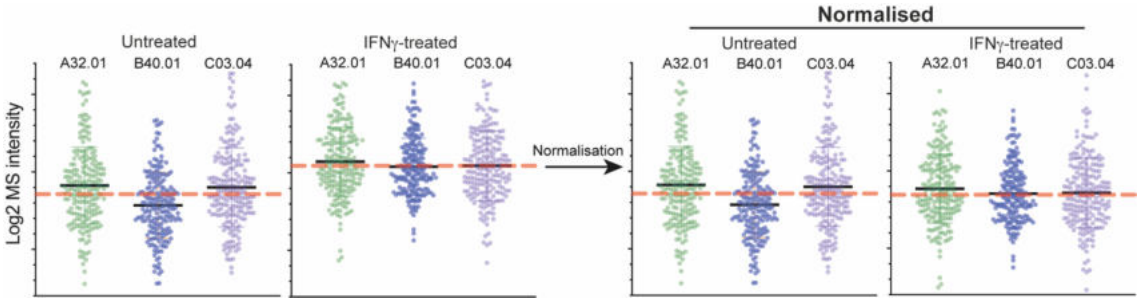


Figure 2.3: Peptidomics normalization schematic, illustrated on CRC-05. Schematic exemplifying the chosen peptidomics normalisation process with sampled data (n = 200 peptides per HLA) from CRC-05 (which only has three HLA-I allotypes): the total peptide intensities of untreated and IFNγ-treated

sample pairs were normalized to be equal. This enabled analysis of how individual peptides changes in abundance within the entire HLA-I-presented peptide population with IFNγ treatment. The black line indicates the mean peptide intensity for each HLA allotype, and the red dotted line indicates the mean peptide intensity across all peptides in each condition.

I next went on to analyse the effect of changing source protein abundance, under the influence of IFNγ, on the presentation of their derived peptides on HLA-I. When combining the immunopeptidomic and proteomic datasets, I found that 21.71%, 21.09%, and 19.96% of the MS-detected peptides for CRC-01, CRC-04, and CRC-05 respectively, were derived from proteins which were not detected by TMT-MS. This may have been due to limited sensitivity of the MS assay, though the proteomics analysis achieved a depth similar to previous MS studies (Komov *et al.*, 2018; Goncalves *et al.*, 2021; Olsson *et al.*, 2021). The absence of some of these HLA-peptide source proteins in the proteomics dataset may have also been because they were high-turnover proteins, meaning they were still able to generate HLA-peptides despite low abundance in the cell. To ascertain whether these were high turnover proteins, I compared the turnover rates of the TMT-MS-detected and non-detected peptide source proteins. For this, I utilised existing data on protein turnover ratios from (Larance *et al.*, 2013). It should be noted that the study used the osteosarcoma cell line U2OS, which may have had differences in protein stabilities from our CRC PDOs. There was data available for 59.25% of the MS-detected peptide source proteins which somewhat limited the depth of the analysis. The analysis showed the proteins found to be absent in proteomics but present in immunopeptidomics, to have an 80.8% lower stability value (mean = -0.235) than the proteins present in proteomics (mean = -0.105). This placed the proteomics-absent proteins in the 35th percentile of all FC data, and the proteomics-present peptides to be in the 58th percentile of all FC data. This indicated the proteomics-absent proteins were far less stable than the proteomics-present proteins. Furthermore, whilst small differences in median protein length can be seen between groups, the Mann-Whitney test found no significant difference in median protein length. This suggested that large differences in protein lengths between groups were not responsible for the significant differences in protein stability. Therefore, the

difference in the turnover ratios of the proteomics-absent vs proteomics-present peptide source proteins may explain some of the disparity (Figure 2.4).

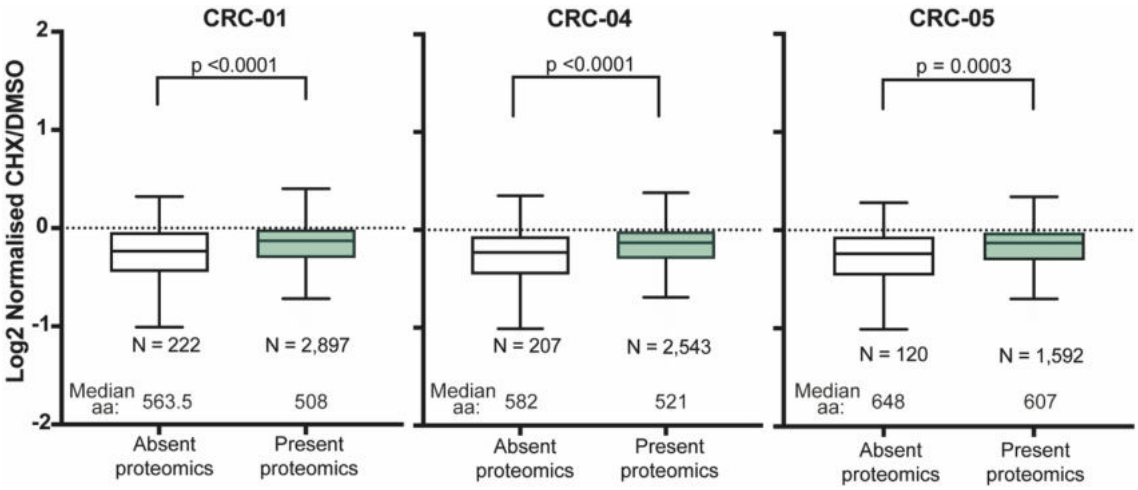


Figure 2.4: Stability data for TMT-detected and non-detected proteins. Protein turnover ratios obtained from (Larance *et al.*, 2013). The analysis assessed the changed in protein abundance following treatment with protein synthesis inhibitor cycloheximide (log2 CHX/DMSO). Turnover ratios compared between TMT-detected and non-detected proteins, with the Mann-Whitney test used for statistical analysis. Median aa represents median protein length of each group.

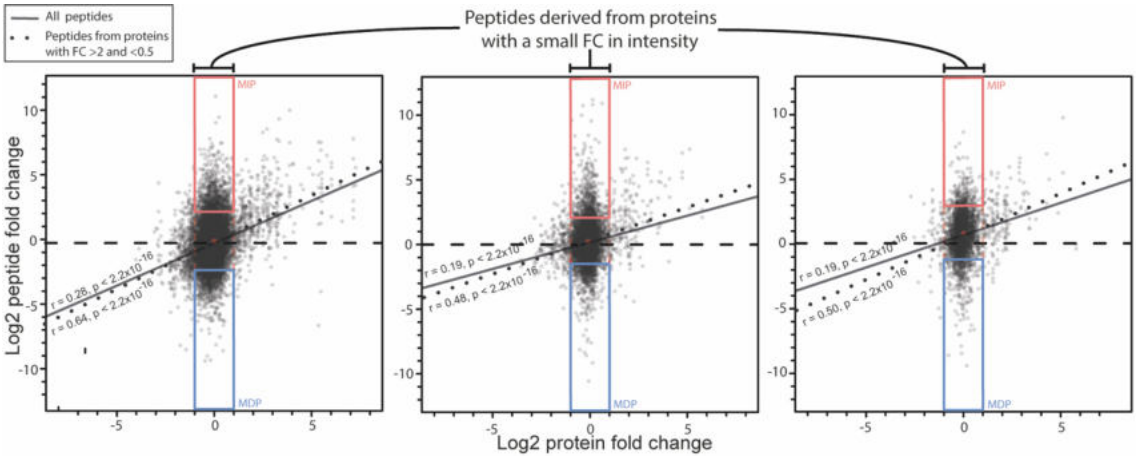


Figure 2.5: Influence of protein abundance and HLA expression changes on immunopeptidome remodelling. Correlation of protein FC between untreated and IFNγ-treated conditions, against normalised immunopeptidomics FC. Regression lines for all peptides indicated by a solid black line, regression line for peptides from proteins with <0.5 or >2 FC indicated with a dotted black line. Most increasing peptides (MIPs - top 10th percentile peptide FC) and most

decreasing peptides (MDPs - bottom 10th percentile peptide FC) derived from low fold-change proteins (0.5-2X fold-change) are highlighted with red and blue boxes, respectively.

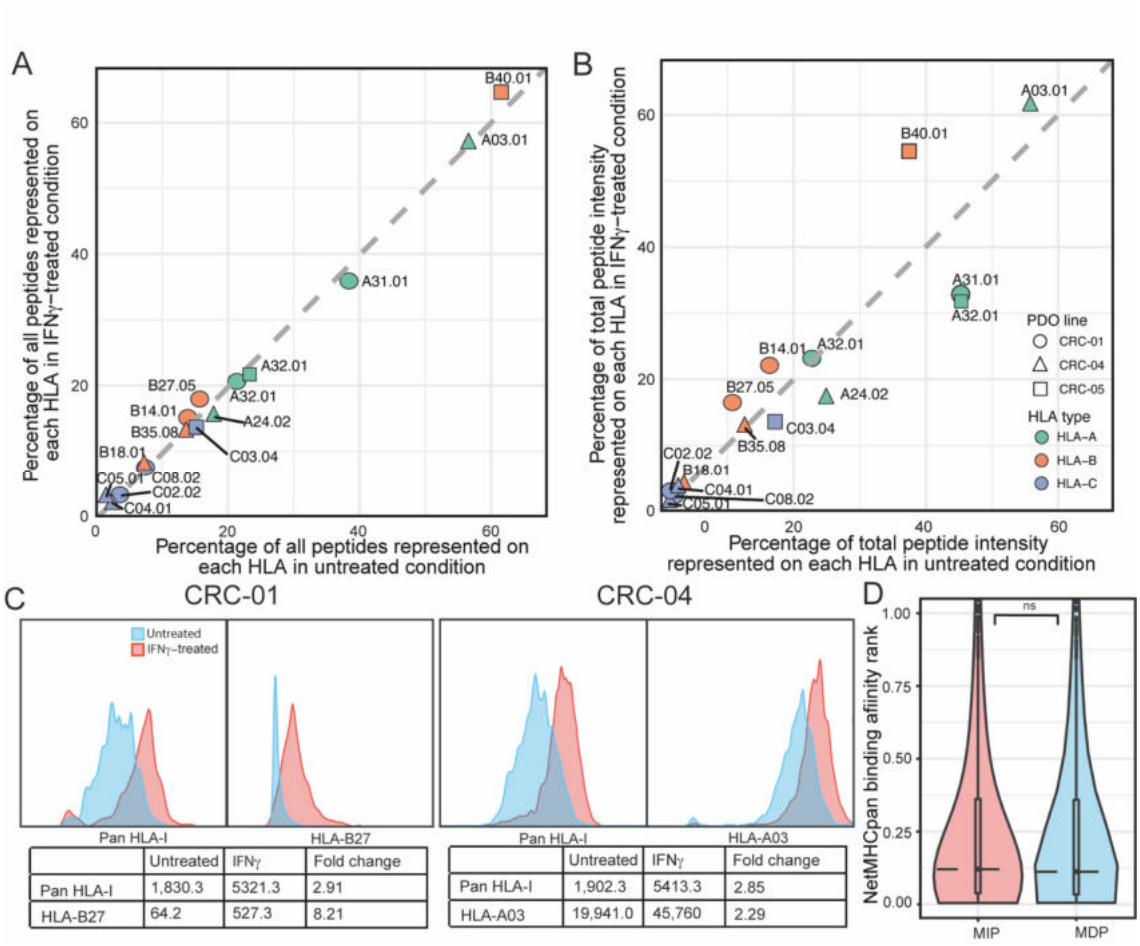
Following the analysis of protein turnover ratios of proteomics-absent and proteomics-present peptide source proteins, I focused on peptides which were derived from proteomics-present peptide source proteins. This was because this allowed for analysis of the relationship between proteomics and immunopeptidomics abundance changes under IFNγ exposure. To enable this analysis peptide abundance normalisation was performed similar to the approach used for RNA expression and protein abundance data: for each PDO, the individual peptide intensity values were normalised to make the total peptidome intensities in untreated and IFNγ-treated conditions identical (Figure 2.5). With IFNγ treatment, I saw an increase in HLA-I abundance (in Chapter 3) which enabled for increased copies of peptides to be presented, indicated by increasing total peptide intensity values. Therefore, normalisation between the conditions enabled assessment of peptide changes beyond the absolute HLA-I increase.

After normalisation, the FCs in normalised protein abundance between untreated and IFNγ conditions were plotted against the FC in normalised abundance of HLA-I peptides (Figure 2.5). This analysis displayed two groups of peptides: those that were up- or down-regulated in concordance with their source protein under IFNγ exposure (showing weak, but significant, correlation between protein FC and peptide abundance FC, similar to a previous study (Goncalves *et al.*, 2021)), and those that were up- or down-regulated despite derivation from proteins with no, or little change in abundance (defined as log2 FC -1 to +1). This illustrated that in some cases an increase in source protein abundance could increase the availability of that substrate for proteasomal breakdown. Furthermore, an increased rate of translation may also be linked with an increase in peptides derived from defective translation or nonsense mediated decay (Yewdell, Antón and Bennink, 1996; Reits *et al.*, 2000; Apcher *et al.*, 2011). Conversely, in other cases protein abundance may be high due to high stability fostering a lack of proteasomal breakdown which would provide fewer peptides for HLA-I presentation. Therefore, changing protein abundance is one of the important drivers of peptidome remodelling.

4.2.4 Protein-independent sources of peptide abundance changes under IFNγ

4.2.4.1 Impact of HLA allotypes on peptide abundance changes

Next, additional analyses were performed to ascertain the ways in which peptides that changed in abundance independent of changes in their source protein were regulated under IFNγ exposure. The first hypothesis was that IFNγ upregulates different HLA alleles to different extents, and that due to their differing anchor residue preferences, this may affect the diversity or abundance of the immunopeptidome. Peptides were assigned to the HLA allotype for which they were predicted to have the lowest BA rank by NetMHCpan4.1 (Reynisson *et al.*, 2020). Plotting the percentage of the total unique peptide count assigned to each HLA allele showed only small differences between each HLA allele after IFNγ treatment, illustrating that no particular HLA allele presented a dramatically larger proportion of the total peptide number after IFNγ exposure (Figure 2.6A). The total peptide intensity detected by mass spectrometry in a fixed number of cancer cells can be utilised as a surrogate measure of the abundance of peptides presented. Therefore, plotting the percentage of the total peptide intensity presented on each HLA allele allowed for analysis of the total peptide abundance presented on each HLA allele between untreated and IFNγ-treated conditions (Figure 2.6B). In IFNγ-treated conditions, HLA-B presented an increased proportion of the total peptide intensity when compared against HLA-B in the untreated condition, conversely 3 out of 5 HLA-A displayed a decreased proportion of the total peptide intensity. Therefore, whilst the total volume of peptides presented on HLA-A may still increase as all HLA proteins showed upregulation in the proteomics data (Figure 2.1), our results indicated that under IFNγ exposure the proportion of the peptidome expressed on HLA-A decreased relative to that of the untreated condition, and that on HLA-B increased.



	Untreated	IFNγ	Fold change
Pan HLA-I	1,830.3	5321.3	2.91
HLA-B27	64.2	527.3	8.21

	Untreated	IFNγ	Fold change
Pan HLA-I	1,902.3	5413.3	2.85
HLA-A03	19,941.0	45,760	2.29

Figure 2.6. Influence of differential HLA expression changes on immunopeptidome remodelling. A: Percentage of each PDO's immunopeptidome that was attributed to each HLA allotype by NetMHCpan4.1 in untreated vs. IFNγ-treated conditions. **B:** Percentage of the total peptide intensity, per PDO, represented on each HLA allotype (attributed by NetMHCpan4.1) in untreated vs. IFNγ-treated conditions. **C:** Expression of total surface HLA-I and single HLA-I allotypes in organoid lines CRC-01 (HLA-B27) and CRC-04 (HLA-A03), assessed as median fluorescence intensity by flow cytometry. **D:** Log2 NetMHCpan4.1-predicted binding affinity ranks for MIPs vs MDPs. The median is marked with a dotted line.

To experimentally validate these findings, I performed flow cytometry to analyse abundance changes of total HLA-I and two HLA allotypes for which specific antibodies were available (HLA-A03 and HLA-B27). To match the treatment conditions of the immunopeptidomics analysis, PDOs were either untreated or treated with IFNγ for 48-hours. For CRC-01, pan-HLA-I antibody staining revealed in a 2.91-fold increase in total surface HLA-I in IFNγ conditions compared against

untreated conditions, whilst HLA-B27 experienced an 8.21-fold increase (a 2.82-fold larger increase than total HLA-I) (Figure 2.6C). Furthermore, CRC-04 displayed a 2.85-fold increase in total HLA-I, whilst HLA-A03 only exhibited a 2.29-fold increase, which was only 0.83-fold of the total HLA-I upregulation. This validated the findings from the immunopeptidomics analysis and is consistent with both transcriptional studies (Schmidt *et al.*, 1990; Girdlestone *et al.*, 1993), and more recent immunopeptidomics studies which showed HLA-B, and its corresponding peptide repertoire (Chong *et al.*, 2017; Komov *et al.*, 2018; Javitt *et al.*, 2019; Goncalves *et al.*, 2021; Olsson *et al.*, 2021), experiences a stronger upregulation than HLA-A under IFNγ exposure.

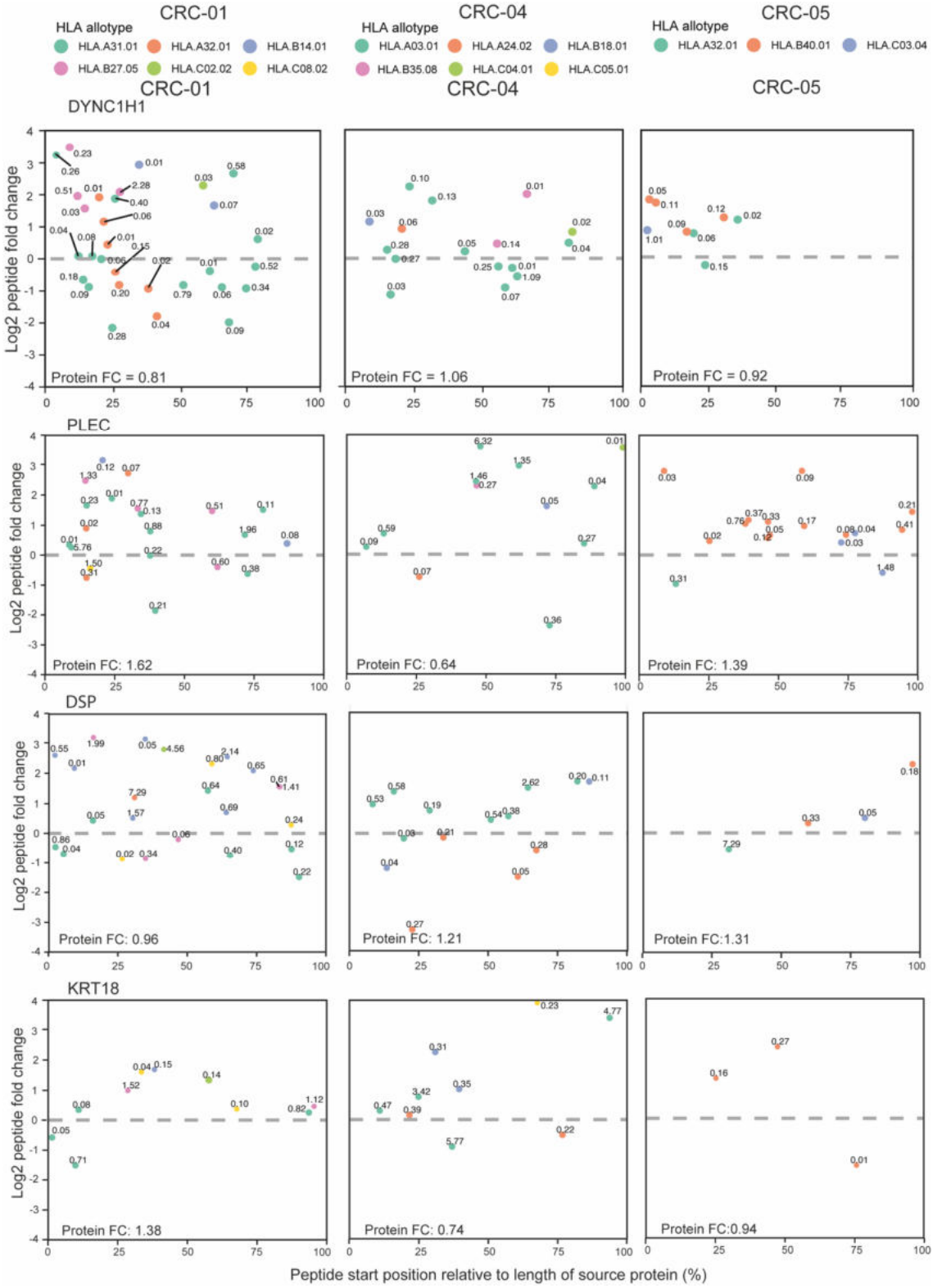
It is known that under normal conditions Tsn abundance is lower relative to peptide abundance. This reduces the peptide editing of Tsn, resulting in lower affinity peptides binding to HLA-I and be presented on the cell surface. Whereas, when there is higher Tsn abundance under IFNγ exposure, the peptide editing capability increases, resulting in the presentation of higher affinity peptides (Boulanger *et al.*, 2022). Therefore, I hypothesised that under the influence of IFNγ treatment, which induced upregulation of APM and PLC genes (including Tsn – gene name TAPBP), the peptides with lower affinity for HLA-I would experience relative downregulation. To scrutinise this, I first analysed the predicted binding affinities of all upregulated and downregulated peptides and found no significant difference between the groups. Following this, I defined the most increasing peptides (MIPs) as the top 10^{th} percentile of peptide FC data, and most decreasing peptides (MDPs) as the bottom 10^{th} percentile peptide data, derived from low fold-change proteins (0.5-2X FC) (Figure 2.6D). It was necessary to analyse the peptides derived from proteins with low FC to be able resolve peptide-specific factors that influenced the peptide abundance changes. I compared the NetMHCpan4.1-predicted BA rank of MIPs and MDPs and found similar distribution of BA ranks, with no significant difference between medians ($p = 0.9708$, Mann-Whitney test). This suggested that peptide binding affinity for HLA did not influence peptide up- or down-regulation through IFNγ.

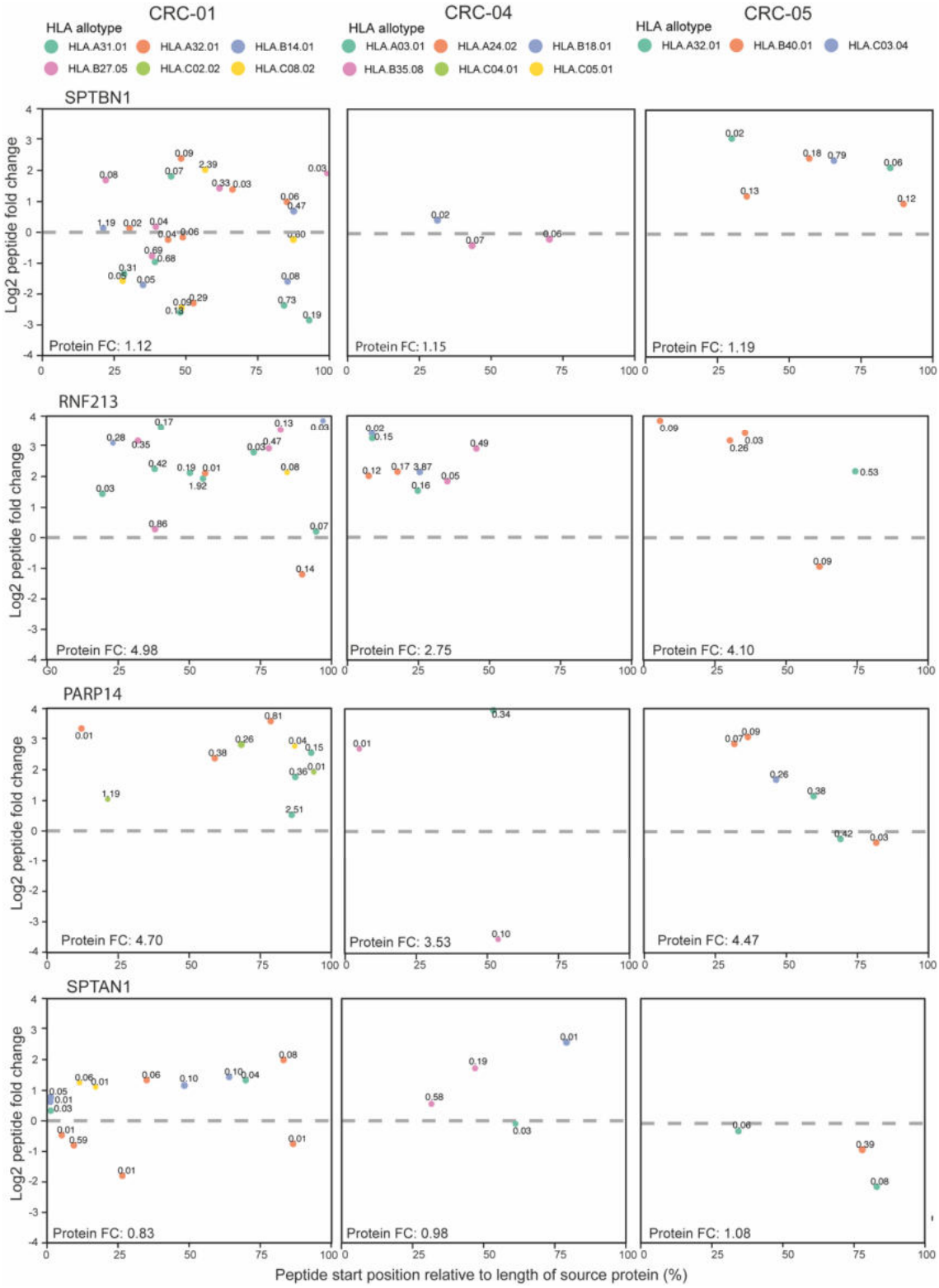
4.2.4.2 Analysis of peptide position protein on peptide abundance changes

The next question was whether the peptide position within a protein could be one of the factors that could influence peptide up- or downregulation. To analyse this, the abundance changes of multiple peptides derived from long proteins were plotted against their relative peptide start position, with 0 depicting the first translated amino acid (Figure 2.7). This was done to control for differences in protein abundance, as any changes in protein manufacture or turnover with IFNγ treatment should affect peptides similarly. However, this approach was limited in that it was not able to control for difference in protein isoforms, which may be relevant for some peptides/proteins. Analysis of 10 long proteins, for which there were the most peptides across the 3 PDOs, showed that some peptides derived from the same protein increased in abundance with IFNγ treatment, whilst others decreased in abundance. This was further seen to be independent of HLA allele and allotype, indicating that differential HLA-A and HLA-B upregulation was not the sole determinant of peptide up- or downregulation. This illustrated that additional peptide-specific factors play a major role in determining peptide up- or downregulation.

Defective ribosomal products (DRiPs) are one of the sources of HLA-I peptidome. It has been defined that DRiPs are rapidly generated from the pioneer round of translation, premature termination of translation, nonsense mediated decay, or protein misfolding. Some publications have specified that, in unperturbed conditions, a large proportion of HLA-I peptides are derived from the N-terminus of proteins, specifically due to the premature termination of translation or nonsense mediated decay (Yewdell, Antón and Bennink, 1996; Reits *et al.*, 2000; Yewdell and Nicchitta, 2006; Apcher *et al.*, 2011; Yewdell, 2011). It is known that IFNγ causes large changes in the translation of proteins, therefore, I sought to investigate if a bias towards the N-terminus of proteins for peptide derivation could be observed, and if this bias is enhanced under IFNγ exposure. The analysis of the 10 long proteins, aside from upregulated peptides clustering close to the N-terminus of DYNC1H1 in CRC-05, showed no systematic bias of peptides being derived from the N-terminus of the protein.

To assess a larger number of data points, the frequency of all MIPs and MDPs were plotted against their relative start position in their source protein (Figure 2.8A). This showed no overrepresentation of MIPs or MDPs originating towards the earlier relative positions in the protein. The plots illustrated more MIPs than MDPs were derived from the first 10-30% of the protein length, and more MDPs derived from 50-90% of way through the protein length, however, these differences were modest which is unlikely to be responsible for such dramatic differences in peptide abundances. To further confirm whether these peptides are derived from very early in the protein length, and could therefore be considered DRiPs, the absolute amino acid start positions for MIPs and MDPs were plotted. This showed only a very small increase in the frequency of MIPs derived from 50-150 and 200-250 amino acids in to the protein length, relative to the MDPS (Figure 2.8B), suggesting that location within the protein had little effect on peptide production between MIPs and MDPs. Thus, peptide location within the protein had little effect on peptide generation between untreated and IFNγ conditions, which agrees with another immunopeptidomics study (Komov *et al.*, 2021). This suggested peptide-specific factors must play a major role in determining whether a peptide is up- or downregulated through IFNγ.


CRC-01
HLA allotype
HLA.A31.01
HLA.A32.01
HLA.B14.01
HLA.B27.05
HLA.C02.02
HLA.C08.02
CRC-04
HLA allotype
HLA.A03.01
HLA.A24.02
HLA.B18.01
HLA.B35.08
HLA.C04.01
HLA.C05.01
CRC-05
HLA allotype
HLA.A32.01
HLA.B40.01
HLA.C03.04
CRC-01
CRC-04
CRC-05
DYNC1H1
PLEC
DSP
KRT18
Log2 peptide fold change
Protein FC = 0.81
Protein FC = 1.06
Protein FC = 0.92
Protein FC: 1.62
Protein FC: 0.64
Protein FC: 1.39
Protein FC: 0.96
Protein FC: 1.21
Protein FC:1.31
Protein FC: 1.38
Protein FC: 0.74
Protein FC:0.94
Peptide start position relative to length of source protein (%)


CRC-01
HLA allotype
HLA.A31.01
HLA.A32.01
HLA.B14.01
HLA.B27.05
HLA.C02.02
HLA.C08.02
CRC-04
HLA allotype
HLA.A03.01
HLA.A24.02
HLA.B18.01
HLA.B35.08
HLA.C04.01
HLA.C05.01
CRC-05
HLA allotype
HLA.A32.01
HLA.B40.01
HLA.C03.04
SPTBN1
RNF213
PARP14
SPTAN1
Log2 peptide fold change
Protein FC: 1.12
Protein FC: 1.15
Protein FC: 1.19
Protein FC: 4.98
Protein FC: 2.75
Protein FC: 4.10
Protein FC: 4.70
Protein FC: 3.53
Protein FC: 4.47
Protein FC: 0.83
Protein FC: 0.98
Protein FC: 1.08
Peptide start position relative to length of source protein (%)

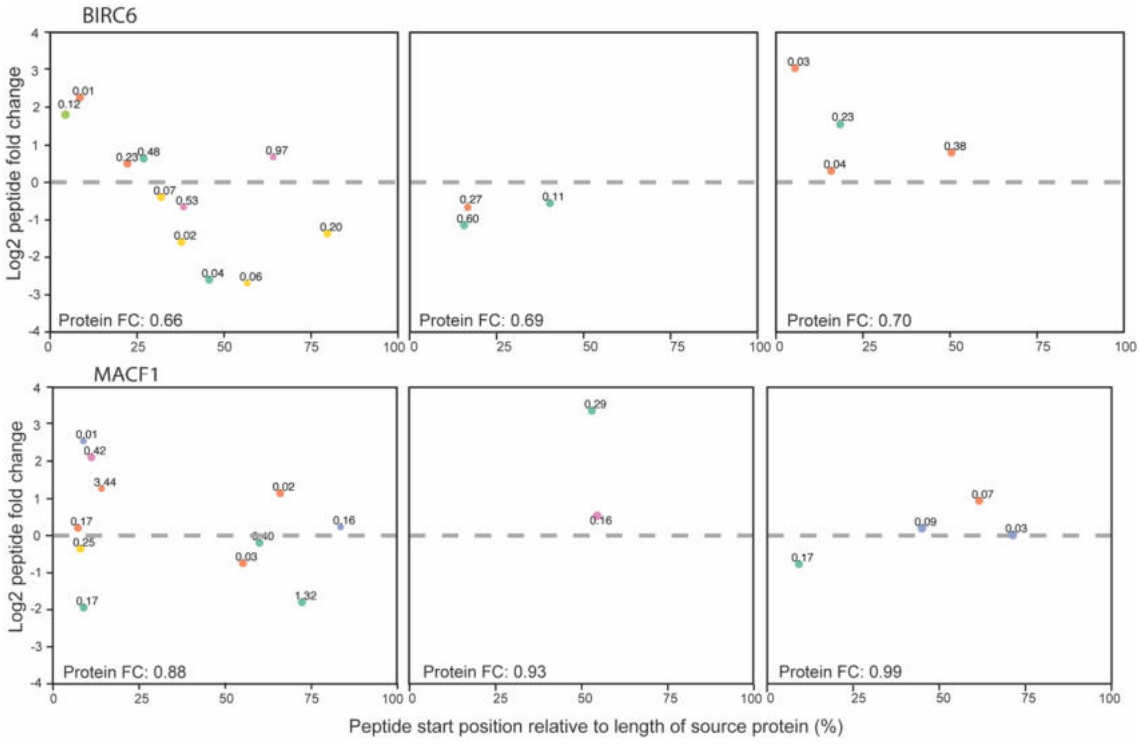


Figure 2.7: Relationship between peptide position in protein and peptide abundance change in long proteins. Log2 change in peptide intensity between untreated and IFNγ-treated conditions for peptides derived from long proteins, plotted against the relative position of the peptide in the protein. Peptides are coloured according to their NetMHCpan4.1-predicted source HLA, with the NetMHCpan4.1 predicted binding affinity rank annotated above. The FC of the protein in each PDO is noted at the bottom of each plot. Proteins analysed were: DYNC1H1, PLEC, DSP, KRT18, SPTBN1, RNF213, PARP14, SPTAN1, BIRC6, MACF1.

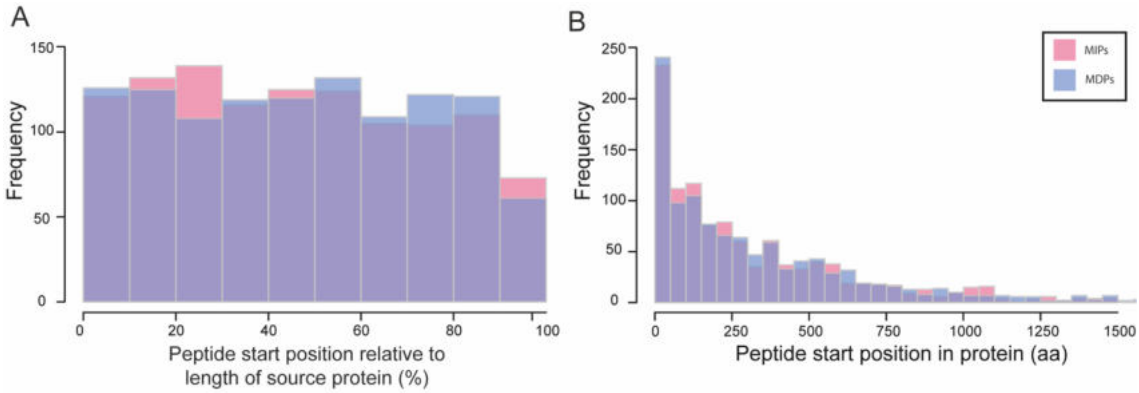


Figure 2.8. Relationship between peptide position in protein and peptide abundance change in MIPs and MDPs. A: Frequency distribution of peptide relative start positions for all MIPs against all MDPs. **B:** Frequency distribution of peptide absolute start positions for all MIPs against all MDPs.

4.2.4.3 *Influence of peptide amino acid composition on peptide abundance changes under IFNγ*

I then focused on the analysis of peptide-specific factors by considering whether the amino acid composition of MIPs and MDPs were associated with the observed changes in abundance under IFNγ conditions. As shown above, different HLA allotypes can be upregulated by different magnitudes. Therefore, to ensure resolution of differences in peptide characteristics which were independent of changes in HLA allotype expression, peptides were first separated in to HLA allotypes based on their NetMHCpan4.1 prediction. Then, for each HLA allotype separately, the intensities of individual peptides were normalised by the FC in total peptide intensity for that allotype, between untreated and IFNγ treated conditions (Figure 2.9A). Next, MIPs and MDPs (again defined as the top and bottom 10th percentile of the immunopeptidomics FC values, from low FC proteins) were selected from each HLA allotype (Figure 2.9B), and the individual MIP and MDP groups were then combined to generate a model list. This approach ensured the same number of up- and downregulated peptides were selected from each HLA allotype (as depicted in Table 4.2), ensuring the amino acid composition of each anchor reside were essentially 'neutralised' and so would not generate a difference value, allowing resolution of differences in MIP and MDP characteristics that are independent of HLA-specific changes (Figure 2.9).

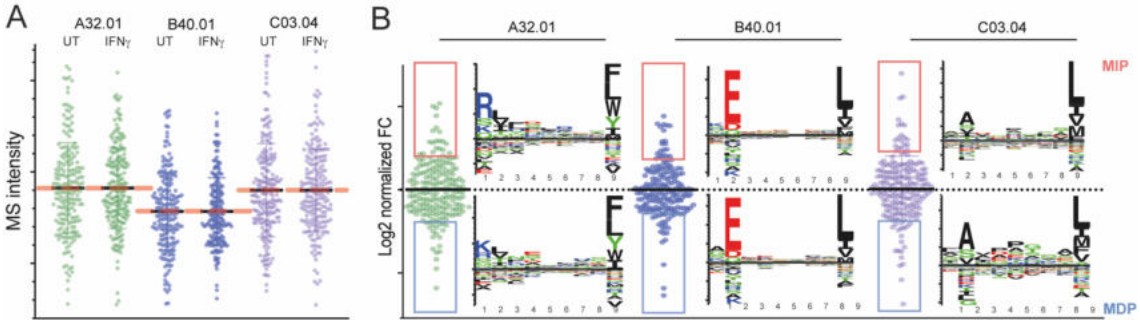


Figure 2.9: Peptidomics normalization and peptide selection schematic, illustrated on CRC-05. A: For the analysis of the amino acid composition of highly up- or downregulated peptides, peptide intensities were further normalized within each HLA allotype; intensity of each of the IFNγ peptides was divided by the FC in total peptide intensity for that allotype, between untreated and IFNγ-

treated conditions. **B:** A schematic of the peptide selection process on the 3 allotypes of CRC-05. Peptides were allocated to the HLA allotype for which they had the lowest BA rank. To select Most Increasing Peptides (MIPs) and the Most Decreasing Peptides (MDPs), the top and bottom 10th percentile of the immunopeptidomics FC data, from low FC proteins, for each allotype. This ensured there was the same number of increasing and decreasing peptides for each allotype so the anchor resides at position 2 or 9 would be neutralised, as indicated by the MixMHCp2.1 plots of each peptide subset. The MIP and MDP lists were then combined, and the differences in peptide sequence analysed.

Category	PDO line	Source HLA	Peptide count	Median FC	Mean FC
MIP	CRC-01	HLA.A31.01	199	11.35	41.92
MIP	CRC-01	HLA.A32.01	125	15.28	34.32
MIP	CRC-01	HLA.B14.01	116	18.74	30.45
MIP	CRC-01	HLA.B27.05	98	19.23	33.88
MIP	CRC-01	HLA.C02.02	25	13.31	27.31
MIP	CRC-01	HLA.C08.02	57	9.58	12.84
MIP	CRC-04	HLA.A03.01	184	3.09	11.54
MIP	CRC-04	HLA.A24.02	79	3.88	13.93
MIP	CRC-04	HLA.B18.01	31	3.61	13.01
MIP	CRC-04	HLA.B35.08	56	5.03	40.67
MIP	CRC-04	HLA.C04.01	11	14.61	23.21
MIP	CRC-04	HLA.C05.01	10	4.58	5.21
MIP	CRC-05	HLA.A32.01	37	27.71	35.17
MIP	CRC-05	HLA.B40.01	90	17.62	36.11
MIP	CRC-05	HLA.C03.04	37	19.13	28.91
MDP	CRC-01	HLA.A31.01	197	0.29	0.28
MDP	CRC-01	HLA.A32.01	125	0.34	0.33
MDP	CRC-01	HLA.B14.01	115	0.38	0.36
MDP	CRC-01	HLA.B27.05	98	0.41	0.40
MDP	CRC-01	HLA.C02.02	24	0.22	0.24
MDP	CRC-01	HLA.C08.02	57	0.28	0.25
MDP	CRC-04	HLA.A03.01	182	0.14	0.13

MDP	CRC-04	HLA.A24.02	78	0.13	0.12
MDP	CRC-04	HLA.B18.01	31	0.19	0.17
MDP	CRC-04	HLA.B35.08	56	0.19	0.16
MDP	CRC-04	HLA.C04.01	12	0.14	0.14
MDP	CRC-04	HLA.C05.01	10	0.27	0.27
MDP	CRC-05	HLA.A32.01	37	0.63	0.60
MDP	CRC-05	HLA.B40.01	90	0.66	0.61
MDP	CRC-05	HLA.C03.04	37	0.52	0.49

Table 4.2. Summary statistics from the peptides which most increase (MIPs) and decrease in intensity (MDPs), derived from proteins with a -1 to +1 log2 fold change, separated by NetMHCpan4.1-attributed HLA: unique peptide count, alongside mean and median of allotype-normalized immunopeptidomics FC.

Once the composite peptide lists for MIPs and MDPs were assembled, the frequency of every amino acid at every position of the 9-mers was calculated, represented as percentage of peptides containing indicated amino acid in each position. Then, the MDP values were subtracted from the MIP values. Studies have shown the specific amino acid composition of the sequences bracketing the peptide at the N-terminus and C-terminus (termed N- and C-terminal extensions) may affect peptide generation (Abelin *et al.*, 2017; Hongo *et al.*, 2019). This is because the presence of particular cleavage motifs facilitates proteasome or peptidase processing, so I also assessed the 9aa N- and C-terminal extensions (Figure 2.10A).

The large number of peptides assessed allowed for a clear difference signal to be observed. To define the amino acids which were most highly changed, I set a threshold of a Z-score of >2.5 – indicating the difference in amino acid abundance was 2.5 times larger than the standard deviation of all the difference values. Most differences across the 27aa sequence were small but the analysis identified 12 positions where the amino acid abundance crossed the defined Z-score threshold (Figure 2.10A – depicted as dotted outlines).

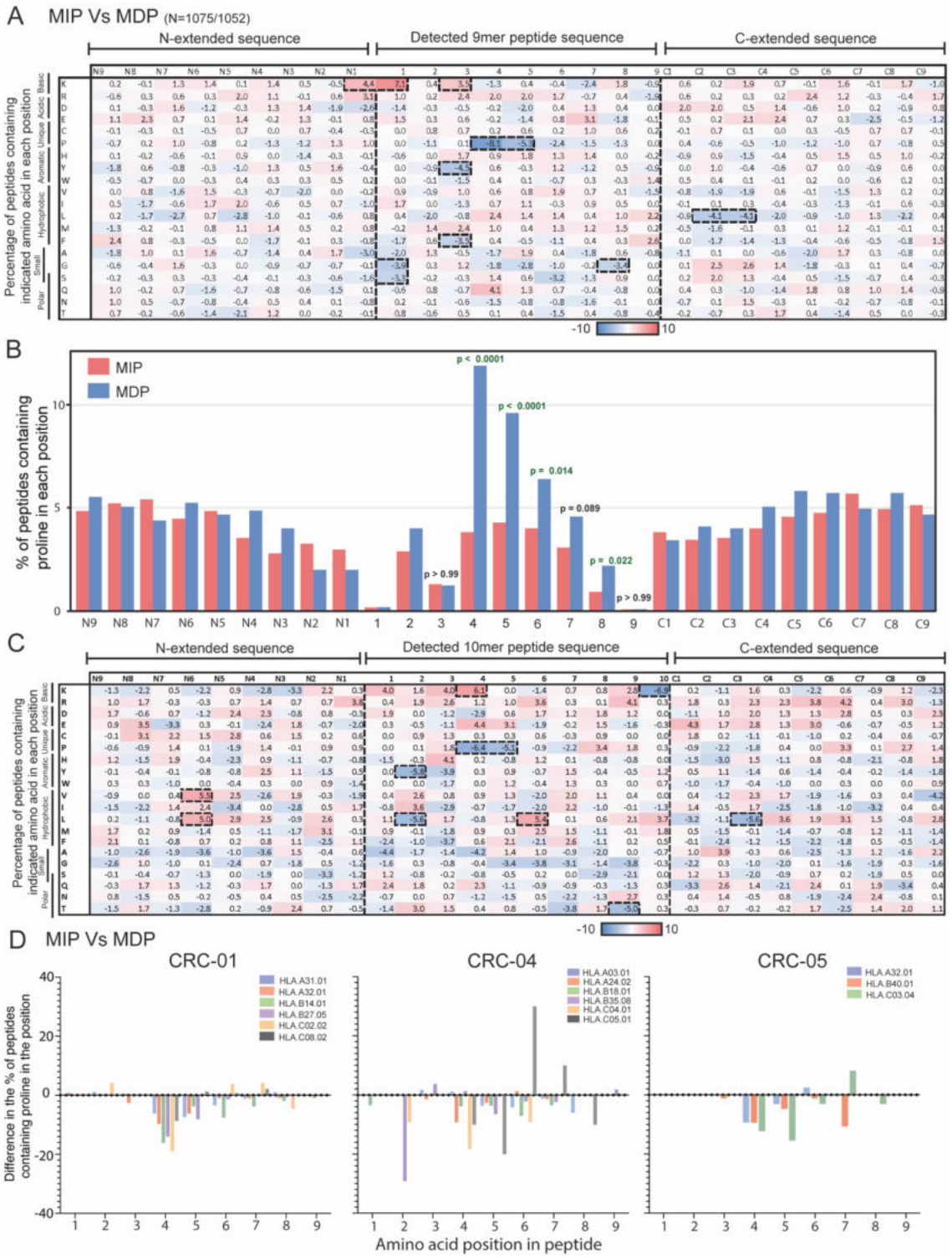


Figure 2.10: Amino acid composition of MIPs Vs MDPs. A: Heatmap of the amino acid composition differences between 9-mer MIPs and MDPs, alongside N- and C extensions. Percentage of peptides with highlighted amino acid in each position were calculated for each group, then the percentage values for the MDPs (N= 1,052) were subtracted from the MIPs (N=1,075). **B:** Percentage of peptides with proline in each position for MIPs (N= 1,075) and MDPs (N=1,052). Statistical analysis for the difference in proline abundance values was performed by Fisher's exact test, with significant results are indicated in green. **C:** Heatmap of the amino

acid composition differences between 10-mer MIPs and MDPs, alongside N- and C extensions. Percentage of peptides with highlighted amino acid in each position were calculated for each group, then the percentage values for the MDPs (N= 333) were subtracted from the MIPs (N= 330). **D:** A graph depicting the change in the percentage of peptides with proline in each position between MIPs and MDPs, split by PDO and by NetMHCpan4.1-predicted source HLA (peptide numbers detailed in Table 4.2).

In position N1 of the N-terminal extension, there was an overrepresentation of the basic amino acid lysine, which had a Z-score of 3.4 (Figure 2.10A). Also notable in the same position, was the less marked overrepresentation of arginine, another basic amino acid (which fell just below the Z-score threshold, with a score of 2.4) – similar to findings which showed activity of immunoproteasomes is higher against basic residues in position N1 (Driscoll *et al.*, 1993). However, although basic residues were enriched in position N1 of the MIPs, which would suggest an increase in tryptic cleavage activity with IFNγ, there was also overrepresentation of lysine in position 1 of the MIPs (with a Z-score of 5.6). This cannot be explained be increased tryptic-like activity of the proteasome, as tryptic activity cleaves to the C-terminus, not the N-terminus, of a basic residue, therefore suggesting these differences may be in enacted by aminopeptidases which cleave away additional amino acids from the N-terminus of the precursor peptide (Abelin *et al.*, 2017). One other possible candidate responsible for overrepresentation of lysine in the MIPs, is aminopeptidase ERAP2 which, as highlighted in the introduction, has a strong preference for cleaving away basic residues from the N-terminus of an HLA-peptide precursor. The transcriptomic and proteomic data showed ERAP2, an IFNγ-responsive gene, increased in expression 2.1-fold, and in protein abundance 2.3-fold. This data therefore suggested that increased ERAP2 abundance may increase the liberation of HLA peptides from their precursor sequences which contain lysine. To confirm this, assays using enzymatic digestion and HPLC analysis of peptides with different amino acids in positions across the peptide would need to be performed (Evnouchidou *et al.*, 2008).

Also, in position 1 of MIPS, the small amino acids glycine and serine were underrepresented (with Z-scores of -3.0, and -2.6, respectively). However, these

changes been may have been due to activity of aminopeptidases; yet, there were no changes in the expression/protein abundance of aminopeptidases with known relevant specificities which could explain these changes in the N-terminus.

In Chapter 3, the data showed increased chymotryptic activity at the C-terminus of peptides in IFNγ conditions, which was displayed through the increased intensity of peptides with a chymotryptic-like cleavage site at their C-terminus, and this was attributed to the immunoproteasome. To further assess the impact of the immunoproteasome, I investigated whether the chymotryptic-like C-termini were enriched amongst the MIPs. 3 of the 7 amino acids preferentially cleaved by chymotryptic-like activity (A, F, I, L, M, V, or Y in position 9) increased in representation amongst the MIPs (42.9%), whereas only 2 out of 13 (15.4%) of the remaining amino acids in position 9 increased. The net increase in the representation of the chymotryptic-like cleavage substrates C-terminal amino acids in position 9 was 1.8%, compared against the net decrease in the tryptic-like cleavage substrates (K or R) of 2.8% which is consistent with the lower abundance, and therefore productivity, of the constitutive proteasome. I further analysed the properties of the amino acids which changed in abundance under IFNγ. The main amino acids underrepresented in P9 were alanine and valine which are small, hydrophobic amino acids. The amino acids overrepresented in P9 were leucine (a larger, branched-chain, hydrophobic amino acid) and phenylalanine (a bulky, aromatic, hydrophobic amino acid). This exchange in preferred amino acids in P9 could be explained by the exchange of PSMB5, which has a preference for small, hydrophobic amino acids, for PSMB8, which has a preference for bulky, hydrophobic amino acids (Huber *et al.*, 2012). This further supports the relative increase in immunoproteasome function under IFNγ.

The only amino acid differences in the C-terminal extension highlighted by the Z-score were the underrepresentation of leucine in the positions C2-C3 (with Z-scores -3.2 in both positions). It is currently understood that the C-terminus of HLA-peptides is exclusively generated by the cleavage activity of the proteasome, with no C-terminal peptidases being described to be involved in antigen processing. Therefore, one could possibly conclude that leucine in positions C2 and C3 could impede peptide production by the

immunoproteasome, although the specific mechanism for this is currently unclear.

The amino acid that displayed the most distinct differences between MIPs and MDPs was proline, which was underrepresented in positions 4 and 5 of the MIPs, with Z-scores of -6.3 and -4.2 (Figure 2.10A). Underrepresentation of proline was also seen in consecutive positions 6-8, but this was not as strong and did not cross the Z-score threshold (-1.9, -1.2, and -1.0 respectively).

To understand how the abundance of proline differs in each position for MIPs and MDPs, the percentage of all peptides containing proline was plotted and statistical testing applied (Figure 2.10B). This showed proline abundance peaked at 12% in the MDPs, and decreased in abundance in consecutive positions 5-8, not being tolerated in position 9, whereas proline stays below 5% in every position of the MIPs. This led to a significant difference for proline abundance in positions 4-6 and 8, which supported the Z-score analysis and highlighted that these are most likely the most biologically relevant differences. As an additional measure, I analysed all 10-mer MIPs and MDPs by the same method to see if the findings from 9-mers could be reproduced. Underrepresentation of proline in the MIPs was again seen in positions 4-7 of the 10-mers, with positions 4 and 5 showing the strongest differences and crossing the Z-score threshold (Figure 2.10C).

To evaluate whether the differences in proline abundance were independent of HLA-allotype, I next analysed difference in proline abundance across every position of the 9-mer MIPs and MDPs for every HLA allotype, in each PDO (Figure 2.10D). The data showed a consistent underrepresentation of proline across all PDOs and HLA allotypes in positions 4 and 5 of the 9-mer sequence, whilst any differences across other positions were less consistent. HLA.B35.08 on CRC-04 was the only case in the 3 PDOs where proline was well tolerated in position 2, due to the importance of proline as a position 2 anchor residue for HLA.B35.08. The MIPs assigned to HLA.B35.08 also displayed a large underrepresentation of proline in position 2, relative to the MDPs. The notable difference in proline in P2 suggested there is the potential for proline depletion across the length of the MIPs, but it is only perceivable in positions where proline is well tolerated (as shown in Figure 2.10B). Altogether, this data suggested that

peptides with proline in positions 4-5 are more likely to be downregulated through IFNγ exposure, with more modest differences also seen at positions 6-8. However, the data from HLA.B35.08 suggested it is possible that proline could be underrepresented in the MIPs, in any position where proline can be well tolerated. The heatmaps also demonstrated that as proline decreased in abundance in the MIP population, there was no direct equivalent increase in another amino acid, instead small increases were dispersed amongst multiple amino acids. Together, this indicated the decrease in proline abundance in the MIPs was a specific consequence of IFNγ exposure.

There were other differences in amino acid abundances which passed the Z-score threshold: tyrosine, phenylalanine, and lysine in P3 with Z-scores -3.5, -2.7, and 2.8 respectively; and glycine in P8 with a Z-score of -2.7. However, as these findings were not reproduced in the original positions, or +/- 1 in the 10-mer analysis (Figure 2.10C), these differences were discounted as potential noise.

I hypothesised, if this underrepresentation of proline could be detected in the MIPs, it is likely that peptides which were exclusive to the IFNγ condition may share this feature. Therefore, I compared the amino acid composition of untreated-exclusive peptides (UEPs) and IFNγ-exclusive peptides (IEPs), derived from proteins with a low FC (log2 -1 to +1 FC). With this analysis HLA abundance could not be controlled for because there were different numbers of UEPs and IEPs attributed to each HLA allotype, producing a relative overrepresentation of HLA-B peptides in the IEPs, and of HLA-A peptides in the UEPs (Table 4.3). The result was that differences in amino acid abundance in anchor positions 2 and 9 were likely due to the differences in HLA abundance (Figure 2.11A).

PDO line	Source HLA	UEPs	IEPs
CRC-01	HLA.A31.01	166	186
CRC-01	HLA.A32.01	97	136
CRC-01	HLA.B14.01	42	212
CRC-01	HLA.B27.05	25	234
CRC-01	HLA.C02.02	18	33
CRC-01	HLA.C08.02	95	113
CRC-04	HLA.A03.01	231	305
CRC-04	HLA.A24.02	155	119
CRC-04	HLA.B18.01	73	91
CRC-04	HLA.B35.08	141	126
CRC-04	HLA.C04.01	43	60
CRC-04	HLA.C05.01	25	63
CRC-05	HLA.A32.01	38	60
CRC-05	HLA.B40.01	52	175
CRC-05	HLA.C03.04	28	51

Table 4.3. Unique peptide count for untreated-exclusive peptides (UEPs) and IFNγ-exclusive peptides (IEPs) from our 3 CRC PDOs, separated by NetMHCpan4.1-attributed HLA.

The heatmap of IEPs Vs UEPs mirrored the underrepresentation of proline in position 4-7 of the IEPs when compared against the UEPs (Figure 2.11A). The difference only crossed the Z-score threshold in position 4, however, differences in positions 4, 6, and 7 were deemed statistically significant by Fisher's exact test (Figure 2.11B). The distribution of proline abundance in the 27aa sequence of the UEPs appeared highly similar to the MDPs, adding confidence to the findings. To confirm these findings could be found across the different PDOs and HLA allotypes, and also to control for differential HLA allotype representation, the difference in the percentage of proline in each position was plotted for the condition-exclusive peptide groups, split in to their source HLA (Figure 2.11C). Due to the lower peptide numbers for some HLA allotypes there was more noise within the data, but IEPs showed an underrepresentation of proline abundance in position 4 across the HLA allotypes. Furthermore, there was also underrepresentation of the percentage of amino acids containing proline in

position 2 of the IEPs attributed to HLA.B35.08, consistent with the findings from the MIPs on HLA.B35.08.

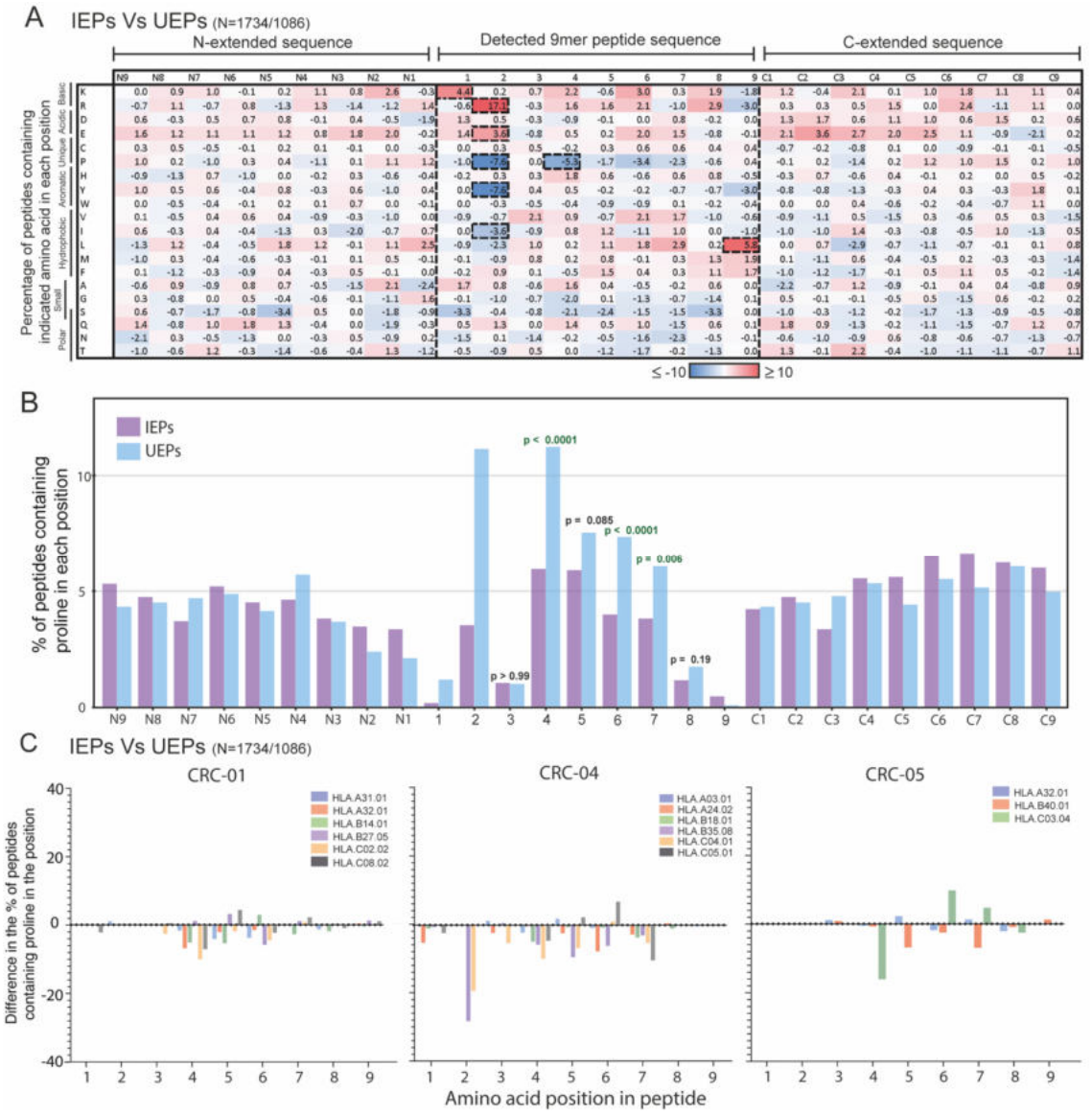


Figure 2.11: Amino acid composition of IEPs Vs UEPs. A: Heatmap of the amino acid composition differences between 9-mer IEPs and UEPs, alongside N- and C extensions. Percentage of peptides with highlighted amino acid in each position were calculated for each group, then the percentage values for the UEPs (N= 1,066) were subtracted from the IEPs (N=1,734). **B:** Percentage of peptides with proline in each position for IEPs (N=1,734) and UEPs (N= 1,066). Statistical analysis for the difference in proline abundance values was performed by Fisher's exact test, with significant results are indicated in green. **C:** A graph depicting the change in the percentage of peptides with proline in each position between IEPs and UEPs, split by PDO and by NetMHCpan4.1-predicted source HLA (peptide numbers detailed in Table 4.3).

Overall, the analysis approach used to analyse peptide-specific changes in the immunopeptidome was able to highlight that exposure of CRC PDOs to IFNγ resulted in a higher abundance of peptides with lysine and arginine in position N1, and lysine in position 1, and a decreased abundance of peptides with small amino acids in position 1. Furthermore, the most striking finding was decrease of proline in the core of the peptide, most strongly in positions 4 and 5 where proline was most well tolerated across all HLA allotypes, allowing this pattern to be evident across all 3 PDOs and their cognate HLAs. The robustness of the finding suggested that the proline depletion effect is conserved across biological models.

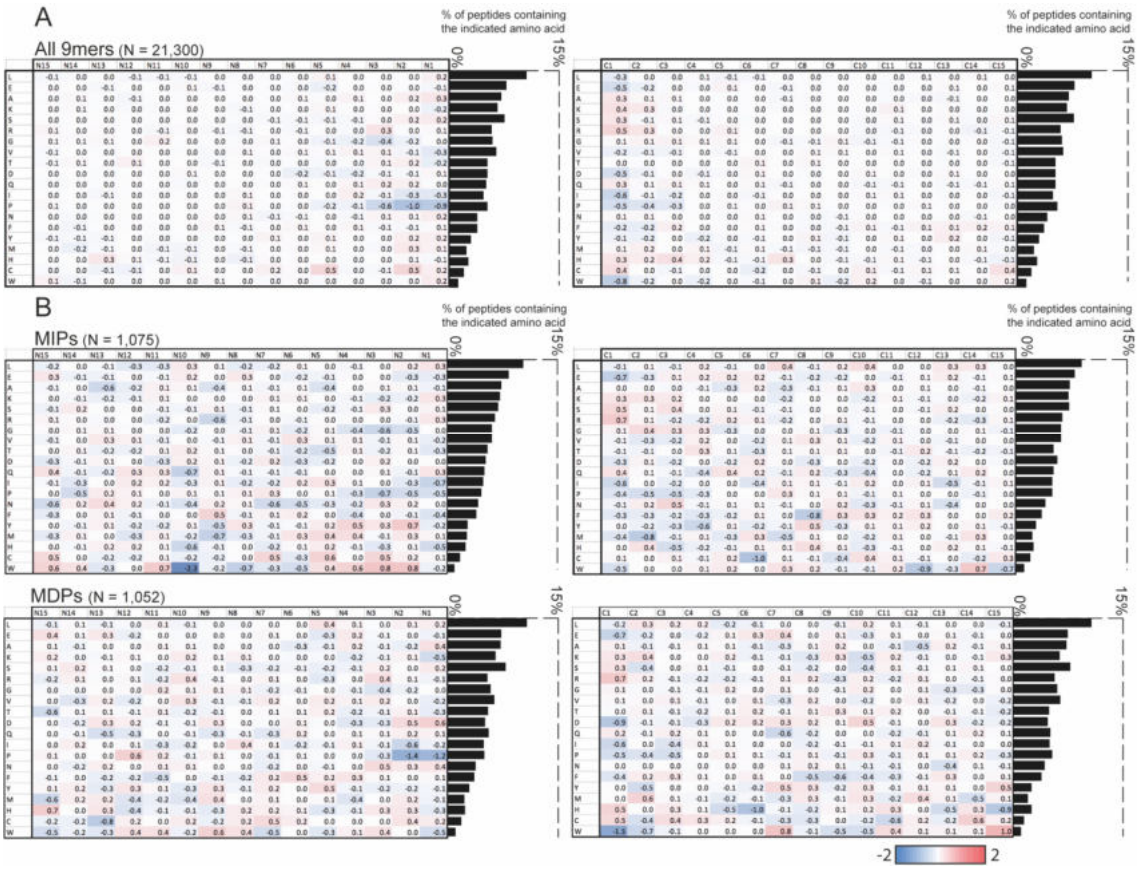


Figure 2.12: Enrichment/depletion of N-terminal or C-terminal extended amino acids in MIPs Vs MDPs. A heatmap depicting the enrichment/depletion of each amino acid relative to the median value of the abundance of each amino acid. This was calculated by dividing the % of peptides containing the indicated amino acid in the indicated position by the median % abundance across the 15-mer peptide extension. The bar chart represents the median % abundance value of each amino acid across the 15-mer peptide extension, with the dotted line demarcating 15%. **A:** All 9-mers. **B:** MIPs and MDPs.

The initial approach of analysing the N- and C-terminal extended sequences showed very few differences in amino acid abundance between the MIPs/MDPs and IEPs/UEPs. Studies from other groups have highlighted the importance of the surrounding amino acid context for peptide cleavage (Abelin *et al.*, 2017; Hongo *et al.*, 2019), therefore an additional approach was trialled to assess for differences in peptides which were either highly up- or down-regulated under IFNγ conditions. To do this, separately for each N- and C-terminal extended 15aa sequences: the percentage of peptides with each amino acid in each position was calculated, then the median abundance value was taken across the 15aa sequence. To calculate the log2 enrichment/depletion score for each amino acid, each individual abundance value was divided by the median abundance value for that amino acid, and the log2 value taken of that.

I first wanted to analyse how the enrichment/depletion profile of each amino acid appeared on all MS-detected 9-mers (Figure 2.12A). This showed proline to be most strongly depleted in positions N1-3 and C1-3, with the depletion approximately 2-times stronger at the N terminus. The repertoire of HLA peptides is, in part, defined by biases in peptide processing, with many peptidases having a bias against proline in P1-3 (Vanhoof *et al.*, 1995; Shimbara *et al.*, 1998; Hongo *et al.*, 2019), and biases in transport to the ER, as proline is not well tolerated in P1-3 by TAP1/2 (Uebel and Tampé, 1999; Lehnert and Tampé, 2017). The stronger depletion of proline at the N-terminus than C-terminus, which is thought to be generated exclusively through the action of the proteasome (Abelin *et al.*, 2017; Chong *et al.*, 2017), suggested that peptidases may be more sensitive to proline presence, and that the effect of this bias may be compounded by the biases of TAP1/2. Isoleucine is also depleted in positions N1-2 and C1 relative to the other positions in the 15-mer, suggesting both proline and isoleucine may impede peptide cleavage. Arginine and lysine were slightly enriched in positions C1-2, suggesting these amino acids may positively influence peptide cleavage by the proteasome. These findings were highly similar to that of a previous study investigating positional amino acid biases in MS-detected peptides of cell lines MKN45, KHM2B, SW480, DU145 (Hongo *et al.*, 2019).

Next, I analysed the MIP and MDP datasets to see whether these biases changed between IFNγ-induced up- or downregulated peptides (Figure 2.12B). It should be noted that these plots have slightly more noise than all 9-mers as they are of a smaller sample size (21,300 peptides vs 1,075/1,052 for MIPs/MDPs). The depletion of isoleucine and the enrichment of arginine and lysine and was similar between all 9-mers, the MIPs, and MDPs. However, the depletion of proline was different between the 9-mers, the MIPs, and MDPs. The N-terminal depletion of proline in the MIPs was lesser than the all 9-mers, with -0.7, -0.5, and -0.5 compared against -0.6, -1.0, -0.9 in N3-N1. In contrast, the C-terminal depletion of proline in the MIPs was similar to the all 9-mer analysis, with -0.4, -0.5, and -0.5 compared against -0.5, -0.4, -0.3 in C1-C3. Furthermore, the depletion of proline in the MDPs was slightly stronger at the N-terminus than the all 9-mer analysis, with -0.3, -1.4, and -1.2, but again consistent at the C-terminus with -0.5, -0.4, and -0.5 C1-C3. This data illustrated that only the proline abundance in the N-terminal extension influences peptide abundance changes in MIPs Vs MDPs. The MIPs displayed the lowest bias against proline in positions N1-3, and the MDPs displayed the highest bias against proline in those positions. Because this bias affected only the N-terminal extension, it is likely not to be explained by changes in the cleavage specificities between the constitutive and immunoproteasome.

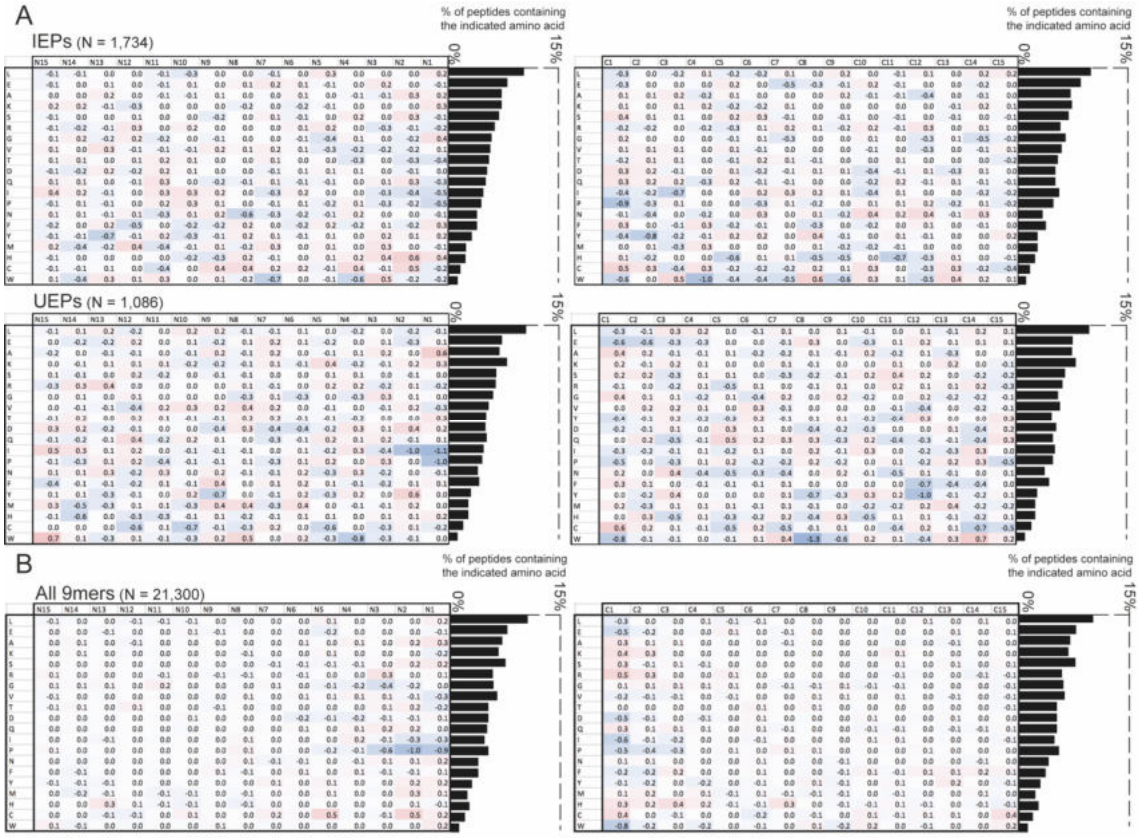


Figure 2.13: Enrichment/depletion of N-terminal or C-terminal extended amino acids in IEPs Vs UEPs. A heatmap depicting the enrichment/depletion of each amino acid relative to the median value of the abundance of each amino acid. This was calculated by dividing the % of peptides containing the indicated amino acid in the indicated position by the median % abundance across the 15-mer peptide extension. The bar chart represents the median % abundance value of each amino acid across the 15-mer peptide extension, with the dotted line demarcating 15%. **A:** IEPs and UEPs. **B:** All 9-mers.

I next wanted to confirm whether this pattern held true for the IEPs Vs UEPs, as in the previous analysis these groups had similar characteristics (Figure 2.13A). In positions N1-3 proline depletion was weakest in the IEPs (-0.5, -0.2, and -0.3), compared against that of the UEPs (-1.1, -1.0, and -0.4) and all-9mers (-0.9, -1.0, and -0.6). In positions C1-3 proline depletion appeared patchier, but was similar in the UEPs and the all 9-mers analysis. In the IEPs the proline depletion was higher in position C1, but more similar in position C2.

Collectively, this data indicated that although proline remained depleted across the first 3 N- and C-terminal amino acids for both MIPs/MDPs and IEPs/UEPs, the bias against proline in position N1-3 is less strong for the MIPs and IEPs. This

illustrated that proline can be tolerated more in positions N1-3 in IFNγ condition, demonstrating increased peptide liberation from N-terminal proline-adjacent sequences. Combining the 15aa extension enrichment/depletion analysis with the full 27aa abundance difference heatmaps showed that for the peptides which increase most intensity, or were presented exclusively in IFNγ conditions, there is a marginally decreased bias against proline in the positions N1-3, but a stronger bias against proline within the 9-mer peptide.

To further validate the finding of proline depletion within the peptide sequence of the most upregulated peptides, I performed the identical 9-mer analysis on an independent dataset. I selected the immunopeptidomics and proteomics datasets from the breast cancer cell line MDA-MB-231 as these cancer cells had also been treated with IFNγ for 48 hours, which was important to keep consistent as peptidome remodelling is likely a dynamic process linked to translation and protein degradation (Figure 2.14A). The same MIP Vs MDP analysis was applied to the 9-mers and this confirmed the underrepresentation of proline in positions of 4-6 and 8 of the MIPs compared against the MDPs, and also in positions 2 and 3 (Figure 2.14B). The differences in positions 2-6 and 8 crossed the Z-score threshold, and Fischer's exact test showed differences in positions 1-6 and 8-9 to be statistically significant (Figure 2.14C). There was also subtle overrepresentation of basic amino acids lysine and arginine, and modest underrepresentation of small amino acids in position 1 of the MIPs, however, none of these differences reached the Z-score threshold. Next, the identical UEP Vs IEP analysis was performed, and it again showed an underrepresentation of proline across the 9-mer UEPs, in this case including position 7. However, only positions 4-6 crossed the Z-score threshold (scores of -5, -3, and -3.8 respectively). The less distinct differences in positions 2-3 and 7-8 may be due to the small sample number of 96 UEPs, compared against 10,699 IEPs. In conclusion, the validation of proline underrepresentation in the peptides which most intensely increase in abundance supported the hypothesis that peptides with proline residues in positions 4-5 are specifically downregulated under IFNγ exposure. The validation dataset also supported the theory that proline presence across the entire peptide length could be responsible for peptide downregulation in IFNγ conditions, but the link is most apparent where proline is best tolerated in the peptide binding grooves of the different HLA allotypes.

4.2.4.4 *Validation of proline influence on peptide abundance changes on external dataset*

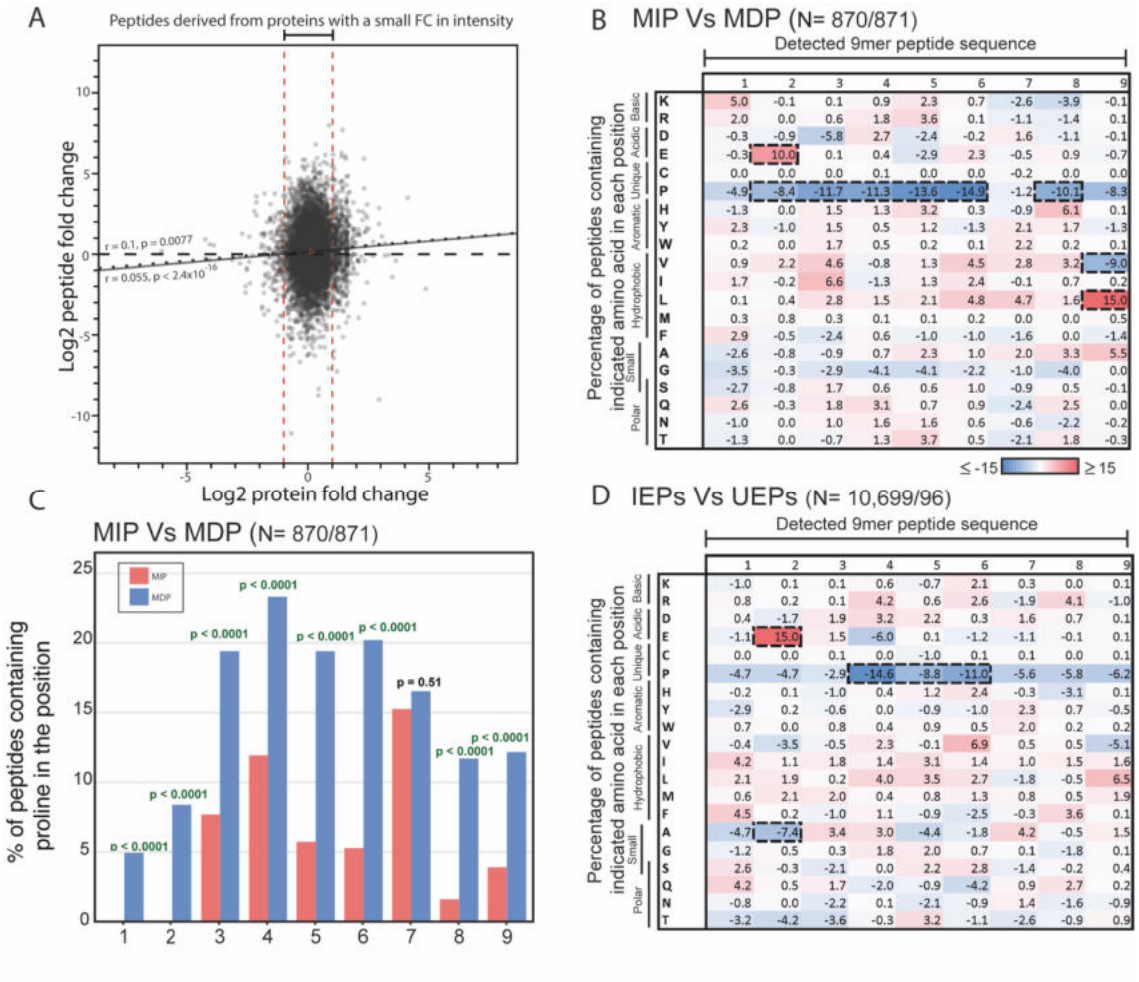


Figure 2.14: Validation in differences in amino acid preferences between untreated and IFNγ-treated conditions using the datasets from *Goncalves et al*. A: Correlation of protein FC between untreated and IFNγ-treated conditions, against normalised immunopeptidomics FC. Regression lines for all peptides indicated by a solid black line, regression line for peptides from proteins with <0.5 or >2 FC indicated with a dotted black line. **B:** Heatmap of the amino acid composition differences between 9-mer MIPs and MDPs. Percentage of peptides with highlighted amino acid in each position were calculated for each group, then the percentage values for the MDPs (N= 871) were subtracted from the MIPs (N=870). **C:** Percentage of peptides with proline in each position for MIPs (N=870) and MDPs (N= 871). Statistical analysis for the difference in proline abundance values was performed by Fisher's exact test, with significant results are indicated in green. **D:** Heatmap of the amino acid composition differences

between 9-mer IEPs and UEPs. Percentage of peptides with highlighted amino acid in each position were calculated for each group, then the percentage values for the UEPs (N= 96) were subtracted from the IEPs (N= 10,699).

Once the results had been validated on an external dataset, I analysed data from our PDOs to critique if the proline depletion within MIPs and IEPs, and increased peptide liberation from N-terminal proline-containing sequences, could be linked to changing expression/abundance of prolyl peptidases. Proline is unique in that it is the only amino acid with a cyclic side chain which is connected to the protein backbone twice; therefore, proline inclusion imparts remarkable conformational rigidity on protein sequences, and can confer resistance to many peptidases (Vanhoof *et al.*, 1995; Männistö and García-Horsman, 2017). Prolyl peptidases therefore have a specialist function in the cleavage of proline-containing peptide sequences. DPP9 is an dipeptidyl aminopeptidase that cleaves off Xaa-Proline dipeptides from the N-terminus of proteins/peptides, that has been shown to play a role in antigen processing (Geiss-Friedlander *et al.*, 2009; Justa-Schuch *et al.*, 2016). Prolyl endopeptidase (PREP) is a cytoplasmic endopeptidase that cleaves within a <30aa peptide, to the C-terminus of proline. Although it is thought to be involved in the digestion of neuropeptides, its role is not yet fully understood (Brandt, Scharpé and Lambeir, 2007; García-Horsman, Männistö and Venäläinen, 2007; Männistö and García-Horsman, 2017). For these peptidases to have caused the underrepresentation of proline in the MIPs, it would be expected that there would have been increased expression of the peptidases to cause boosted proline dipeptide removal or internal cleavage of proline-containing peptides, facilitating peptide destruction. The expression analysis of DPP9 and PREP showed both genes to increase in expression, but they both decreased in protein abundance, meaning it is unlikely that prolyl peptidases are responsible for the underrepresentation of proline in the MIPs. Furthermore, the activity of proline peptidases can also be restricted by the cis/trans-isomerisation state of proline within the peptide chain (Vanhoof *et al.*, 1995), therefore I analysed the protein abundance of proline isomerases PPIA, PPIB, PPIC, PPID, and PIN1. There was no distinct pattern of up- or downregulation of any of these proteins across the 3 PDOs that could explain changes in proline cleavage (data not shown). Although protein abundance is not always linked to total amount of enzymatic activity, the lack of patterns in gene expression and protein abundance

data for these proline peptidases and isomerases, suggested that these factors may not be the reason for downregulation of proline-containing peptides under IFNγ exposure.

4.2.4.5 Analysis of proline impact of peptide affinity to, and stability with, HLA

As mentioned above, Proline has a cyclic side chain which bestows the amino acid with conformational rigidity, inducing a 'kink' of the bracketing amino acids away from the proline residue (Vanhoof *et al.*, 1995). Therefore, proline is known as a 'helix breaker' as it can destabilise secondary structures like alpha helices and beta pleated sheets (Perez *et al.*, 2019). With this, I chose to investigate whether proline substitution would affect peptide affinity to, or stability with, the PDOs encoded HLA allotypes. Although peptide-HLA binding affinity and stability are intrinsically associated, the stability of the peptide-HLA complex may be more important, as many studies have shown that stability is correlated with immunogenicity (Lipford *et al.*, 1995; Nicholls *et al.*, 2009; Spierings *et al.*, 2009). Peptide-HLA stability allows for successful transport to the cell surface and a longer interaction with T-cell receptors at the immune synapse, which is important to enable sufficient signalling to facilitate full T-cell activation.

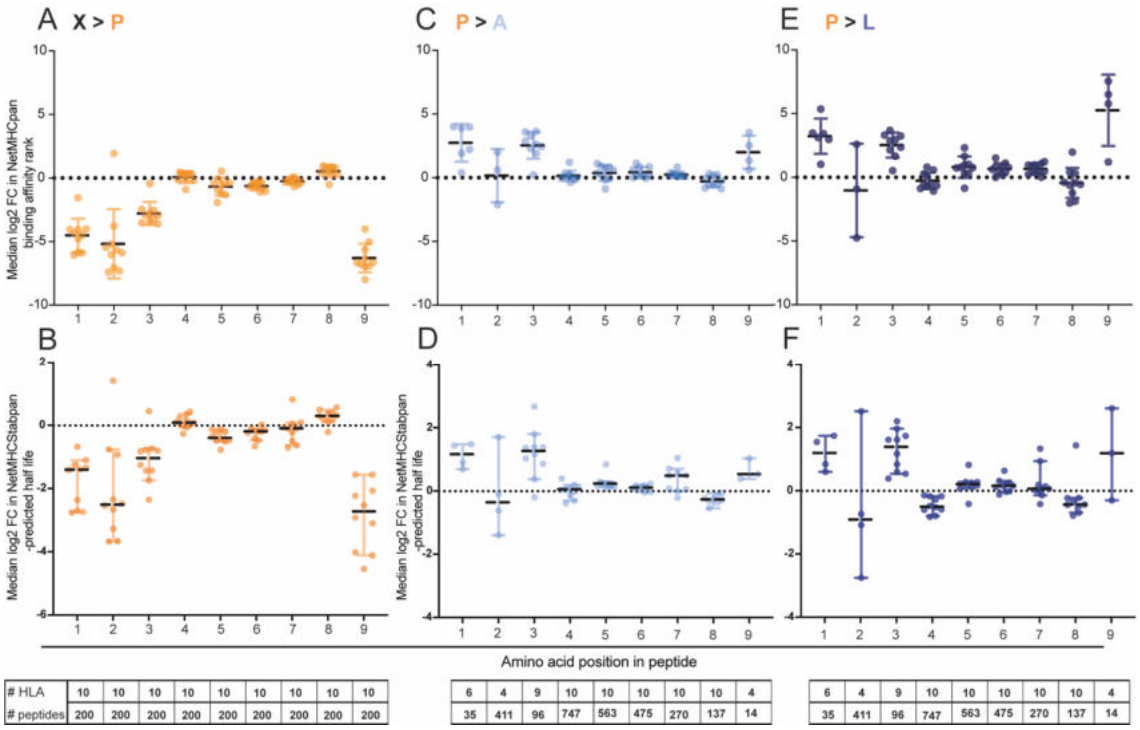


Figure 2.15: ***Simulating the impact of amino acid replacements on peptide affinity and binding stability to their cognate HLAs.*** **A:** Median log2 FC value of NetMHCpan4.1-predicted binding affinity rank for peptides from each HLA-A and -B from each PDO, when every single individual amino acid, in positions 1-9, from detected peptides, is exchanged with proline (N = 200 sampled peptides per HLA). **B:** Median log2 FC value of NetMHCStabpan1.0-predicted half-life for peptides from each HLA-A and -B from each PDO, when every single individual amino acid, in positions 1-9, from detected peptides, is exchanged with proline (N = 200 sampled peptides per HLA). **C:** Median log2 FC value of NetMHCpan4.1-predicted binding affinity rank for peptides from each HLA-A and -B from each PDO, when proline is exchanged for alanine (in peptides with a single proline) - individual sample numbers annotated on the plots. **D:** Median log2 FC value of NetMHCStabpan1.0-predicted half-life, for peptides from each HLA-A and -B from each PDO, when proline is exchanged for alanine (in peptides with a single proline) - individual sample numbers annotated on the plots. **E:** Median log2 FC value of NetMHCpan4.1-predicted binding affinity rank for peptides from each HLA-A and -B from each PDO, when proline is exchanged for leucine (in peptides with a single proline) - individual sample numbers annotated on the plots. **F:** Median log2 FC value of NetMHCStabpan1.0-predicted half-life, for peptides from each HLA-A and -B from each PDO, when proline is exchanged for leucine (in peptides with a single proline) - individual sample numbers annotated on the plots.

I began by analysing the PDO immunopeptidomics data to understand how proline inclusion in peptide sequences influenced the NetMHCpan4.1-predicted BA rank (Reynisson *et al.*, 2020) to or NetMHCstabpan1.0-predicted stability (Jørgensen *et al.*, 2014; Rasmussen *et al.*, 2016) to their HLA. For each PDO, a sample of 200 MS-detected 9mers was randomly-selected, and every individual amino acid was sequentially replaced with proline, and the impact on NetMHCpan4.1-predicted BA rank and NetMHCstabpan1.0-predicted half-life studied. NetMHCpan4.1 and NetMHCStabpan1.0 predictions indicated replacing amino acids with proline would negatively affect both peptide-HLA binding affinity rank and binding half-life in the anchor residue positions 1-2 and 9, and to a lesser extent, position 3 (Figure 2.15A-B). One noteworthy exception was at position 2 of peptides from HLA-B35.08 on CRC-04, where replacing amino acids with proline increased the binding affinity and stability. This was because proline acts as an anchor residue in P2 on HLA-B35.08, so its addition actually heightened binding affinity and stability. In P4-P8, substituting amino acids for proline had no strong effect on predicted binding affinity and stability, which highlighted there was little amino acid specificity in these amino acid positions in our PDOs.

I next performed the inverse analysis, whereby I selected all peptides containing a single proline and replaced it with alanine or leucine to examine the effect on binding affinity rank and binding half-life (Figure 2.15C-F). Alanine was selected as it doesn't form side-chain interactions and eradicates the conformational characteristics of the proline side chain. Leucine holds similar properties, but it is a larger amino acid, so it is used where size may be an important factor. For most HLA allotypes across the 3 PDOs, proline was infrequently detected in P1-3 and P9, as it is less well tolerated in those anchor positions for most HLA-I. Here again, B35.08 was the exception, providing 98% of the MS-detected peptides with proline in P2 due its preference for proline as its anchor residue. Therefore, proline replacement led to a large decrease in binding affinity and stability in B35.08. In positions 4-8 replacement of proline with alanine or leucine caused only small changes in the binding affinity and stability across all the HLA allotypes. This data reinforced the conclusion that proline in positions P4-P8 does not have a detrimental effect on peptide binding affinity and stability for their

predicted HLA, which therefore cannot explain the disparity in peptide abundance change under IFNγ seen in MIPs and MDPs.

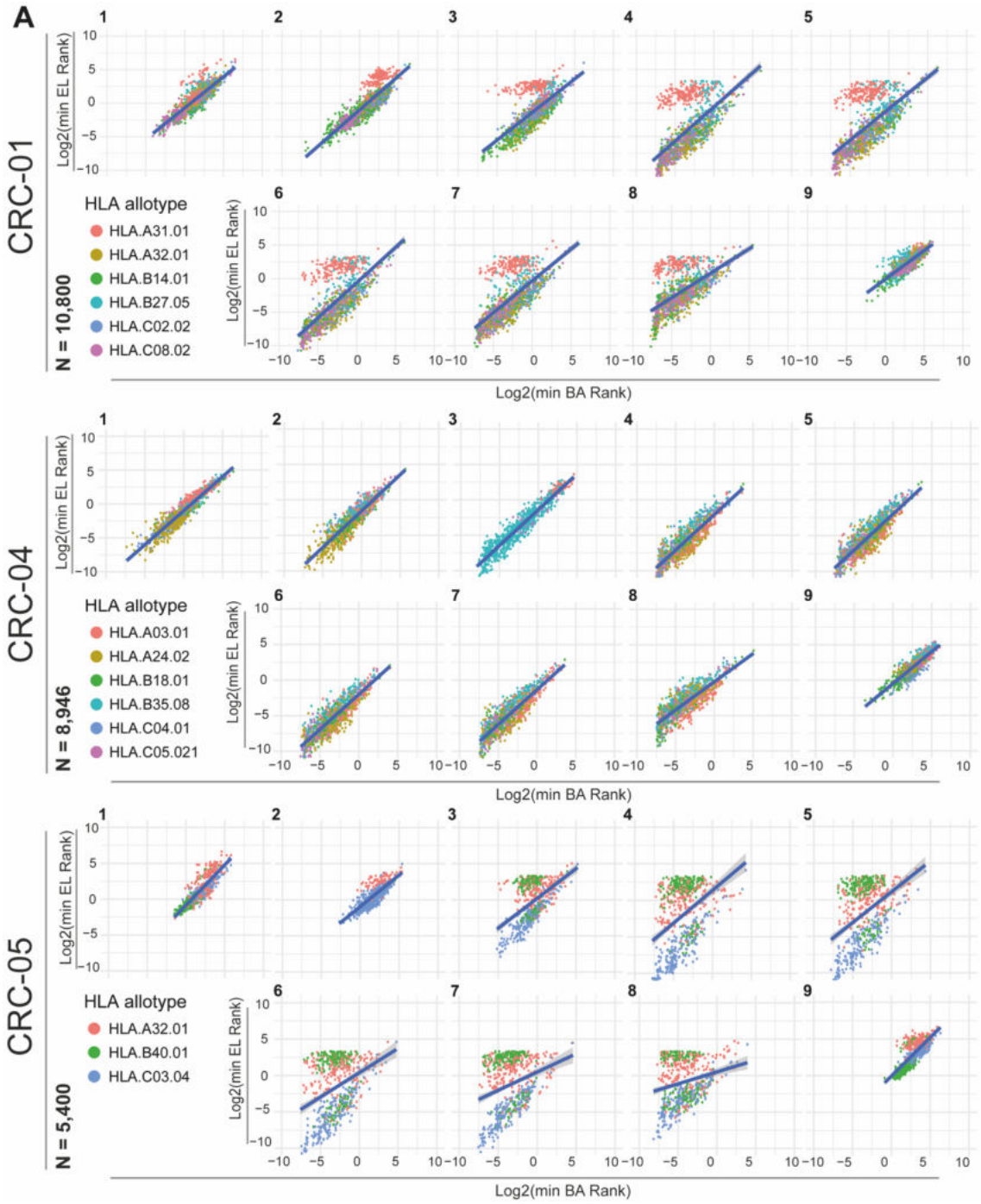


Figure 2.16: Analysing the difference between BA rank and EL rank for proline-containing peptides. A: Sample of up to 200 MS-detected 9-mer peptides per HLA allotype for HLA-A, B and C, where every single individual amino acid in positions 1-9, is exchanged with proline. Log2 NetMHCpan4.1-predicted BA rank plotted against log2 NetMHCpan4.1-predicted EL rank. Peptides coloured according to their NetMHCpan4.1-predicted HLA allotype

(assigned to the HLA allotype with the lowest rank); only peptides predicted to bind the same HLA allotype by both BA rank and EL rank analysed.

I also investigated the relationship between BA rank and EL rank in proline-containing peptides, to understand the difference between predicted binding affinity and the likelihood of eluting the ligand, based on the MS-training datasets (Reynisson *et al.*, 2020). I utilised the same computationally-mutated peptide list as Figure 2.15, whereby a sample of up to 200 MS-detected peptides per HLA allotype had every single amino acid in positions 1-9 exchanged with proline. For CRC-01, the predicted BA and EL rank were higher for positions 1 and 9, indicating low predicted affinity to their cognate HLA, which is explained by the HLA allotypes having a preference against proline in the anchor residue positions 1 and 9. In positions 3-8 the HLA-A31.01 and HLA-B27.05 have a higher EL rank than BA rank, suggesting that although the peptides are predicted to have a higher affinity for the HLA, they are less likely to be detected by MS. CRC-04 displayed a high correlation between BA and EL rank for all peptides, from all HLA allotypes. CRC-05 also demonstrated a high predicted BA and EL rank for positions 1 and 9, but also for position 2. For positions 3-8, most strongly in positions 4-7, most peptides from HLA-A32.01 and HLA-B40.01 had a higher EL rank than BA rank. This again suggested that, despite higher predicted affinity, peptides with proline predicted to bind HLA-A32.01 and HLA-B40.01 are less likely to be detected by MS analysis. The divergence between BA rank and EL rank suggested that for some HLA allotypes, proline in positions 2-8 may supress display of peptides, for reasons beyond HLA affinity. The reasons for such are yet unknown, and are unlikely to be related to MS biases, as it was not seen for every HLA allotype.

4.3 Discussion

When considered together, this work has shown that whilst the quantitative changes in the immunopeptidome seen under IFNγ are modest, the remodelling of the immunopeptidome characteristics is immensely complex and regulated by a multitude of different mechanisms. I first saw that for some peptides, their up- or downregulation under IFNγ exposure was correlated with abundance changes of their source proteins. However, a larger proportion of the peptides experienced changes in cell surface abundance, despite little or no changes in source protein abundance (within log2 -1 to +1). A small portion of this change could be attributed to increased chymotryptic activity from the increased abundance of immunoproteasomes as shown in previous studies (Chong *et al.*, 2017; Newey *et al.*, 2019). Furthermore, the data showed that IFNγ exposure upregulated HLA-B expression and surface abundance to a higher extent than that of HLA-A, which led to an increase in the proportion of peptides presented on HLA-B. Analysis of multiple peptides from long proteins showed that peptides derived from the same protein could show a mix of up- or downregulation, and that this was neither dependent on their bound HLA allotype, HLA binding affinities, nor their location within the protein. This eliminated many possible variables and drove the focus towards resolution of peptide-specific characteristics that affected their abundance under IFNγ exposure.

The most noteworthy and novel finding was the underrepresentation of proline within the core of the peptides which most increase in intensity under IFNγ conditions, or were exclusive to IFNγ conditions. Although proline was most strongly underrepresented in positions 4-6 in our CRC PDOs, the downregulation of proline in P2 of HLA35.08 and across the peptide length of MDA-MB-231 intimated that proline is likely to be underrepresented in every position of the MIPs where proline presence is tolerated in that HLA allotype. To confirm this, immunopeptidomics datasets from other untreated and IFNγ-treated cell lines or PDO lines with more HLA allotypes should be analysed; in this case there were limited choices of appropriate datasets, as I decided to only analyse 48-hour IFN-γ treated datasets as immunopeptidome remodelling is a dynamic process.

The highest tolerance for proline being in position 4 can be explained by the biochemical structure of HLA. HLA-I molecules have 6 binding pockets named A-F which define their binding specificities by accommodating particular amino acid side chains (Saper, Bjorkman and Wiley, 1991); amino acid positions P1 and P2 are defined by pockets A and B, and for some HLA there is a secondary anchor position in P3 or P5/P6, defined pockets C or F, respectively (Nguyen, Szeto and Gras, 2021). This may be why there was an observable decrease in peptide binding affinity and stability present in P3, but with a magnitude of change being lesser than that of P2. Furthermore, the HLA allotypes of our PDOs lacked a P5/6 peptide anchor, shown because replacing amino acids in these positions with proline had little effect on binding affinity and stability, which generated peptide repertoires with no amino acid biases against proline in P4-8. This enabled the HLA to tolerate peptides containing proline in those positions. Previous studies have shown proline presence in the core of a peptide sequence protects that peptide from internal cleavage by endopeptidases and proteasomal cleavage (Shimbara *et al.*, 1998; Abelin *et al.*, 2017). However, known prolyl endopeptidases PREP, DPP3, DPP8, DPP9 were not upregulated by IFNγ treatment in our proteomics dataset to cause increased internal cleavage. Furthermore, prolyl cis/trans-isomerases, which could change the isomerisation of proline to make the peptide more susceptible to cleavage by prolyl peptidases, were also not upregulated. However, it is possible that their enzymatic activity level may have increased under IFNγ treatment instead.

This data informed the hypothesis that the described protective effect of proline was less pertinent under IFNγ conditions. Studies have shown that up to 99% of all proteasome-derived peptides are degraded by peptidases (Reits *et al.*, 2003). In non-perturbed conditions the abundance and activity of APM proteins, alongside supply of peptide-receptive HLA molecules (Komov *et al.*, 2018), are the limiting factors in surface HLA-I presentation. The lower baseline expression of such molecules results in peptides spending more time in the ER before being transported to the cell surface on HLA-I, giving peptides longer exposure times to peptidases, increasing the probability of internal cleavage. Therefore, increased abundance of APM and HLA-I molecules under IFNγ conditions enables the acceleration of peptide production, processing, and rapid loading on to HLA-I which may enable peptides without the protective internal proline to

escape degradation. This would consequently dilute proline-containing peptides within the peptide pool, explaining why proline-containing peptides decrease in relative abundance, or drop out from detection under IFNγ exposure. This theory is supported by the evidence that in cells overexpressing cytosolic peptidase THOP1, the low peptide supply meant an inability to form stable peptide-MHC complexes, preventing MHC-I surface expression. However, in the presence of IFNγ, where peptide supply and abundance of MHC-loading machinery increased, MHC expression was restored (Kim *et al.*, 2003). This indicated the balance of peptides to cytosolic peptidases, alongside the time spent in the cytosol, are important factors regulating peptide presentation.

I also found conversely that proline, while typically depleted in the N1-3 and C1-3 in the peptide extended sequences, experienced reduced depletion in the N1-3 of MIPs and IEPs. The specificity of this change to the N-terminal extended sequence suggests that this change is affected by peptidases, and not the immunoproteasome. However, this again could not be explained by peptidase expression, suggesting peptidase activity or abundance of unknown peptidases may more relevant for this question. It is important to reference that whilst the cause of this change is currently unknown, it is not certain that the differences in proline seen internally and externally of the peptide have the same cause; this is because due to the unique structural nature of proline, it affects many different biological processes (Vanhoof *et al.*, 1995; Shimbara *et al.*, 1998; Perez *et al.*, 2019). Understanding this better could be incredibly important in providing more information for peptide prediction tools. Therefore, biochemical assays confirming positional biases in peptide cleavage and determining the source of it, would be important future work.

The novel analysis workflow undertaken in this study allowed for specific determinants of peptide abundance changes with IFNγ exposure to be revealed. This new information could be incorporated into peptide prediction algorithms or peptide selection tools, which could help to improve cancer vaccine or TCR therapy design. For example, this work has shown peptides presented on HLA-B allotypes are more likely to increase in intensity when tumour cells are exposed to IFNγ than HLA-A, which may make those preferable targets. Furthermore, the analysis showed that proline in positions 4 and 5, and possibly throughout the

peptide length, caused peptide downregulation, or even peptide loss under IFNγ exposure. This suggested that tumour antigens containing proline could be excluded from peptide vaccines, to allow inclusion of candidate peptides with features that do not influence their downregulation under IFNγ exposure. Furthermore, if IFNγ can be attributed to the increased liberation of peptides from sequences with preceding proline, it could help to mark those peptides as better candidates for interventions like peptide vaccines. It could also be beneficial to produce an entirely separate peptide prediction algorithm training dataset, consisting of only peptides from IFNγ-treated cells, to allow better predictions of how peptide presentation can change on a larger scale. To what extent the IFNγ-induced relative changes in HLA expression and remodelling of the immunopeptidome are clinically relevant, and how the improved understanding could benefit patients, needs to be further researched.

Chapter 5 Optimisation of endoplasmic reticulum isolation for peptidomics of the endoplasmic reticulum.

The data from Chapter 3 showed that only 2 of the 196 mutations in expressed genes that had been predicted to generate a neoantigen actually produced an MS-detectable HLA-I-presented neoantigen. One further mutation generated an MS-detectable neoantigen but computational binding prediction did not classify this as a strong or weak binder. This dramatic drop from 196 predicted neoantigens, to 2 actually detected, suggested that there is a poorly understood bottleneck in the antigen processing and presentation pathway. This may be due to the failure to generate peptides encompassing the predicted neoantigen by the proteasome, or rapid degradation of proteasome-generated peptides before they are loaded onto HLA-I. It is furthermore possible that binding prediction is not always accurate, and some HLA-bound peptides may be unstable or not be detectable by MS due to technical biases. Some components of this bottleneck are already understood. For example, proteins need to be expressed and turned over to provide peptides, peptides are then selected for/against at several stages of antigen processing and transport, including TAP1/2 transport, and proteasomal and peptidase cleavage (Beninga, Rock and Goldberg, 1998; Harris *et al.*, 2001; Saric *et al.*, 2001; Van Endert, 2011; Huber *et al.*, 2012; Lehnert and Tampé, 2017; Boulanger *et al.*, 2022). However, detailed information of the peptide repertoire that is available for loading onto HLA-I in the endoplasmic reticulum (ER) is not available. My hypothesis was that a direct analysis of the ER peptidome by mass spectrometry (MS) could give an insight into which peptides are made available for loading on to HLA in the ER. This could elucidate the extent to which peptides had already been selected from the theoretically possible peptidome that would be generated by the proteasome. Therefore, it was prudent to design a method to isolate the ER so that the peptidome of the ER could be interrogated, to better understand the pool of peptides that are generated by proteasome and peptidase cleavage, survive degradation, and are able to be transported in to the ER by TAP1/2.

It should be noted that immunopeptidomics data is generated by cell lysis and immunoaffinity purification (IP) of HLA; this study focuses on the IP of HLA-I. The

peptides are eluted from purified HLA-I and analysed by LC-MS/MS. Because the HLA-I are purified from whole cell lysates, the HLA-I immunopeptidomics data cannot differentiate between ligands eluted from HLA-I on the cell surface, and those from eluted from HLA-I in the ER (Gfeller and Bassani-Sternberg, 2018). Therefore, when comparing concurrently prepared ER peptidomics samples with HLA-I immunopeptidomics samples, the peptides that are exclusive to the ER are likely those that did not get loaded on to HLA-I. In this text, I will refer to the HLA-I immunopeptidome as the HLA-I-eluted peptidome, and HLA-I-eluted peptides. In this study, the ER peptidome was compared against an existing HLA-I-eluted peptidome. Recognising that the samples were not prepared concurrently, I was cautious not to draw definitive conclusions from my analyses.

5.1 Subcellular proteomics

The cell is compartmentalised into different subcellular organelles which provides a specific environment and context for optimal protein function, and allows multiple different biological processes to occur concurrently. Therefore, understanding subcellular localisation of proteins, and how it can be affected in disease, can be very important for better understanding cell function (Thul *et al.*, 2017a). However, approximately 50% of the proteome is estimated to be localised to multiple organelle compartments, which adds complexity to the understanding of protein network functions (Thul *et al.*, 2017a). Many techniques have been developed to deepen the understanding of subcellular protein localisation, including: immunofluorescence (Thul *et al.*, 2017a), proximity labelling (Lee *et al.*, 2016; Roux *et al.*, 2018), and biochemical fractionation (Dunkley *et al.*, 2004; Foster *et al.*, 2006; Itzhak *et al.*, 2016).

Whilst the technique of immunofluorescence can work well in the assigning of proteins to subcellular localisations, it cannot be used to elucidate the peptidome of an organelle such as the ER. Furthermore, proximity labelling techniques such as APEX labelling, which relies on the labelling of tyrosine/tryptophan-containing sequences (Lee *et al.*, 2016), and BioID, which relies on the labelling of lysine-containing sequences (Roux *et al.*, 2018), is more suitable to the labelling of full-length proteins due to the labelling of only select amino acids (aas). Proximity labelling therefore could only label a subset of possible peptides within a

peptidome. Further analysis of the HLA-I-eluted peptidome examined in Chapter 3 and Chapter 4 showed that APEX labelling could only represent an average of 40.3% of peptides across the PDOs, and BioID labelling could only characterise 68.5% (Figure 3.1A-B). This representation of the ER peptidome would be biased and too sparse to be able characterise its features. Therefore, I concluded the most appropriate approach to take for the characterisation of the ER peptidome would be the subcellular fractionation of cells.

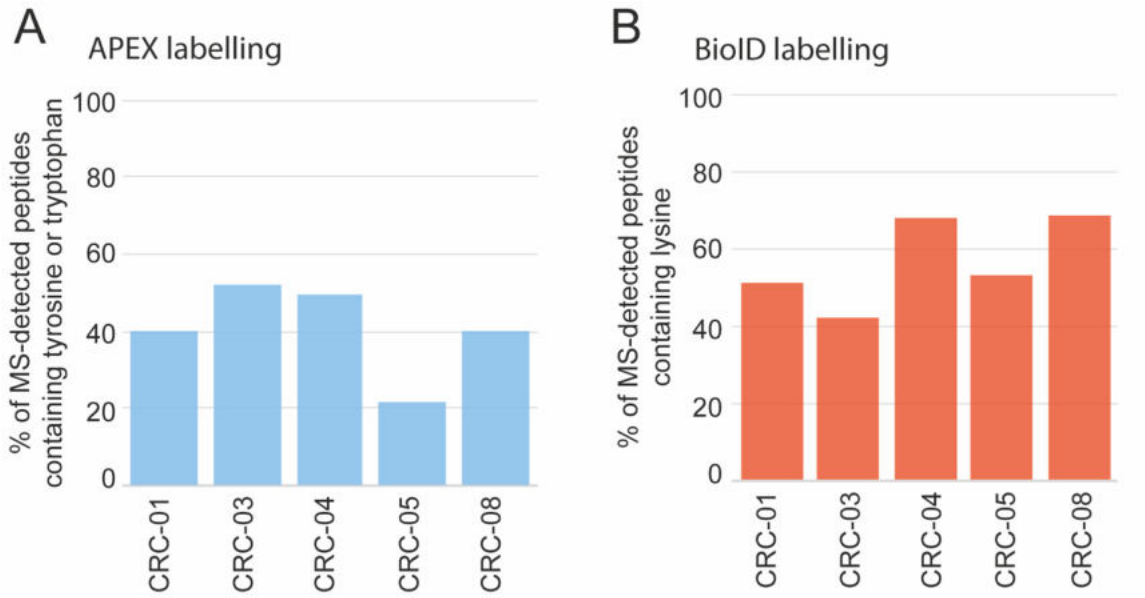


Figure 3.1. The feasibility of using APEX labelling or BioID to label the ER peptidome. **A:** The percentage of MS-detected HLA-I peptides containing tyrosine or tryptophan, used for APEX labelling**. B:** The percentage of MS-detected HLA-I peptides containing lysine, used for BioID labelling**.**

Preparation of a pure organelle compartment, free of contamination from other compartments, poses a significant challenge due to the similar densities of the different organelles (Table 5.1), and the constant cycling of membranes and proteins between different organelles (Dunkley *et al.*, 2004). Furthermore, the ER is a continuous membrane network system which stretches throughout the cell (Voeltz, Rolls and Rapoport, 2002), and when cells are lysed the ER breaks and reforms in to smaller vesicles known as microsomes. Obtaining high purity ER microsomes and qualifying the ER peptidome were important objectives of this study. To be able to use the system of high-purity ER isolation and MS peptidomics to answer biological questions, such as interrogation of the

characteristics of the ER peptidome, I first needed to define a protocol which gave sufficient ER organelle yield and purity, and ER peptide yield.

Organelle/compartment	Diameter (μm)	Density (g/cm^3)
Nuclei	3-12	1.4
Mitochondria	0.5-2	1.15
Golgi apparatus	-	1.08
Lysosomes	0.5-0.8	1.2
Peroxisomes	0.5-0.8	1.25
Plasma membrane	0.5-3	1.15
Smooth endoplasmic reticulum	0.5-3	1.16
Nucleic acid	0.03	1.7
Ribosomes	0.03	1.6
Soluble proteins	0.001-0.002	1.6

Table 5.1. Organelle size and densities taken from (Michelsen and von Hagen, 2009).

5.2 Understanding the bottleneck in neoantigen presentation

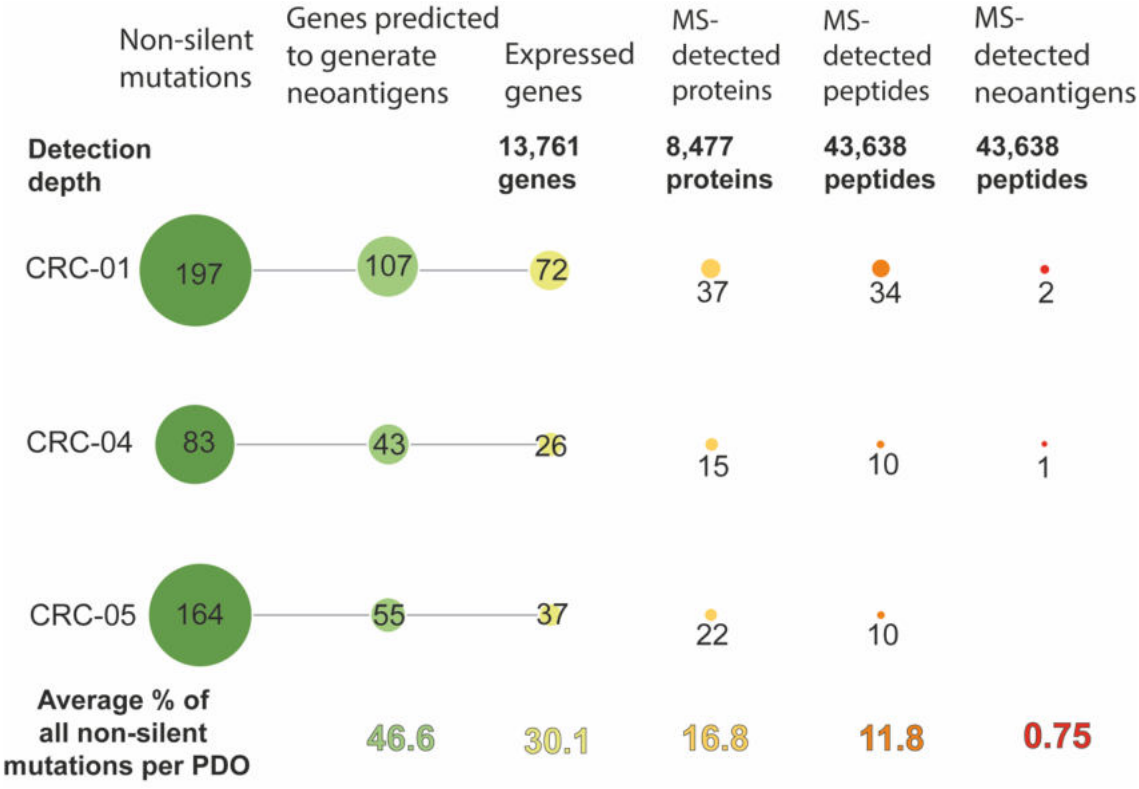


Figure 3.2. The bottleneck of neoantigen presentation. Number of mutations that encode for amino acid changes (missense, frame-shift and stop-loss mutations), genes predicted to generate strong binders predicted by NetMHCpan4.0 (defined as percentile rank below 0.5%), strong binder-generating genes that are expressed at the mRNA level, strong binder-generating genes that are expressed at the protein level, and strong binder-generating genes that generate non-mutated HLA-I antigens, compared against MS-detected neoantigens.

My initial work on the mass spectrometry (MS) immunopeptidomics of HLA-I-eluted peptides showed there was a disparity between the number of predicted neoantigens from expressed genes, and the number of detectable HLA-I-presented neoantigens. To help understand the nature of the bottleneck of neoantigen presentation, I wanted to examine if the number of detectably-expressed proteins were a contributing factor to the bottleneck, as it is well described that RNA expression is not a fully accurate depiction of protein abundance. Therefore, I was able to use previously-generated tandem-mass-tag (TMT)-MS data, described in Chapter 4, to detect the abundance of the mutated

neoantigen source genes as proteins. Compared to 13,761 genes detected as expressed by RNA sequencing, only 8,477 proteins were detected (Figure 3.2). This disparity may be a limiting factor in this analysis. Whereas 30.1% of the mutated genes predicted to generate a neoantigen were detectable as expressed by RNA sequencing, only a mean of 16.8% were detectable by TMT-MS. Thus, suggesting protein abundance is an additional limiting factor in neoantigen generation, as this provides protein for turnover. However, proteins that are high-turnover may not be detectable by MS, but can still provide peptides for HLA-I presentation. This is why I next plotted the number of MS-detected peptides derived from non-mutated regions of the predicted neoantigen source proteins – a mean of 11.8% of the mutated genes predicted to generate a neoantigen produced MS-detectable non-mutated HLA-I peptides.

This suggests that, in our dataset, under 1/3 of predicted neoantigen source genes were able to generate HLA-I-compatible peptides. This is certainly a major reduction in the number of viable sources of neoantigens, suggesting the turnover and processing of mutation-containing proteins is a limiting factor in neoantigen presentation. However, there was still a further disparity between the number of predicted neoantigen source genes that generated nonmutated HLA-I peptides, and the number of neoantigens that were detected on HLA-I. Therefore, there were still multiple additional factors constraining the production of neoantigens in our MMRp CRCs.

5.3 Assessing the quality of a commercial ER enrichment kit

I first tested whether a commercial ER enrichment kit (MinuteTM Enrichment Kit, Invent Biotechnologies) would be a viable technique to generate a sufficiently pure extract of ER for MS peptidome analysis. In this protocol, and all protocols following, all the lysis reagents and all the buffers contained protease inhibitors (cOmplete^{TM}, EDTA-free protease inhibitor cocktail, Roche) and 1μM peptidase inhibitor PMSF.

PDOs from CRC-05 were cultured in Advanced DMEM:F12 media supplemented with 10% FCS and 2% matrigel. 35x10^6 cells were harvested for this protocol; this was the highest end of the manufacturer-recommended source cell number. This was used in order to obtain a high yield of final ER product. Cells were harvested and pellets were snap frozen on dry ice with 100% ethanol, and stored in the -80°C freezer overnight. This freeze/thaw cycle was recommended by the manufacturers to enhance the disruption of outer cell membranes to encourage release of subcellular organelles. The protocol was executed according to the manufacturer's instructions (Figure 3.3A). The protocol generated discard pellets of 1) un-lysed cells (ULC) and nuclei, and 2) mitochondria, cell debris and plasma membrane. The final pellet, pellet 3, should have contained enriched ER, and was therefore called the ER pellet.

To assess the purity of the final ER pellet, and the composition of the discard pellets, I processed them for western blot (WB) using RIPA buffer. RIPA buffer was selected for this protocol as the harsher conditions of the buffer are better at disrupting all organellar membranes.

5.3.1 Results of ER enrichment kit WB analysis

To assess whether the kit enriched for ER microsomes, I performed a western blot for the ER marker proteins GRP78 and CalR. GRP78 (also known as BiP) was selected because it is a highly abundant ER-luminal chaperone protein (Lee, 2007; Lang *et al.*, 2017; Huang *et al.*, 2022). Calreticulin was selected because it is an ER protein that is part of the PLC which is critical for loading peptides on to HLA. GAPDH was selected as a marker of cytoplasmic contamination.

The WB showed decreasing intensity of the GRP78 and CalR bands across the ER isolation procedure, highlighting ER proteins were likely being lost to the discard pellets (Figure 3.3B). The high presence of ER markers in the ULC/nuclei pellet suggested that cell lysis had been insufficient, resulting in much of the ER remaining sequestered within the cell membrane and being pelleted in the ULC/nuclei centrifugation. Furthermore, densitometry using ImageJ software showed the ratio of the ER marker GRP78 to the cytoplasmic marker GAPDH actually worsened across each of the stages, decreasing from 1.11 in the ULC/nuclei pellet, to 0.58 in the mitochondrial pellet, to 0.38 in the final ER pellet (Figure 3.3B). Therefore, according to the results of this WB, the kit did not enrich for the ER microsomes, and therefore was not suitable for analysis of the ER peptidome.

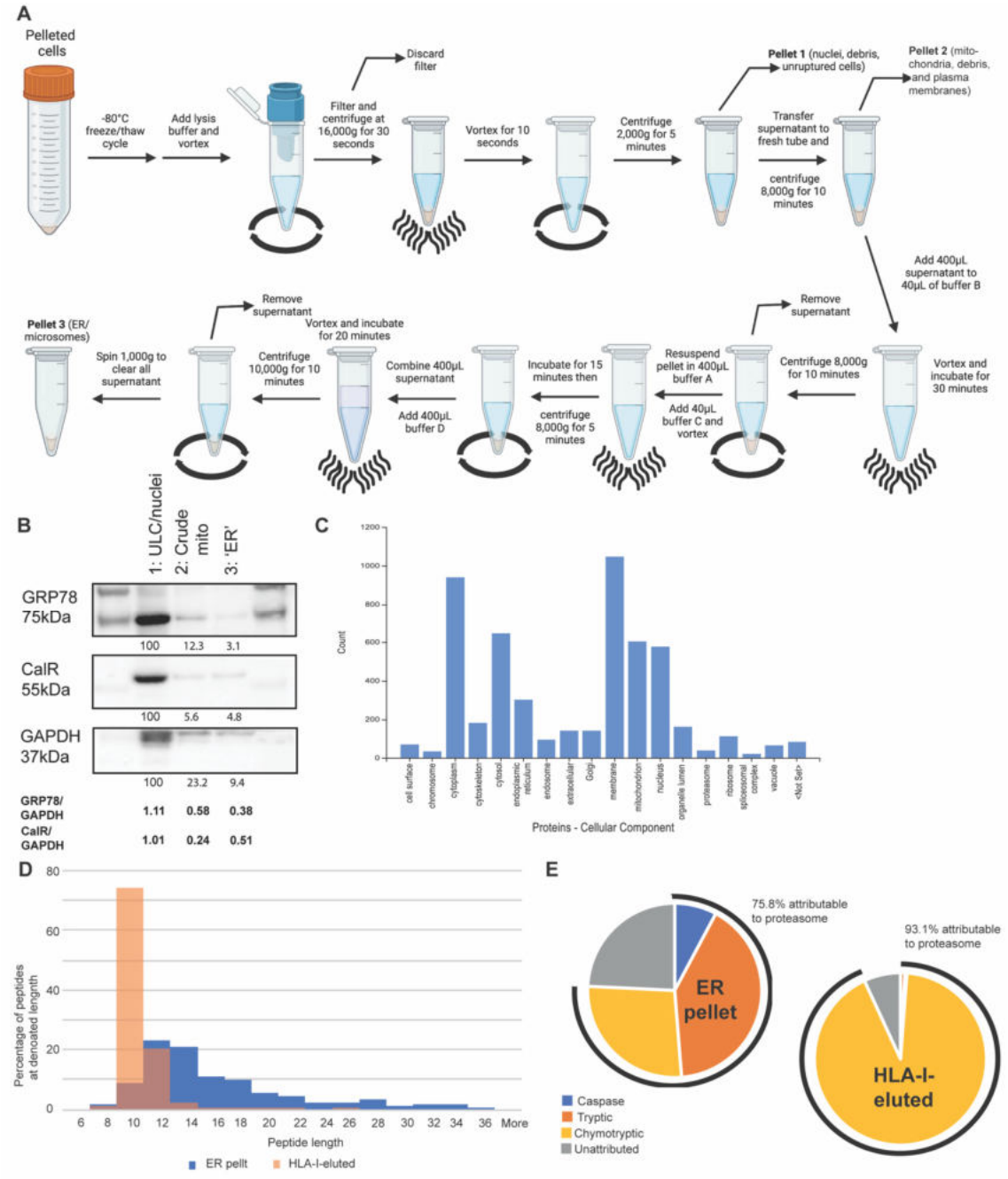


Figure 3.3. ***Characterising the Minute 'ER enrichment' kit.*** **A:** Schematic depicting the process of ER enrichment. **B:** WB illustrating the yield of ER and cytosolic proteins across the isolation process. **C:** Gene ontology (GO)-allocated cellular compartments of proteins yielded from ER pellet. **D:** Peptide length distribution of MS-detected peptides obtained from 'ER pellet' from the isolation performed in CRC-05, compared against MS-detected HLA-I-eluted peptide lengths**. E:** Characteristics of the C-terminus of the MS-detected HLA-I-eluted and ER pellet peptides**.**

5.3.2 Results of ER kit MS analysis

Despite the WB showing the kit could not strongly enrich for the ER, the ER pellet was also submitted for analysis by MS to help ascertain whether mass spectrometry could detect small peptides from organelle preparations, and whether these exhibited features similar to those expected from the ER.

The ER pellet was permeabilised by probe sonication and prepared for MS as described in Materials and Methods. The mixed protein and peptide solution was applied to an Amikon 3K size exclusion filter. The filter should only allow peptides under ≈ 27 amino acids in length to pass through, and the solution left inside the filter contained proteins. These proteins were then recovered, reduction-alkylated, and trypsin digested. MS spectra for the proteome and the peptidome were searched against UniProt annotated reference proteomes of *Homo Sapiens* in Proteome Discoverer v2.4. For the peptidome search, to reduce the search space and increase the power, only peptides 7-36 amino acids long were analysed.

I first analysed the TMT-MS results for the proteome of the ER pellet to further analyse its protein composition (Figure 3.3C). The pellet showed considerable contamination with proteins attributed to other compartments (for example, there were a large number of cytoplasmic proteins, suggesting the cytoplasm had not be sufficiently depleted).

The Amikon 3k filter flow-through yielded 163 unique MS-detectable peptides from PDO CRC-05 (Figure 3.3D). The yield of peptides was remarkably low, likely due to the low yield from the ER enrichment kit, and also some suspected peptide loss in the size exclusion filter. I next analysed the characteristics of these peptides. I wanted to assess if these peptides had length distributions which could suggest they could be derived from the ER, or show features which suggested they had been generated by the proteasome. I also compared the ER pellet peptide results against the HLA-I-eluted peptidomics data for CRC-05. As mentioned in the introduction, HLA-eluted peptidomics data comes from both ligands eluted from HLA on the cell surface, alongside those from eluted from HLA in the ER. This analysis therefore should highlight the differences between

peptides that were loaded on to HLA-I, and those that likely weren't. Plotting the peptide length distribution of the ER pellet-derived peptides against the HLA-I-eluted peptides showed the ER pellet peptides to have a wider range of peptide lengths. 95.1% of the ER pellet peptides were over 9aa long, compared to just 45.5% of the HLA-I-eluted peptides (Figure 3.3D).

It is widely accepted that the C-terminus of HLA-I-presented peptides can be used to determine the cleavage specificity of the peptide (Chong *et al.*, 2017; Sarkizova *et al.*, 2020). Because, as referenced in the introduction, ER peptides are subject to selection by TAP1/2 which exhibits similar binding preferences to HLA-I (Lehnert and Tampé, 2017), it was appropriate to perform this analysis on both the HLA-I-eluted peptides and ER pellet-derived peptides. Therefore, I analysed the amino acid composition of the C-terminus of the ER pellet-derived peptides to see if the amino acids matched the cleavage specificities of the proteasome. This showed 75.8% of the ER pellet-derived peptides to have likely been generated by chymotryptic, tryptic, or caspase cleavage activity, supporting that they had likely been generated by the proteasome (Figure 3.3E). The attribution of peptides to different cleavage specificities was quite different between the HLA-I-eluted and ER pellet-derived peptides; in the ER pellet-derived peptides only 27% of peptides were attributed to chymotryptic cleavage activity, compared against 92% in the HLA-I-eluted peptides. The presence of tryptic- and caspase-like C-termini in the ER pellet-derived peptides did not guarantee that these were contaminant peptides however, as the HLA allotypes presented by CRC-05 only tolerated chymotryptic ligands at the C-terminus (as illustrated in Chapter 3).

I next analysed for peptide overlaps between the HLA-I-eluted peptides and the ER pellet-derived peptides to see if it was possible to isolate possible HLA-precursor peptides (longer versions of peptides that had not been yet cleaved by ERAPs for HLA loading) from organelle preparations. There were multiple exactly matched peptide sequence overlaps between the HLA-I-eluted peptides and the ER pellet-derived peptides (N=10). Additionally, there were 13 cases of peptides with an extended N-terminus, 4 cases of peptides with an extended C-terminus, and 2 cases of peptides extended at both the N- and C-terminus (Figure 3.4). All the peptides with an extended C-terminus had lysine at their C-terminus, showing

them to be possible tryptic-like ligands, and therefore possible alternative proteasomal ligands.

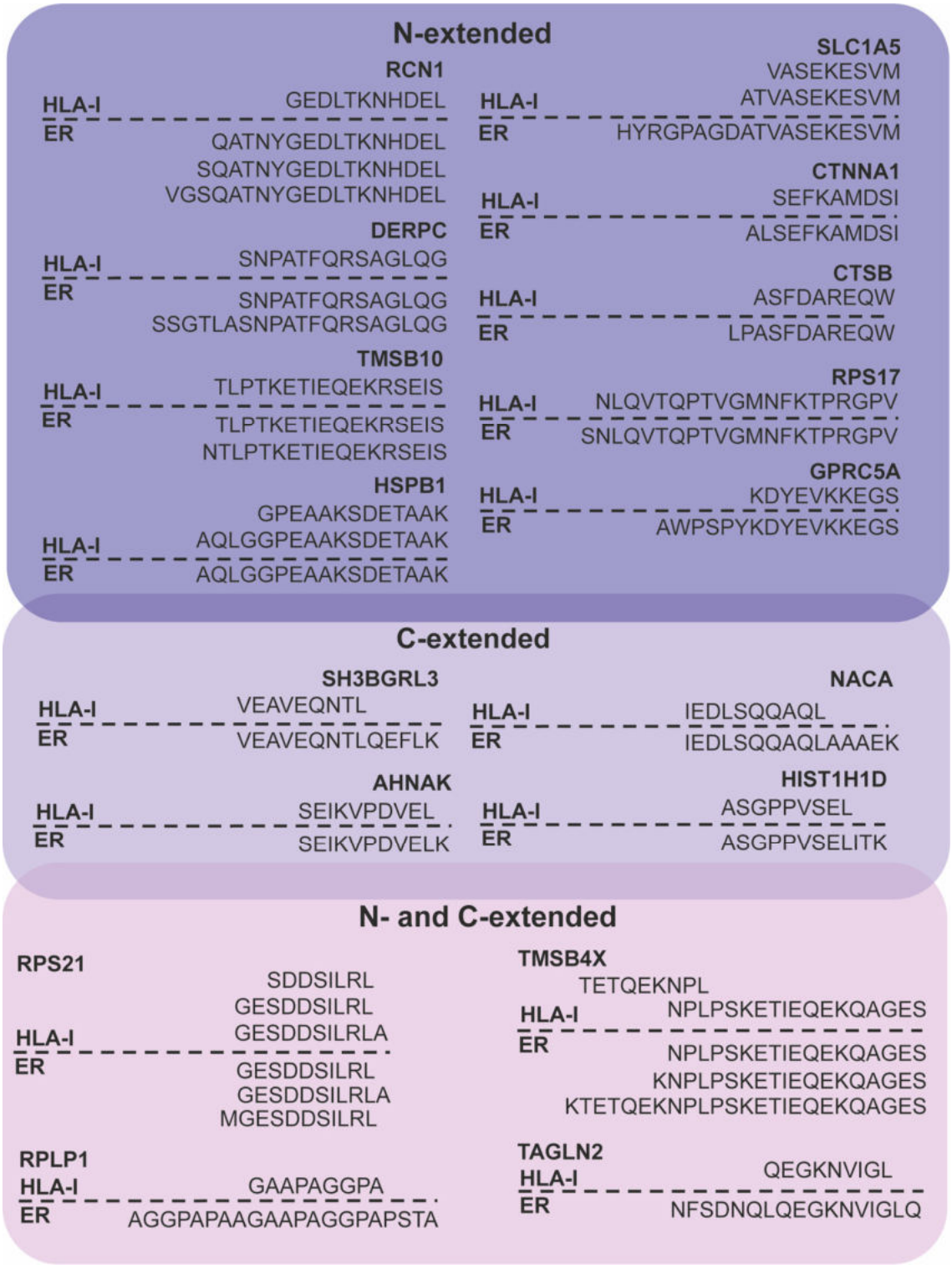


Figure 3.4 Alignment of all N- and C-extended HLA-I-eluted and gradient-derived peptides from CRC-05.

This data indicated it was possible to detect peptides with lengths similar to those expected from the ER. This gave me confidence that it was possible to obtain and technically analyse peptides from organelle preparations by mass spectrometry. However, all the data together suggested that the ER enrichment kit was not compatible with my goal to analyse the ER peptidome due to the insufficient ER protein yield, poor purity, and poor peptide yield.

5.4 Testing common ER isolation protocols

To summarise the approaches trialled to reach the final protocol for ER isolation: I began with two differential centrifugation steps to remove nuclear and mitochondrial contamination, followed by a rate-zonal centrifugation with a discontinuous sucrose gradient. The main focus was reducing ER loss in the mitochondrial exclusion stage (depicted in Sections 5.4.1-5.4.4). After discovering the original sucrose gradient system was not suitable for the removal of lysosomal contamination, the gradient system was changed for a Histodenz-based discontinuous gradient (Sections 5.5.1-5.5.2). The final protocol, which functioned better at enriching for ER whilst eliminating more organelle compartments, is outlined in Section 5.5.2.

To commence the ER isolation optimisation process, I began with commonly utilised ER isolation protocols based on ultracentrifugation. Most ER protocols utilised two different centrifugation techniques sequentially: differential centrifugation for the elimination of larger and heavier organelles which have faster sedimentation rates, followed by rate-zonal centrifugation for the further separation of organelles (Burdon, van Knippenberg and Sharpe, 1988; Frei, 2012).

5.4.1 Sucrose gradient centrifugation approach 1

The first method I used was based upon a combination of ER isolation by differential centrifugation from (Chen *et al.*, 2010) combined with the gradient system from (Williamson *et al.*, 2015). The gradient system from (Chen *et al.*, 2010) was not used because during a preliminary test the 1.28M and 1.35M

sucrose layers were not sufficiently different in their densities to form discrete layers.

5.4.1.1 Method

Because these protocols required large numbers of cells (at least $100x10^6$ per condition), I decided to use cells that grew faster than our PDOs, and did not require the use of expensive matrigel. For the selection of the candidate cell line, I examined a published proteomics dataset of 39 CRC cell lines to find an MMRp CRC cell line with a high abundance of PLC and HLA genes in the panel (Roumeliotis *et al.*, 2017). Cancer cell lines can have defects in the antigen processing and presentation pathway (Anderson *et al.*, 2021; Dhatchinamoorthy, Colbert and Rock, 2021), therefore the objective was to identify a line which expressed APM proteins, and was more likely to have an intact antigen presentation pathway. From the panel, I found the cell line NCI-H747 to have highest total abundance of PLC and HLA components. NCI-H747 was cultured as described in the Materials and Methods section, in RPMI-1640 medium supplemented with 1x Glutamax and 10% FCS.

The protocol for cell harvesting, cell lysis, and centrifugation is described in detail in Materials and Methods, and the schematic of the protocol is shown in Figure 3.5A. To summarise, cells were harvested, snap frozen, and stored in the -80°C freezer overnight. The pellet was then resuspended in isotonic sucrose buffer (ISB) and homogenised with 40 strokes of a Dounce homogeniser using the tight pestle. I used Dounce homogenisation instead of sonication, as Dounce homogenisation is gentler than sonication (Williamson *et al.*, 2015; Itzhak *et al.*, 2016), and preserving as much organelle membrane integrity as possible was important for this protocol, to reduce the risk of peptide loss. The lysate was then subjected to 2 rounds of differential centrifugation: 800g for 10 minutes to remove ULC and nuclei, and 12,000g for 20 minutes to remove mitochondria. The post-mitochondrial supernatant was then pelleted, resuspended in ISB and layered over the sucrose gradient before ultracentrifugation.

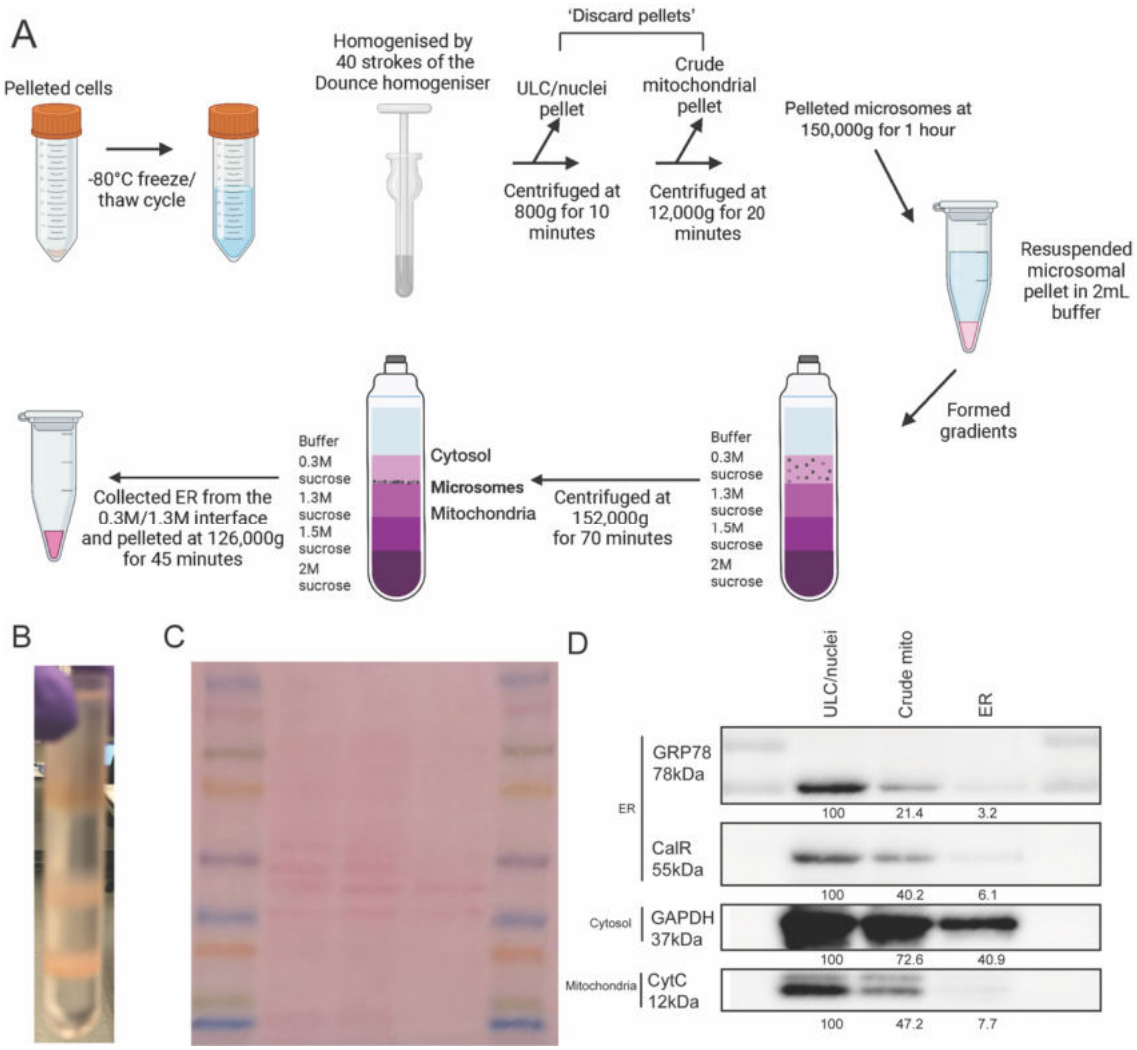


Figure 3.5. ***Characterising the success of a published sucrose-based ER isolation method.*** **A:** Schematic depicting the process of ER enrichment. **B:** Photograph of the sucrose gradient post ultracentrifugation. **C:** Ponceau stain of the WB of the yielded pellets from across the ER isolation process. **D:** WB illustrating the yield of ER, cytosolic, and mitochondrial proteins across the isolation process**.**

5.4.1.2 Results

The discard pellets (ULC/nuclei and mitochondrial pellets) and the final ER pellet were prepared for WB, executed as described in the Materials and Methods section.

I evaluated GRP78, CalR, GAPDH, and Cytochrome C (CytC). CytC was added to panel of markers to assess the mitochondrial content at each stage of the ER enrichment process. The blot (Figure 3.5C-D) illustrated a poor final yield of ER proteins. It appeared as if slightly fewer ER proteins (GRP78 and CalR) were lost

in the ULC/nuclei pellet than with the ER enrichment kit, however, very little of each ER marker was detectable in the final ER pellet. ImageJ densitometry analysis showed the signal for both ER markers decreased dramatically in each step, for example GRP78 signal in the ER lane was only 3% of the signal in the ULC/nuclei lane, and CalR only 6% (Figure 3.5D). However, this protocol did succeed in reducing mitochondrial contamination in the ER pellet, with the densitometry values of the mitochondrial marker CytC decreased to 7.7% of the intensity of the ULC/nuclei lane. The still large amount of ER protein in the ULC/nuclei pellet, suggested the efficiency of the cell lysis was poor. Therefore, I decided it was necessary to trial a new cell lysis approach. The WB also demonstrated a reduction of GAPDH in the final ER fraction, illustrating reduced cytoplasmic contamination across the pellets. However, a notable amount of GAPDH did remain, suggesting I was not stringent enough harvesting the ER layer with a syringe.

Despite these limitations, this protocol did succeed in removing mitochondrial contamination in the ER pellet. However, the loss of ER proteins to the mitochondrial pellet in-fact suggested the spin was too harsh, and that the speed could be lowered to reduce ER loss whilst sustaining mitochondrial removal.

5.4.2 Sucrose gradient centrifugation approach 2

The lessons learnt from run 1 informed the changes applied in the next attempt. In this protocol, I also took an aliquot of the total lysate to provide a comparator to help inform the purity of the final product.

5.4.2.1 Method

The full protocol is described in detail in Materials and Methods, and the schematic of the protocol is shown in Figure 3.6A.

In order to improve cell lysis, a harsher cell lysis protocol was used: cells were incubated in hypotonic buffer for 20 minutes prior to manual Dounce homogenisation as described (Sigma-Aldrich, 2006; Itzhak *et al.*, 2016). The hypotonic buffer should swell cells to make them more prone to lyse when

subjected to the shearing forces of Dounce homogenisation. After the ULC/nuclear exclusion centrifugation I noted that the pellet was smaller than in the previous protocol which didn't utilise the hypotonic buffer. This was promising as it suggested there were fewer ULCs left.

The mitochondrial exclusion centrifugation was reduced from 12,000g for 20 minutes in the previous attempt, to 10,000g for 10 minutes. This was intended to reduce ER loss whilst maintaining the ability to eliminate mitochondrial contamination.

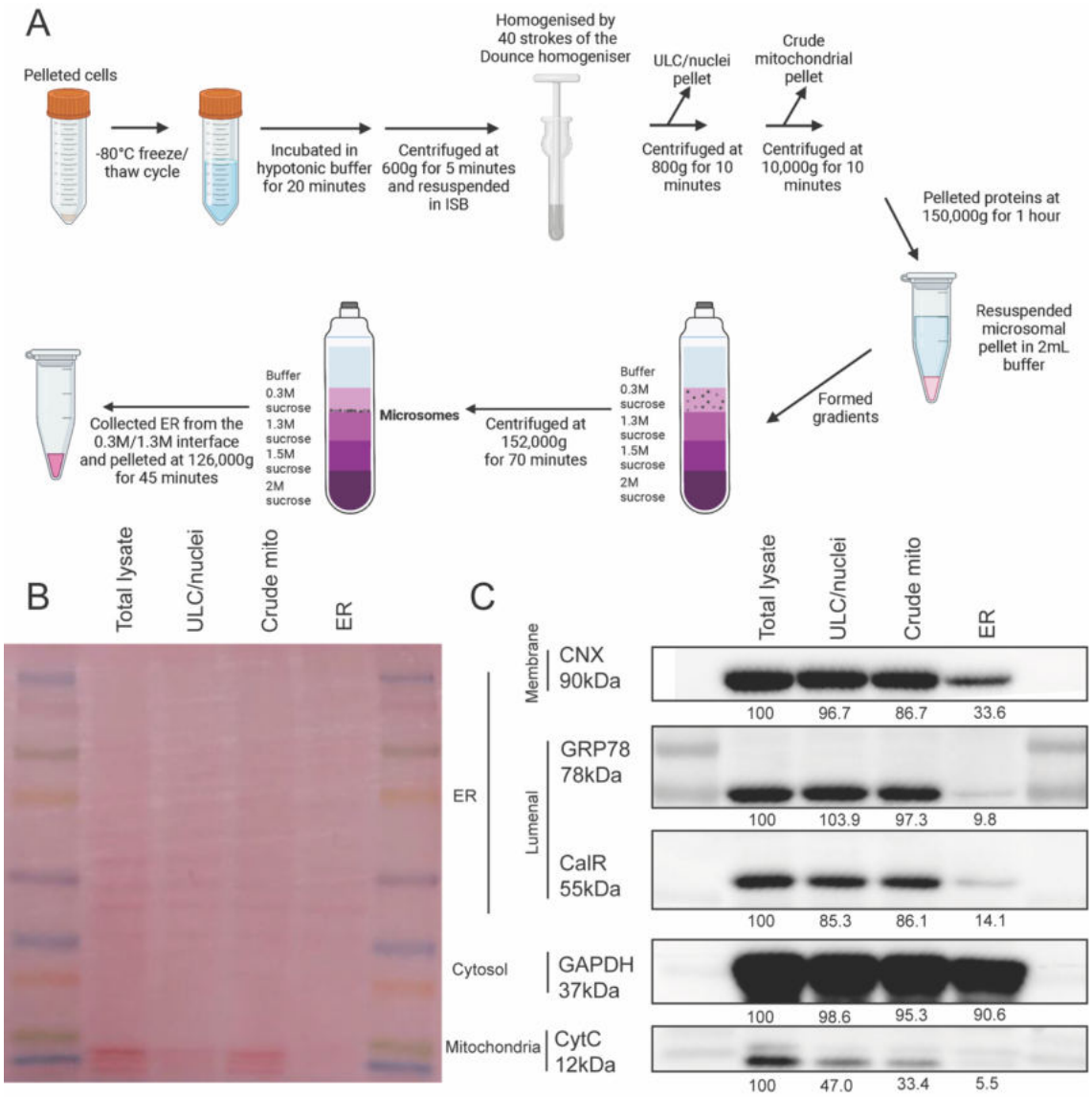


Figure 3.6. ***Characterising the success of a modified sucrose-based ER isolation method.*** **A:** Schematic depicting the process of ER enrichment. **B:** Ponceau stain of the WB of the yielded pellets from across the ER isolation

process. **C:** WB illustrating the yield of ER, cytosolic, and mitochondrial proteins across the isolation process**.**

5.4.2.2 Results

The Ponceau stain (Figure 3.6B) suggested the ER lane may be underloaded compared to the first 3 lanes. The blot could not be repeated due to the limited yield of the ER pellet and, conclusions therefore needed to be drawn with caution.

In the western blot, I also evaluated the ER membrane-integral protein Calnexin (CNX), a chaperone protein which is a component of the PLC (Figure 3.6C). This was to help illustrate if there was a difference in enrichment of ER membrane and luminal proteins.

The cell lysis was more successful as the ImageJ densitometry values of the ER markers were similar between the ULC pellet and the mitochondrial pellet. For example, GRP78 and CalR were no longer dramatically enriched in the ULC/nuclei pellet (illustrated by the similarity in CNX and GRP78 in the ULC/nuclei and crude mitochondrial lanes). This is in contrast to the previous protocol where there was a 78.6% and 59.8% decrease in GRP78 and CalR band intensity between the ULC/nuclei and crude mitochondrial lanes. However, the ER markers were still depleted in the ER pellet, with ER-luminal proteins GRP78 and CalR decreased to 9.80% and 14.1%, and the membrane protein CNX decreased to 33.6%, of the total lysate signal. The decreasing level of ER markers across the protocol indicated ER proteins are being lost by being pelleted in the differential centrifugation steps, again suggesting the centrifugation speeds may have still been too harsh. Furthermore, the disparity between the ER membrane and ER luminal protein retention signified loss of ER membrane integrity. Further to this, the GAPDH marker showed there was high cytosolic contamination in the ER fraction, indicating the need for stricter isolation of the ER layer. However, even with the slower centrifugation and shorter duration, mitochondrial contamination was still depleted in the final ER pellet.

5.4.3 Sucrose gradient centrifugation approach 3

5.4.3.1 Method

After the results from the second iteration of the protocol, the speed of the mitochondrial exclusion step was reduced to 5,000g for 10 minutes in an attempt to reduce ER loss. The full protocol detailed in Materials and Methods and depicted in Figure 3.7A.

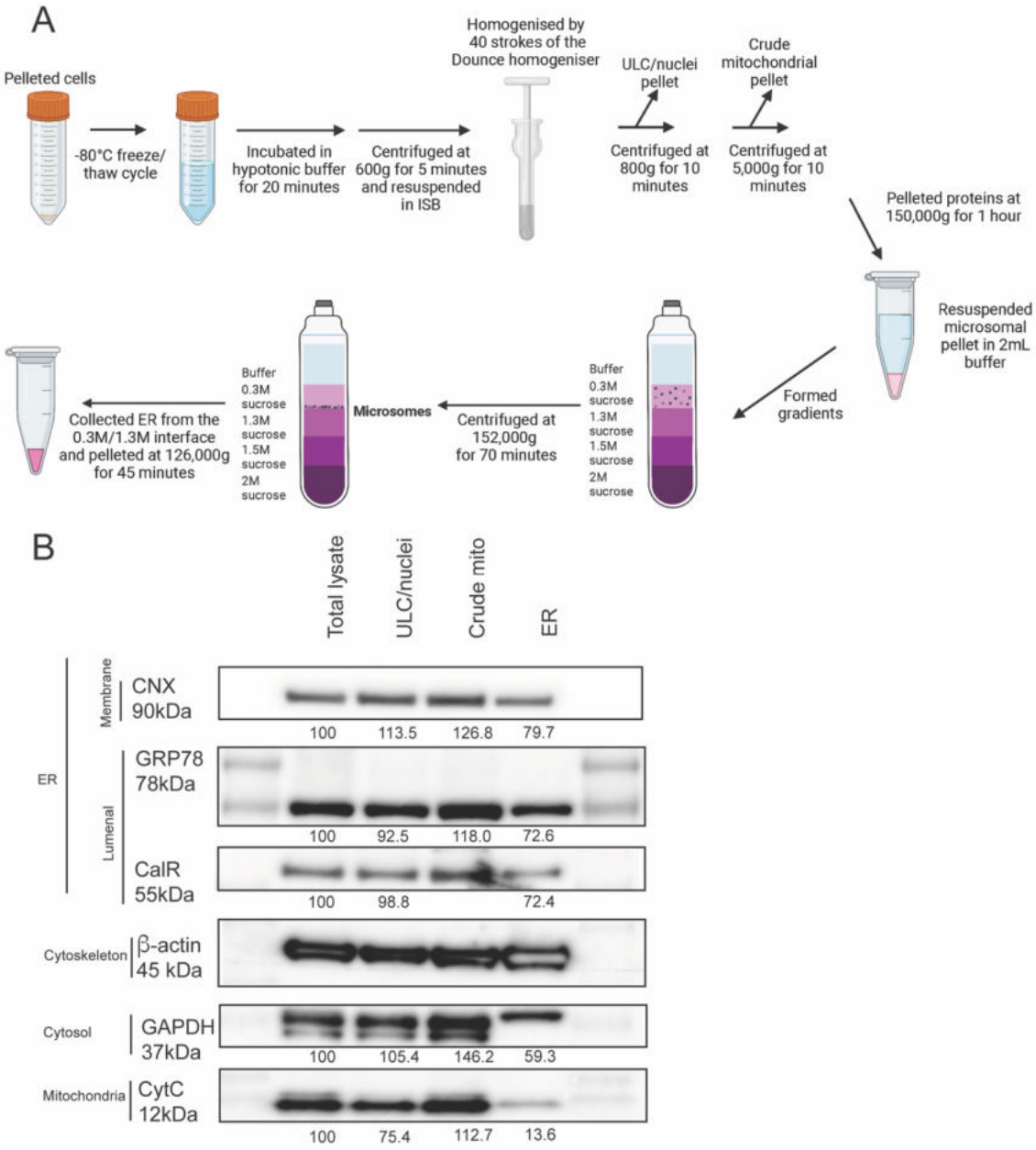


Figure 3.7. ***Characterising the success of a further modified sucrose-based ER isolation method.*** **A:** Schematic depicting the process of ER enrichment. **B:**

WB illustrating the yield of ER, cytosolic, and mitochondrial proteins across the isolation process.

5.4.3.2 Results

The results of the WB (Figure 3.7B) showed significantly better ER protein retention than in previous protocols. The signal of the luminal ER proteins (GRP78 and CalR) in the ER fraction only decreased to 72.6% and 72.4% of the total lysate signal, showing less ER was being lost. Furthermore, the ER membrane protein CNX decreased to 79.7% of the total lysate signal. The similarity in ER luminal and membrane protein retention indicated that fewer ER luminal proteins were being lost, and therefore ER membrane integrity was preserved better.

In this run I used a more stringent approach to collect the gradient layers in this attempt, by collecting a reduced volume around the ER layer. This was successful in reducing cytoplasmic contamination as indicated by the GAPDH marker, which reduced to 59.3% of the total lysate signal. However, this level of GAPDH retention was still quite high, so additional measures needed to be applied to reduce this further in the next protocol.

There was a small increase of mitochondrial contamination across this isolation protocol, retaining 13.6% of the CytC marker signal of the total lysate, against 7.7% and 5.5% in the previous protocols. This indicated that the purity of the final product, in relation to mitochondrial content, was very slightly affected by the decreasing speed of the mitochondrial exclusion spin. Therefore, I concluded that whilst the protocol could tolerate a further small decrease in centrifugation speed, in the interest of balancing ER yield and mitochondrial depletion, it should not be decreased much further.

5.4.4 Sucrose gradient centrifugation approach 4

The data from the previous runs illustrated that this protocol needed additional rounds of optimisation, making it unlikely that it would be possible to perform the optimised protocol on the PDOs for which there was immunopeptidomics data.

Therefore, I decided to test the rapidly growing MMRd cell line HCT116 for which I also had access to immunopeptidomics data. I analysed NCI-H747 alongside HCT116 in this iteration of the ER isolation protocol.

5.4.4.1 Method

The results of the WB from approach 3 highlighted the need to reduce the ER loss by decreasing the mitochondrial exclusion spin even further. The publication (Itzhak *et al.*, 2016) exhibited use of a 3,000g spin for mitochondrial separation, and so this was tested (Figure 3.8A).

Additionally, the article (Itzhak *et al.*, 2016) detailed that a differential centrifugation step after mitochondrial exclusion, a spin of 5,400g, could achieve a high level of ER isolation whilst leaving endosomes, lysosomes, and the Golgi apparatus in the supernatant. Therefore, alongside the 3,000g-only centrifugation from the previous experiment (titled Version 1 (V1)), I tested the effect of the sequential 3,000g and 5,400g centrifugations (titled V2) on final product purity (Figure 3.8A). Protocol detailed in full in the Materials and Methods Section.

Further to the findings from approach 3, I also chose to discard roughly 80% of the top layer, containing cytoplasmic content, prior to needle collection of the gradient layers. This was done to try and additional remove sources of cytoplasmic contaminants from the final preparation.

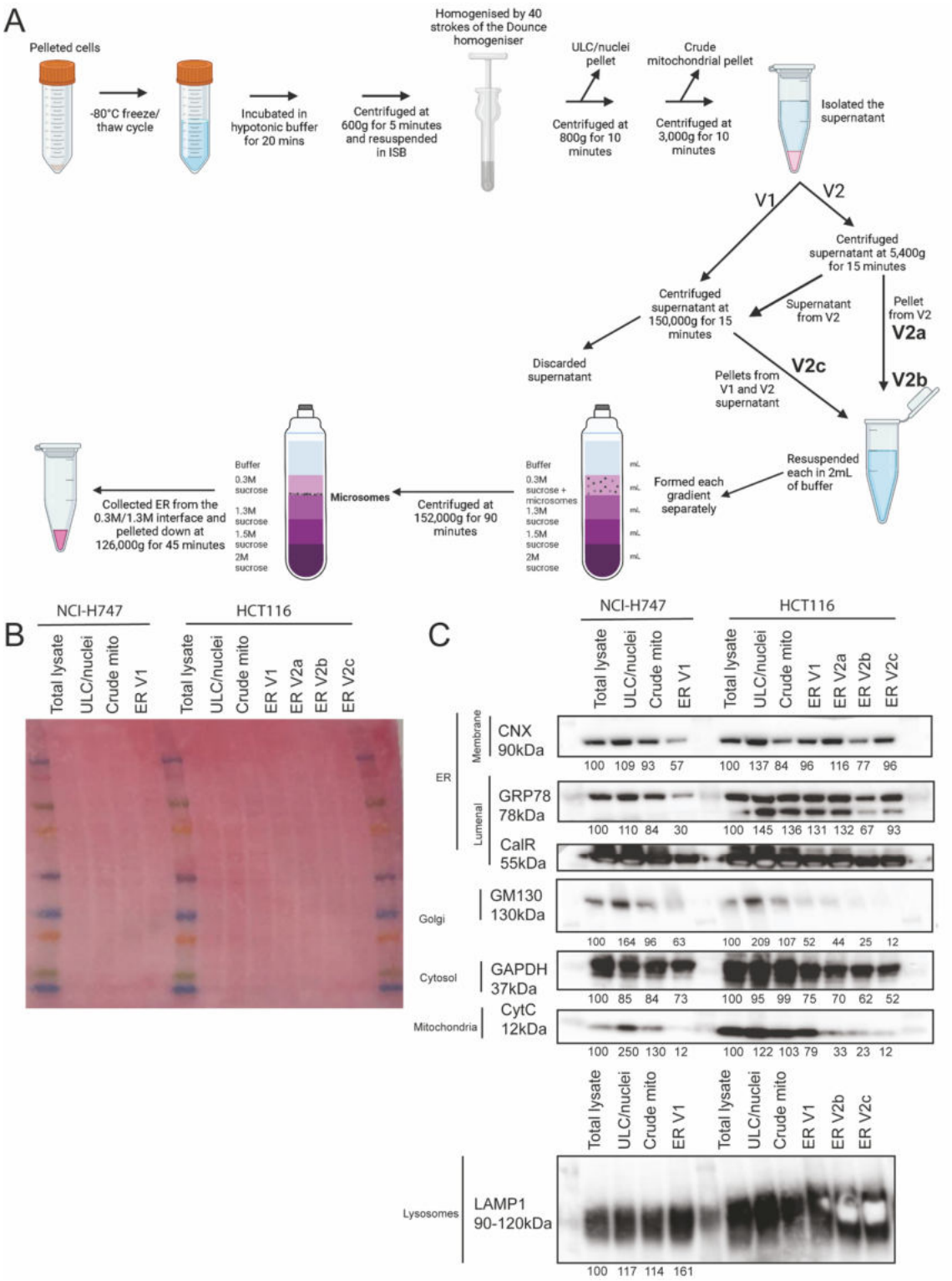


Figure 3.8. ***Characterising the success of a further modified sucrose-based ER isolation method.*** **A:** Schematic depicting the process of ER enrichment. **B:** Ponceau stain of the WB of the yielded pellets from across the ER isolation process. **C:** WB illustrating the yield of ER, Golgi, cytosolic, mitochondrial, and lysosomal proteins across the isolation process**.**

5.4.4.2 Results

Lane ER V1 represented the gradient product that was generated by layering the post-mitochondrial (3,000g spin) supernatant over the gradient. Lane ER V2a depicted the post-5,400g spin pellet, not subjected to a gradient; ER V2b represented the post-5,400g spin pellet which was resuspended and layered over a gradient; and ER V2c represented the post-5,400g spin supernatant which was pelleted, resuspended, and put over a gradient (detailed on schematic in Figure 3.8A). The WB (Figure 3.8B-C) of the sucrose approach 4 lysates for NCI-H747 did not show an increased retention of ER luminal proteins CalR and GRP78 in the final ER V1 lane. Unfortunately, the CalR signal was not quantifiable due to overexposure, but in the ER V1 lane the CalR band was visibly smaller, and the GRP78 marker only had 30% of the band intensity of the total lysate. Furthermore, in the ER V1 lane the ER membrane protein CNX appeared to have decreased to 57% of the total lysate intensity.

In contrast to the NCI-H747 cells, in HCT116 the ER proteins CalR, GRP78, and CNX were better retained in the candidate ER layers (ER V1, ER V2b, and ER2c). I previously hypothesised that, if the data presented in (Itzhak *et al.*, 2016) held true for the cell lines I was using, that ER V2b, would contain more ER, rather than ER V2c. This was because the publication had indicated that majority of the ER was expected to pelleted in the 5,400g centrifugation step (Itzhak *et al.*, 2016). The ER appeared most highly enriched in the ER V2a (with GRP78 and CNX at 32% and 16% higher intensity than the total lysate). For the post-gradient lanes, more ER appeared to be obtained from the post-5,400g spin supernatant (marker GRP78 at 93% for V2c compared against 67% for V2b). A much higher similarity of ER luminal protein retention against ER membrane protein retention was also achieved in HCT116. This supported the notion that different cell lines have different sensitivities to lysis protocols (Pryor, 2015). This highlighted that whilst it was appropriate to use HCT116 for the optimisation, the cell lysis protocol may need to be adjusted for use in other cell lines.

In terms of cytoplasmic contamination, similar to the previous run, a 52% reduction of GAPDH was achieved in the HCT116 ER V2c lane. Furthermore, despite reducing the mitochondrial exclusion spin speed to 3,000g, in the ER V2b

lane the CytC band was only 23% of the intensity of the total lysate, showing this protocol improved ER enrichment whilst achieving mitochondrial exclusion. This amount of mitochondrial removal was considered acceptable when balanced with the increased retention of the ER markers, so I chose not to further adjust the mitochondrial exclusion step.

Lysosomal content of the final ER preparations was not initially evaluated as I was first focused on reducing ER loss, and none of the founding protocols had investigated the lysosomal content of the final pellets (Chen *et al.*, 2010; Williamson *et al.*, 2015). Further to inspiring the testing of an additional 5,400g spin, the (Itzhak *et al.*, 2016) publication highlighted that it was highly likely that the post-3,000g supernatant would contain lysosomes. Therefore, I chose to perform a second WB for LAMP1, a lysosomal marker, on lysates from V1, V2b, and V2c. I focused on V1, V2b and V2c preparations as they had been subjected to the sucrose gradient system and therefore contained less GAPDH, GM130, and CytC contamination than the pre-gradient fraction V2a. Unfortunately, all ER preparations showed insufficient lysosomal removal, with all protocols co-enriching for the ER and lysosomes (Figure 3.8C). Contaminating peptide signal from lysosomes could have confounded any peptide signal from the ER peptides (Roche and Furuta, 2015), making it difficult to discern the ER peptidome. Therefore, removal of lysosomes was important for the end product of the protocol to be applicable for ER peptidomics. The co-enrichment of ER and lysosomes indicated that the densities of the gradient system were not compatible with the removal of lysosomes. Therefore, I chose to pursue a different gradient system to hopefully achieve ER enrichment with acceptable purity.

5.5 Repurposing a gradient intended for mitochondrial enrichment

Particle density and sedimentation coefficients for the endoplasmic reticulum and lysosomes are very similar, making them hard to separate (Ridge, 1978). However, one study (Zhang *et al.*, 2000) showed that lysosomes can be successfully purified from mitochondria utilising a modified (Storrie and Madden, 1990) metrizamide discontinuous gradient. The results also showed a

corresponding enrichment of ER at the 17/35% metrizamide interface (using CNX as the ER WB marker), where it had good separation from lysosomes (using LAMP1 as the lysosomal WB marker). This data suggested this gradient system was appropriate for the separation of the ER from the lysosomes, following the mitochondrial exclusion step to reduce the chance of co-enrichment.

With Metrizamide no longer commercially available, it was replaced by Histodenz as an alternative gradient material with similar features. Histodenz was appropriate for use as it displayed a number of desirable features: 1) low toxicity, 2) an osmolality close to that of mammalian tissues (112-299 mOsm in 10-30% Histodenz (Rickwood, Ford and Graham, 1982), 290-320 mOsm in mammalian tissues (Gagné, 2014)) which is important for the maintenance of organelle features, 3) and high density. Histodenz was furthermore viable for subsequent MS, as it could be removed from samples by dilution, filtration, or TCA precipitation of the sample (Rickwood, Ford and Graham, 1982).

5.5.1 Histodenz gradient centrifugation approach 1

This approach was intended to improve on previous iterations by isolating the ER with higher purity, whilst removing contaminating lysosomes, and sustaining good elimination of mitochondria. For this protocol, the cell line HCT116 was used. In this run, aliquots of the supernatant were taken every stage, both before and after centrifugation to inform the success of ER enrichment.

5.5.1.1 Method

The protocol is illustrated by a schematic in Figure 3.9A, and detailed description in Materials and Methods.

The steps of cell lysis, nuclear exclusion and mitochondrial exclusion were kept constant with the previous protocol. After sucrose gradient approach 4, I decided to re-attempt the post mitochondrial exclusion 5,400g ‘ER-pelleting’ step, to allow me to definitively evaluate whether this step could help to yield ER with fewer contaminants for layering over the gradient.

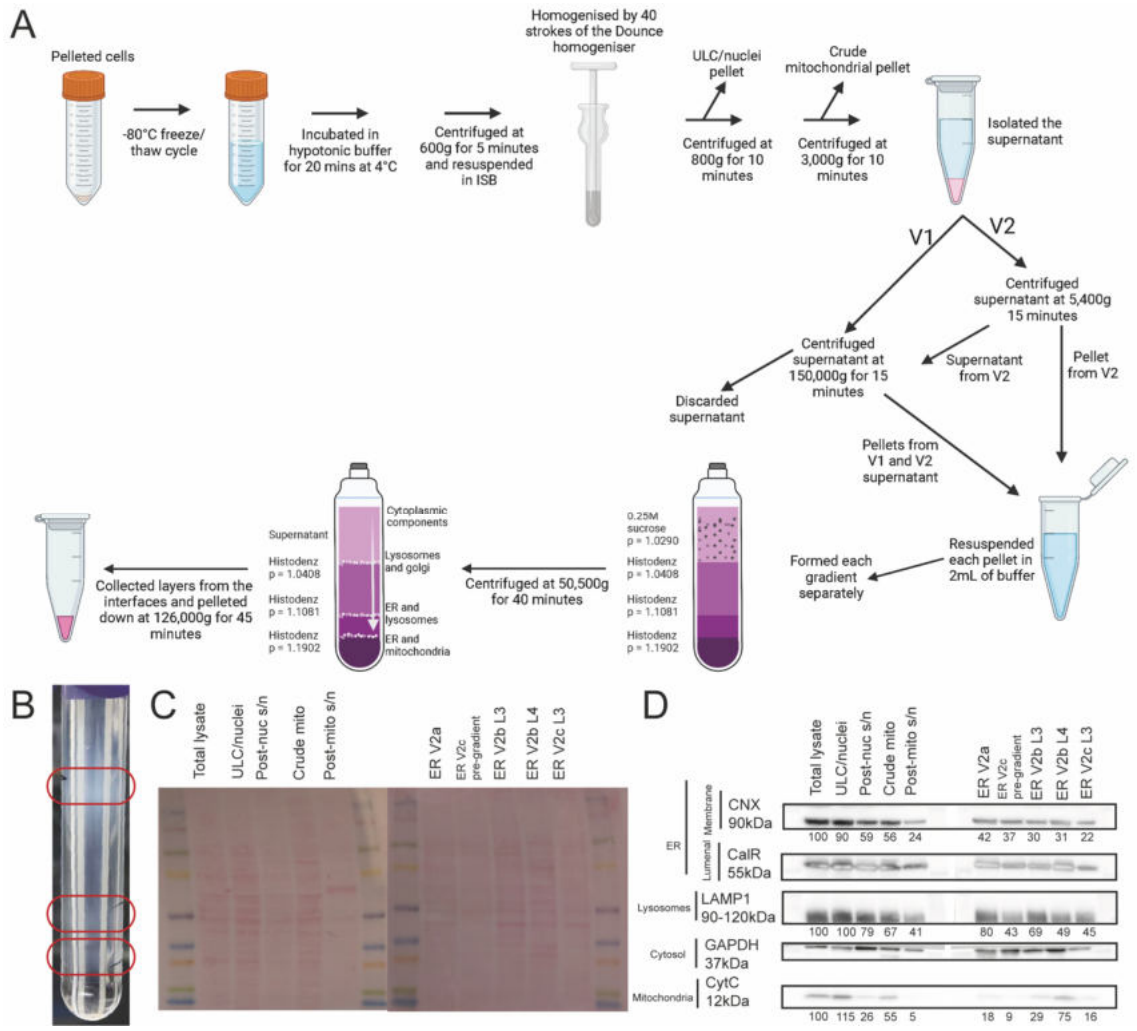


Figure 3.9. ***Characterising the success of a Histodenz-based ER isolation method.*** **A:** Schematic depicting the process of ER enrichment. **B:** Photograph of the Histodenz gradient post ultracentrifugation. **C:** Ponceau stain of the WB of the yielded pellets from across the ER isolation process. **D:** WB illustrating the yield of ER, cytosolic, mitochondrial, and lysosomal proteins across the isolation process**.**

5.5.1.2 Results

Lane ER V2a depicted the post-5,400g spin pellet, not subjected to a gradient; ER 2b represented the post-5,400g spin pellet which was resuspended and layered over a gradient; and ER V2c represented the post-5,400g spin supernatant which was pelleted, resuspended, and put over a gradient (detailed on schematic in Figure 3.9A).

As the lysates were of a low concentration, they were run on 10-well gels to allow for a higher loading volume. Unfortunately, there was not enough protein content from condition V2c layer 4 (L4) to add to the WB (the post-5,400g spin supernatant, collected from L4 of the Histodenz gradient), however, the WB provided enough information to help decision-making on how to further optimise the protocol (Figure 3.9C-D).

This WB displayed highest enrichment of ER protein CalR in the ER V2a lane and the ER V2b L4 lane. This showed that more ER was pelleted in the 5,400g spin than left in the supernatant. However, the densitometry for CNX highlighted a large amount of ER protein was still left behind in the supernatant (ER V2c pre-gradient displayed 37% of total lysate intensity, only 5% less than the post-5,400g spin pellet). Furthermore, the addition of the 5,400g spin did not sufficiently decrease contamination of cytosol, lysosomes, or mitochondria in the pellet (ER V2a). This data suggested to eliminate the 5,400g spin as it had a negative effect on ER yield, and a minor effect on ER purity.

Despite apparent ER enrichment in L4 of the Histodenz gradient, it still showed a large amount of contaminating proteins, including cytosolic contamination. One advantage of this gradient was that I was no longer co-enriching for LAMP1, with the abundance clearly depleting across the WB (with pellet V2b L4 LAMP1 signal intensity decreasing to 49% of the total lysate signal), but the extent of the lysosomal removal was unsatisfactory. These results suggested that the details of the protocol needed further refinement to improve ER enrichment, and decrease the amount of contamination. This informed the following approach.

5.5.2 Histodenz gradient centrifugation approach 2

As previous run indicated that the 5,400g spin did not adequately enrich for the ER, this step was removed from the protocol, and the post-mitochondrial supernatant was simply layered over the Histodenz gradient.

From the results shown above, I hypothesised there could be three main issues for the suboptimal purification of the ER. 1) The use of frozen cells with possibly compromised organelle membranes was leading to loss of ER in lighter fractions.

2) The cell lysis approach, whilst improved by the addition of an incubation with hypotonic buffer, may have caused uneven lysis of the cells and led to increased contamination of the final preparations (Suski *et al.*, 2014). 3) The method of making Histodenz preparations by the recommended simple mass by volume calculation may not achieve precise reagent densities. The equation published by (Weast, 1986) was subsequently used to generate more accurate densities.

$$V_y = V_i \frac{(P_i - P)}{(P - P_y)}$$

V_y = Volume of diluting medium in mL.
V_i = Volume of medium to dilute in mL.
P_i = Density of medium to dilute in g/mL.
P_y = Density of diluting medium in g/mL.
P = Density of diluted solution desired in g/mL.
(Weast, 1986)

To further address the issue of possible uneven cell lysis when using Dounce homogenisation, I chose to trial nitrogen cavitation. Nitrogen cavitation works by dissolving nitrogen in to the cytoplasm of intact cells under high pressure conditions, so when the pressure is released, nitrogen bubbles form and burst the outer cell membrane. This method lyses the cells in a uniform way, which may generate microsomes of more uniform sizes.

5.5.2.1 Method

I found a paper employing a nitrogen cavitation cell lysis protocol with the aim of purifying the ER (Hammond *et al.*, 2012), with conditions gentler than many standard cavitation protocols. Nitrogen cavitation was performed at 450psi for 8 minutes to obtain cell lysate with intact ER microsomes (illustrated in their publication by electron microscopy). The data provided suggested this method could address issues of uneven cell lysis and provide a good compromise between lysate yield and retaining organelle membrane integrity.

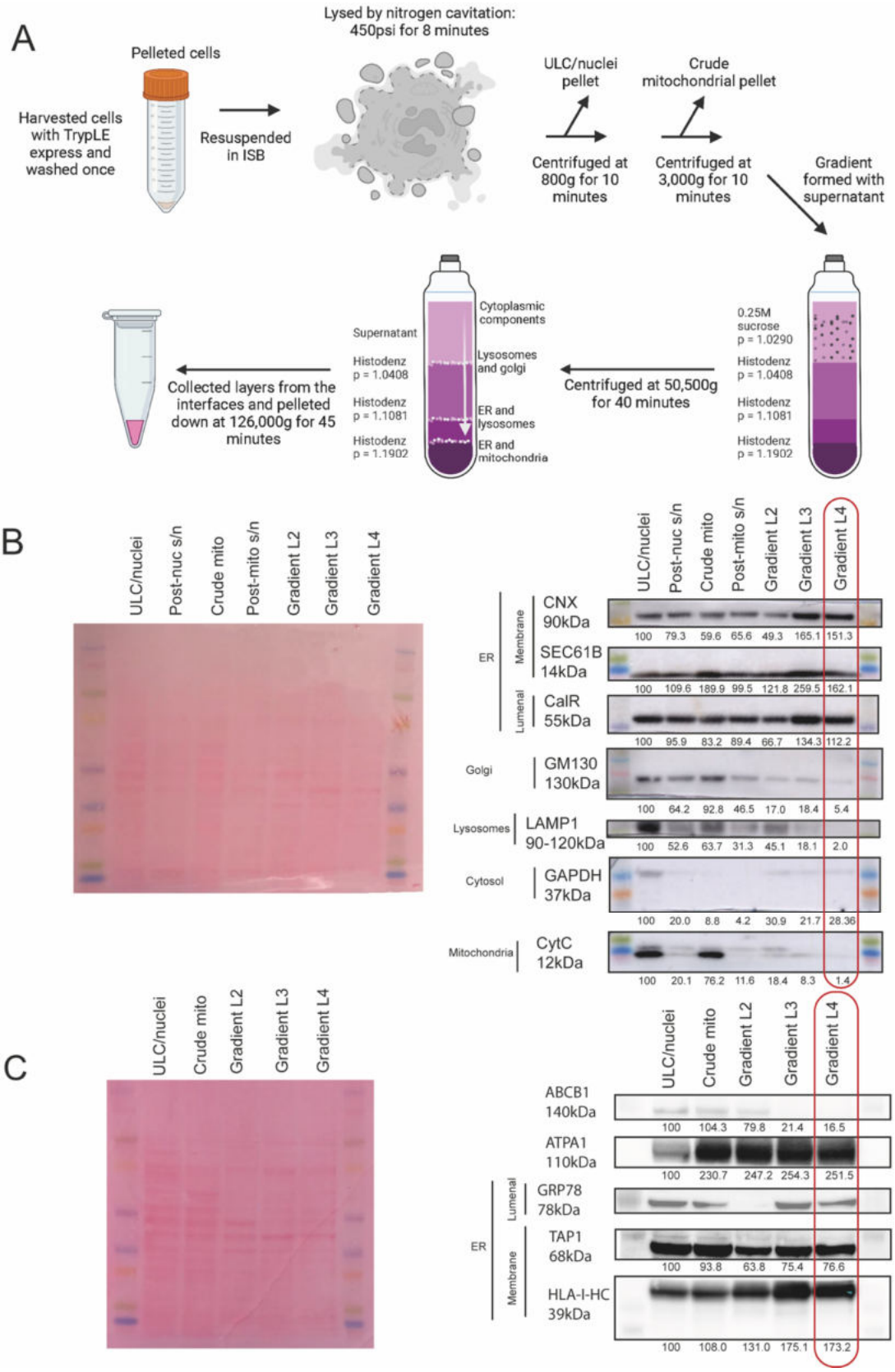


Figure 3.10. ***Characterising the success of an adjusted Histodenz-based ER isolation method.*** **A:** Schematic depicting the process of ER enrichment. **B:** Ponceau stain of the WB and analysis of enrichment for markers for the ER,

Golgi, lysosomal, cytosolic, and mitochondrial compartments across the isolation process. **C:** Ponceau stain of the WB and analysis of enrichment for markers for the ER and plasma membrane compartments across the isolation process.

5.5.2.2 Histodenz run 2 results

This iteration of the ER isolation protocol yielded a much better enrichment of ER proteins in the final product (gradient L4). In L4 ER proteins CalR, CNX, and SEC61B had a band intensity of 112.2, 151.3, and 162.1% of the ULC/nuclei band, showing good enrichment of ER components (Figure 3.10B). The ER luminal protein GRP78 was notably depleted in L2 compared against L4 (Figure 3.10C), whilst still retaining the ER membrane protein TAP1. Retention of both luminal and membrane ER proteins in L3 and L4 suggested provided further support that these were the layers of interest (Figure 3.10C).

Because the WB still showed loss of some ER proteins to the discard pellets and other gradient layers, I also interrogated the presence of HLA-I. HLA-I was confirmed have been lost to discard pellets and other gradient layers throughout the process, however, it was most strongly enriched in L3 and L4 (Figure 3.10C). The high presence of HLA-I alongside ER markers in L3 and L4 added confidence that I was isolating the site of HLA-I-loading, and therefore it was likely to be able to resolve an ER peptidome.

Although L3 and L4 displayed similar ER enrichment, L4 displayed the lowest level of Golgi, lysosomal and mitochondrial contamination. Therefore, the best ratio of ER enrichment to contaminants was apparent in layer L4, so this was confirmed as the layer of interest.

Finally, I assessed for presence of cell surface proteins to establish if the plasma membrane was being depleted, to ensure any peptide signal was not coming from HLA-I from the cell surface. For the cell surface markers, two proteins were assessed for, based on data from the HeLa spatial proteome (Itzhak *et al.*, 2016) and The Human Protein Atlas (HPA) analysis. ATPA1 is considered a cell membrane marker, and the HeLa spatial proteome showed it to be of mostly plasma membrane (PM) distribution, but HPA analysis showed it to be also found

in the ER with a confidence of 4/5. As a result, a second cell membrane marker, ABCB1 was selected. ABCB1 was attributed to the PM by the HeLa spatial proteome and assigned to the PM with high confidence (5/5) by HPA analysis. The WB showed sequential depletion of protein ABCB1 throughout the protocol, but enrichment of ATPA1 in the gradient layers. This data therefore suggested that ABCB1 was a true PM marker, and ATPA1 was also present in the ER, as this correlated with the enrichment of other ER proteins. However, to have absolute confidence of PM depletion, this needed to be further investigated when MS was been performed.

5.5.3 Assessing reproducibility of the Histodenz approach 2

As organelle isolation by rate-zonal centrifugation can be a very delicate process, it was important to determine if this protocol could be robust and yield reproducible results. The 2nd Histodenz approach produced a satisfactory level of ER enrichment and lysosomal depletion, therefore the methodology was replicated.

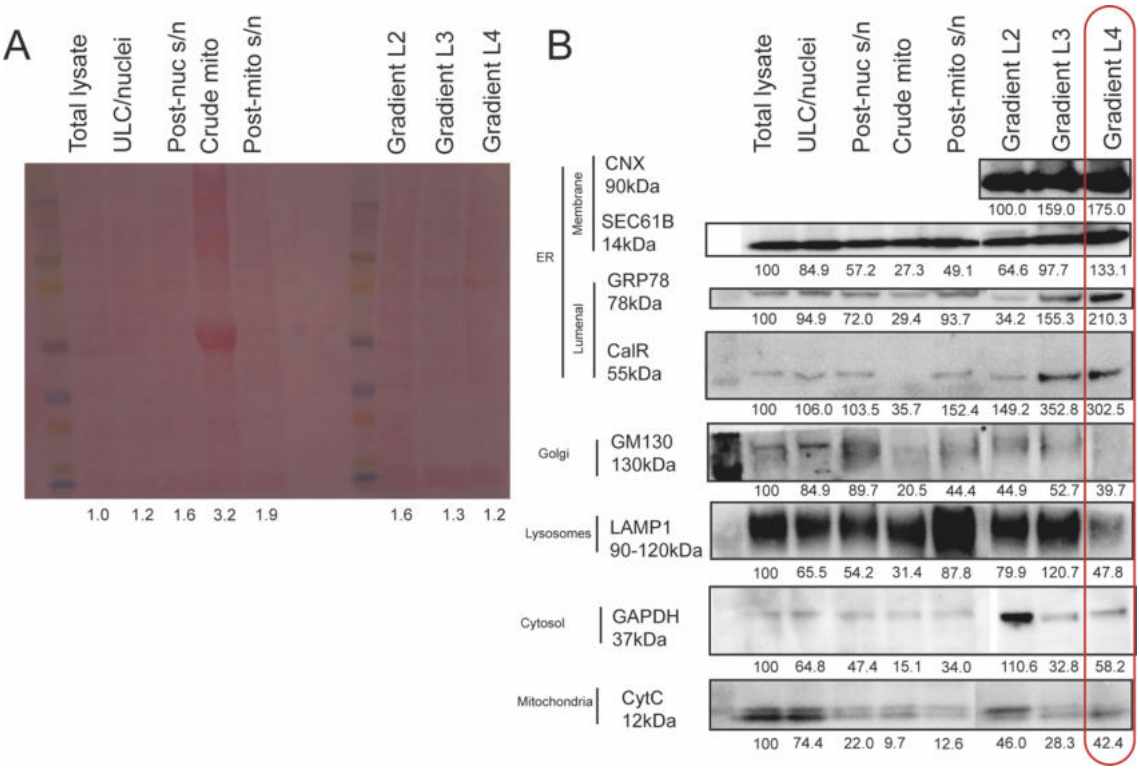


Figure 3.11. ***Assessing reproducibility of the adjusted Histodenz-based ER isolation method.*** **A:** Ponceau stain of the WB of the yielded pellets from across the ER isolation process. **B:** WB analysis of enrichment for markers for the ER,

Golgi, lysosomal, cytosolic, and mitochondrial compartments across the isolation process.

Figure 3.11A showed the accidental overloading over the crude mitochondrial lane which could have impeded the quantification and comparison of proteins across the lanes. I therefore quantified the Ponceau stain on the Amersham Western Blot Imager 600 to be able to more precisely scrutinize loading. The Ponceau staining across each lane was used to normalise the intensity of the probed bands. This analysis was performed on all future WB where possible.

Analysis of the marker proteins confirmed the good ER enrichment and contaminant depletion seen in the previous run (Figure 3.11B). The gradient layer with the strongest enrichment of ER proteins, and the strongest depletion of contaminants was again layer 4, consistent with the previous run. This confirmed this technique was therefore suitable for the first attempt of peptide isolation from the enriched ER.

5.5.4 Assessing the purity of the technical duplicate of the samples to be sent for MS

Another ER preparation using the protocol developed in Section 5.5.2 was performed to yield samples for MS. Technical duplicates were generated so that one sample could be analysed for subcellular markers in a Western Blot, and the second one by MS. After each of the layers were harvested from the duplicate gradients, L2, L3, and L4 from both gradients were combined during the wash step, then split in half for processing. This was done to ensure the WB would be entirely representative of the sample being sent for MS. Layers were lysed with RIPA buffer and lysates were quantified with the Pierce Rapid Gold BCA assay. The yields for L2, L3 and L4 were 240.7μg, 42.3μg, and 86.3μg respectively. This was consistent with the visual size of these layers (Fig 3.12A).

The discard pellets and resulting supernatants from each stage were prepared as duplicates in loading buffer (aside from L3, where there was not high enough yield for two blots). The first half of the preparation was loaded in to a Bis-Tris 4-12% gel, assessed for loading with a Ponceau stain (Figure 3.12B) and blotted for different subcellular markers (Figure 3.12C). The Ponceau stain appeared

smudgy for this transfer, so it was not appropriate for loading normalisation. The blot depicted a satisfactory removal of Golgi, cytosolic, lysosomal, and mitochondrial markers. However, unlike the other the blots from Section 5.5.2.2 and 5.5.3, it appeared as if there was enrichment of CNX in L2. The other blots previously showed a defined depletion of CNX in L2 when compared against in L4. This loss of CNX to L2 may have resulted from increased organelle membrane lysis, despite the cells being lysed under the exact same conditions as before.

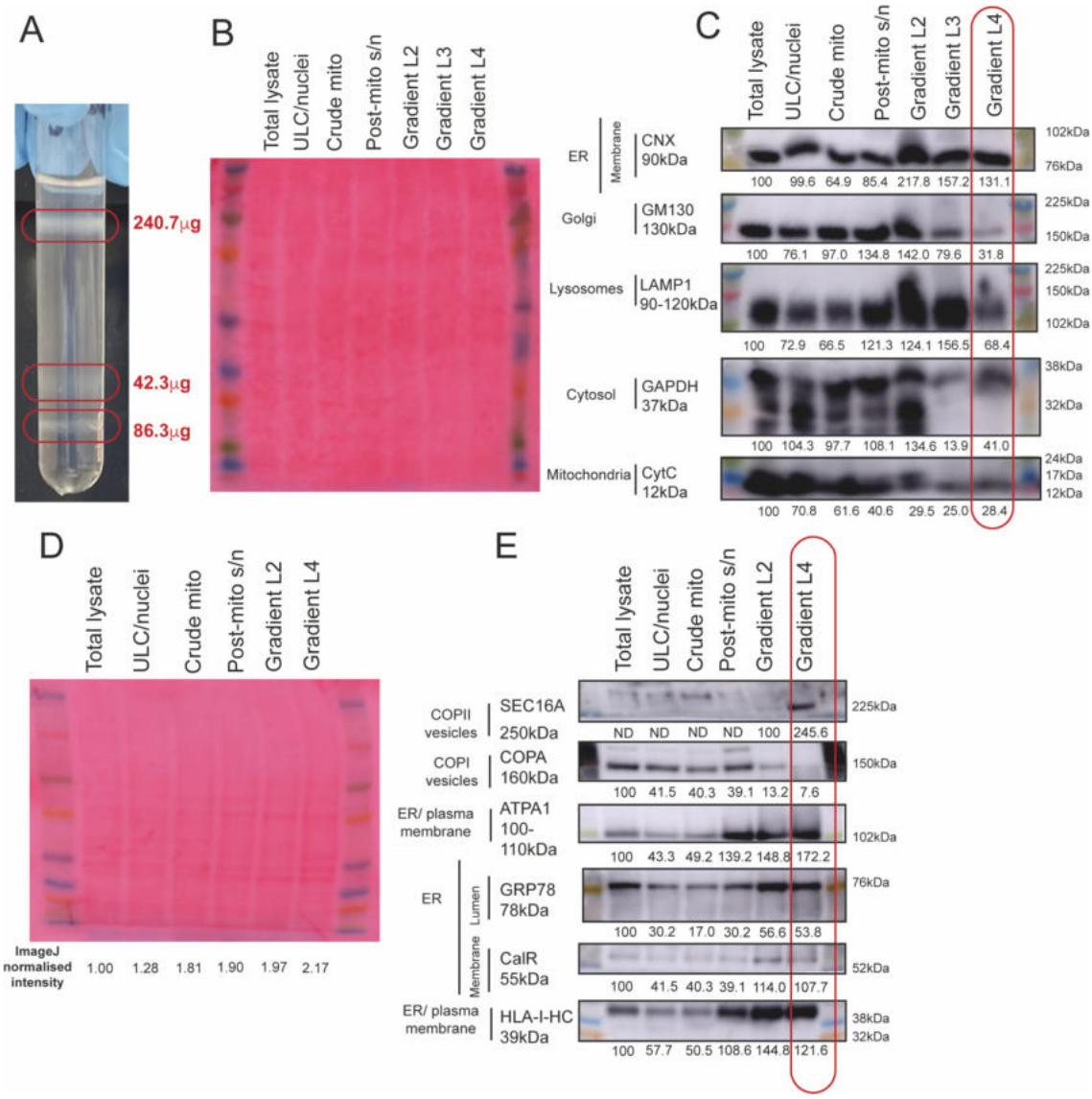


Figure 3.12. ***WB analysis of purity of yield from Histodenz-based ER isolation method.*** **A:** Photograph of the Histodenz gradient post ultracentrifugation**. B:** Ponceau stain of the Bis-Tris WB of the yielded pellets from across the ER isolation process. **C:** Bis-Tris WB analysis of enrichment for markers for the ER, Golgi, lysosomal, cytosolic, and mitochondrial compartments

across the isolation process. **D:** Ponceau stain of the Tris-Acetate WB of the yielded pellets from across the ER isolation process. ImageJ normalised lane intensity annotated underneath (intensity/intensity of lane with minimum intensity). **E:** WB analysis of enrichment for markers for the ER, COPI, and COPII compartments across the isolation process. Ponceau-normalised ImageJ intensities listed underneath.

To achieve better resolution of the higher molecular weight proteins and to scrutinize additional ER markers, another WB was performed utilising 3-8% Tris-Acetate gel in Tris-Acetate buffer, the Ponceau stain depicted in Figure 3.12D. I analysed for enrichment of ER proteins GRP78, CalR, and ATPA1. In each case, there appeared to be similar amounts of ER proteins in L2 and L4; this was certainly a poorer enrichment than the previous 2 runs. Ideally, this run would have been repeated, but due to time constraints, it was important to assess these pellets further for suitability for MS. However, L4 was still considered the optimal layer due to the depletion of lysosomal, cytosolic and Golgi proteins.

The WB from the previous run showed HLA-I-HC to be present in all the fractions of the gradient despite the different densities of the Histodenz layers. There were two main possible reasons for HLA-I-HC occurring in every fraction of the gradient: 1) the nitrogen cavitation caused uneven lysis of the ER, breaking it in to smaller and larger microsomes, or 2) some of the HLA-I-HC that is found in the lighter layers may actually be part of vesicles that are moving between the ER, the Golgi, and the cell surface. I approached this question in 3 ways: 1) looked for enrichment of COPI or COPII vesicles by western blotting for vesicle markers, 2) looked in-depth at the proteomics profile of each fraction, 3) looked at the peptide lengths within each fraction. I predicted peptides to be longer in the heavier fractions of the gradient, as this was be predicted to contain the ER which would contain non-cleaved HLA-precursor peptides. Therefore, I expected the shorter peptides to occur in the lighter fractions where the correct peptide length had been loaded on to HLA-I, and the complex released from the ER.

I first blotted for markers of retrograde and anterograde transport: COPA for COPI, and SEC16A for COPII (Figure 3.12E). COPA protein was most enriched in the total lysate, and it was totally undetectable in the L4 lane (the mark that is

seen below is seen due to being imaged too close to another membrane piece). This data suggested that HLA-I-HC signal in L2 may be partly explained by the presence of COPI vesicles in L2. This also supported that this technique yields a L4 which is successfully depleted of COPI transport membranes. Initially SEC16A was recommended as a COPII marker, however, further research showed SEC16A to be a known part of the membrane budding protein complex at ER exit sites (ERES), and therefore is likely to indicate the sites in the ER where protein export and vesicle budding is occurring (Maeda *et al.*, 2019). Two protein bands were visible, the higher molecular weight (MW) band being depleted, as the lower MW band became enriched for. A repeat WB with a 250kDa and 225kDa MW ladder illustrated the lower band to be at 250kDa, confirming it to be a single isoform of SEC16A. The higher band was therefore either SEC16A with post-translational modifications, or it was non-specific staining that disappears as enrichment of ER proteins occurs throughout the assay. The strong enrichment of SEC16A initially supported that the assay achieved specific enrichment of the ER due to the presence of SEC16A at ERES.

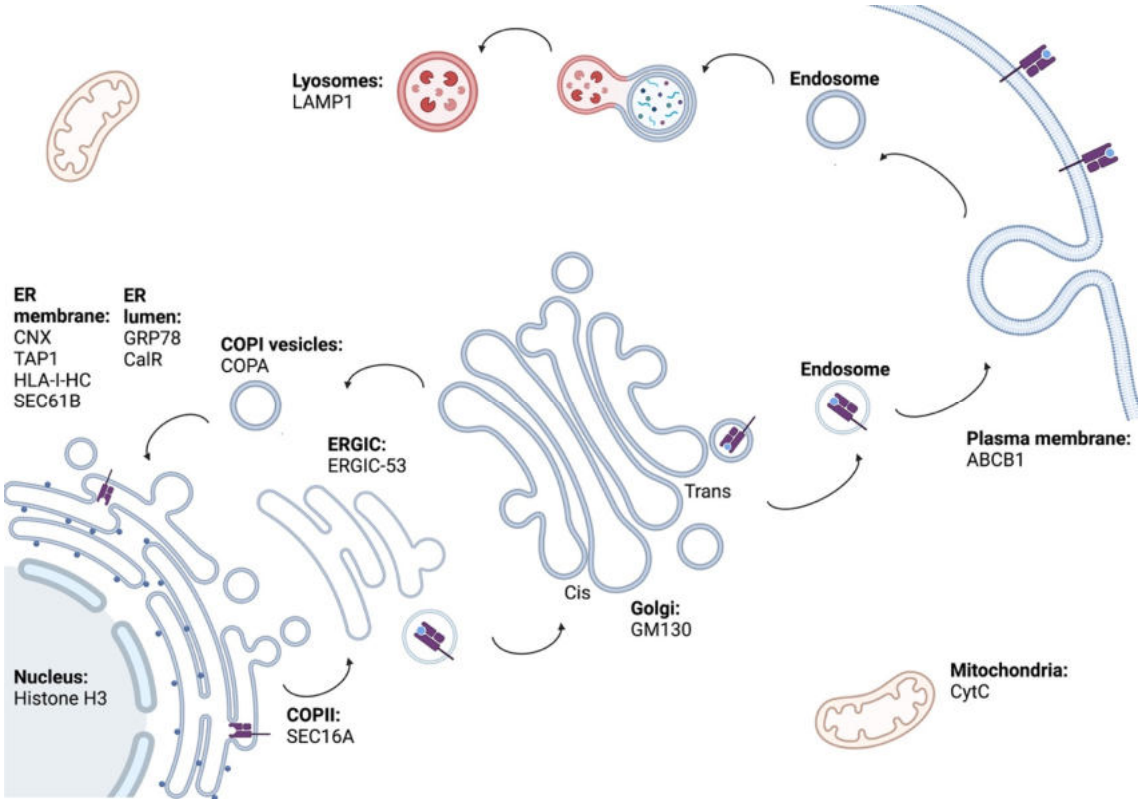


Figure 3.13. Subcellular localisation map of markers interrogated by WB.

Across 3 replicate western blots (Figures 3.10 – 3.12), I was able to analyse success of the depletion or enrichment of a large number of subcellular markers

through the Histodenz ER enrichment method (Figure 3.13). This helped to confirm strong enrichment of the ER proteins CNX, GRP78, CalR, HLA-I-HC, SEC61B (mean 1.86x enriched across the protocol), alongside COPII protein SEC16A which defines ERES. Furthermore, I have shown that L4 is successfully depleted of plasma membrane with marker ABCB1; of Golgi apparatus with marker GM130; of mitochondria with marker Cytochrome C; and of lysosomes with marker LAMP1. I believe this could be a useful panel of markers for the future interrogation of success of subcellular fractionation.

Although this gave confidence that this isolation protocol could retain the ER whilst successfully depleting the contaminants, the next step was to analyse the proteomics of the yielded pellets to allow for more detailed analysis of layer purity. Peptidomics was also performed concurrently to analyse the number of peptides that could be yielded from a single gradient of a single 5-layer T175. This data was important as allowed for assessment of the feasibility of peptidomics from the current input, and would indicate if higher input to a single gradient, or increased gradient numbers, would be required.

5.6 Preparation of gradient fractions for MS analysis

To begin processing for MS, the samples were thawed on ice before pelleting at 16,100g for 15 minutes. Samples were then processed as described in the Materials and Methods section (Schematic in Figure 3.14). To summarise, the pellets were resuspended in lysis buffer containing sodium deoxycholate (SDC) and disrupted by sonication. The quantifications of the duplicate preparation informed the amount of lysis buffer that was utilised, so the larger pellet from L2 was lysed in 4X the volume of lysis buffer than the L3 and L4 pellets. The samples were then boiled, sonicated further, and the proteins precipitated out by addition of acetonitrile (ACN). The samples were then spun down, allowing the peptide-containing supernatant and the protein pellet to be separated and processed independently.

The supernatant was dried by vacuum centrifugation. Peptides were then resuspended in triethylammonium bicarbonate (TEAB) buffer and TMT-labelled. The protein pellet was resuspended in lysis buffer containing reduction and

alkylating agents, further disrupted by sonication, then left to incubate for an hour. Following this, the proteins were trypsin-digested overnight, then TMT-labelled.

The full volume of the peptide suspensions from each layer were loaded to help determine the peptide yields from each fraction. For proteins, the volume of lysate loaded was normalised to the input of the lowest abundance sample. This was achieved by injecting a 1:1:1 1μL pre-mix in to the MS to ascertain the relative concentrations of each lysate. This was important as it enabled the same mass of protein to be loaded for each layer, allowing assessment of relative protein abundances to occur.

Database searching was performed by Zuza Kozik for our collaborator Professor Jyoti Choudhary. MS spectra were searched against UniProt annotated reference proteomes of *Homo Sapiens* (January 2023) in Proteome Discoverer v2.4. For the peptidomics, only peptides 8-25aa were considered.

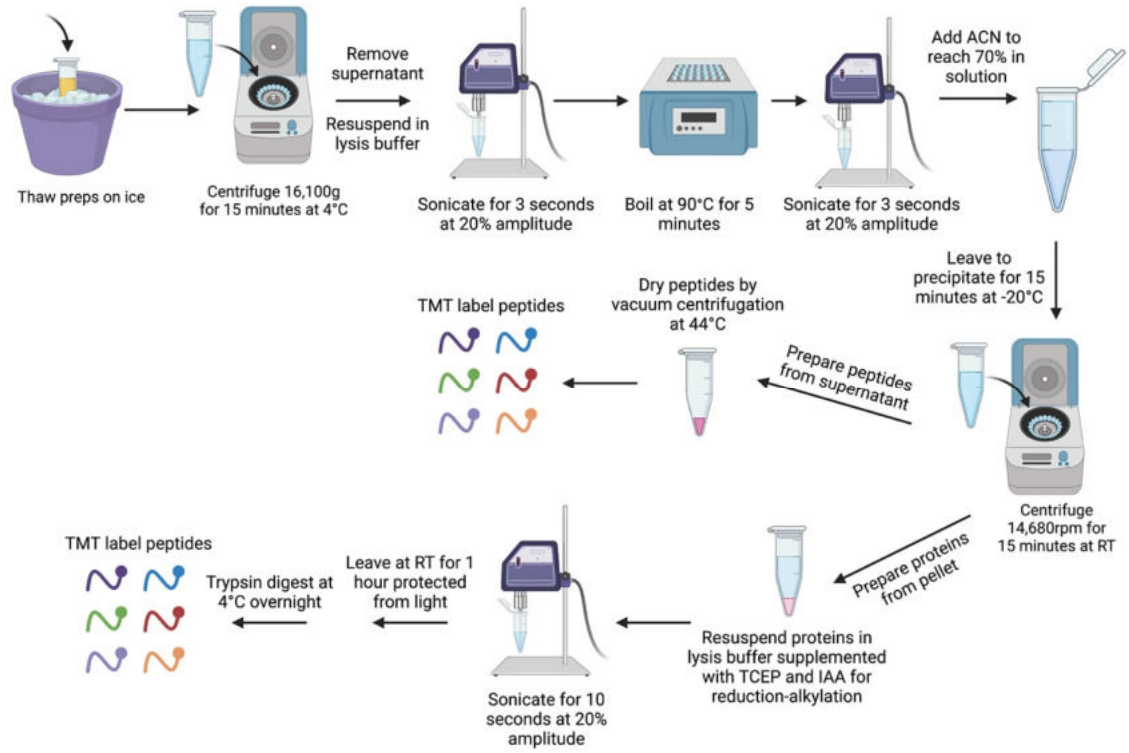


Figure 3.14. ***Schematic demonstrating the parallel preparation of peptides and proteins for MS from a single sample***.

5.6.1 MS proteomics of subcellular fractions

To get more detail on the proteins that remained in the post-mitochondrial supernatant and formed the 3 discrete layers in the gradient, I began by analysing all the proteins that were detected across 3 layers together. Across the 3 layers, I expected to see cytosolic and Golgi (due to expected presence in L2), lysosomal (due to expected presence in L3), and ER proteins (due to expected presence in L4). I had also expected to detect low levels of nuclear and mitochondrial proteins due to the exclusion by differential centrifugation. Following this, I will analyse the enrichment between layers. I used The Human Protein Atlas (HPA) organelle gene lists to assign proteins to specific organelle(s) to analyse the components of each layer (Thul *et al.*, 2017b; Uhlén *et al.*, 2019). Across the 3 layers the MS data detected a high number of nuclear and cytosolic proteins, with 2,029 and 1,889 proteins respectively, and a lower number of proteins from other compartments. For example, 552 mitochondrial proteins, and 250 proteins from the ER were detected (Figure 3.15A). However, the nuclear and cytosolic protein pools are far larger, with 7,345 and 5,8657 proteins respectively, and approximately 23.1% of the proteins that are annotated as ER are also annotated as nuclear or cytosolic in The HPA (Figure 3.15B). I therefore calculated the percentage of all proteins annotated for each of the different compartments in The HPA, that had actually been detected by MS across the three layers to determine the representation level of each organelle proteome across the gradient layers. This showed that the mitochondria had the largest proportional representation of all the compartments, followed closely by the ER (Figure 3.15C). This suggested that the mitochondrial and ER compartments had lost the least of their protein pool to the discard pellets. Whilst this was a positive finding for the ER, this was a negative finding regarding the mitochondria, as the 3,000g mitochondrial exclusion spin was intended to deplete a large proportion of the mitochondrial compartment.

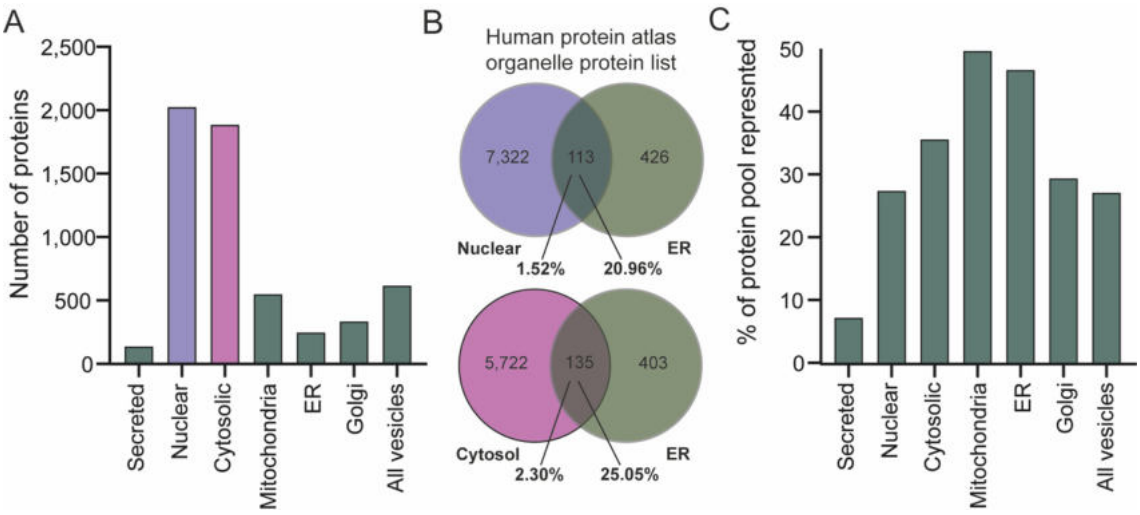


Figure 3.15. ***Subcellular locations of MS-detected proteins, according to The Human Protein Atlas (HPA)*****. A:** Number of proteins attributed to all subcellular compartments. **B:** Overlap of HPA nuclear and ER, or cytosol and ER, proteins, according to HPA gene lists. **C:** The percentage of all known proteins for each compartment detected by MS.

I next examined the relative enrichment of well-described individual organelle marker proteins (including those assessed in WB and summarised in Figure 3.13) across L2-L4 of the gradient system by comparing protein intensity values (Zhang *et al.*, 2000; Watson *et al.*, 2006; Suski *et al.*, 2014; Gomes *et al.*, 2019; Maeda *et al.*, 2019; Ilik *et al.*, 2020; Cavazza *et al.*, 2021). Since the proteomics data of all detected proteins showed a high number of nuclear proteins, markers for multiple nuclear compartments were assessed for.

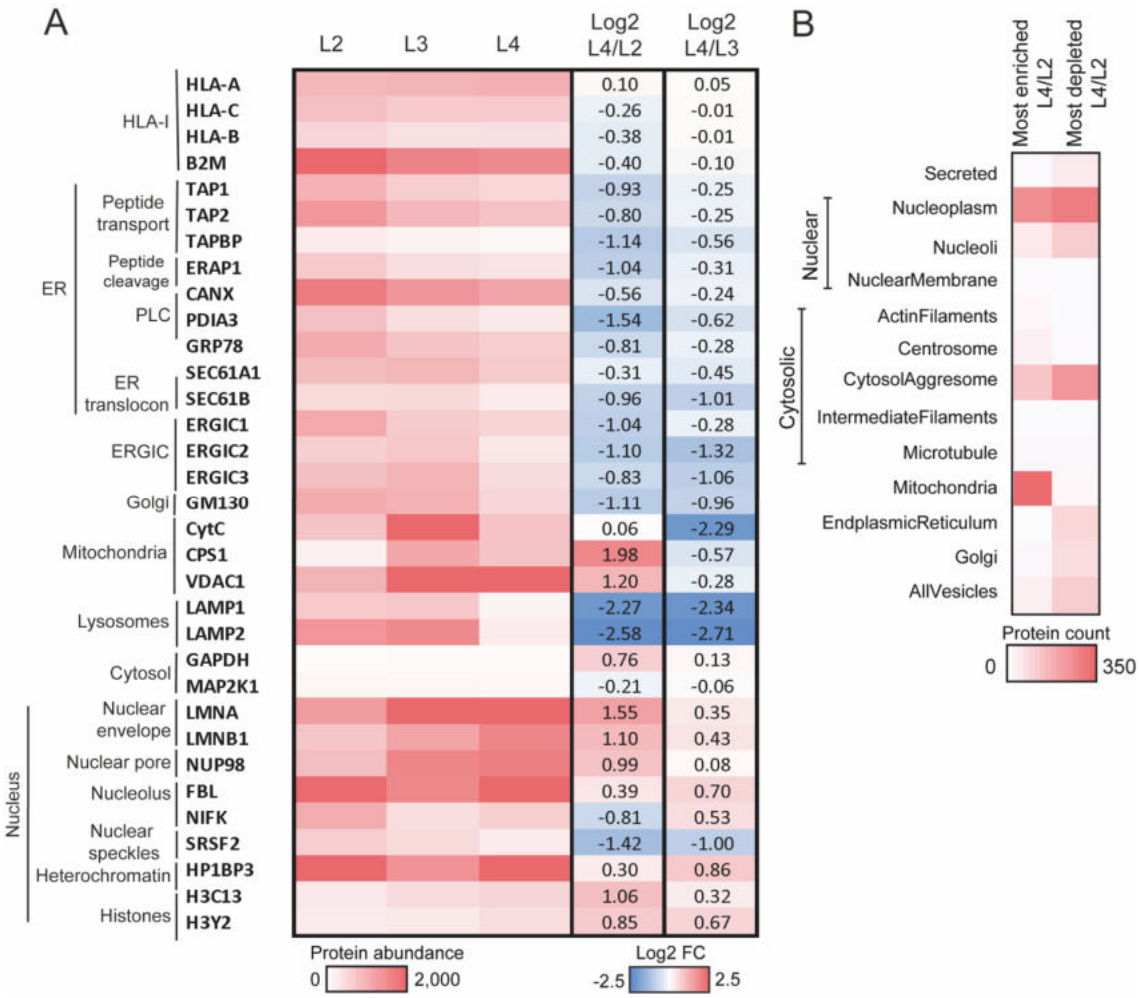


Figure 3.16. ***Relative enrichment of TMT-MS-detected organelle proteins across each gradient layer*****. A:** TMT-MS abundance values of known organelle marker proteins. **B:** Number of proteins attributed by the HPA to each organelle in most enriched and most depleted 20% of proteins (L4 relative to L2).

This analysis showed high presence of nuclear, mitochondrial, and B2M proteins (Figure 3.16A). This indicated that nuclear and mitochondrial proteins had not been well depleted, which did not correlate with the reduction of mitochondrial proteins observed in the WB. However, it may have simply been that the smaller amount of mitochondrial content evident in the L2-L4 WB lanes could still contribute a large number of proteins to the gradient due to the large size of the mitochondrial proteome. The data also showed very low presence of the cytosolic proteins GAPDH and MAP2K1 across all 3 layers (with mean intensity values of 51.4 and 100 respectively). This indicated that cytosolic proteins had been well depleted, which did correlate with the WB which showed depletion of cytosolic proteins in the gradient layers compared against the total lysate.

To analyse the relative enrichment of proteins in L4 relative to L2 or L3, I calculated the log2 fold change of the protein abundances. This analysis indicated that there was a reduced relative intensity of the ER and HLA proteins in L4 relative to L2 and L3 – an average of a -0.69 and -0.31 log2 FC in L4 relative to L2 and L3, respectively. CNX and GRP78 were most enriched in L2, which did correlate with that seen in the WB in Section 5.5.4. This was at odds with the WBs in Sections 5.5.2 and 5.5.3 which used the same protocol.

There was also depletion of the ERGIC, Golgi, and lysosomal compartments in L4 relative to L2 and L3. Seeing a successful depletion of lysosomal markers was a promising indication that the function of this gradient, in terms of lysosomal depletion, was working. L2 and L3 mostly had the same direction of FC for all the compartments, aside from the mitochondria which was enriched in L4 relative to L2, but depleted in L4 relative to L3. This was likely due to the larger size and density of mitochondria, allowing it to pass through the 5.5% Histodenz concentration, to join the L3 and L4 bands at the 5.5%/17% and 17%/35% Histodenz interfaces. The nuclear proteins were enriched in L4 relative to L2 and L3, with the only exception being nuclear speckle marker SRSF2, which saw a -1.42 log2 FC. These results indicated in L4, the layer most of interest due to depletion of lysosomes, there was also relative depletion of ER, ERGIC, Golgi and conversely enrichment of mitochondria and the nucleus, specifically the nuclear envelope most strongly. The relative depletion of the ER observed in L4 was in contrast with previous WB from Sections 5.5.2 and 5.5.3, which showed clear enrichment of the ER in L4 compared against L2. This highlighted that this run did not proceed as normal, and a repeat of the protocol would be required to confirm this.

To avoid any unintentional selection bias from choosing individual marker proteins, I next went on to compare the number of different organelle proteins that were most enriched, and most depleted, in L4 of the gradient system relative to L2. To define the most enriched proteins, I selected the 20% of proteins with the largest abundance increases in L4 relative to L2, and to define the most depleted proteins, I selected the 20% with the largest abundance decreases. I analysed the number of proteins assigned to each organelle compartment by The HPA (Figure 3.16B). Again, there appeared to be depletion of ER proteins in L4

against L2, which supported the findings from Figure 3.16A. This again was not the desired result, and showed this run of the experiment did not perform as it did in Sections 5.5.2 and 5.5.3. However, the data also showed a reduced number of secreted proteins, cytosolic, Golgi, and vesicle proteins in L4, which was the desired outcome. There was presence of nuclear proteins in both the most-enriched subset and most-depleted subset, showing nuclear proteins could be found across the layers of the gradient. The only organelle compartment that appeared to be strongly enriched, was the mitochondrial compartment. When designing this protocol, I understood that the gradient could not separate the ER and the mitochondria, therefore, it would be expected for any residual mitochondria to be enriched in L4 relative to L2. However, I was disappointed at the amount of residual mitochondrial proteins that remained in the post-mitochondrial supernatant; I had expected more mitochondrial proteins to be removed in the mitochondrial exclusion step. This illustrated more mitochondrial removal was required. This indicated that perhaps the mitochondrial exclusion spin could be increased to 4-5,000g to reduce this.

Additional Gene Ontology (GO) analysis showed only 11 cell surface proteins to be found in the most enriched protein subset, and 91% of those proteins could also be attributed to other organelles, mainly the nucleus (73%) and mitochondria (63.6%). This suggested that there was no, or very minor amounts of, cell surface membrane proteins present in L4 of the gradient, illustrating that any peptides detected will not have come from surface HLA-I.

5.6.2 MS peptidomics of subcellular fractions

As the proteomics run showed L4 to not strongly enrich for ER proteins, in contrast to what had been seen in previous runs, I was cautious to make too many conclusions about the L4-enriched peptidome. Therefore, I first focused on all detected peptides across all 3 layers. These will be referred to as ‘gradient-derived’ peptides.

All peptidomics samples were prepared as in Figure 3.14. One important feature of this protocol was that a size-exclusion filter was not used, which facilitated optimum peptide yield, and further meant that the MS results were not

biased/selected by length. For this analysis, I normalised this peptidomics data as all others, normalising individual absorbance intensities such that the total absorbance for each condition was equal. I will first describe the features of the HCT116 gradient-derived peptidome, and then compare against the HCT116 HLA-I-eluted peptidome.

3,009 unique peptides were detected from the gradient fractions, derived from 1,635 unique source proteins. The length of the gradient-derived peptides ranged from 8-25aa, with 59% of the peptides 8-10aa long, 35% of all peptides being 9mers (Figure 3.17A).

I next analysed the C-terminal ends of the gradient-derived peptides to resolve the proteasomal cleavage specificity by which they were created, as applied in Section 5.3.2 (Figure 3.17B). 75.2% of the gradient-derived peptides had C-termini that could be attributed to proteasomal activity, with the largest proportion of the peptides being attributed to chymotryptic cleavage activity. This provided some confidence that the gradient-derived peptidome contained peptides that had been likely generated by the proteasome. This illustrated that they had features consistent with ER-derived HLA-bound or HLA-precursor peptides. Analysing the HPA organelle annotations of the gradient-derived peptidome source proteins showed the nuclear and cytosolic compartments to be represented by the most proteins. The mitochondria, ER, and Golgi compartments were represented similarly (Figure 3.17C).

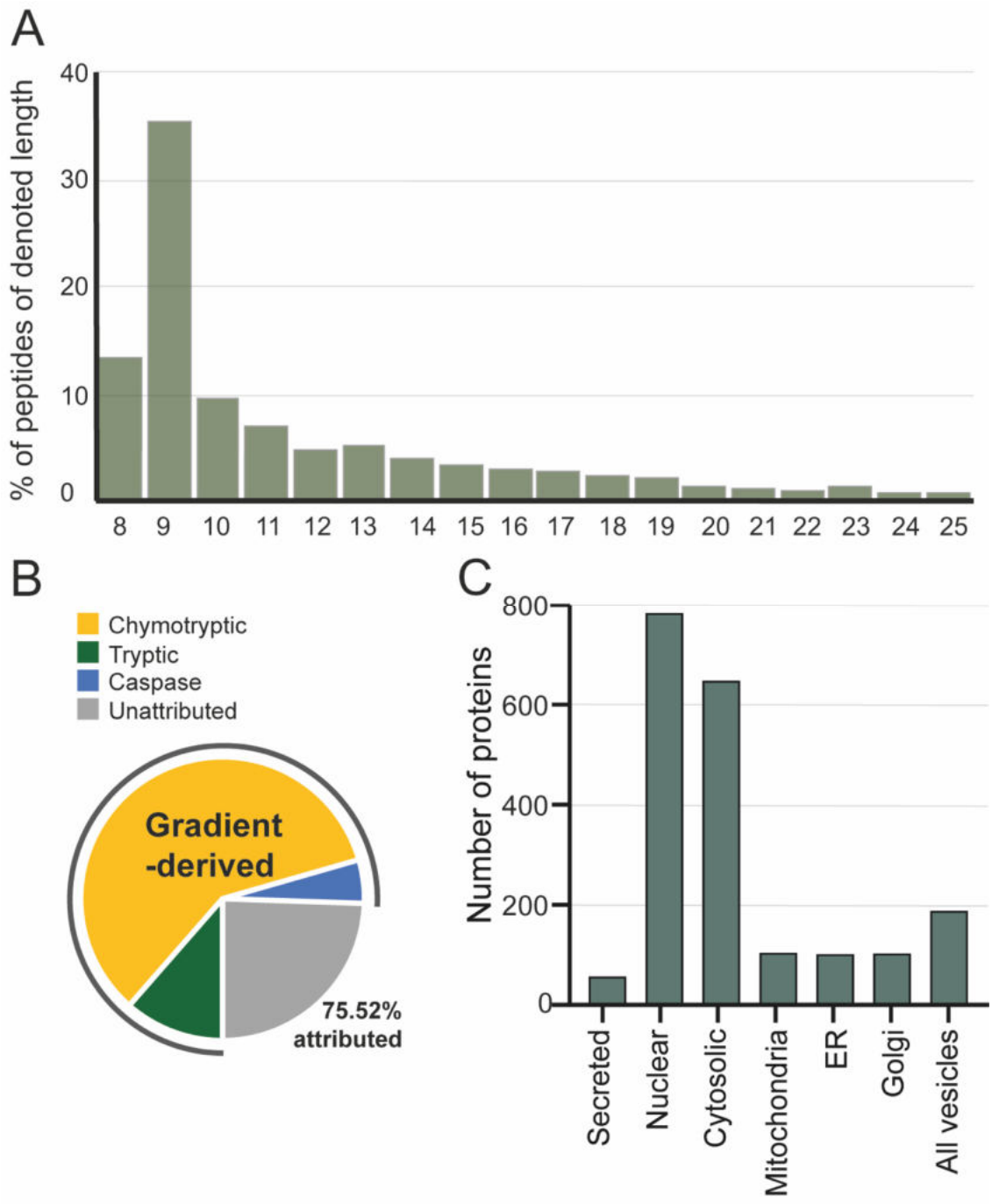


Figure 3.17. Peptide characteristics of gradient-derived peptides. **A:** Peptide length distribution of gradient-derived peptides from HCT116. **B:** C-terminal characterisation of gradient-derived peptides. **C:** Number of MS-detected peptide source proteins attributed to cellular compartments by HPA analysis.

I next compared the gradient-derived peptides against the HLA-I-eluted peptides from HCT116, also treated with IFNγ for 24H. Between the 11,897 peptides eluted from HLA-I of HCT116, and the 3,009 peptides detected from the gradient, 594 peptides were found in both. This represented 19.74% of the gradient-derived peptides, but only 4.99% of the HLA-I peptides (Figure 3.18A). Between the 6,868

HLA-I-eluted 9-mers and the 1,072 gradient-derived 9-mers, 402 were shared. This was 37.5% of the gradient-derived 9-mers; this degree of overlap suggested these may have been HLA-compatible peptides found in the ER. Furthermore, gradient-derived peptides originated from 1,635 unique source proteins and HLA-I-eluted peptides from 5,095 source proteins. Comparing source proteins of gradient-derived and HLA-I-eluted peptides, 1,094 of the source proteins contributed peptides to both peptide pools. Thus, 70.28% of the gradient-derived peptide source proteins were also represented on HLA-I. This high degree of similarity suggested that there may be a bias in the representation of these proteins as peptides, possibly due to higher protein turnover rates. I next hypothesised that ER-derived peptides would have a peptide length distribution skewed towards longer peptides, when compared against HLA-I-eluted peptides. I therefore compared the peptide length distributions of gradient-derived peptides and HLA-I-eluted peptides; this analysis was limited to peptides 8-15aa long, due to the database searches for the HLA-I-eluted peptidome only considering peptides 8-15aa long. This showed gradient-derived peptides included less peptides 9-10aa in length, and more peptides 11-15aa in length than the HLA-I-eluted peptides (Figure 3.18B).

Next, I compared the C-terminal characteristics of the gradient-derived peptides and the HLA-I-eluted peptides. In comparison to the 75.2% of gradient-derived peptides being attributed to proteasomal activity, 90.21% of the HLA-I-eluted peptides could be attributed to proteasomal activity (Figure 3.18C). As there are currently no known C-terminal ER peptidases, this difference may suggest some presence of non-proteasomally-derived peptides in the gradient-derived peptides. In both the gradient-derived and HLA-I-eluted peptides, chymotryptic activity represented the highest proportion of peptides (58.2% and 84.9% respectively), and caspase activity represented the least (4.7% and 1.4% respectively). The disparity between the proportion of peptides attributable to chymotryptic activity could be explained by the binding preferences of the HCT116 HLA, as all 6 HLA allotypes have a preference for chymotryptic-like amino acids at the C-terminus (Figure 3.19). To compare the representation of organelle compartments as peptides in the gradient-derived peptides and the HLA-I-eluted peptides, the proportional representation of all the compartments was compared, scaled for peptidome size. This illustrated overarching similarities

in distribution, but with the gradient-derived peptides exhibiting higher representation of the endoplasmic reticulum and slightly lower of nucleoplasm, actin filament, centrosome, intermediate filament, and mitochondrial proteins.

The strong similarities in peptide characteristics, and protein representation as peptides, illustrated that the gradient-derived peptides possessed features that would be expected of possible ER-derived peptides. However, I acknowledge more validation steps are required to confidently conclude this.

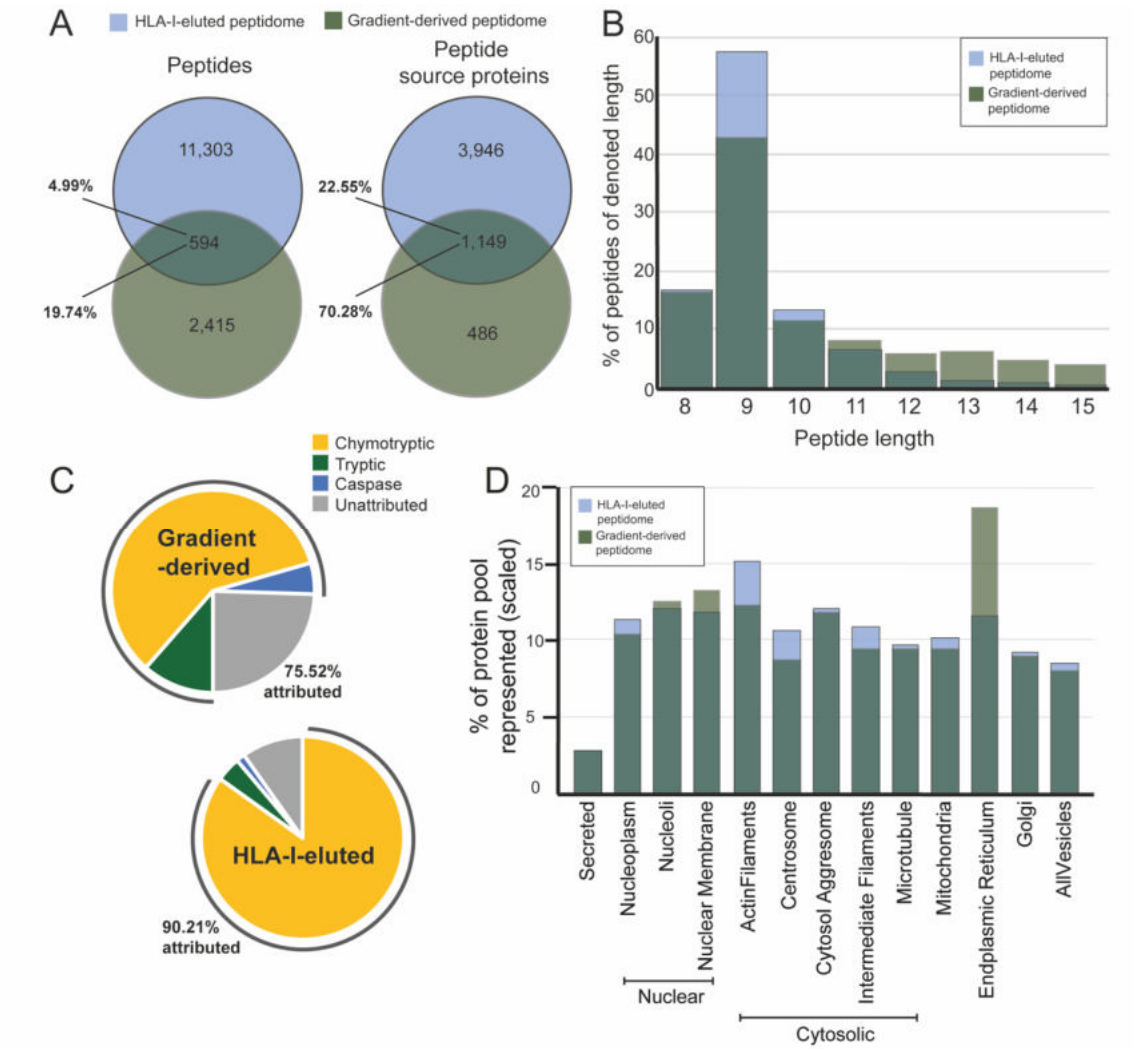


Figure 3.18. ***Peptide characteristics of gradient-derived and HLA-I-eluted peptides.*** **A:** Overlap of peptide sequences and peptide source proteins between HLA-I-eluted and gradient-derived peptides from 24H IFNγ-treated HCT116. **B:** Peptide length distribution of HLA-I-eluted and gradient-derived peptides from HCT116. **C:** C-terminal characterisation of HLA-I-eluted and gradient-derived peptides. **D:** Percentage of the organelle proteome represented as gradient-derived and HLA-I-eluted peptide source proteins, scaled for peptidome size.

Next, I went on to analyse the NetMHCpan4.1 HLA-I binding affinity predictions (BA rank) for the gradient-derived peptides to better understand the HLA-I compatibility of these peptides. For this analysis peptides 8-15aa long were selected to make them most comparable to the HLA-I-eluted peptidome. A BA rank of ≤0.5% represented peptides considered strong binders (SB), and of ≤2% considered strong and weak binders (SB+WB). There was a large difference in the percentage of 8-15mers predicted to bind HLA-I between the gradient-derived and HLA-I-eluted peptides – 48.8% and 55.6% of gradient-derived peptides had a rank of ≤0.5% and ≤2% respectively, compared against 73.0% and 86.3% of the HLA-I-eluted peptides. This illustrated that the gradient-derived peptidome contained approximately 45% of peptides that were not predicted to HLA-I. Some of this may have been have been HLA-precursor peptides which are yet to be cleaved. Because I hypothesised the ER peptidome may contain HLA-precursor peptides which have not yet been cleaved to be compatible with HLA-I, I limited the next stage of the analysis to 9-mers, which represented the highest proportion of HLA-I peptides. Among these, the percentage of peptides predicted to bind HLA-I strongly or weakly (BA rank ≤ 2%) increased to 86.80%, bringing it closer to the values of the HLA-I-eluted peptides at 93.78%. Some differences in binding rank would align with the hypothesis that a higher proportion of the true ER peptides are to be of a lower affinity for HLA-I than HLA-I-eluted peptides, due to the selection of optimal affinity peptides by HLA-I and the PLC.

	Gradient-derived peptides (BA rank)		**HLA-I eluted peptides (BA rank)**	
	≤ 0.5%	**≤ 2%**	**≤ 0.5%**	**≤ 2%**
All peptides	48.84	55.59	73.03	86.26
9mers	81.26	86.80	83.11	93.78

Table 5.2. Percentage of MS-detected gradient-derived and HLA-I-eluted peptides predicted to be strong or weak HLA-I-binders according to NetMHACpan4.1 BA rank.

I next wanted to compare the NetMHCpan4.1-predicted HLA-I allotype distribution of both HLA-I-eluted and gradient-derived peptides, to assess whether there was a bias of peptides attributable to a particular HLA. Peptides

were assigned to the HLA for which they had the lowest BA rank. I compared the distribution of only strong and weak-binding peptides (Figure 3.19A). The gradient-derived peptides displayed a lower percentage of peptides assigned to HLA.A01.01, HLA.A02.01, and HLA.C05.01, but a higher percentage to HLA.B18.01, HLA.B45.01 and HLA.C05.01. The biggest difference was in the percentage of peptides attributed to HLA-B45.01 and HLA-B18.01. HLA-B45.01 has a preference for Alanine and Valine at the C-terminus, meaning it prefers chymotryptic-like ligands at the C-terminus. Therefore, the difference in peptides attributed to HLA-B45.01 is unlikely to be due to the C-terminal characteristics of the peptide pools, as the HLA-I-eluted peptide pools had a higher percentage of chymotryptic-like ligands at the C-terminus (Figure 3.19B). Whilst the peptide motifs of the WB and SB peptides displayed a high level of similarity, the gradient-derived peptides lacked a motif for HLA-C07.01 due to the low numbers of HLA-C07.01-compatible peptides. Of the 1,095 9-mers detected from the gradient, only 149 had a BA rank >2% (13.6%). Not a single peptide motif could be determined from these peptides. The same was true for the HLA-I-eluted peptides, of the 6,868 9-mers detected, only 427 had a BA rank >2% (6.2%), and this peptide pool also could not generate a single peptide motif. The similarity between the HLA-I-eluted and gradient-derived peptide motifs, illustrated that a portion of the gradient-derived peptides were likely HLA-compatible. This would be expected for a portion of the peptides that were sourced from the ER.

Because, as referenced in the introduction, the peptides exclusive to the gradient were likely to be those that were not bound to HLA-I, I hypothesised that gradient-exclusive peptides would have a higher BA rank than the HLA-I-exclusive peptides. Therefore, I analysed the BA rank distribution of peptides that were found only in the gradient-derived peptidome, only in the HLA-I-eluted peptidome, or were found in both. Peptides of each length were analysed separately so peptide length effect on BA rank could not confound the analysis. Only peptides of length 8-12aa were analysed due to sufficient numbers of matched peptides. For every peptide length, this analysis displayed a significant difference in BA rank between the gradient-exclusive peptides and the HLA-I-eluted peptidome-exclusive peptides, and the matched peptides (Figure 3.19C). This did support my hypothesis, however, I was hesitant to make firm assertions from this

analysis, as it was performed on all gradient-derived peptides which may have included some peptides from other sources.

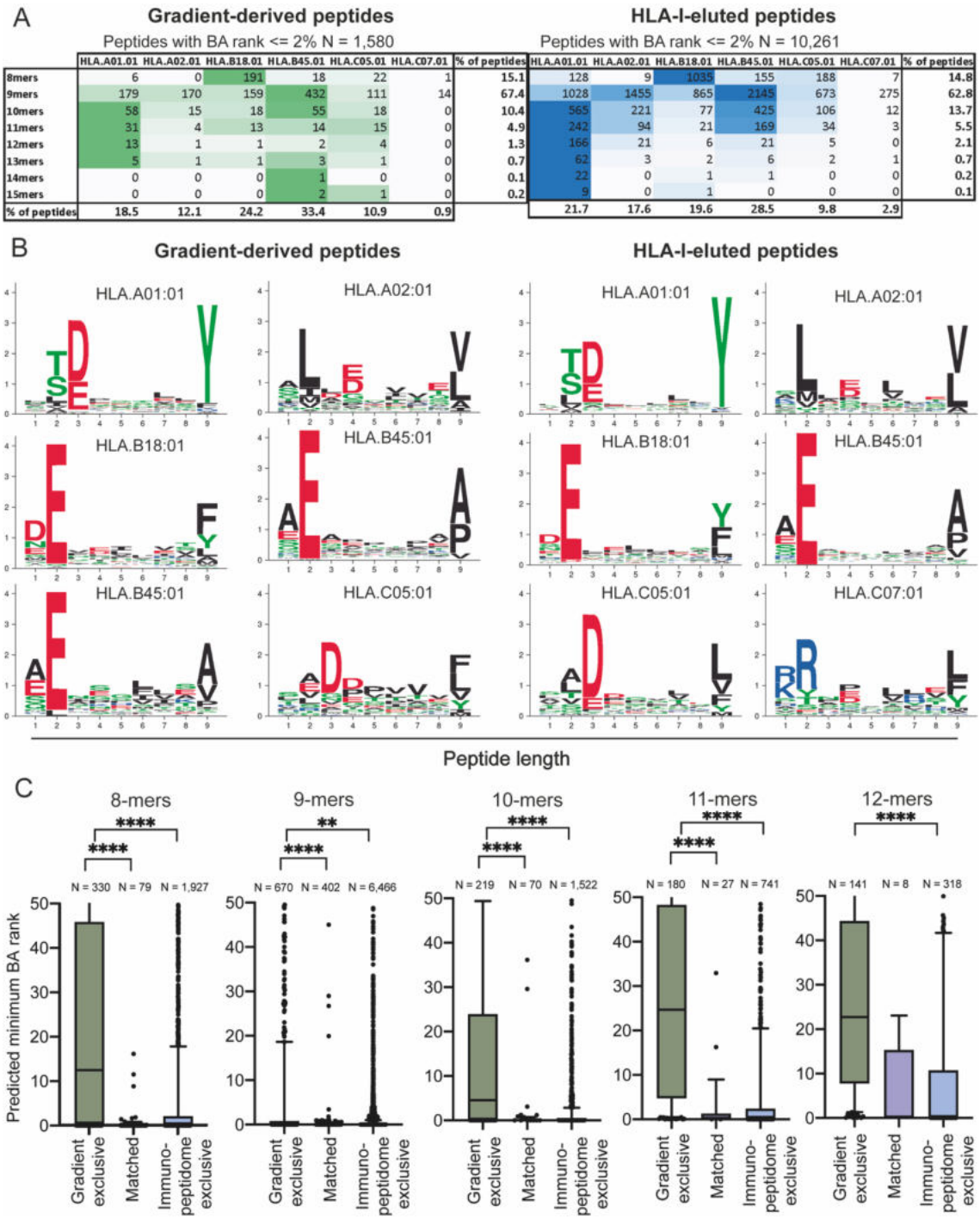


Figure 3.19. ***NetMHCpan4.1-predited HLA attribution of gradient-derived and HLA-I-eluted peptides from HCT116*****. A:** A plot of the number of weak and strong-binding peptides attributed to HLA-I allotypes (defined by BA rank <= 2%), according to the HLA which had the minimum NetMHpan4.1-predicted BA rank. **B:** MixMHCp2.1 peptide sequence motifs of the weak and strong-binding 9-mer

peptides from gradient-derived and HLA-I-eluted peptides. **C:** Minimum BA rank distributions of peptides that were exclusive to the gradient-derived peptidome, matched between the gradient-derived and the HLA-I-eluted peptidome, and the peptides that were exclusive to the HLA-I-eluted peptidome. Whiskers represented the 10-90th percentile of data. The Mann-Whitney test was used for statistical analysis.

To investigate the hypothesis that the ER-peptidome would contain HLA-I peptides and HLA-I precursor peptides which were extended at the N-terminus, I then performed alignment of peptides with partial matches between HLA-I-eluted and gradient-derived peptides. The first critical finding was that 100% of peptide alignments contained a complete match between the HLA-I-eluted peptides and the gradient-derived peptides (Figure 3.20). In addition, 59.6% of the partial peptide matches were extended at the N-terminus, 19.2% were extended at both the N-terminus and the C-terminus, and 21.2% were extended just at the C-terminus. The highest percentage of peptides being extended at the N-terminus correlates with the understanding that there are multiple ER N-terminal peptidases which shape the HLA-peptidome (Van Endert, 2011). Furthermore, 85.7% of the C-extended peptides were either chymotrpytic, tryptic, or caspase ligands, which may indicate that these were alternative proteasomal product variants. This showed that not only did the gradient-derived peptidome contain likely HLA-I binders, it also contained the precursors of some HLA-I bound peptides that were extended at the N-terminus. This added to my confidence that the gradient-derived peptidome contained peptides with features that would be expected of possible components of the ER peptidome.

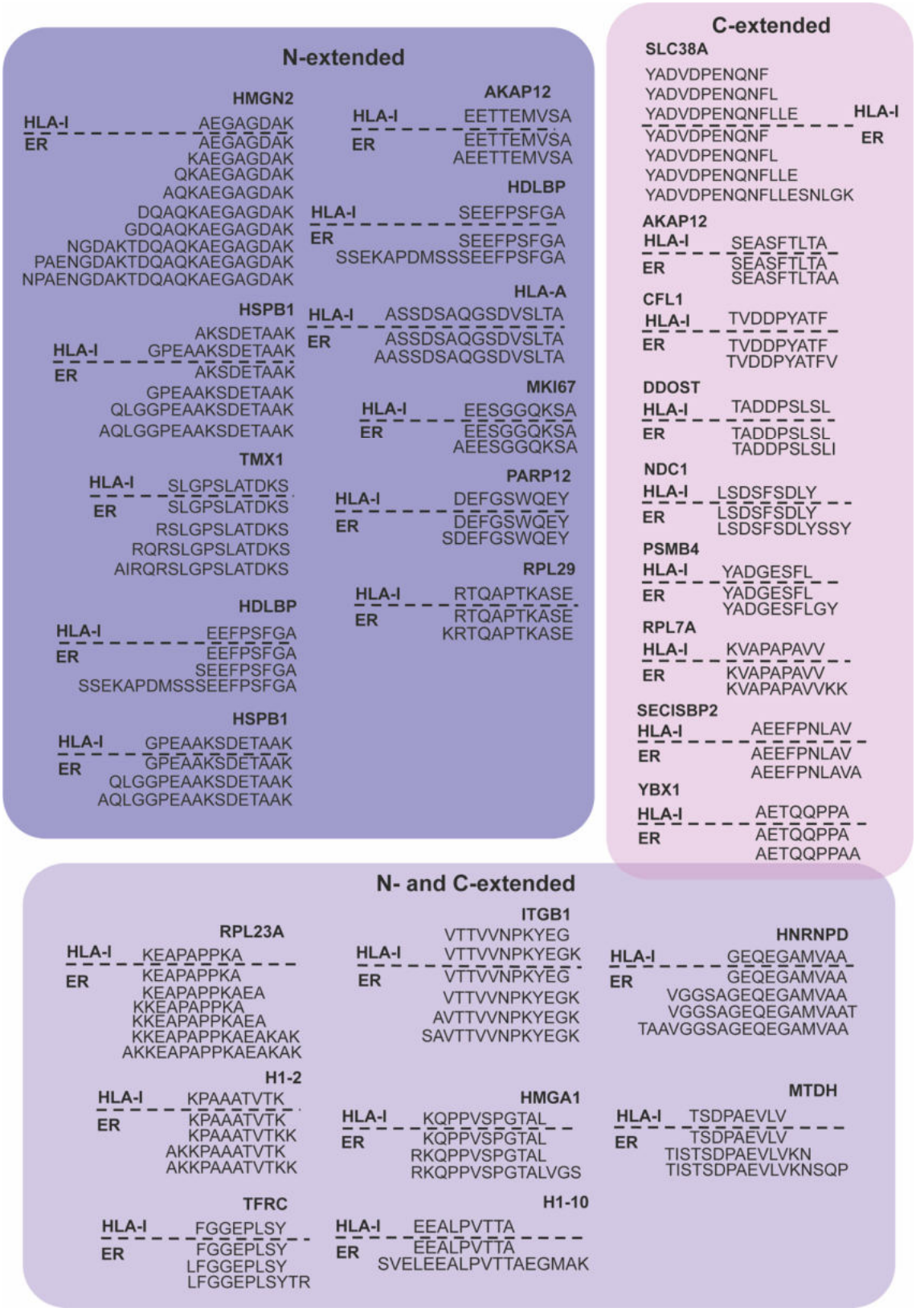


Figure 3.20. Alignment of all N- and C-extended HLA-I-eluted and gradient-derived peptides from HCT116.

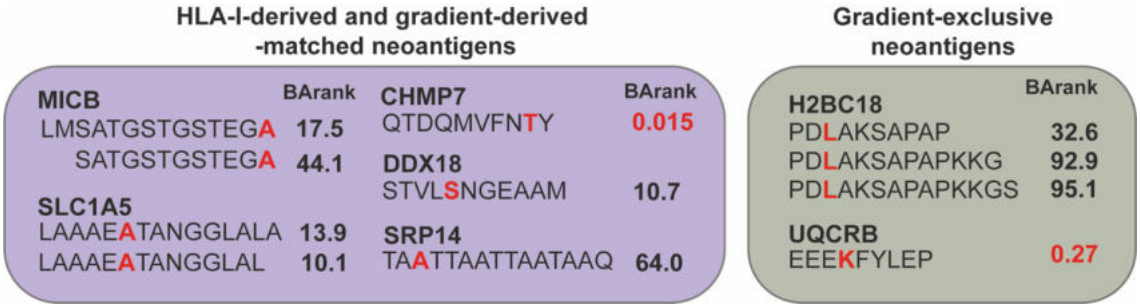


Figure 3.21. Alignment of all neoantigens detected from gradient-derived peptides, matched against HLA-I-eluted neoantigens from HCT116. Point mutation highlighted in red. Peptides predicted to bind HLA-I (BA rank ≤ 2) highlighted in red.

To address the initial question of using ER peptidomics to scrutinise the ER as a site of neoantigen loss, I analysed the overlap of neoantigens between the gradient-derived peptidome and the HLA-I-eluted peptidome (Figure 3.21). Details of the neoantigens exclusive to the HLA-I-eluted peptidome could not be shared as they at the time of submission they were unpublished and therefore embargoed. From this analysis, I was able to detect 11 neoantigens, 7 shared with the HLA-I-eluted peptidome, and 4 exclusive to the gradient-derived peptidome. It is possible these peptides had been assigned high predicted BA ranks, higher than the median predicted ranks of the 8-12mer peptides (in Figure 3.19), because they were longer in length, as longer peptides are predicted to bind less well to HLA-I. The 4 neoantigens exclusive to the gradient-derive peptidome were mapped to 2 genes: *H2BC18* and *UQRB*. Only 2 of the 11 neoantigens were predicted to bind to HLA-I, one exclusive to the gradient-derived peptidome, and one shared between the HLA-I-eluted peptidome and the gradient-derived peptidome. In a higher-confidence ER preparation, it could be interesting to interrogate if there are multiple predicted HLA-I binders that are restricted to the ER, and the reasons behind this. Through this piece of analysis, I have shown that it is possible to resolve neoantigens from organelle preparations for the comparison against the HLA-I peptidome.

Finally, I wanted to interrogate my initial hypothesis from Section 5.5.4, that the peptides derived from the ER, predicted to be enriched in L4, were longer than the peptides in L2, which appeared to contain lysosomes and transport vesicles containing HLA-I (indicated by higher COPA abundance in L2 of the WB in Figure

3.12E). Although the enrichment of ER was not as strong in L4 as seen in previous runs, I still theorised that peptides derived from ER preparations without the presence of transport vesicles would be longer than those with the presence of vesicles. For this analysis I plotted the length distribution of most enriched and most depleted 20% of peptides (L4 relative to L2) (Figure 3.22). The results supported this hypothesis, with the most enriched peptides consisting of less 8-12mers, and more 13-25mers. These findings indicated the peptides enriched in L4 are more likely to be HLA-I-precursor peptides, without signal from HLA-I-transport vesicles.

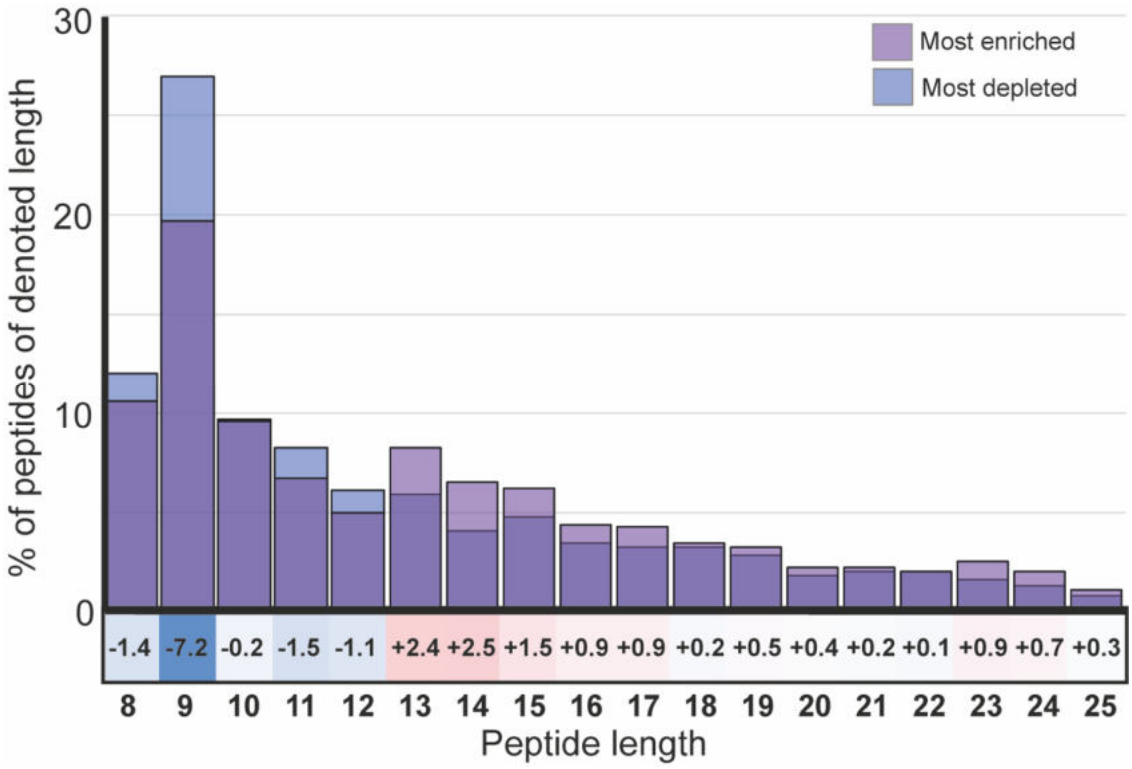


Figure 3.22. ***Peptide length distribution of L4-enriched and L4-depeted peptides from HCT116***. The most enriched and most depleted 20% of peptides, L4 relative to L2, were included in this analysis (N = 602 peptides per group).

5.7 Discussion

Over multiple iterations I have developed a method to yield an ER preparation with reproducible depletion of Golgi, lysosomal, secreted, and cell surface proteins, confirmed by WB and TMT-MS. The combination of methodologies that facilitated this was lysis by nitrogen cavitation; a 10-minute 800g nuclear exclusion centrifugation; a much less stringent mitochondrial exclusion step than many protocols, with a 10-minute 3,000g centrifugation; and a modified Histodenz gradient which was originally designed for the isolation of mitochondria.

From the small amount of starting material yielded from the gradient, TMT-MS was able to detect a relatively high number of peptides (3,009 peptides). This is also a scalable format where increasing cellular input or gradient number could be used to increase peptide number. Unfortunately, broken ultracentrifuge equipment and scheduling issues with the remaining equipment precluded me from repeating the proteomics and peptidomics analysis.

The key findings of the peptidomics were the similarities and differences of certain peptide features between the gradient-derived and HLA-I-eluted peptides. For example, the TAP1/TAP2 heterodimer can transport peptides 8-40aa long, but most efficiently binds to and transports peptides 8-16 residues long (Koopmann *et al.*, 1996). Once in the ER, peptides which are too long for the HLA-I peptide binding groove can still be cleaved at the N-terminus by ERAPs to facilitate loading (Saric *et al.*, 2002; Chang *et al.*, 2005; López de Castro, 2018). Therefore, I had predicted the ER to contain a mix of HLA-compatible peptides 8-12aa long, and also HLA-precursor peptides which would be longer than 8-15aa in length. The peptide length distribution was very similar between the two peptide pools, but with more gradient-derived peptides being >11aa long which supported this hypothesis. There were also similarities in peptide C-terminus characteristics, with chymotryptic ligands being the most common in both peptide pools. Furthermore, 594 peptides, and specifically 402 9-mers were exact matches between the HLA-I-eluted and gradient-derived peptides, representing 19.74% of all gradient-derived peptides and 37.5% of the gradient-derived 9-mers. The high similarity of the 9-mer peptide motifs between the gradient-derived and HLA-I-

eluted peptides illustrated that I was able to isolate HLA-compatible peptides. Furthermore, the higher proportion of N-extended peptides than C-extended peptides corresponded with the known N-terminus specificity of ERAP cleavage (Hearn, York and Rock, 2009). This illustrated that the peptides detected from the gradient had features similar to that expected of an ER peptidome. This provided some optimism that it could be possible to isolate an ER peptidome, even though the ER isolation run did not function well enough to define a true ER peptidome on this occasion. I was also able to show I could resolve neoantigens from the gradient-derived peptidome: 7 of which were shared with the HLA-I-eluted peptidome, and 4 which were exclusive to the gradient-derived peptidome. This also illustrated that it could be possible to use this technique to find ER-exclusive neoantigens on preparations with a higher confidence of ER enrichment. Furthermore, the 20% most enriched peptides in L4 relative to L2, with the lowest level of lysosomal and vesicle contamination, showed a longer peptide length distribution. A lower proportion of the most-enriched peptides were 8-12aa long, and a higher proportion of peptides were 13-25aa long. This may represent the ER peptidome without contamination from lysosomes and transport vesicles, but more work would need to be done to confirm this.

Despite the knowledge that the outer nuclear membrane is continuous with the first cisternae of the ER (Cooper, 2000), I had expected a larger proportion of the nuclear contamination to have been removed from the gradient-derived proteins in the 800g and 3,000g centrifugations. This was based on the number of protocols that include a nuclear removal step at 500-1,400g (Marzluff, 1978; Suski *et al.*, 2014; Pryor, 2015; Williamson *et al.*, 2015; Itzhak *et al.*, 2016; Rousselle *et al.*, 2022). The fact that some nuclear content stayed in suspension even after a 3,000g centrifugation did suggest that at least some of this nuclear content may have come from the synthesis of nuclear proteins in the ER, or from the areas of the nucleus which are continuous with the ER. This is something that may be harder to eliminate, therefore controls become immensely important to understand the impact of any contamination on the ER peptidome.

This research showed there is a pressing need to have a negative control to be able to properly define the ER peptidome. *TAP1* gene knockout was selected as the intended negative control as the TAP1/2 heterodimer is the main mode of

transport of peptides in to the ER, therefore the ER of the TAP1 KO would be depleted of the majority of peptides. This would enable assessment of the contribution of the ER to any peptide pools that were detected from the isolation process. With the help of postdoctoral researcher Dr Marta Buzzetti multiple efforts were made to generate TAP1 knockouts (KOs) through the use of CRISPR-Cas9 technology. First, a doxycycline inducible Cas9 construct was selected, and 3 TAP1 CRISPR guides were trialled. However, despite survival of transduced clones through antibiotic selection and optimisation of doxycycline dose, no significant KO of TAP1 could be detected by WB in the cells containing any of the 3 guides. This was likely explained by poor Cas9 transduction rate and Cas9 activity, as Cas9 signal on a WB was weaker than in other Cas9 cell lines. Therefore, future work should include the ultracentrifugation concentration of Cas9 virus to improve transduction rate. This should be combined with single cell cloning of the Cas9-transduced cells to generate a homogenous population of highly Cas9-expressing cells which are likely to have higher activity when transduced with the TAP1 guides. Another alternative negative control would be treating cells with the TAP-inhibiting Herpes Simplex Virus protein ICP47 to temporarily prevent the transport of peptides in to the ER (Oldham, Grigorieff and Chen, 2016). Further to negative controls, Endoglycosidase H (EndoH) sensitivity testing of HLA-I could be used to test the purity of the ER preparation. If HLA-I has not entered the trans-Golgi network, then it would be expected to be sensitive to EndoH, and therefore derived from the ER (Bailey *et al.*, 2015).

I acknowledge that whilst the first two of replicate WBs showed similar reproducibility of ER enrichment, the run from which peptidomics and proteomics samples were derived saw reduced ER enrichment. Although the technique did reproducibly deplete the Golgi, lysosomal, secreted, and cell surface contaminants, the enrichment of the ER appeared less robust in the 3rd attempt. Subcellular fractionation is known to be a very delicate process due to the numerous components that need to be controlled. The evidence indicated that the most likely step that confounded the third repeat was the cell sensitivity to lysis by nitrogen cavitation. This was my conclusion because the acceptable clearance of the less dense contaminants was sustained, suggesting the presence of more ER proteins in L2 may have come from additional ER lysis on this occasion. The only complicating factor in this conclusion was the specific

presence of SEC16A in solely the L4 lane of the WB in Section 5.5.4, which I initially interpreted as L4 possibly containing the highly enriched ER. However, more evidence pointed to the depletion of the ER in L4, so on balance I still concluded that the ER was lysed unevenly on this occasion. Even though all variables were tightly controlled and the developed protocol was followed to the letter, small differences in factors like cell density or time in suspension before nitrogen cavitation could affect membrane sensitivity to lysis. This is something that would have to be continually scrutinised in the future. It would be useful to define a panel of proteins (like those in Figure 3.13 and Figure 3.16A) and acceptable L4/L2 enrichment ratios to compare future runs against, to ensure the optimal amount of ER protein was retained and the right conclusions about the peptidomics can be drawn.

It is also highly important for future work to include replicates to allow this run to be scrutinised and to allow for analysis of reproducibility. With this information I could be more confident that I was successfully sampling a representative portion of the ER.

I believe that whilst this research requires further replicates to accurately define the qualities of each gradient layer, I have shown that I have developed a protocol that is able to yield ER with reproducible depletion of Golgi, lysosomal, secreted, and cell surface proteins. I also believe the similarities in peptide length distribution, C-terminus features, peptide overlaps, and protein representation between the gradient-derived peptides and the HLA-I-eluted peptides suggested I have been able to sample HLA-I and HLA-precursor peptides. With defined controls and protein purity thresholds, I believe this process has the ability to be a platform to answer biological questions, including developing the understanding behind low neoantigen presentation in some cells.

Chapter 6 Co-culture of MMRp CRC / GOA PDOs with immune cells

6.1 Introduction

6.1.1 Utilising organoids to study cancer immunology

As referenced in the introduction section, the ability of PDO cultures to more faithfully reproduce genetic alterations seen in the tumour enables better representation of the biology of the tumour. The PDO experimental model is particularly powerful in the precision medicine setting, advancing the understanding of individual patient tumours and using the information to determine their treatment plan. For example, studies in PDOs have shown that organoids can accurately predict the response of patients to chemotherapeutic or targeted treatments, and could be particularly advantageous for patients with treatment-refractory disease (Vlachogiannis *et al.*, 2018; Wensink *et al.*, 2021). PDO models can also be compatible with culture of immune cells, providing a platform for assessment of tumour immunogenicity, susceptibility to immune cell attack, mechanisms of immune cell evasion, and interventions to circumnavigate this (Chalabi *et al.*, 2020; Booij, Cattaneo and Hirt, 2022).

There are two approaches for PDO-immune co-cultures, the holistic approach, and the reductionist approach (Bar-Ephraim, Kretzschmar and Clevers, 2020). In the holistic approach tumour biopsies, containing all cells types found in the tumour containing immune cells and CAFs, are cultured in an air-liquid interphase culture to facilitate the expansion of tumour-specific CTLs. In the reductionist approach organoids are grown from chopped or dissociated biopsies and combined with immune cells derived from autologous PBMCs. Whilst the holistic approach allows for a complete representation of the TME, the format of the model prevents biobanking or cryopreservation which precludes long-term or diverse experimental approaches. This is particularly important when trying to generate replicates from the same limited material, or where multiple experimental techniques are required to better understand tumour and immune cell characteristics. For instance the reductionist approach was successfully employed in a 2019 study using PDO-immune co-cultures as a pre-clinical platform to select optimal target antigens for CAR-NK therapy (Schnalzger *et al.*,

2019). This was achieved by assessing the specificity of PDO killing in matched cancer and healthy organoids, to reduce the risk of the therapy causing off-target effects.

6.1.2 Alternative targeting mechanisms for MMRp tumours.

An alternative approach to targeting or boosting neoantigen load in MMRp CRCs would be to redirect T-cells to CRC cells in a TCR-independent fashion. A bispecific antibody (BsAb) is an asymmetrical molecule which binds to an epitope of a protein expressed on a tumour cell on one end, and to CD3, a T-cell co-receptor, on the other. Simultaneous binding of CD3ε on the T-cell and the target epitope activates TCR-independent signalling in T-cells and imposes formation of a pseudo-immune synapse to enable efficient tumour cell killing. Strong binding of BsAbs to target protein and CD3ε induces CD3 clustering causing strong T-cell activation; this bypasses the need for an HLA-peptide-TCR interaction, enabling polyclonal T-cell activation. Multiple BsAbs have shown clinical efficacy in haematological cancers and a few have since been approved for use in the clinic. Once example is blinatumomab, a BsAb targeting CD19 for the treatment of lymphoblastic leukaemia. Blinatumomab showed a significantly higher survival rate than chemotherapy (7.7 months compared against 4.0 months) and a higher remission rate (34% compared against 16%) (Kantarjian *et al.*, 2017). There are also multiple more different BsAbs that have reached late-phase clinical trials or have been granted accelerated approval through early stages of clinical trials due to their predicted clinical benefit (Wang *et al.*, 2021). For example, in December 2022, mosunetuzumab (a BsAb targeting CD20) received accelerated FDA approval for use in relapsed or treatment-refractory follicular lymphoma after showing an 80% objective response rate and 60% complete response rate.

CEA-TCB is an IgG-based BsAb that has been developed to treat CEA (CEACAM5)-positive CRC. CEA-TCB targets CEA on the cell surface. CEA is a carcinoembryonic antigen that is overexpressed on CRC cells, amongst other cancers (Hammarström, 1999). Although CEA is expressed on some normal tissues at a low level, CEA-TCB has two CEA-binding arms, meaning it will only bind bivalently to cells with a higher density of CEA, therefore selecting for the

tumour cells with aberrant CEA expression (Bacac *et al.*, 2016). CEA-TCB efficacy is correlated well with number of target CEA molecules on the cell surface. When CEA-TCB is expressed >10,000 molecules/cell CEA-TCB was able to induce immune synapse formation, leading to release of lytic granules causing death of CEA-expressing tumour cells. The activated T-cells also secreted IFNγ, TNF, IL-2, IL-6, and IL-10 and underwent proliferation. When used with a human carcinoma xenograft model CEA-TCB caused statistically significant tumour regression, and the immune activity also induced expression of intratumour PD-L1 (Bacac *et al.*, 2016).

6.2 Allogeneic T-cell-PDO co-culture for the purpose of generating bispecific and T-cell-resistant PDOs.

6.2.1 Results

6.2.1.1 Assay design for the long-term CEA-TCB bispecific PDO-allogeneic CD8+ T-cell co-culture

This study was used to assess the modes of resistance acquired by PDOs to develop insensitivity to killing by activated CD8+ T-cells induced by BsAb CEA-TCB treatment. CEA-TCB is a novel IG-based BsAb, one end targets CEA on cancer cells, and the other end binds CD3 on T-cells which activates TCR-independent signalling in T-cells. This treatment operates well on CEA-high expressing cells (Bacac *et al.*, 2016), but unpublished clinical data suggested that resistance can be acquired over time. I utilised a simple co-culture system with PDOs and pre-activated allogeneic-CD8 T-cells to see how resistance to T-cell cytotoxic activity can develop over time.

PDO cells were treated with CEA-TCB and anti-CD3/anti-CD28 bead-activated CD8 T-cells repeatedly over multiple weeks until regrowth of resistant populations occurred. Pre-activated cells were used to ensure rapid and efficient killing upon addition to the co-culture. For every long-term assay the same donor, pre-validated for killing efficacy, was used across all treatments. Treatment response was evaluated by confluence tracking on the Celigo Imaging Cytometer. The high-throughput microscope exposed the GFP-expressing PDO cells to a 488nm laser to make them fluoresce so they could be imaged and have confluence assessed. The GFP labelling of the cells allowed the confluence of live PDO cells to be analysed without being confounded by debris or T-cells.

I first assessed the feasibility of performing long-term PDO and CD8 T-cell co-culture assays by assessing whether PDOs could grow consistently, and stably express CEA when cultured in RPMI medium and in the absence of matrigel. This was important because RPMI media was required for optimal T-cell function, and the 2D setting was important to enable new T-cells to access the PDOs when applied for retreatment.

To be systematic, I began by investigating the impact of cell density on CEA expression. This was essential as with the experimental design, the cells would be seeded somewhat sparsely, then would grow in confluence over time, so I wanted to ensure that the culture conditions did not cause CEA loss. $1x10^5$-$1x10^6$ PDO cells were seeded in a 12 well plate and cultured for 5 days before antibody staining and flow cytometric analysis. The data showed cell density did not have a large or systematic effect on the percentage of cells expressing CEA, or the MFI of the stain, which suggested the surface molecule number did not change much. I then analysed the effect of different media on surface CEA abundance: DMEM alone, DMEM and 2% matrigel, RPMI alone, RPMI and 2% matrigel. During the culture period, the different wells were also imaged to assess the effect on cell morphology. There was not a visible difference in cell morphology between DMEM and RPMI media, but in both the wells with the absence of 2% matrigel the cells clustered together in dense patches, rather than being evenly spread over the culture surface. This was likely due to cells binding together to retain extracellular contacts. This was in fact beneficial for the assay, as it meant the cells grew clustered in the centre of the well, which made it easier to track confluence increases by imaging on the Celigo. Flow cytometric analyses showed the percentage of cells expressing CEA remained relatively stable in CRC-07, but reduced slightly in CRC-05 in conditions without matrigel. CRC-07 experienced a reduction in MFI when cultured with DMEM and matrigel, and CRC-05 had reduced MFI in conditions without matrigel. Although CRC-05 did exhibit reduced MFI without matrigel, I settled on using RPMI without matrigel as this is provided the best conditions for T-cell function and wasn't detrimental to CEA expression.

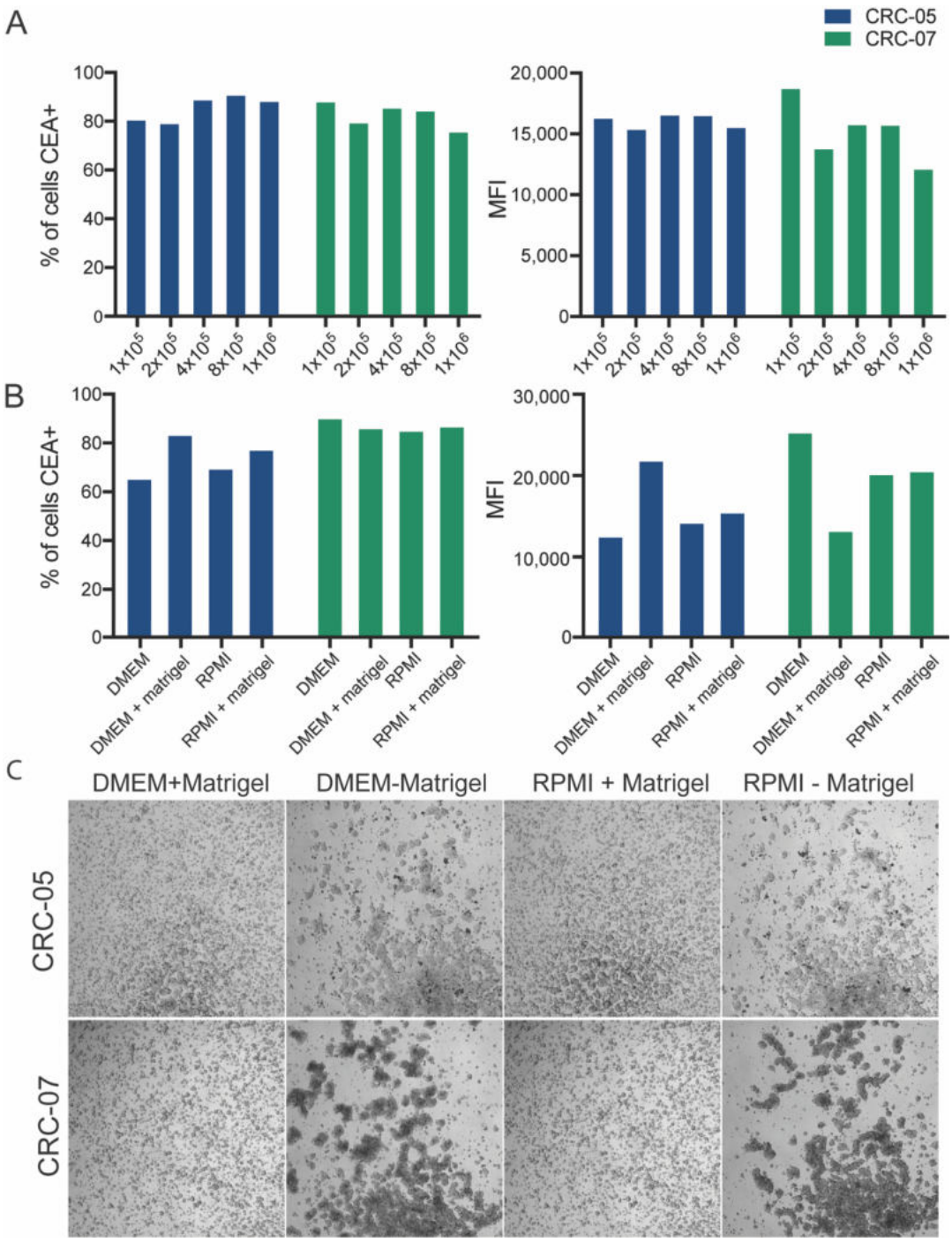


Figure 4.1: Establishing culture conditions for the CEA-TCB long-term assay. A: Percentage of the live cell population expressing CEA, and the MFI of the CEA stain after being cultured at different cell densities. **B:** Percentage of the live cell population expressing CEA and the MFI of the CEA stain after being exposed to different media conditions. **C:** Images of cell morphology differences when cultured in different media conditions.

Once I had determined seeding density and media conditions, I then I optimized the effector to target (E:T) ratio to ensure good tumour cell control, but to an

extent where PDO regrowth was possible, after the initial period of growth inhibition. Cells were seeded on D-1, then imaged on D0 before treatment for confluence assessment, and then regularly throughout the culture duration (illustrated by dots on the confluence graphs). The control-TCB-treated wells showed the growth rate of PDOs not targeted by cytotoxic activity of the activated CD8+ T-cells, as the bispecific did not target a cell surface molecule on the PDO.

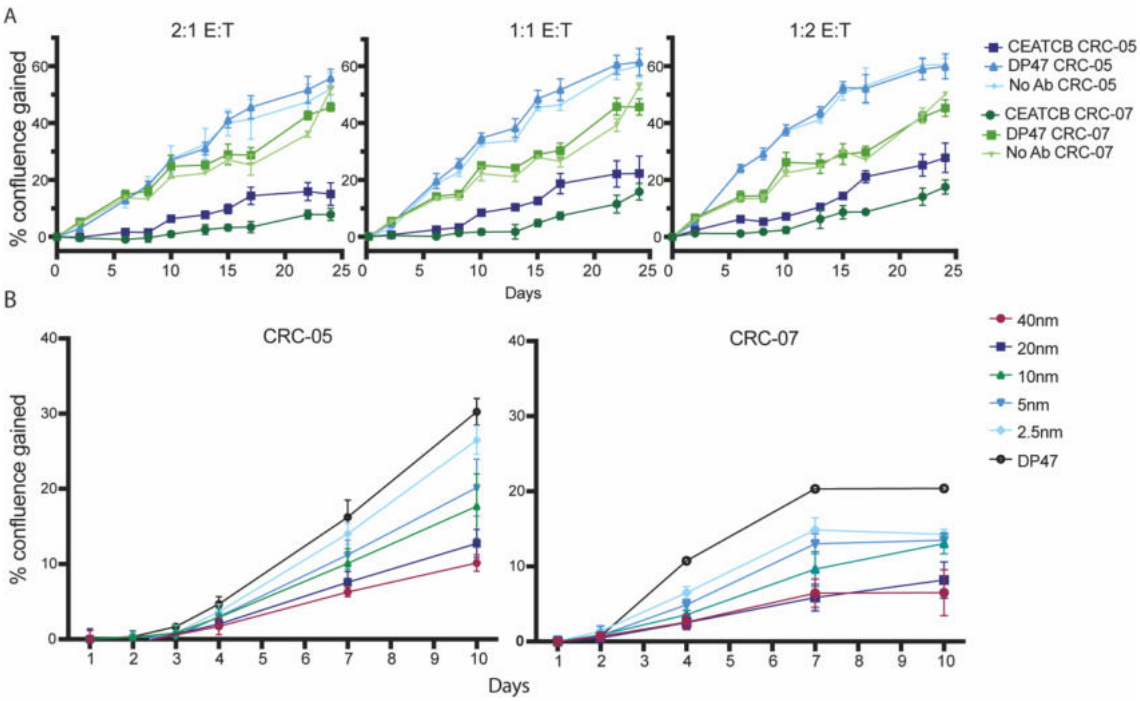


Figure 4.2: Establishing the experimental setup for CEA-TCB long-term assay. A: Growth pattern of CRC-05 and CRC-07 when treated with 20nm CEA-TCB and different E:T ratios. **B:** Growth pattern of CRC-05 and CRC-07 when treated with different concentrations of CEA-TCB made by serial dilution.

Therefore, the difference between untargeted PDOs and CEA-TCB-targeted PDOs showed the efficacy of cell killing induced by CEA-TCB treatment. An E:T of 1:2 did not sufficiently inhibit growth of CRC-05, shown by smaller differences in growth rate between the control-TCB and CEA-TCB-treated wells. This confirmed it this was not a strong enough selection method to enable resistance acquisition of the PDOs. An E:T of 2:1 did exhibit tumour control, but also a small killing in CRC-07. This was a concern, because the assay needed to be robust by the way that enough PDOs would survive to be able to regrow for retreatment. Therefore, I decided to use a 1:1 E:T ratio so cells could grow back. Once an E:T ratio of 1:1 was decided, a titration of CEA-TCB was performed to ascertain the most effective dose of the bispecific in the 1:1 E:T setting. In the seminal paper

describing CEA-TCB functionality, 20nm was defined as optimal (Bacac *et al.*, 2016). However, for other drug treatments performed on our PDOs, they have required higher concentrations to see proper drug efficacy, therefore I also tested a 40nM dose. CEA-TCB showed a dose-dependent effect for doses 2.5nm-20nm, however in both CRC-07 and CRC-05 there was not a significant difference in growth between 20nm and 40nm. Therefore, to be most economical with the use of the drug the PDOs continued to be treated with 20nm CEA-TCB. This provided me with enough information about the desired experimental setup, so I went on to use the RPMI and 1:1 E:T setup, with 20nm CEA-TCB.

6.2.1.2 Generation of CEA-TCB-resistant PDOs

PDO cells were seeded at cell number which gave a 5% confluence on D0, then treated with 20nm CEA-TCB or 20nm control-TCB with a 1:1 E:T ratio with allogeneic CD8 T-cells. Half of the CEA-TCB-treated wells also had 2µg/mL Atezolizumab added at the same time, to ascertain if inhibition of the PD-1/PD-L1 axis could restore cytotoxic activity of the CD8+ T-cells. This dose of Atezolizumab was selected due to research suggesting it was most effective in receptor neutralization testing. PDOs were retreated with fresh T-cells and fresh antibody-supplemented T-cell media when they had grown 1.25-fold from the previous treatment. Furthermore, E:T ratio was also increased with each re-treatment: 1.5:1 E:T for the first retreatment, then 3:1 for the second retreatment.

All 3 PDOs saw tumour control through CEA-TCB and T-cell exposure compared against control-TCB and T-cell exposure. In the cases of CRC-05 and CRC-07, additional treatment with Atezolizumab did not improve killing or growth inhibition compared with CEA-TCB and T-cells alone. CRC-07 and CRC-05 were retreated at D10 and D16 due to faster growth; and CRC-01 at D14. The growth of CRC-01 did not recover after the retreatment, suggesting that the treatment was too harsh for these more treatment-sensitive cells. This illustrated that every PDO used in this assay would need to be optimized separately to make sure enough cells survive for re-growth and re-treatment. Optimisation of CEA-TCB dose performed for the next round revealed 10nm to be the optimal dose for tumour control without killing in CRC-01.

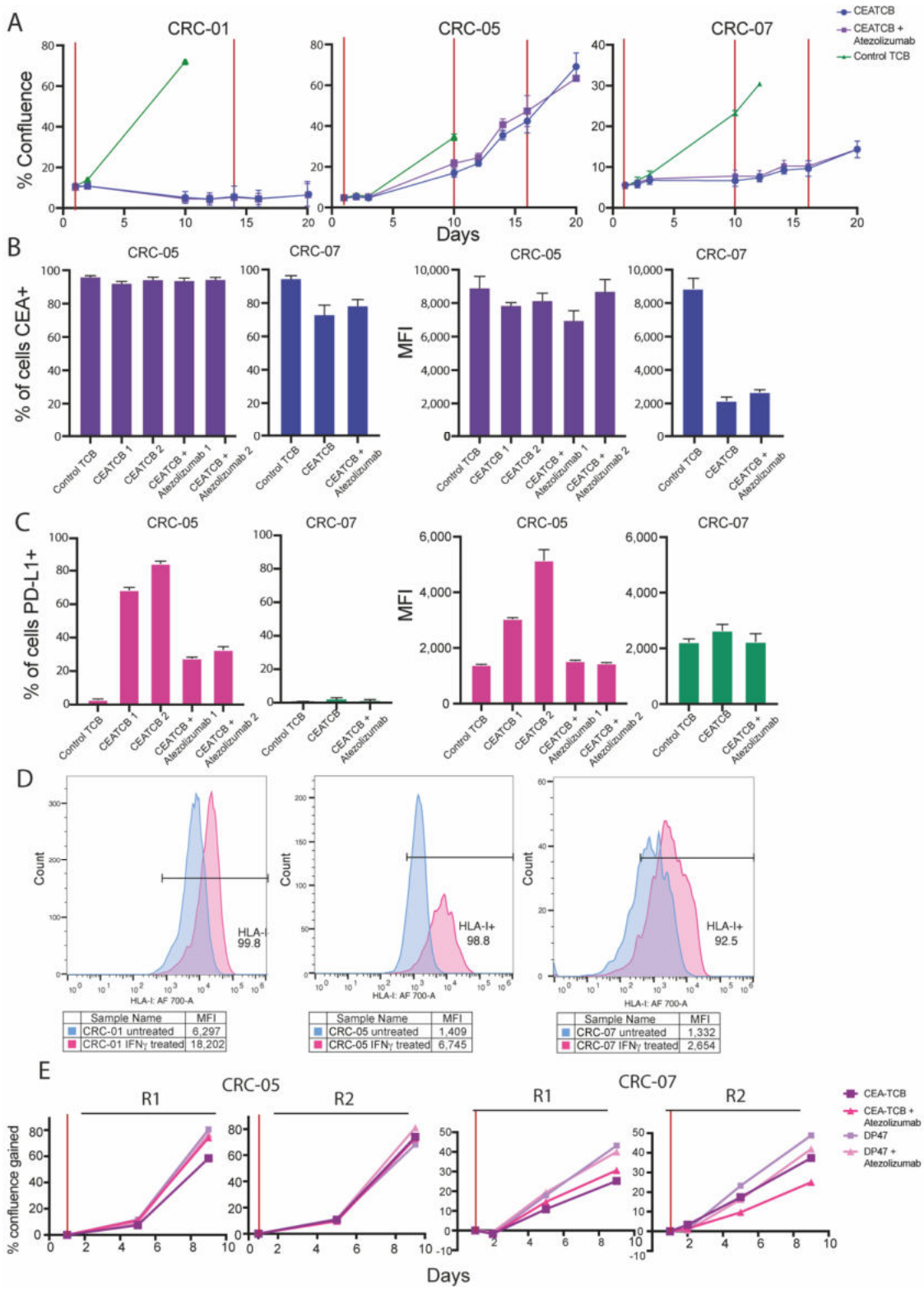


Figure 4.3: First findings from CEA-TCB resistance culture. Characteristics of PDOs after being treated with 20nm CEAC-TCB alone, 20nm CEA-TCB and 2µg/mL Atezolizumab, or with 20nm DP47. **A:** Growth rates of PDOs tracked over time by imaging on the Celigo (N of 3 for each treatment combination). Initial treatment with an E:T ratio of 1:1, first retreatment with an E:T ratio of 1.5:1, then the second retreatment with an E:T ratio of 3:1. **B:** Percentage of the live cell

population expressing CEA and the MFI of the CEA. CRC-07 had 2 wells pooled for staining, and CRC-05 had 2 wells from each condition stained separately. **C:** Percentage of the live, GFP+ cell population expressing PD-L1 and the MFI of the PD-L1. CRC-07 had 2 wells pooled for staining, and CRC-05 had 2 wells from each condition stained separately. **D:** Flow cytometry histogram overlays of untreated and 10ng/mL IFNγ-treated PDOs stained with anti-HLA-A,B,C. MFI figures below are averages of two technical replicates. **E:** Growth rates of PDOs expanded for 4 weeks after being harvested from the killing assay at day 20, treated with 20nm CEA-TCB or with 20nm DP47.

The main modes of acquired treatment resistance are known to be: 1) alterations of the target, 2) alterations within, or connected to, the signalling pathway of the target molecule, 3) phenotypic alterations in the tumour (e.g. epithelial-to-mesenchymal transition), and 4) loss of the target (Aldea *et al.*, 2021). CRC-07 and CRC-05 were harvested at D20 of the assay and stained for CEA and PD-L1 staining to examine if target downregulation or PD-L1 upregulation could explain the rapid acquisition of treatment resistance. To ensure any residual T-cells were not evaluated for PD-L1 staining, the PDOs were gated on by their GFP expression. Furthermore, it was possible to stain for CEA and PD-L1 after CEA-TCB and atezolizumab co-culture because the antibodies were confirmed to target different epitopes on the proteins. CRC-07 displayed a small drop in the percentage of cells expressing CEA when compared against the untargeted control TCB, but also showed a much more dramatic drop in MFI (approximately 4.5-fold downregulation) (Figure 4.3B). This suggested that whilst only a small percentage less cells expressed CEA, the amount of CEA expressed on the surface of each cell declined dramatically. This may have been significant, as molecular abundance has been shown to be correlated to drug sensitivity (Bacac *et al.*, 2016). CRC-05 did not experience a dramatic drop in percentage of cells expressing CEA, or CEA MFI. This suggested that the mechanism of resistance was distinct from antigen loss. PD-L1 staining showed CRC-05 experienced significant increases in PD-L1 expression in the CEA-TCB conditions against the control TCB. Curiously, the PD-L1 expression was not increased as much in the culture including atezolizumab. This was not due to epitope competition, as testing of atezolizumab incubation before PD-L1 staining did not show a decrease in PD-L1 binding. For CRC-07, there appeared to be nearly no induction of PD-

L1 expression, aside from a small increase in PD-L1 MFI in the CEA-TCB condition. I hypothesised that CRC-07 may be IFNγ-resistant due to the lack of significant PD-L1 upregulation, therefore I treated the 3 PDOs with 10ng/mL IFNγ for 48hrs and assessed for changes in HLA-I surface expression as this is considered a mark of the IFNγ response (Zhou, 2009). CRC-07 did respond to IFNγ treatment, but it upregulated HLA-I the least in across the 3 PDOs, 1.99-fold compared to 2.89-fold in CRC-01, and 4.79-fold in CRC-05 (Figure 4.3 D).

2 wells from each treatment condition of CRC-05 and CRC-07 were re-grown without CEA-TCB treatment for 4 weeks and then retreated with 20nm CEA-TCB. The data suggested that CRC-07 retained some sensitivity to CEA-TCB, but had a much faster growth rate. Conversely, CRC-05 had only very small differences in growth between control-TCB-treated and CEA-TCB-treated conditions, but shared the accelerated growth rate.

6.3 Discussion of allogeneic T-cell-PDO CEA-TCB bispecific co-culture

Utilising the 2D co-culture model I executed a long-term co-culture of PDOs with allogeneic-CD8 T-cells and CEA-TCB to develop treatment-resistant PDOs. I first investigated whether target downregulation was a mechanism employed in our PDOs to evade immune cell recognition. Flow cytometric analysis showed no significant changes in percentage of PDOs expressing CEA in CRC-05, although there were some variations in surface molecule number suggested by changes in MFI. However, CRC-07 displayed a small drop in the percentage of cells expressing CEA, paired with a large decrease in MFI (approximately 4.5-fold downregulation). This suggested CRC-07 exhibited target downregulation. This could have been a major source of resistance in this PDO. Understanding how this downregulation occurred through future investigation of RNAseq data may improve the understanding for the reasons behind the plasticity of CEA observed in our PDOs, and possibly support the involvement of the WNT/β-catenin pathway (Gonzalez-Exposito *et al.*, 2019).

Due to ICI against PD-L1 showing great success in the clinic I also treated some wells with a combination of 20nm CEA-TCB and 2.5μg/mL atezolizumab to see if blocking PD-L1 could restore effective tumour cell killing. In this assay I saw no significant benefit of treating the co-culture with atezolizumab, suggesting PD1 activation on T-cells is not the major source of CEA-TCB resistance. Flow cytometric analysis was used to ascertain if PD-L1 upregulation is occurring in our PDOs and only saw notable PD-L1 upregulation in CRC-05 treated with CEA-TCB alone, and a smaller increase in CEA-TCB in the combination treatment. CRC-07 did not strongly upregulate PD-L1, which was attributed to a reduced IFNγ sensitivity, indicated by the smaller upregulation of HLA-I than the other PDOs. It was also possible that the concentration of IFNγ used in the experiment exceeded the concentration of IFNγ reached in the co-culture, which may explain the apparent disparity in IFNγ-response. To address this question, it may have been beneficial to stain CRC-05 control-TCB and CEA-TCB-treated cells for HLA-I so the level of IFNγ response in the assay could be characterised. A study by the Ribas group which utilised *JAK1* and *JAK2* knockout, showed that loss of JAK1 or JAK2 caused insensitivity to IFNγ, and that this was a mode of resistance to immune checkpoint inhibitor treatment (Torrejon *et al.*, 2020). A recent CRISPR screen highlighted loss of *JAK1*, and other protein in the JAK/STAT pathway, to be an important factor in bispecific antibody resistance (Liu *et al.*, 2021). Further studies have shown that downregulation of *JAK2* expression by decreased H3K27 acetylation could also cause resistance to BsAb targeting HER2 (Arenas *et al.*, 2021). This is unlikely to have had an impact in CRC-07 due to its intrinsic low sensitivity to IFNγ, but the downregulation of JAK/STAT components could explain rapid onset of resistance in CRC-05. This should be one of the focuses of future work, alongside analysis on IFNγ-independent CEA-TCB resistance apparent on CRC-07.

Finally, 4 weeks post-culture I re-tested the cells yielded from the re-growth of the killing assay for sustained CEA-TCB resistance. CRC-07 showed some sensitivity to CEA-TCB from the time out of selection, through a bigger discrepancy in growth between CEA-TCB and DP47 wells. In contrast, CRC-05 showed little sensitivity to CEA-TCB treatment. It should be noted that whilst the growth rates of both PDOs appeared to have had accelerated dramatically, this cannot be confidently concluded without comparison against untreated PDOs.

6.3.1 Future plans curtailed by COVID-19

The first iteration of the T-cell-PDO culture illustrated the rapid acquisition of resistance to BsAb and T-cell treatment. Furthermore, the flow cytometric data illustrated possibly different modes of resistance developed by the two PDOs which survived treatment: CRC-05 and CRC-07. This showed a lot more information was required to help develop a deeper understanding of the mechanisms involved. The plan was to perform both bulk RNA sequencing and single cell RNA sequencing (scRNAseq). For the bulk sequencing, it was decided that the T-cells would be removed with anti-CD8 magnetic beads before sample preparation. The further advantage of the scRNAseq is the mix of PDOs and T-cells could be sent for RNA sequencing and deconvoluted, and this therefore would give information about both PDO and T-cell transcriptional changes over time. Figure 4.4 depicts the experimental plan for the scRNAseq experiment planned to be performed in collaboration with Roche. Duplicate plates would be prepared, and then each plate would be harvested at every timepoint.

The first iteration of this experiment was in process when the institute was closed during COVID-19 lockdown, so the work had to cease and the cultures discarded. Once the institute re-opened there was a 1-person maximum capacity in the tissue culture room and preparation of clinical trial samples took precedence. Because this experiment took weeks of preparation and a lot of time at the tissue culture hood it was not possible to continue with this technique upon return to work. Furthermore, with the concern of additional lockdowns, it was not practical to try and initiate a long-term resistance assay.

This work showed that use of allogeneic T-cell-PDO co-cultures can be a suitable an effective model for the investigation of acquired immunotherapeutic resistance. Once findings in this model system were made, additional immune cell subsets could be analysed to compare if the modes of resistance appeared to be reproduced. Furthermore, historical clinical sample data could be analysed to see if patient data replicated the transcriptomic findings from the *in vitro* setting. I do believe that this approach has potential to elucidate some of the mechanisms of targeted immunotherapy resistance in CRC. Furthermore, some of the

findings, aside from target downregulation which is likely to be pathway-specific, may even be applicable to resistance to other BsAbs, and in other tumour types.

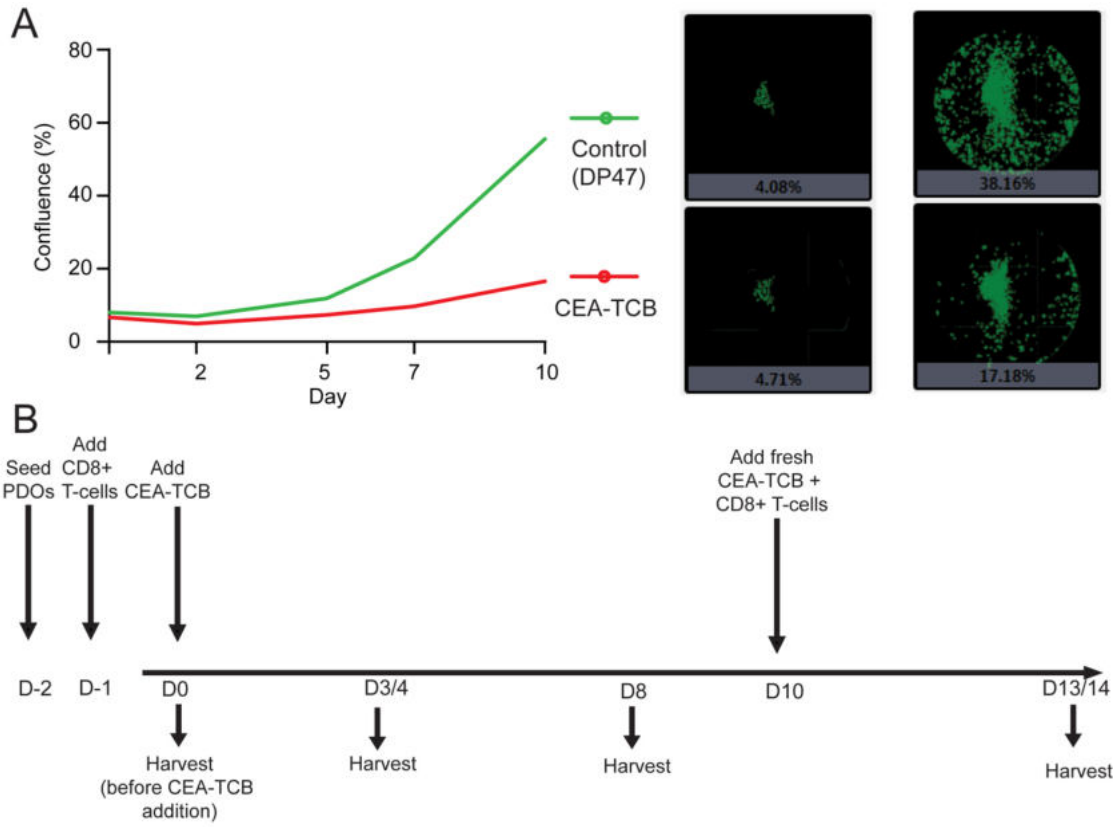


Figure 4.4. Details of planned scRNAseq experiment. A: A plot of confluence changes over time with CEA-TCB treatment or DP47 treatment, and how that growth appears on D0 and D7 on the Celigo imaging cytometer. **B:** Schematic of experimental plan for scRNAseq with 4 harvesting timepoints to enable dynamic analysis of transcriptomic changes of PDOs and T-cells.

6.4 Results for autologous PBMC-PDO co-culture

6.4.1 The foundation of the autologous PBMC-PDO co-culture for the expansion of tumour-specific CTLs

For this study I used MMRp GOA PDOs because matched PDO and PBMC samples were only available for these patients. With these MMRp GOA PDOs I developed a model system to interrogate recognition of PDO-presented antigens by autologous T-cells. This could be followed by the assessment of T-cell stimulating potential of MS-detected and predicted neoantigens to help reveal the most clinically relevant antigens. This could then be trialled on other GI cancers in the future.

The co-culture in this study was inspired by a study which utilised the reductionist approach to PBMC-PDO co-culture, whereby MMRd CRC and NSCLC PDOs were combined with autologous PBMCs to expand tumour-reactive T-cells (Dijkstra *et al.*, 2018; Cattaneo *et al.*, 2020). The resultant CTLs provide another way of examining tumour immunogenicity as it provides samples for TCR sequencing, immune activation readouts, tumour killing assays, and assessment of novel drug combinations. The group then further utilised this system as a complementary model for translational studies of clinical trials investigating immune checkpoint inhibitor (ICI) combinations (Chalabi *et al.*, 2020).

The same PBMC-PDO co-culture approach was also taken to examine whether T-cell response in the co-culture could be correlated with patient response to therapy (Chalabi *et al.*, 2020). This study mainly focused on MMRd CRC due to their high mutational load; their cohort harboured 795-2877 non-silent mutations per tumour, and showed high response rate to anti-PD-1 therapies. However, this study did also include some MMRp CRCs. A single patient (CRC-13) provided both MMRd and MMRp CRC PDOs, and they were able to show a T-cell response to the MMRd organoids, but no response to MMRp organoids. The study also showed that it is possible to generate tumour-reactive CTLs from PBMCs in MMRp CRC patients, detected in the only MMRp responder to immunotherapeutic treatment (patient N26). This patient responded much more strongly in the IFNγ axis than in the degranulation/cytotoxic (CD107a) axis. Other

patients also exhibited their activation through disparate markers, therefore I noted both intracellular and extracellular staining is incredibly important for most sensitive detection of anti-tumour reactivity.

This tool showed good specificity in that there was an absence of T-cell response for the non-responders, however it showed poor sensitivity in that there was also an absence of response in 2 of the 5 HLA-I-competent treatment responder PDOs. This is thought to be due patients receiving anti-PD-1 and anti-CTLA-4 combination treatment and only anti-PD-1 being included in the co-culture setup (Chalabi *et al.*, 2020). This illustrated that whilst this co-culture platform could be a useful tool for correlative analysis, experimental design is key to achieving the best sensitivity.

6.4.2 Establishing organoids from GOA biopsies

This study focused on material from the ICONIC (Peri-operative **I**mmuno-**C**hemotherapy in **O**perable oesophageal gastr**I**c **C**ancer) clinical trial, which investigated the efficacy of PD-L1 inhibition with avelumab in combination with FLOT (5-FU, oxaliplatin, and docetaxel) as a peri-operative therapy in GOA patients (Mansukhani *et al.*, 2018; Athauda *et al.*, 2021). The inclusion criteria was for non-metastatic, treatment naïve GOA tumours.

Needle core biopsies were taken before any treatment was undertaken (referred to as a baseline biopsy). According to the ICONIC laboratory manual, biopsies were rapidly transported to the ICR and digested in to single cells. The dissociated cells were then plated in to 3D long-term culture with matrigel and media supplemented with numerous growth factors, as devised by Professor Clevers' group (Sato *et al.*, 2011) (listed in Materials and Methods section). The first signs of organoid growth would be visible within 1-2 weeks, and cell growth up to the required number could vary dramatically between 1-4 months. Some cells would be allocated for cryopreservation to develop an operating biobank, some for exome and RNA sequencing, and some used for the PBMC-PDO co-culture. Blood samples were taken at baseline and before every cycle of treatment. PBMCs were isolated from this fresh blood draw within 24H, then cryopreserved long-term in liquid nitrogen for future use. All members of the

Gerlinger group trained in tissue-culture contributed to the biobanking efforts, and I personally contributed by establishing GOA PDO lines and processing blood for PBMCs (examples in Figure 4.5A and B).

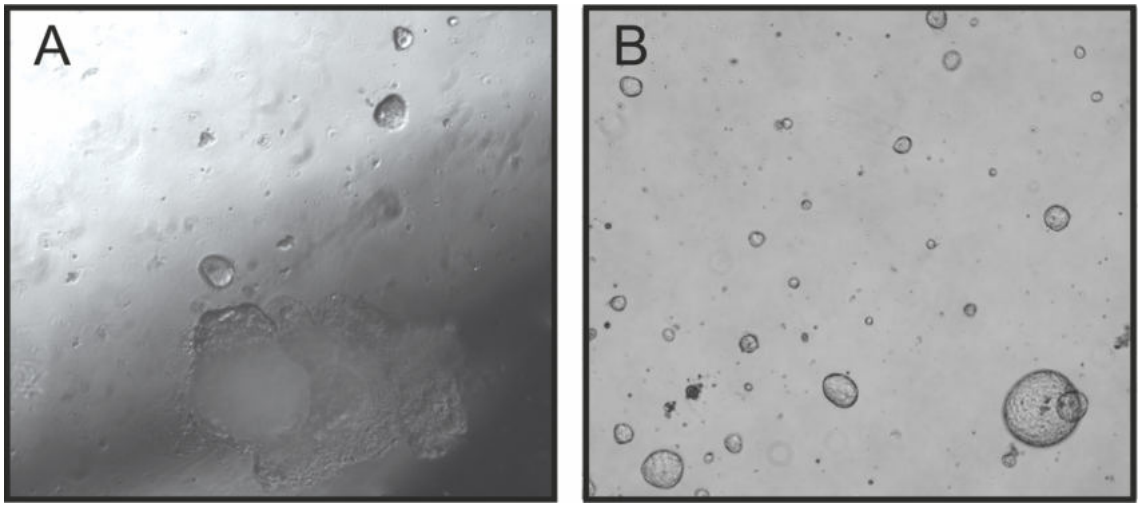


Figure 4.5. Representative images of establishing GOA organoids from pre-treatment patient biopsies. Images of the appearance of newly established organoids, taken on the Zeiss Observer Microscope. **A:** Very early organoids from patient GOA02 which had a biopsy of their GA cancer. Imaged at 5x magnification. **B:** Early, but more developed organoids from patient GOA03 which had a biopsy of their gastro-oesophageal junction cancer. Imaged at 5x magnification.

6.4.3 Establishing a robust positive control for PBMC-PDO co-culture T-cell activation assay

Because the original publication stained for extracellular and intracellular markers under separate protocols with different incubation periods, I first assessed extracellular and intracellular markers similarly.

Testing for tumour reactivity in GOA tumours provided multiple experimental challenges due to the low expected TMBs. As referenced in the introduction, 66-94.4% of GOA cancers are of the MMRp subtype which are likely to exhibit a low TMB and sparse neoantigen landscape. This made it immensely important to have a robust positive control for T-cell activation so any lack of signal could confidently be attributed to the lack of reactivity between the PBMCs and PDOs, and not from a technical experimental failure.

6.4.3.1 Establishing extracellular readout

To assess T-cell activation with extracellular markers, the costimulatory molecule CD137 (4-1BB), and the extracellular/intracellular cycling marker of degranulation CD107a (LAMP-1) were chosen. Previous experiments of my own had shown that treatment of PBMCs with PMA/Ionomycin can result in dramatically different levels of upregulation of activation markers in T-cells between patients. Therefore, to ensure I could be confident that the PMA/Ionomycin stimulation had worked for the positive control for each trial participant, I tested PBMCs from 5 different lymphocyte reduction cone system (LRS) donors to obtain a range of activation marker upregulation that I could refer to, to demonstrate successful T-cell stimulation.

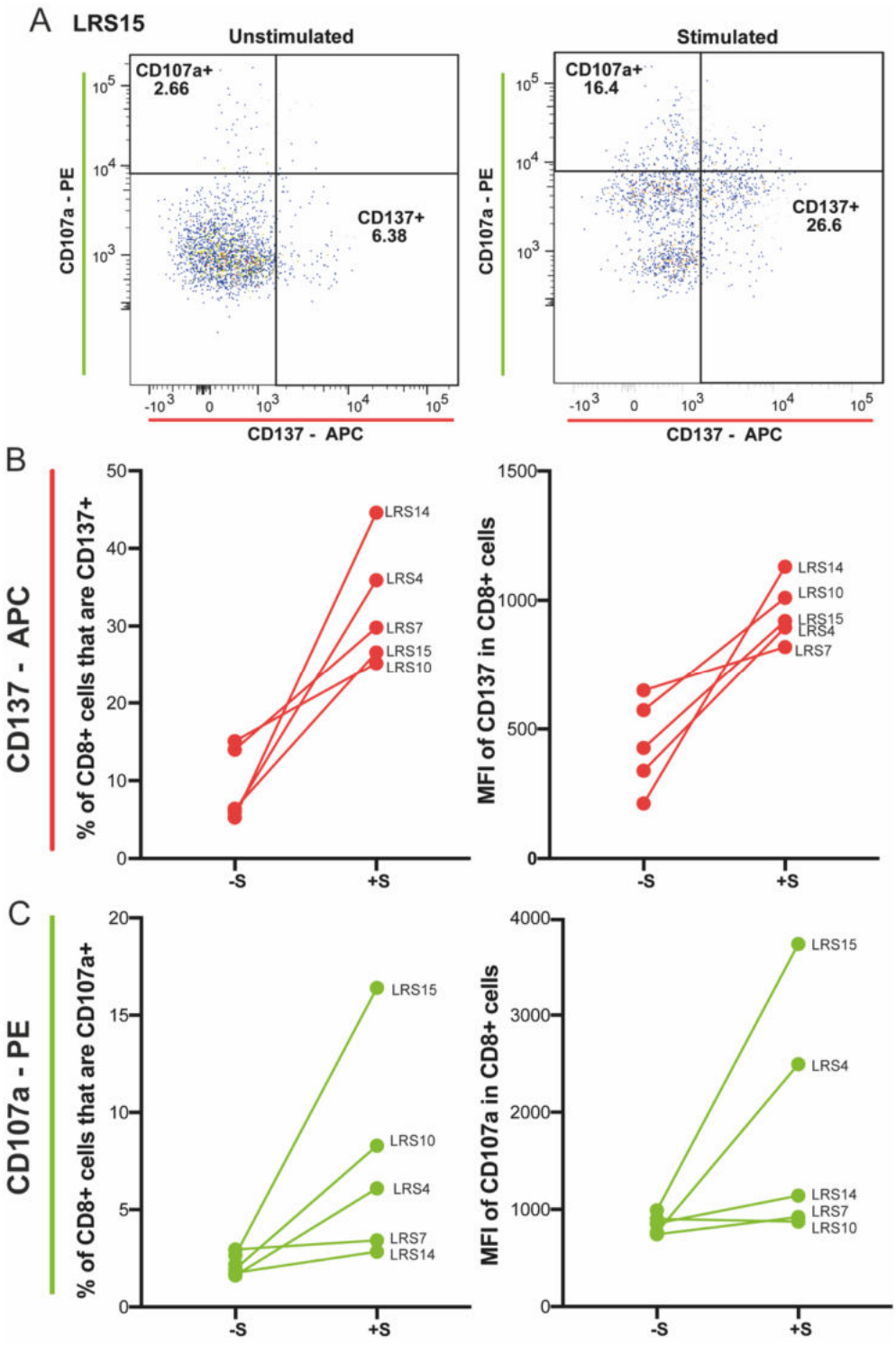


Figure 4.6. Testing suitability of extracellular readout for T-cell activation by PMA/Ionomycin over 24 hours. A: Representative scatter plot of CD107a and CD137 in unstimulated and PMA/Ionomycin-stimulated cells (from LRS15). **B:** Percentage of CD8+ cells positive for CD137, and the MFI of CD137 in CD8+ cells within PMA/Ionomycin stimulated PBMCs. **C:** Percentage of CD8+ cells positive for CD107a, and the MFI of CD107a in CD8+ cells within PMA/Ionomycin stimulated PBMCs.

To begin, I assessed the PBMCs for extracellular marker CD137, and the extracellular/intracellular cycling marker CD107a after 24H stimulation with PMA/Ionomycin. In this experiment, only CD3+ and CD8+ populations within the PBMCs were stained for.

The data from PBMCs from 5 LRS cones showed a large variation in the percentage of cells expressing CD137 (Figure 4.6A), but still with an increase in the number of cells expressing CD137 by 10.10-23.08%. This showed CD137 to be a marker which can clearly signal activation of all 5 LRS cones. The MFI increases of CD137 in CD8+ cells was not completely in line with the percentage of cell expressing CD137, with LRS10 having the second highest MFI after PMA/Ionomycin, despite having the lowest percentage of cells positive for CD137. This illustrated that even in cells expressing CD137, the number of molecules expressed can vary between different donors. CD107a expression was induced to a lesser extent - the magnitude of the increases in CD137 expression was on average 3x higher than that of CD107a. The highest increase was seen in LRS4, where the number of cells expressing CD107a increased by 18.48%. Two donors, LRS14 and LRS7, only experienced small increases of 0.47% and 1.08%, respectively. This showed that it is essential to use multiple markers to assess T-cell activation, as cells could respond in one axis and not another.

6.4.3.2 Establishing intracellular readout

I next assessed the PBMCs for intracellular marker IFNγ to assess cytostatic/cytotoxic cytokine expression, after only 6H stimulation with PMA/Ionomycin stimulation. GolgiPLUG and GolgiSTOP were added to the culture after 1H, and incubated for the final 5H of the assay. This was done to enable secreted factors to be retained in the Golgi apparatus and ER, making it possible to stain for them with fluorescent antibodies. I also re-tested the extracellular/intracellular cycling marker CD107a to see how it differed after a shorter incubation with PMA/Ionomycin. In this experiment staining of the CD4+ population was added.

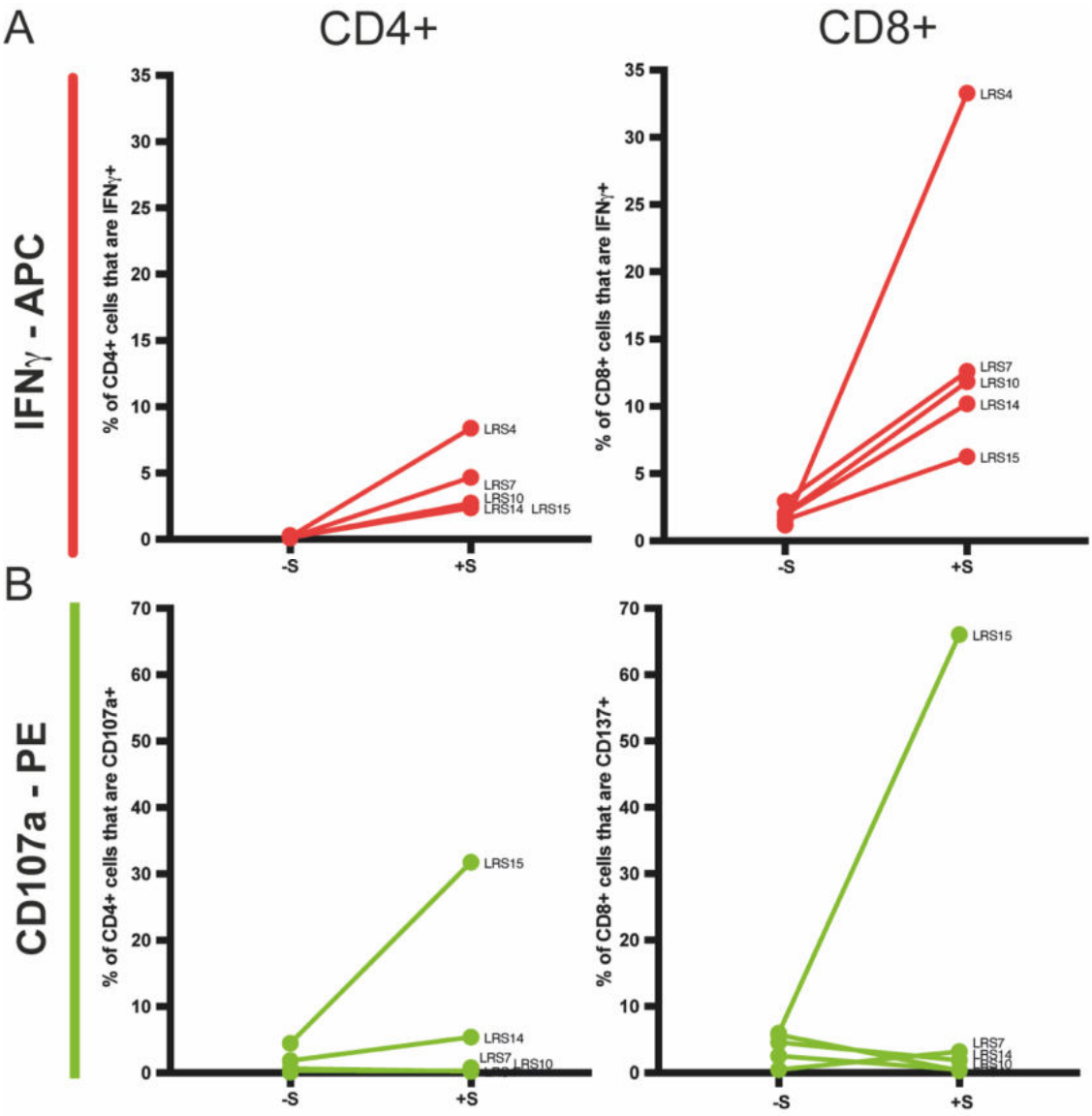


Figure 4.7. Testing suitability of intracellular readout for T-cell activation by PMA/Ionomycin over 6 hours. **A:** Percentage of CD4+ and CD8+ cells positive for IFNγ. **B:** Percentage of CD4+ and CD8+ cells positive for CD107a.

Over a 6H incubation with PMA/ionomycin treatment, the expression of IFNγ increased in both CD4+ and CD8+ T-cells, but much more strongly in CD8+ T-cells. The T-cells from LRS4 increased expression of IFNγ more dramatically than the other donors, but in contrast to this, it experienced the smallest increase in CD107a. For all donors aside from LRS15, 6H was not sufficient to generate a strong signal by CD107a incorporation during the culture. LRS15 was the donor which most increased expression of CD107a, which agrees with the previous CD107a stain from the 24H incubation. However, paradoxically, LRS15 experienced the smallest increase in IFNγ expression. This again supported the

thinking that it is important to test multiple components of T-cell activation to increase sensitivity.

6.4.4 Combined intracellular and extracellular protocol

Because I wanted to be able to preserve as many cells as possible for future analysis it was important to find a mode of readout which was efficient in the number of cells required, so as many as possible could be cryopreserved for additional analysis. Therefore, I designed a protocol for combined extracellular and intracellular readouts, which required a different incubation period to enable detection of both markers. I also added a stain for Granzyme B, to illustrate T-cell cytotoxic potential. Cells were incubated in the presence or absence of PMA/Ionomycin stimulation for a total of 21H, with GolgiPLUG and GolgiSTOP being added to the culture at the 16H mark, and incubated for the final 5H of the assay.

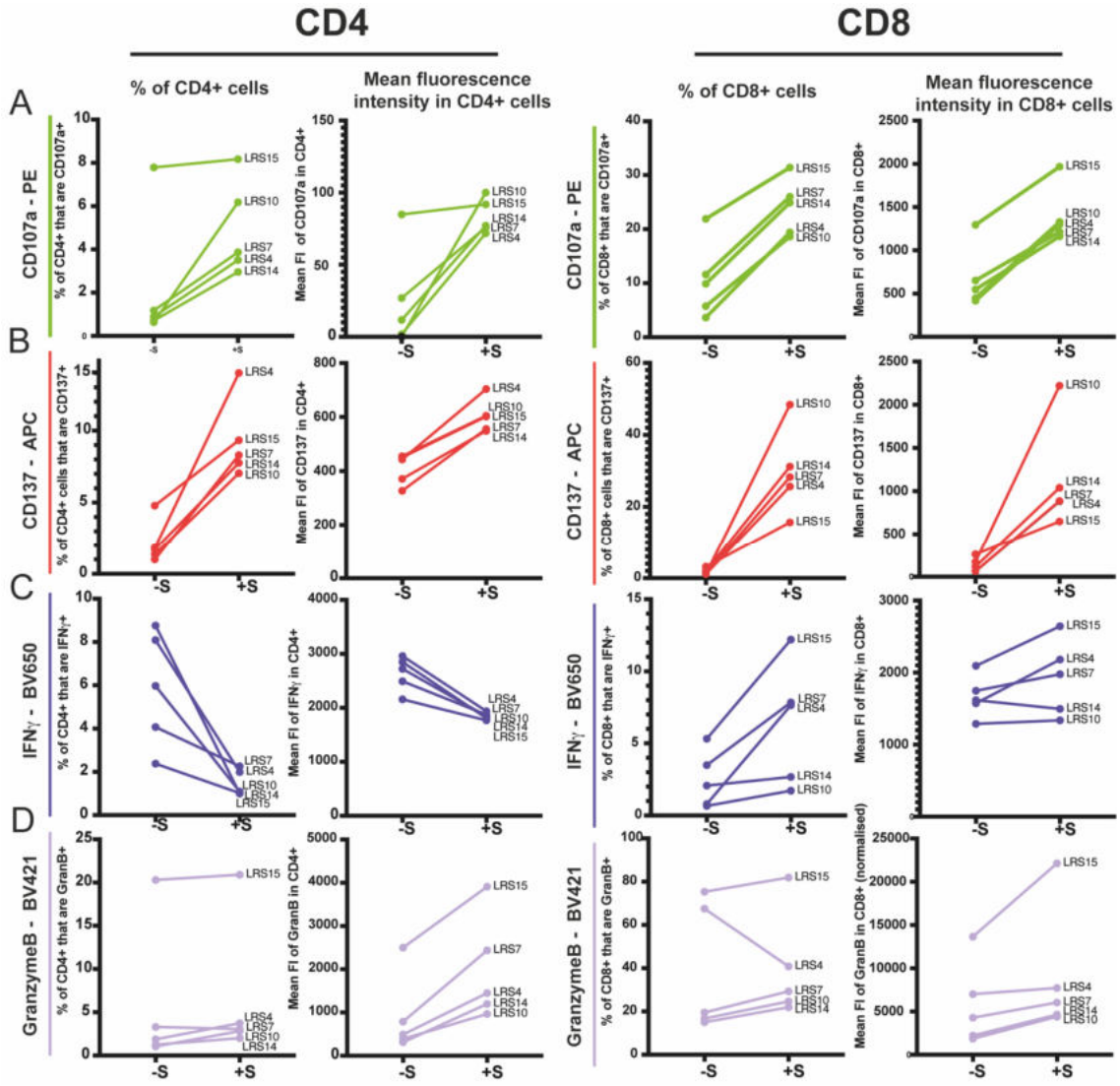


Figure 4.8. Testing suitability of the combined extracellular and intracellular readout for T-cell activation by PMA/Ionomycin over 21 hours. Percentage of CD4+ and CD8+ cells positive for designated marker, and the MFI of the designated marker in CD4+ and CD8+ cells within PMA/Ionomycin stimulated PBMCs. **A:** CD107a. **B:** CD137. **C:** IFNγ. **D:** Granzyme B.

Through this method the magnitudes of activation maker increase were lesser than those from the specifically timed cultures. For example, the percentage of T-cells expressing IFNγ was approximately 50% lower in the combined protocol compared against the intracellular protocol. However, despite this, in all cases, aside from IFNγ in CD4+ cells, the total 21H PMA/ionomycin treatment (which included a 5-hour GolgiPLUG and GolgiSTOP treatment) was able to stimulate an increase in the expression of the chosen activation markers. The reason for the downregulation of IFNγ in CD4+ T-cells could not be elucidated in this case. The markers were more strongly upregulated in CD8+ T-cells than CD4+ T-cells,

for example the percentage of cells expressing CD137 increased by an average of 28.13% in CD8+ T-cells, but only 7.14% in CD4+ T-cells. The corresponding changes in MFI were also smaller in CD4+ T-cells. Critically, this readout served its purpose at detecting T-cell activation over multiple parameters, and would be used in the future to test the tumour-reactivity of the T-cells expanded out from the autologous patient PBMCs.

6.4.5 Exome sequencing of patient GOA01

The time restrictions meant I could not obtain the exome sequencing data of the GOA01 PDOs to directly assess PDO mutations. However, Dr Louise Barber in our group had obtained the exome sequencing from a macrodissected FFPE biopsy taken at the same time, at the same location in the primary lesion. This is a limitation of this assessment, yet, studies have shown 90% of mutations can be shared between tumour tissue and tumour organoids (Weeber *et al.*, 2015). Macrodissection was performed on all the biopsies in the trial by the clinical pathologist Dr Ben Challoner due to visibly low tumour content in some of the biopsies.

Of the 600 mutations that were detected, 128 passed the most stringent filtering steps. There was 1 frameshift deletion, 4 frameshift insertions, 2 non-frameshift deletions, 8 stopgains, and 107 nonsynonymous point mutations. 66-94.4% of GOA cancers are of the MMRp subtype. Given the low TMB and absence of mutations in MSH2, MSH6, and PMS2, and MSH1, it is likely that this biopsy represents a MMRp tumour. Of those mutations, 6 were present in the 612 cancer-associated gene panel published in (Cai *et al.*, 2019): CDKN2A, GNAQ, MMP8, RYR1, SMAD4, and TP53. The patient GOA01 harboured a TP53 and SMAD4 mutation which was seen in 59.09% and 7.79% of the gastric cohort studied, respectively (Cai *et al.*, 2019). This illustrated this GOA tumour harboured some classical GOA mutations. The same study found WT TP53 was enriched in the high TMB tumours, and interestingly this holds true with our findings in GOA01: the tumour has mutant TP53 and displays a lower TMB. Tumours with low TMB tend to have a smaller pool of mutated proteins to present as peptides on HLA-I/HLA-II, which often means the tumours are non-immunogenic and do not benefit from ICI therapies. Therefore, it was also

possible that there would be no observable tumour-reactive CTLs in the T-cells yielded from the co-culture, once again making the sensitivity of the positive control important.

6.4.6 Characterising HLA-I expression on GOA01 under IFNγ conditions

Although the patient harboured no mutations in HLA-I/HLA-II and the APM pathway which meant it would highly likely to be HLA-I-proficient, it was also important to assess surface HLA-I abundance. This is because without surface HLA to express neoantigens, it is unlikely to be able to generate tumour-reactive CTLs from the co-culture.

After PDO dissociation with TrypLE express prior to co-culture assembly, PDO cells were also stained for flow cytometric quantification of HLA-I expression. I again utilised a pan-HLA-I antibody and QIFIKIT bead set to quantify the abundance of HLA-I on the surface. This approach allowed me to not only ascertain that organoids are HLA-I proficient, but the quantification of HLA-I against a defined bead set also provided a mode of normalisation between separate runs. This meant HLA-I expression on multiple PDO lines could be assessed and properly compared over time, which was important for the ongoing characterisation of all the GOA PDOs. This would not be the case if MFI was the sole readout of HLA-I expression due to the day-to-day variability in the lasers of flow cytometric equipment. QIFIKIT analysis showed IFNγ-treated GOA01 PDOs to express a median of 495,450 HLA-I molecules on the cell surface (Figure 4.9A and 9B); this was similar to that of HLA-I expression seen in the 5 MMRp CRC PDOs in Chapter 3 (330,100-496,000).

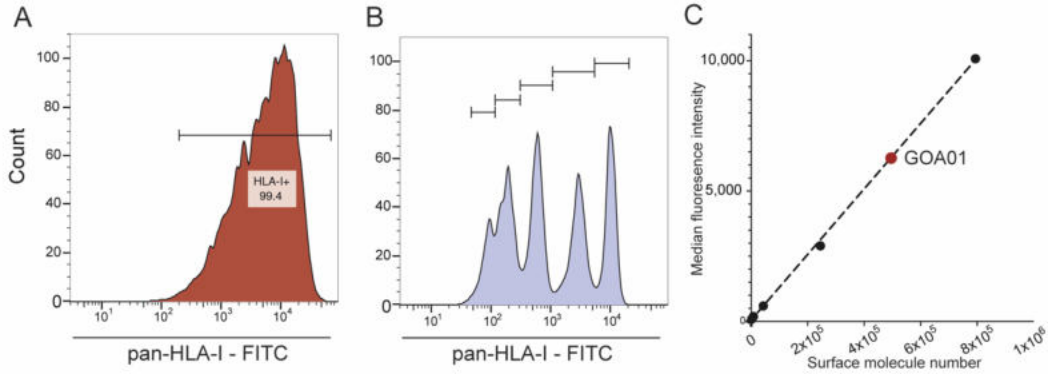


Figure 4.9. Quantification of surface HLA-I abundance on GOA01. A: Histogram of the FITC pan-HLA-I stain on GOA01. **B:** Histogram of the FITC stain on the QIFIKIT quantification beads. **C:** MFI of HLA-I relative to the standard curve created by the QIFIKIT beads.

6.4.7 Running the autologous PBMC-PDO co-culture on patient GOA01

The T-cell expansion protocol is described in detail in the materials and methods section. To summarise the protocol (illustrated in Figure 4.10A), PDOs were established from biopsies of clinical trial patients, and matching PBMCs were isolated. Co-culture maintenance and antibody staining was performed by Dr Marta Buzzetti. Once the PDOs had grown to a sufficient number, a well of PDOs (approximately $3x10^5$ cells) were removed from matrigel by an incubation with dispase (D-2), and cultured in 3D organoid media supplemented with ROCKi overnight to prevent cell death from the detachment from matrigel. The next day (D-1), the PDOs were treated with IFNγ to enhance the surface expression of HLA, and the PBMCs were thawed and placed in to high-dose IL-2 T-cell media. The co-culture was then assembled (D0) by combining the matched patient PBMCs with TrypLE Express-dissociated PDO cells at a ratio of 20:1 in T-cell mediated supplemented with high-dose IL-2 and anti-PDL1 antibody. The co-culture was incubated for a week, with media changes every 2-3 days, then refreshed with more PDO cells (D7). By D14, the expanded T-cells were harvested and used for interrogation in a tumour-reactivity assay.

At each stage of the co-culture preparations, the morphology of the PDOs were closely monitored as it was important to determine how the PDOs were affected removal from the 3D culture substrate Matrigel (Figure 4.10B). The D-1 image illustrated that the PDOs tolerated overnight culture out of Matrigel and retained their morphology and looked bright, illustrating the PDOs were still alive. After a further 24H of culture, and with the addition of IFNγ, some of the cells clumped together and loosely attached to the plastic culture dish. There was some visible cell death, with small clusters of cells broken away from the PDO spheroids; this was confirmed when the cells were dissociated and counted, as the PDO cells only had a viability of 53%. However, the surviving PDOs mainly maintained their morphologies with their lumen still visible. After PBMCs were mixed at 20:1 with

the dissociated PDOs, live PDO single cells and small cell clusters were visible, with the PBMCs harder to visualise due to their small size and suspension in media. This illustrated that the PDOs were able mostly to tolerate culture without matrigel for the 48H preceding addition to the co-culture, making them viable for addition to the co-culture with PBMCs.

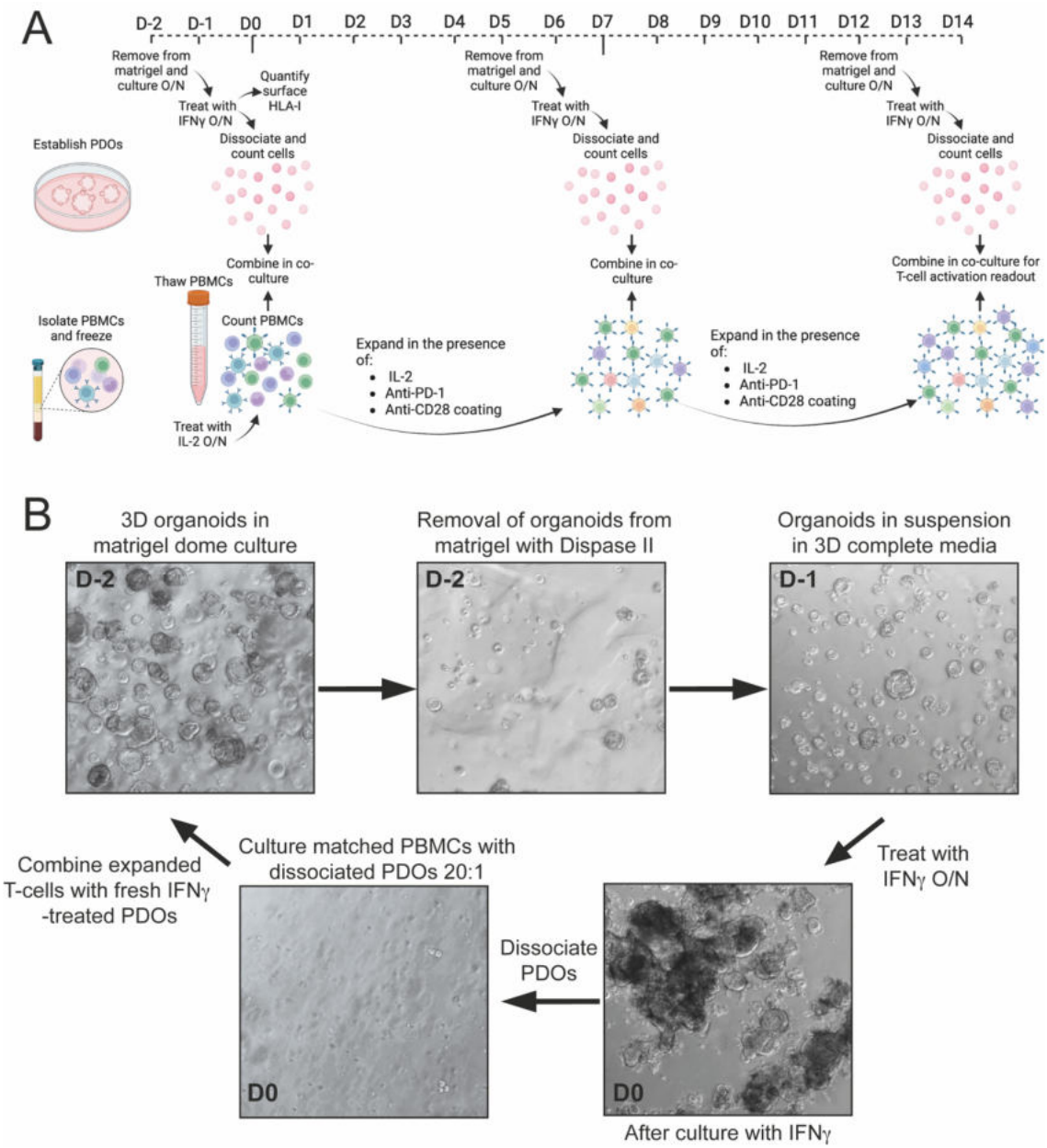


Figure 4.10. Seeding of the autologous PBMC-PDO T-cell expansion co-culture. A: A schematic depicting the protocol for autologous tumour-T-cell expansion from PBMCs from clinical trial patients. Inspired by the publications (Dijkstra *et al.*, 2018; Cattaneo *et al.*, 2020). **B:** A microscope of the cell cultures throughout the co-culture preparation process. Images taken on a Zeiss Axiovert 135 microscope at 5x magnification.

6.4.8 The T-cell activation readout from the autologous PBMC-PDO co-culture on patient GOA01

This was the last experiment performed in my final few weeks as a PhD student at the ICR. Due to delays in the order of the optimised PE anti-CD107a antibody, a CD107a on an alternative fluorophore had to be borrowed (fluorophore AF647) which necessitated the re-composition of the antibody panel used for the assay. Due to the lack of an anti-CD137 antibody with a compatible fluorophore, CD137 assessment could not be performed in this run. The long-term intention of this assay is to analyse both extracellular and intracellular markers, and once the PE anti-CD107a antibody arrives, the future readouts will be reverted back to the panel that was optimised in Figure 4.8.

For the tumour reactivity co-culture, PDOs were prepared as in the T-cell expansion protocol. However, for this assay, the expanded T-cells were combined with the fresh PDOs cells at a ratio of 2:1 in another anti-CD28 pre-coated plate. The two-week co-culture produced $1x10^6$ cells from a vial containing $1.4x10^6$ PBMCs, which enabled 10 wells of the co-culture to be performed. In every condition expanded T-cells and PDOs were combined under different conditions: 1) no immunotherapy, 2) atezolizumab (20μg/mL), 3) nivolumab (20μg/mL), 4) HLA-I-blocked-PDOs with T-cells (prior 200μg/mL incubation as detailed in Materials and Methods) and 20μg/mL atezolizumab, and 5) PMA and ionomycin (50ng/mL and 1μg/mL, respectively). HLA-I-blocked PDOs were intended to serve as the negative control, and PMA and ionomycin-treated cultures were intended to serve as the positive control for staining. For each condition, duplicate cultures were performed.

All cells and treatments were combined and left to incubate at 37°C for 16 hours. After the 16H incubation, GolgiSTOP and GolgiPLUG were added to the T-cell culture medium, and left to incubate for a further 5 hours. After the incubation, cells were harvested and stained for flow cytometric assessment.

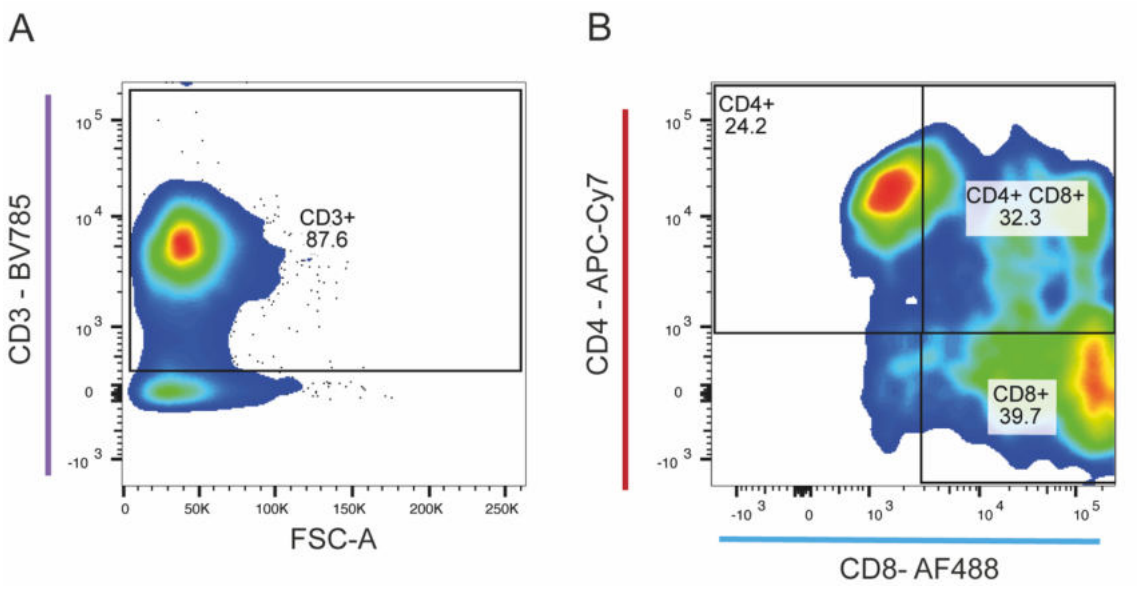


Figure 4.11. Subsets of co-culture-expanded T-cells. A: CD3 assessment of live cells. **B:** CD4, CD8 assessment of CD3+ live cells**.** Single plots representative of the 10 test wells.

First, I analysed the characteristics of the cells that were expanded from the PBMC-PDO culture. Across the 10 samples a mean of 88.8% of live cells were CD3+ (SD of 3.2). This is very closely in line with the findings from the original paper (Cattaneo *et al.*, 2020), showing the co-culture system functioned well in supporting the expansion of T-cells from the PBMCs. I then went on to analyse the CD4/CD8 positivity of the expanded T-cells. Across the 10 samples a mean of 21.4% of T-cells were solely CD4+ (SD of 7.54), and 43.6% of T-cells were solely CD8+ (SD of 6.56). A mean of 29.7% of the CD3+ subset was both CD4+ and CD8+ (SD of 4.38). In healthy individuals CD4+ CD8+ T-cells only constitute 1-2% of circulating T-lymphocytes (Clénet *et al.*, 2017), however, double positivity has been recorded in multiple disease settings including urological cancer, autoimmune, and chronic inflammation diseases (Parel and Chizzolini, 2004; Bohner *et al.*, 2019). Initial PBMC subset analysis was not performed out of concern for keeping as many of the limited clinical trial PBMCs for future analysis. However, in the future it would be beneficial to observe whether the ratio of CD4+, CD8+, and CD4+CD8+ cells was similarly abundant in the source PBMCs, as this could indicate whether the culture may have influenced this.

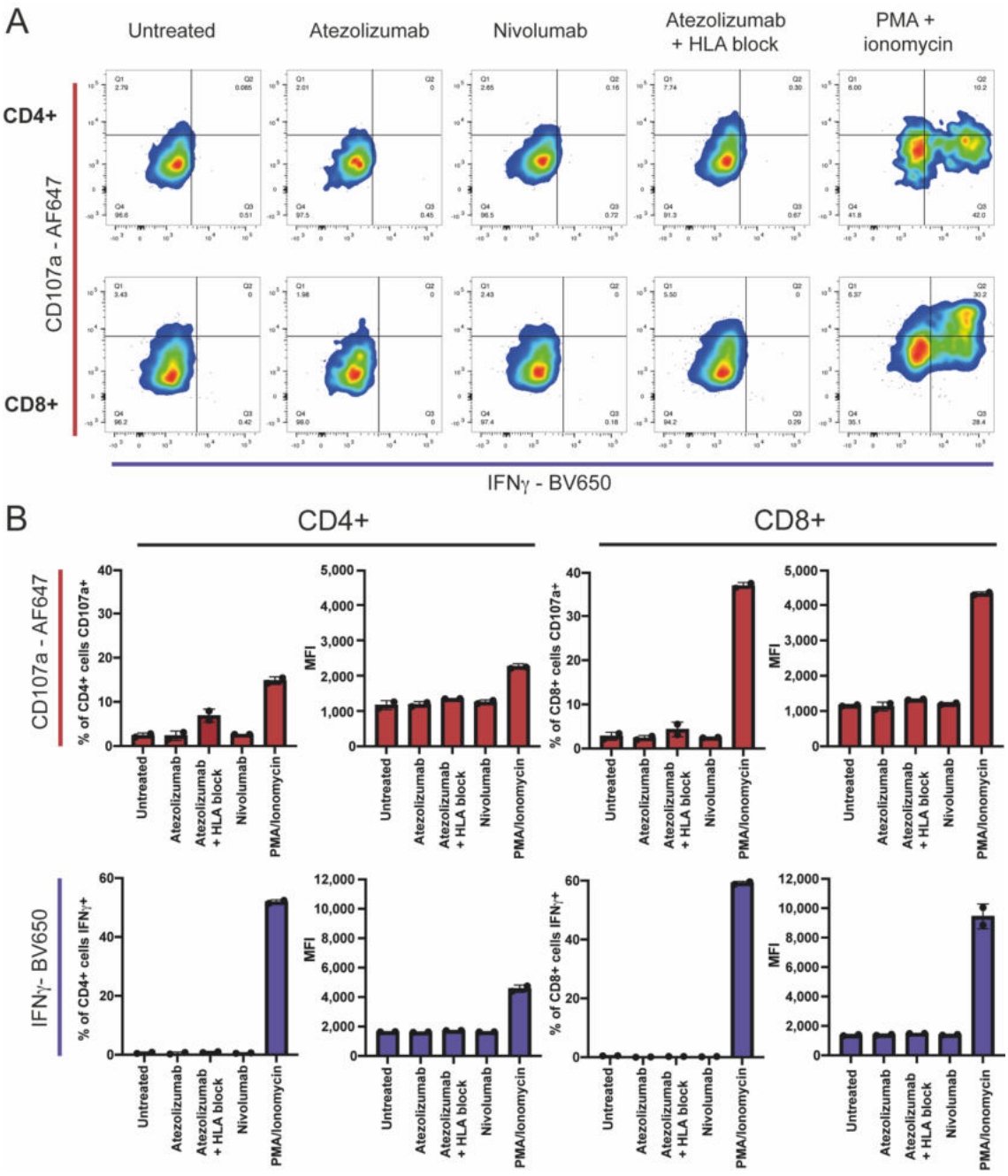


Figure 4.12. Flow cytometric evaluation of CD107a and IFNγ in T-cells expanded from PBMC-PDO co-culture. A: A representative scatterplot of CD107a and IFNγ on live, CD3+ T-cells, CD4+/CD8+ T-cells. **B:** Plots of percentage of CD4+ and CD8+ T-cells expressing, and MFI of, CD107a and IFNγ. N = 2 biological replicates.

Next, to evaluate whether autologous T-cell-PDO co-culture was able to stimulate T-cell activation, the CD4+ and CD8+ T-cells were analysed for CD107a, IFNγ, and Granzyme B expression. None of the untreated, Atezolizumab-treated, or Nivolumab-treated PDO and T-cell co-cultured populations displayed an increase

in CD107a or IFNγ positivity or brightness when compared against the negative control HLA-blocked cultures (Figure 4.12A-B). In fact, the HLA-blocked wells displayed higher levels of CD107a positivity in both CD4+ and CD8+ T-cell populations, but not IFNγ positivity. It is possible that the mouse anti-HLA-I antibody bound to the PDOs may have triggered a small amount of degranulation in the T-cells. In both CD4+ and CD8+ T-cells PMA/ionomycin treatment strongly induced CD107a positivity (to an average 14.8% and 37.1% respectively) and IFNγ positivity (to an average 51.9% and 59.2% respectively). PMA/ionomycin treatment also strongly increased CD107a and IFNγ brightness in each of the T-cell populations. This illustrated the stimulation timeframe allowed for successful induction of markers for strong antibody staining, and could therefore define the lack of visible T-cell response down to lack of PDO immunogenicity, and not failed staining.

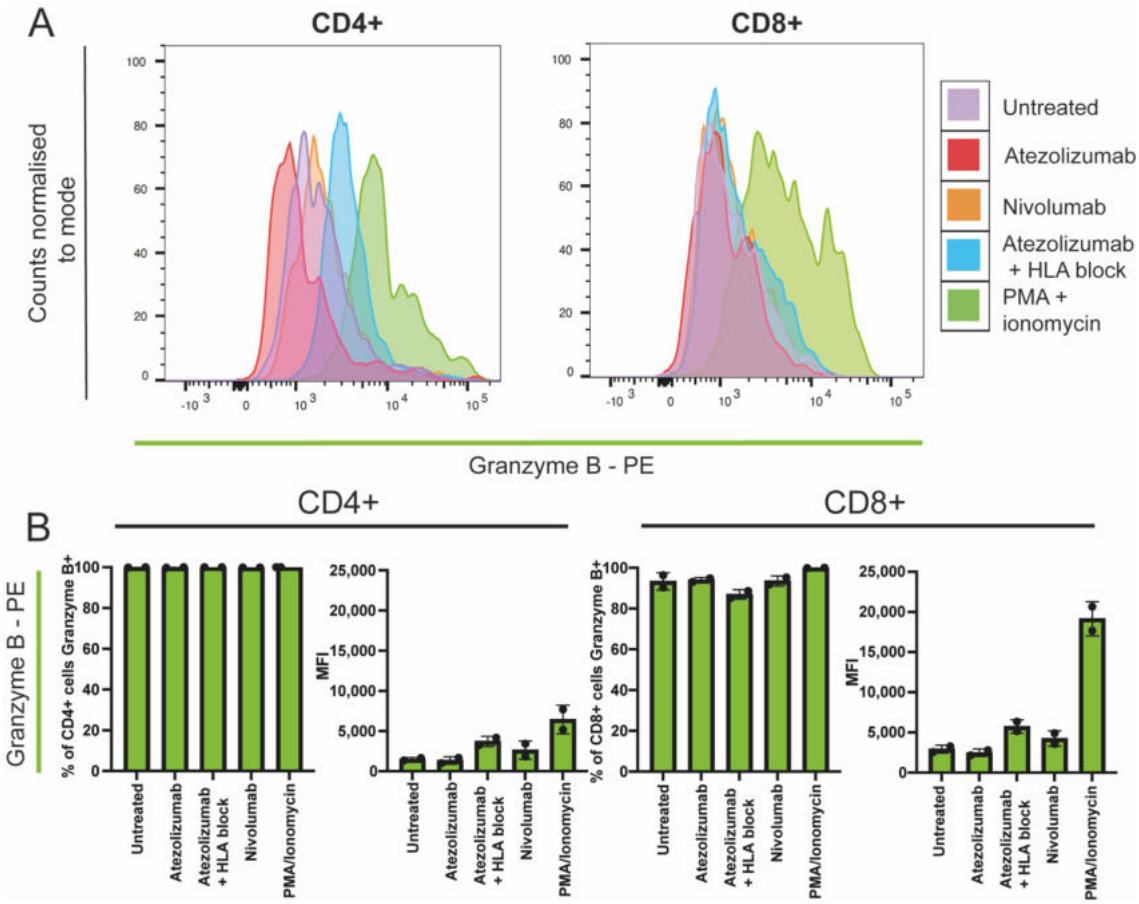


Figure 4.13. Flow cytometric evaluation of Granzyme B in T-cells expanded from PBMC-PDO co-culture. A: A representative histogram of Granzyme B fluorescence on live, CD3+ T-cells, CD4+/CD8+ T-cells. **B:** Plots of percentage

of CD4+ and CD8+ T-cells expressing, and MFI of, Granzyme B. N = 2 biological replicates.

The percentage of cells expressing Granzyme B was very high even in the untreated condition which was likely due to the prolonged exposure to IL-2 (Tamang *et al.*, 2006). There were only small differences in the Granzyme B positivity or brightness between the negative control HLA-blocked cultures and the untreated, Atezolizumab-treated, or Nivolumab-treated PDOs (Figure 4.13A-B). Between 86.0%-100% of the CD4+ and CD8+ T-cells were defined as positive for Granzyme B staining. The atezolizumab and HLA-blocked culture in CD8+ had the lowest positivity and MFI for Granzyme B, yet had the strongest (aside from the positive control) in CD4+ T-cells. The lack of Granzyme B upregulation in the main test wells showed that none of the immunotherapeutic treatments could help in unlocking anti-PDO T-cell activity.

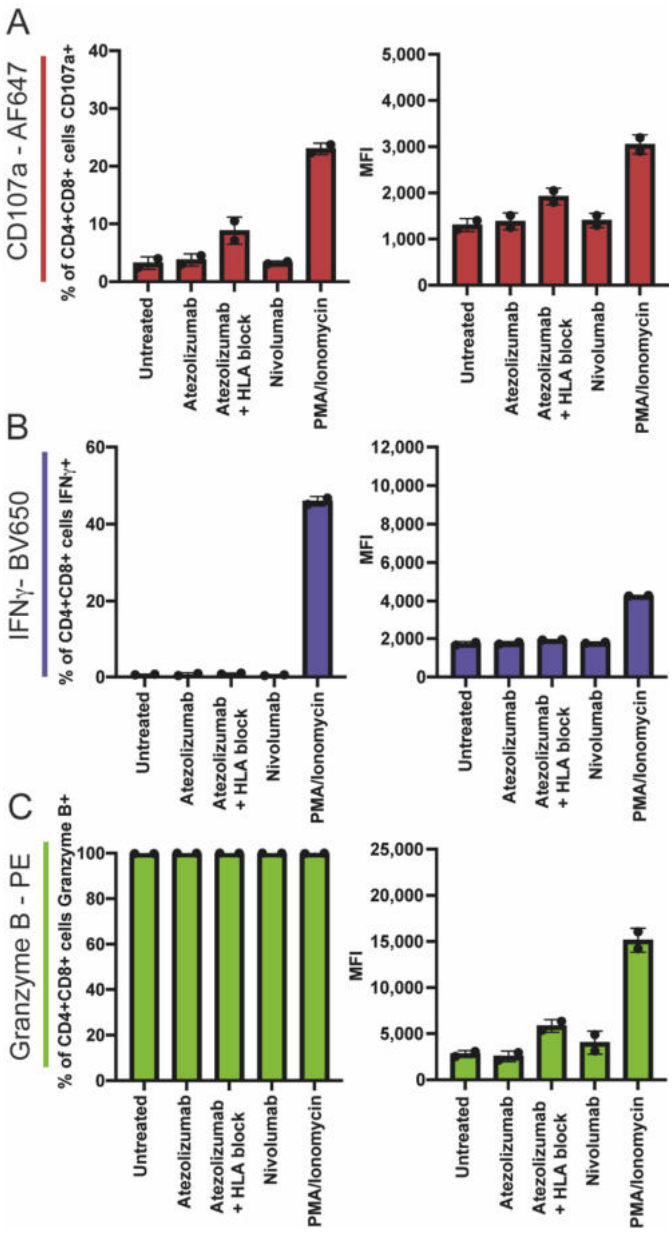


Figure 4.14. Flow cytometric evaluation of T-cells activation markers in CD4+CD8+ T-cells expanded from PBMC-PDO co-culture. Plots of percentage of CD4+CD8+ T-cells expressing, and MFI of, T-cell activation markers. N = 2 biological replicates. **A:** CD107a. **B:** IFNγ. **C:** Granzyme B.

As studies have shown that CD4+CD8+ T-cells have a greater potential to generate IFNγ and granzyme B when compared against single positive T-cells (Clénet *et al.*, 2017), I wanted to observe whether this subset of T-cells responded differently to PDO exposure. Across none of the test conditions did the T-cells experience an increase in CD107a, IFNγ, or Granzyme B. Again, small increases in Granzyme B brightness and CD107a positivity were seen in the Atezolizumab and HLA-blocked samples when compared against the test conditions. The

PMA/ionomycin positive control exhibited lower levels of CD107a and IFNγ upregulated than that seen in the CD8+ T-cells. The CD4+CD8+ T-cells also showed lower Granzyme B upregulation than CD8+ T-cells, but higher than the CD4+ T-cells. This data together showed that the co-culture-derived CD4+CD8+ subpopulation from patient GOA01 did not have a higher capacity to produce cytokines and lytic enzymes when compared against CD8+ T-cells, and furthermore, could not exert anti-PDO T-cell activity.

6.5 Discussion of autologous PBMC-PDO co-culture

The purpose of this platform in MMRp GOA was to test the reactivity of the expanded autologous T-cells against the PDOs. Further to assessing if tumour-reactive cells can be expanded from PBMCs in the PDO co-culture, this co-culture platform could be used to scrutinize the clinical relevance of MS-detected and predicted neoantigens. Immunopeptidomics and neoantigen prediction could be carried out on the PDOs for the generation of synthetic peptides. The synthetic peptides from both sources could then be used to try to expand neoantigen-specific T-cells from PBMCs. The resultant T-cells could then be combined with the PDOs to assess for T-reactivity against the PDO. This could help to validate whether MS-detected neoantigens are more likely to be presented by PDOs and recognised by T-cells, than predicted neoantigens. Therefore, preserving PBMC numbers to enable testing of multiple candidate neoantigens was an important feature of the protocol. Consequently, to try and get as much data as possible from a minimal amount of input, the mix of extracellular and intracellular staining was important for defining PDO reactivity from a single well of a co-culture.

These experiments exhibited that the 2-week autologous PBMC-PDO co-culture platform could successfully support the expansion of viable T-cells from 1.4 $x10^6$ PBMCs, with $1x10^6$ cells yielded (88.8% CD3+). They also highlighted the need for stringent controls; robust positive controls were required to determine success of antibody staining in a system where PDO mutation numbers, and therefore likely neoantigen numbers were low. Testing of positive controls in LRS donors allowed intracellular and extracellular readouts to be successfully combined in the optimisation stages of the experiment. Furthermore, negative controls were limited to solely HLA-I-blocked PDOs in an effort to preserve cell number,

however, a small amount of activation was induced in the control, likely from the presence of mouse antibodies bound to HLA-I on the PDOs. This illustrated that a different HLA-I-blocking antibody may be required, or that T-cells alone are an essential negative control for this system.

Although CD137 assessment was absent from the final co-culture readout, 3 markers of T-cell activation were used, and were able to illustrate a lack of T-cell activation by an absence of upregulation of CD107a, no IFNγ, or granzyme B. The PMA/ionomycin stimulation of the positive control was able to strongly illustrate activation of co-cultures in all 3 markers, providing confidence the lack of CD107a, no IFNγ, or granzyme B signal was not due to failed antibody staining or an insufficient activation duration. However, it may have been beneficial to have a further positive control that would validate the success of the co-culture system in successfully generating PDO-specific T-cells. For example, having a high mutation load PDO and paired PBMCs for autologous culture may have provided an example more likely to stimulate PDO-specific T-cells, and could have therefore provided confidence that the co-culture system could support the expansion of PDO-specific T-cells. However, this was not possible as we did not have access to such material.

The system functioned as expected for a low TMB PDO: generating viable T-cells, but that were not reactive against the matched patient PDOs. Furthermore, these findings were very similar to the papers in MMRp CRC which motivated use of this co-culture platform (Dijkstra *et al.*, 2018; Chalabi *et al.*, 2020). Therefore, I believe that this co-culture platform could be used to characterise other GOA PDOs and PBMCs. For example, patient GOA01 successfully established PDO cultures from 4 different core biopsies, therefore this technique could be used on all 4 cores to help better evaluate intratumour heterogeneity through the lens of immunogenicity. In addition to this, the co-culture platform could be used to scrutinize the clinical relevance of MS-detected and predicted neoantigens. Testing whether exposure to synthetic peptides informed by MS data, or to synthetic peptides informed by neoantigen prediction could better promote T-cell reactivity against PDOs could highlight which approach generated more clinically-useful neoantigens. Understanding this difference could impact the approach of cancer vaccine design.

Chapter 7 Final conclusion and future implications

The goal of my thesis was to advance the understanding of antigen presentation and immunogenicity in MMRp CRC and GOA, and to investigate possible treatments to augment it. This was a clinically pertinent topic to address as the lower tumour mutation load in MMRp cancers, and therefore the lower possible neoantigen load, is a form of intrinsic resistance to ICI therapy which results in poorer response rates than their MMRd equivalents. This question was approached through a number of different angles, utilising multiple complementary techniques.

In the first results chapter (Chapter 3) I investigated the hypothesis that MMRp CRCs exhibit low presentation of neoantigens which could stimulate an immune response. This hypothesis was informed by data illustrating the disparity between predicted neoantigen numbers and low clinical response to ICI therapy (Grasso *et al.*, 2018; Woolston *et al.*, 2019; Glaire *et al.*, 2022). This was the first described immunopeptidomics analysis of CRC PDOs, and provided clinically important findings about the immunopeptidome of MMRp CRCs. Through high-depth MS identification of HLA-eluted peptides, basal presentation of 2,124-16,030 HLA-I unique peptides (UPs) and 6-713 HLA-II UPs could be detected across 5 PDOs. The UP number was modestly correlated with HLA surface abundance, with low peptide diversity in one sample being explained by loss of heterozygosity (LOH) on chromosome 6. Furthermore, another result of LOH was that only HLA allotypes that could present peptides with a chymotryptic-type C-terminus remained. This directly illustrated the importance of HLA diversity for immunopeptidome diversity. Mutation status or expression of antigen presentation machinery (APM) or peptide lading complex (PLC) genes could not further explain the disparity between other PDOs. Only 3 neoantigens were detected across 5 PDOs with a total of 612 non-silent mutations, indicating the neoantigen landscape to be extremely sparse in these MMRp CRCs. This also illustrated the disparity between NetMHC4.0-predicted neoantigens and MS-detectable neoantigens, where from 196 of the expressed genes that were predicted to generate a neoantigen, only 2 were detected, with the third neoantigen not considered to be a strong binder (with binding rank of 5.8%).

Similarly, the repertoire of non-mutated cancer/testis (C/T) antigens amongst the PDOs was scarce, with only 3 C/T antigens identified from the 5 PDOs. This work confirmed the intrinsic low immunogenic potential of these MMRp PDOs. Even when considering the sensitivity limits or intrinsic MS biases against highly hydrophobic peptides, or peptides containing 'reactive' amino acids cysteine and methionine may have caused some loss of peptides, the low detection of neoantigens and CT/antigens was still remarkable. The evidence of low neoantigen presentation in these therapy-resistant advanced MMRp CRCs suggest that their low HLA-presented neoantigen load could contribute to poor sensitivity to ICI therapy. A limitation of this study was that it was not possible to test the immunogenicity of the PDOs by *in vitro* T-cell reactivity assays due to the lack of autologous PBMC samples for these patients. This would be important to include in future work for understanding the immunogenicity of the neoantigens detected by MS. It should also be noted that there may be other mechanisms which contribute to poor ICI sensitivity such as deregulation of immune cell trafficking caused by lack of chemokine expression, or immune cell exclusion by the tumour stroma (Bonaventura *et al.*, 2019; Liu and Sun, 2021). Together, low neoantigen loads, combined with a high prevalence of these other immunosuppressive mechanisms may explain the poor results in the clinic.

Because immunopeptidomics is such a fast-moving field, certain developments in methodology have facilitated higher depth of detection, such as a TMT-MS approach (Pfammatter *et al.*, 2020), or a double reduction protocol to assist resolution of cysteine and methionine-containing peptides (Wang *et al.*, 2019). I believe future studies should test and incorporate these additions in the interest of improving detection depth. One of the limitations of the immunopeptidomics approach used in this thesis was the large number of cells required for high-depth analysis. Collecting these cells was costly and time-intensive. The higher sensitivity of the new adjusted protocols would allow for reduced cell input, which increases the feasibility of using immunopeptidomics for studying biological questions. This means that PDOs could be studied far sooner after establishment, and could therefore best represent the immunopeptidome of the tumour without the risk of accumulation, or loss, of mutations. Studies are just starting to be performed in early-stage MMRp CRCs and are showing promising results, with a 30% response rate to ipilimumab (anti-CTLA-4) and nivolumab

(anti-PD-1) combination treatment (Verschoor *et al.*, 2022). This could also support the investigation of immunoediting by immunopeptidomics; it would be most interesting to take biopsies from early-stage CRC patients at time points through diagnosis, treatment, and possible progression to track the dynamics of neoantigen presentation. Having definitive evidence that neoantigens are lost over time, and fully understanding the impact of different therapies on neoantigen presentation, would support the argument for more clinical studies of immunotherapy in early stage CRCs. This would be particularly important for MMRp CRC patients, as their initial neoantigen load is likely to still be smaller than their MMRd equivalents. A limitation of this study is the lack of analysis of non-canonical antigens present on HLA (derived from defects in antigen processing, fusion genes, de-repressed endogenous retroviruses, transposable elements, post-translationally modified peptides, and peptides from novel open reading frames). It would also be beneficial in future studies to analyse post-translational modifications on MS-detected peptides (Bedran, Gasser, *et al.*, 2023; Bedran, Polasky, *et al.*, 2023), as features like glycosylation are reported to have a significant impact on the production of antigens (Li *et al.*, 2008) and their recognition by T-cells (Housseau *et al.*, 2001; Olvera *et al.*, 2021). These could be particularly relevant in the identification of alternative targets for peptide cancer vaccines or TCR-based therapies.

Further to the investigation of basal neoantigen expression, I also investigated if perturbation with IFNγ or MEK inhibitor treatment could increase the number of neoantigens or C/T antigens. Although IFNγ treatment increased the surface abundance of HLA-I, and expression of APM and PLC genes, only small changes in peptide numbers were seen due to the balance of peptide loss and peptide gain between the two conditions (a mean 5.67% gain across the 4 PDOs). The surface abundance of HLA-I and HLA-II on the CRC PDOs increased under IFNγ exposure, and this resulted in increased HLA-II peptide numbers. Despite this, no new neoantigens or C/T antigens could be detected on HLA-I or HLA-II after IFNγ treatment. Treatment with MEK inhibitor trametinib had more varied results on HLA-I upregulation, but did not significantly increase the number of unique peptides on HLA-I or HLA-II, nor did it result in the presentation of additional neoantigens or C/T antigens. A further treatment of HDAC inhibitor domatinostat was optimised, and assessed for transcriptome remodelling potential for by

RNAseq analysis. HDAC inhibitor treatment did not statistically significantly increase expression of APM or PLC genes, but did result in a median 8.1% gain in the number of expressed nonsynonymous point mutated genes. Because these results indicated only modest changes, given the time restraints imposed by institute lockdown, immunopeptidomics was not performed on domatinostat-treated cells due to the time taken to grow them to the appropriate number. Further to this, later clinical data illustrated the combination treatment of HDAC inhibitor and ICI avelumab (anti-PD-1) did not provide clinical benefit. With the improvements in immunopeptidomics sensitivity described above, it would be possible to test the effects of different treatments more readily. Furthermore, this illustrates the importance of using immunopeptidomics analysis as a pre-clinical validation platform that can test the efficacy of neoantigen-boosting drugs before they are tested on patients. Immunopeptidomics provides the best hope for direct identification of targetable antigens that could be included in cancer peptide vaccines. However, as this is not feasible for all individuals, collating immunopeptidomics data to show the regions of proteins represented as peptides on HLA-I may help to better predict the presentability of neoantigens, as seen in (Bedran *et al.*, 2022).

In Chapter 4, the combination of TMT-proteomics data with the transcriptomic and immunopeptidomics datasets from Chapter 3, enabled me to resolve some of the mechanisms behind immunopeptidome remodelling under IFNγ treatment. The first finding was that the abundance change of some peptides was significantly correlated to the abundance change of the peptide source protein (average $r = 0.54$, $p < 2.2\ 10^{-16}$). However, a larger proportion of the peptides changed in abundance independent of protein abundance change. The analysis indicated a small proportion of the change was likely due to the increased chymotryptic activity experienced with increased immunoproteasome activity under IFNγ treatment. Another portion of the immunopeptidome change could be attributed to the differential upregulation of HLA-B compared against HLA-A. This resulted in the higher upregulation of peptides that were restricted to HLA-B. One key finding from the analysis of multiple peptides from long proteins was that there are peptide-specific characteristics that controlled their abundance – as peptides derived from the same protein could be up- or downregulated

independent of their position in the protein, HLA allotype, and HLA binding affinity.

The most striking peptide-specific characteristic was the underrepresentation of proline within positions 4-6 of the peptides which most increased in intensity under IFNγ treatment, or were exclusive to the IFNγ condition. Data from a PDO expressing an HLA which tolerated proline in position P2 of the peptide, and the validation dataset which tolerated proline in additional positions, also showed proline to also be underrepresented in those same positions of peptides which increase in abundance under IFNγ. This suggested that peptides with proline in any position where it was tolerated on the HLA, could experience a decline in abundance under IFNγ exposure. This could not be explained by effects of proline on peptide HLA-binding affinity, or increases in prolyl peptidase protein abundance. From this, I have hypothesised that the known peptide-protective effect of proline is more pertinent in the untreated condition (Shimbara *et al.*, 1998; Abelin *et al.*, 2017), where peptides have more time in the cytoplasm and the ER being exposed to peptidases before they are then presented on HLA (Kim *et al.*, 2003).

This analysis was limited in that it was only validated on a single other cancer dataset. However, this was necessitated because I only analysed datasets of cancer cell lines that had been treated with IFNγ for 48-hours. This was to ensure resolution of differences was possible, as protein degradation, and therefore resulting peptide presentation, is a dynamic process. I am confident in the validity of these findings, but testing larger numbers of PDOs would be necessary to define whether this is a universal effect across cancer types and different HLAs. Additionally, biochemical validation of the stability of proline-containing HLA-peptides and precursor peptides between untreated and IFNγ-treated conditions would further support this finding. If this finding could be replicated on a large number of datasets, it could be important to use this knowledge to train peptide-prediction algorithms, so this bias could be integrated. Alternatively, this study argues for the creation of a separate peptide-prediction training dataset to be created with IFNγ-treated samples. This is because whilst the data suggests that IFNγ-treatment does not affect peptide affinity to their HLA allotype, the difference in eluted HLA-ligands would affect the eluted-ligand likelihood predictions. The

understanding of features which would impose peptide downregulation in an IFNγ environment could be immensely important for the design of cancer vaccines. Exclusion of peptides which are more likely to be downregulated under IFNγ exposure, such as peptides which contain proline in positions 4-6, would ensure the inclusion of peptides which are more likely to remain presented in both conditions.

Data analysis in Chapter 3 had shown that NetMHCpan does overpredict the number of neoantigens that can be represented on HLA-I, suggesting there may be a bottleneck in the presentation of neoantigens. Further analysis performed in Chapter 5 illustrated that some of the bottleneck in neoantigen presentation came from an absence of mutated proteins, or lack of turnover of the mutated proteins (illustrated by lack of non-mutated peptides present on HLA-I). However, even whilst only 54 of the 444 mutated genes predicted to generate a neoantigen were represented by non-mutation-containing peptides on HLA-I, this was still at odds with the only 3 HLA-I neoantigens detected by MS. I therefore hypothesised that there are more limiting steps to the neoantigen-presentation pathway, which could perhaps be better understood by the analysis of the ER peptidome. I first eliminated the possibility of proximity labelling of the ER peptidome, as analysis of the HLA-I immunopeptidomics data showed it would be unable to represent enough of the peptidome. I further excluded the use of a commercial ER enrichment kit due to poor yield and product purity. Following this, I began the optimisation process for the enrichment of the ER by subcellular fractionation. The first two steps of the protocol were differential centrifugation to remove the nuclei and un-lysed cells first, followed by exclusion of the mitochondria. The centrifugation speeds were optimised to reduce the amount of ER loss to the discard pellets. The next step was the rate-zonal centrifugation of the post-mitochondrial supernatant with a density gradient. I first worked to optimise the use of a sucrose density gradient, but further analysis showed it resulted in the co-enrichment of the ER with the lysosomes which was a potential source of peptide contamination. Following this I trialled the use of a Histodenz-based gradient. WB analysis of this found enrichment of ER coupled with good elimination of Golgi, lysosomal, secreted, and cell surface proteins. The final protocol utilised: cell lysis by nitrogen cavitation; a 10-minute 800g nuclear exclusion centrifugation; a 10-minute 3,000g mitochondrial exclusion

centrifugation that minimised ER loss; and a Histodenz density gradient which was centrifuged at 126,000g before needle fractionation of the gradient layers. This protocol was brought forwards for MS proteomics and peptidomics analysis. Proteomics analysis of the samples derived from the gradient layers confirmed the good depletion of the Golgi, lysosomal, secreted, and cell surface components in the layer of interest.

I began by analysing all MS-detected peptides from across all layers of the Histodenz gradient. TMT-MS yielded 3,009 peptides from across the 3 layers. Critically, the peptide length distribution was similar between the gradient-derived and HLA-I-eluted peptides, but with more gradient-derived peptides 11-15aa long. Additionally, the majority of both the gradient-derived and HLA-I-eluted peptides had a peptide C-terminus defined by chymotryptic activity. Furthermore, 594 peptides were exact matches between the HLA-I-eluted and gradient-derived peptides, with 402 of the matches being 9-mers. This illustrated that 37.5% of the gradient-derived 9-mers were also seen on the cell surface. This high degree of overlap suggested that these could be HLA peptides. When analysing the NetMHCpan4.1 BA rank of the peptides I found a lower proportion of gradient-derived peptides (55.59%) were predicted to bind their attributed HLA than the HLA-I-eluted peptides (86.26%). When this analysis was restricted to 9-mers a much higher proportion of the gradient-derived peptides were predicted to bind to HLA (86.80%), closer to that of the HLA-I-eluted peptides (93.78%). The disparity in BA ranks between the gradient-derived and HLA-I-eluted peptides was expected, as peptides with an insufficient binding affinity would not be loaded on to HLA-I and would be retained in the ER. The peptide motifs of the gradient-derived peptides predicted to bind to HLA-I showed remarkable similarities to those from the HLA-I-eluted peptides.

One other critical finding from the peptidomics analysis of the gradient-derived peptides was the number of peptides which appeared as N- and C-extended versions of the peptides detected in the HLA-eluted peptides. 100% of the peptides which had extended versions detected in the gradient-derived peptidome, also had an exact match between the HLA-eluted peptides and the gradient-derived peptides. 59.6% of extended sequence peptide matches were extended at the N-terminus, 19.2% extended at both the N-terminus and the C-

terminus, and 21.2% were extended just at the C-terminus. This was consistent with the understanding that there are multiple N-terminal-acting peptidases which cleave amino acids from longer precursor peptides (Beninga, Rock and Goldberg, 1998; Van Endert, 2011; Ferro, Gewehr and Navon, 2020). Furthermore, 85.7% of the C-extended peptides were possible chymotrpytic, tryptic, or caspase ligands, which may indicate that these were alternative proteasomal product variants. This was consistent with what would be expected of peptides from the ER. Furthermore, I was able to detect neoantigens from the gradient-derived peptidome: 7 of which were shared with the HLA-I-eluted peptidome, and 4 which were exclusive to the gradient-derived peptidome. Whilst this run did not perform as well as it had previously done, necessitating the less-specific study of the gradient-derived peptidome, I still believe this illustrates the potential of this methodology to study neoantigen loss in the ER.

It is important to acknowledge two of the main limitations of this piece of work. First, I wasn't able to generate any replicates for the MS analysis due to time and ultracentrifuge access issues, which precluded me from being able to analyse the reproducibility of the MS, or critique the quality of this particular run. Furthermore, this protocol lacked a negative control, which was needed to prove that those peptides were derived from the ER. Despite multiple efforts being made to generate a TAP1 KO control, which would have been depleted of peptides, it was not possible to generate a HCT116 line with detectable knockout of TAP1. Because subcellular fractionation is known to be a sensitive protocol, being able to understand the true source of the peptides is essential. Without this, this protocol cannot be used to confidently answer biological questions. Despite these limitations, I believe that with defined controls, and defined protein purity thresholds to make analysis more reproducible, this platform does have the ability to be used for testing biological questions, including evaluation of the bottleneck in neoantigen presentation. Understanding the bottleneck in neoantigen presentation could provide additional clinical targets for the increased presentation of neoantigens.

Finally, in Chapter 6 I highlighted the suitability of PDO-immune cell co-culture for the interrogation of biological questions. In the first half of the chapter I developed a 2D long-term CRC PDO-T-cell co-culture, with pre-activated

allogeneic T-cells, to test resistance acquisition to bispecific antibody (BsAb) treatment. The BsAb treatment used was CEA-TCB, a novel IG-based BsAb which targeted two molecules of CEA on one end, and CD3 on T-cells on the other end, which activates TCR-independent signalling. Through this co-culture setup I was able to illustrate the rapid acquisition of resistance to BsAb treatment in PDOs after a period of tumour growth control. This occurred notwithstanding with the addition of anti-PD-L1 antibody atezolizumab, and despite re-treatment with fresh BsAb and pre-activated T-cells. This resistance even persisted after a period of culture without treatment. Some of the resistance could be attributed to strong target downregulation; CRC-07 displayed an approximately 4.5-fold downregulation in the surface abundance of CEA. This is important as studies have shown insensitivity to CEA-TCB when CEA is expressed at a lower surface abundance (Bacac *et al.*, 2016). The resistance could not be explained in the other surviving PDO. This work illustrated the potential of this co-culture setup to research mechanisms of acquired resistance, and it would have been beneficial to use this platform to generate samples for RNA sequencing. This is a promising approach to reveal mechanisms of CEA downregulation, and alternative resistance mechanisms exhibited by the other PDO. However, the long-term cultures were time-intensive, and the first culture for RNA sequencing had to be discarded due to institute closure due to the COVID-19 lockdown. As a result, the lost time and graduated return-to-work protocol meant that I could not continue with this technique. However, this work was able to show that use of allogeneic PDO-T-cell co-cultures can be an effective model for the investigation of acquired immunotherapeutic resistance in CRC.

Chapter 6 also described use of autologous MMRp GOA PDO-PBMC co-culture for the expansion of T-cells for the further assessment of tumour reactivity. I moved to using MMRp GOA PDOs because matched PDO and PBMC samples were only available for these patients. However, developing a model system through which one could investigate whether predicted neoantigens or MS-detected antigens can be recognised by autologous T-cells, was important to help reveal the most clinically relevant antigens. Furthermore, this model system could be used to interrogate how different treatments could promote or enhance the T-cell response. First, MMRp GOA PDOs were established from patient needle core biopsies, and PBMCs were isolated from patient bloods at screening, before

treatment, and before every new cycle of treatment. For tumour reactivity analysis, it was important to minimise the use of the finite PBMCs and expanded T-cells, to allow for them to be preserved for additional analyses. Conversely, it was important to monitor the upregulation of multiple activation markers, as studies have shown that some tumour-reactive T-cells may only demonstrate activation by upregulation of a single marker protein (Chalabi *et al.*, 2020). Therefore, I established a protocol whereby the expanded T-cells would be exposed to PDOs or activated with PMA/ionomycin for 16 hours to allow for the upregulation of cell surface markers. This was followed by an additional 5-hour incubation with the addition of GolgiPLUG and GolgiSTOP to enable cytokine accumulation in the Golgi apparatus and ER to facilitate intracellular staining. This was validated in PBMCs from 5 LRS cones, and was able to resolve upregulation of CD107a, CD137, IFNγ and Granzyme B in activated CD4+ and CD8+ T-cells.

For the MMRp GOA PDO tested, the mutation load of the source tumour was first assessed. Exome sequencing revealed 128 high-confidence mutations, illustrating the low mutation load of the GOA PDO. The mutations included characteristic GI mutations in TP53 and SMAD4, and showed no mutations in APM, PLC, or HLA genes. HLA-I surface abundance was confirmed by flow cytometry, with quantification indicating 495,450 cell surface HLA-I molecules; this was similar to that of HLA-I expression seen in the 5 MMRp CRC PDOs (330,100-496,000) in Chapter 3. The autologous PBMC-PDO co-culture showed that $1x10^6$ viable cells (88.8% CD3+) could be expanded from 1.4 $x10^6$ autologous patient PBMCs over a 2-week PDO-PBMC co-culture. However, the following T-cell reactivity readout illustrated no tumour-specific reactivity from the autologous T-cells. This could not be credited to failed antibody staining, as the PMA/Ionomycin control strongly stained for all the markers assessed: CD107a, IFNγ, and Granzyme B. The absence of a tumour-specific T-cell response suggested either a lack of tumour-specific T-cells in the PBMCs or lack of intrinsic immunogenicity of the PDOs. Furthermore, treatment with atezolizumab and nivolumab provided no additional tumour-reactivity in the expanded T-cells. This showed that whilst T-cell reactivity to PDOs could not be detected in this co-culture, this platform could be taken forwards to assess T-cell reactivity to, and

therefore clinical relevance of, neoantigens predicted by neoantigen prediction and MS-detected neoantigens.

To conclude, I have taken a multidisciplinary approach to the assessment of antigen presentation in MMRp CRCs and MMRp GOAs, and have been able to confirm a sparse neoantigen landscape in MMRp CRCs which may explain the poor sensitivity of these tumours to ICI. Whilst neither IFNγ or trametinib treatment succeeded in boosting canonical neoantigen presentation, I believe the PDO model system and the methodologies described could be useful in the translational setting for the pre-clinical validation of possible treatments intended to increase neoantigen presentation. Furthermore, I have also developed a system for the isolation of ER for the interrogation of the ER peptidome. The protocol was optimised, and the first MS revealed peptides with characteristics expected for the ER peptidome, but further work will now be necessary to provide insight on how the constrained the ER peptidome may be. I have also highlighted specific differences in the IFNγ and non-IFNγ immunopeptidome which provides an argument for having a specific peptide prediction tool trained on IFNγ immunopeptidomics datasets. This could help better predict which peptides were expected to be present or absent in cells exposed to IFNγ. This would be important for cancer vaccine design, as it allows for selection of peptides which are still presented even when T-cell recognition and activation leads to IFNγ secretion. Further to this, I developed a MMRp GOA co-culture T-cell reactivity readout that used a minimal amount of cells to assess T-cell reactivity to PDOs. This could be used to better understand immunogenicity of MMRp GOA PDOs, and also to scrutinise the clinical relevance of neoantigens predicted by prediction algorithms, or detected by MS. The summation of this work emphasises the importance of developing the understanding of neoantigen presentation in advanced MMRp CRCs and MMRp GOAs, so that they might benefit from immunotherapy in the future.

Bibliography.

Abelin, J. G. *et al.* (2017) 'Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction', *Immunity*. Immunity, 46(2), pp. 315–326. doi: 10.1016/j.immuni.2017.02.007.

Adams, J. M. and Cory, S. (2007) 'The Bcl-2 apoptotic switch in cancer development and therapy', *Oncogene*, 26(9), pp. 1324–1337. doi: 10.1038/sj.onc.1210220.

Aldea, M. *et al.* (2021) 'Overcoming Resistance to Tumor-Targeted and Immune-Targeted Therapies', *Cancer Discovery*. American Association for Cancer Research, 11(4), pp. 874–899. doi: 10.1158/2159-8290.CD-20-1638.

Alvarez-Navarro, C. *et al.* (2015) 'Endoplasmic Reticulum Aminopeptidase 1 (ERAP1) Polymorphism Relevant to Inflammatory Disease Shapes the Peptidome of the Birdshot Chorioretinopathy-Associated HLA-A*29:02 Antigen.', *Molecular & cellular proteomics : MCP*. United States, 14(7), pp. 1770–1780. doi: 10.1074/mcp.M115.048959.

Anagnostou, V. *et al.* (2017) 'Evolution of neoantigen landscape during immune checkpoint blockade in non-small cell lung cancer', *Cancer Discovery*. American Association for Cancer Research Inc., 7(3), pp. 264–276. doi: 10.1158/2159-8290.CD-16-0828/333315/AM/EVOLUTION-OF-NEOANTIGEN-LANDSCAPE-DURING-IMMUNE.

Anderson, P. *et al.* (2021) 'HLA class I loss in colorectal cancer: implications for immune escape and immunotherapy', *Cellular and Molecular Immunology*. Nature Publishing Group, 18(3), p. 556. doi: 10.1038/S41423-021-00634-7.

André, T. *et al.* (2020) 'Pembrolizumab in Microsatellite-Instability–High Advanced Colorectal Cancer', *New England Journal of Medicine*, 383(23), pp. 2207–2218. doi: 10.1056/NEJMoa2017699.

Apcher, S. *et al.* (2011) 'Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(28), pp. 11572–11577. doi: 10.1073/PNAS.1104104108/SUPPL_FILE/PNAS.201104104SI.PDF.

Apps, R. *et al.* (2015) 'Relative Expression Levels of the HLA Class-I Proteins in Normal and HIV-Infected Cells', *The Journal of Immunology*. doi: 10.4049/jimmunol.1403234.

Aranha, M. P. *et al.* (2020) 'Combining Three-Dimensional Modeling with Artificial

Intelligence to Increase Specificity and Precision in Peptide-MHC Binding Predictions', *Journal of immunology (Baltimore, Md. : 1950)*. J Immunol, 205(7), pp. 1962–1977. doi: 10.4049/JIMMUNOL.1900918.

Arellano-Garcia, M. E. *et al.* (2014) 'Interferon-γ induces immunoproteasomes and the presentation of MHC I-associated peptides on human salivary gland cells', *PLoS ONE*. PLoS One, 9(8). doi: 10.1371/journal.pone.0102878.

Arenas, E. J. *et al.* (2021) 'Acquired cancer cell resistance to T cell bispecific antibodies and CAR T targeting HER2 through JAK2 down-modulation', *Nature Communications*. Nature Publishing Group, 12(1). doi: 10.1038/S41467-021-21445-4.

Armaghany, T. *et al.* (2012) 'Genetic alterations in colorectal cancer', *Gastrointestinal Cancer Research*.

Athauda, A. *et al.* (2021) 'Perioperative FLOT plus anti-PD-L1 avelumab (FLOT-A) in resectable oesophagogastric adenocarcinoma (OGA): Interim safety analysis results from the ICONIC trial.', *https://doi.org/10.1200/JCO.2021.39.3_suppl.201*. American Society of Clinical Oncology, 39(3_suppl), pp. 201–201. doi: 10.1200/JCO.2021.39.3_SUPPL.201.

Ayers, M. *et al.* (2017) 'IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade', *Journal of Clinical Investigation*. American Society for Clinical Investigation, 127(8), pp. 2930–2940. doi: 10.1172/JCI91190.

Baban, B. *et al.* (2009) 'IDO activates regulatory T cells and blocks their conversion into TH17-like T cells', *Journal of immunology (Baltimore, Md. : 1950)*. NIH Public Access, 183(4), p. 2475. doi: 10.4049/JIMMUNOL.0900986.

Bacac, M. *et al.* (2016) 'A novel carcinoembryonic antigen T-cell bispecific antibody (CEA TCB) for the treatment of solid tumors', *Clinical Cancer Research*. doi: 10.1158/1078-0432.CCR-15-1696.

Bailey, A. *et al.* (2014) 'Two Polymorphisms Facilitate Differences in Plasticity between Two Chicken Major Histocompatibility Complex Class I Proteins', *PLOS ONE*. Public Library of Science, 9(2), p. e89657. doi: 10.1371/JOURNAL.PONE.0089657.

Bailey, A. *et al.* (2015) 'Selector function of MHC I molecules is determined by protein plasticity', *Scientific Reports 2015 5:1*. Nature Publishing Group, 5(1), pp. 1–15. doi: 10.1038/srep14928.

Balachandran, V. P. *et al.* (2017) 'Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer', *Nature 2017 551:7681*. Nature

Publishing Group, 551(7681), pp. 512–516. doi: 10.1038/nature24462.

Bar-Ephraim, Y. E., Kretzschmar, K. and Clevers, H. (2020) 'Organoids in immunological research', *Nature Reviews Immunology*, 20(5), pp. 279–293. doi: 10.1038/s41577-019-0248-y.

Bassani-Sternberg, M. *et al.* (2015) 'Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation', *Molecular & cellular proteomics : MCP*. Mol Cell Proteomics, 14(3), pp. 658–673. doi: 10.1074/MCP.M114.042812.

Bassani-Sternberg, M. *et al.* (2016) 'Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry', *Nature Communications*. doi: 10.1038/ncomms13404.

Bassani-Sternberg, M. and Coukos, G. (2016) 'Mass spectrometry-based antigen discovery for cancer immunotherapy', *Current Opinion in Immunology*, 41, pp. 9–17. doi: 10.1016/j.coi.2016.04.005.

Bedran, G. *et al.* (2022) 'The immunopeptidome from a genomic perspective: Establishing immune-relevant regions for cancer vaccine design', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2022.01.13.475872. doi: 10.1101/2022.01.13.475872.

Bedran, G., Gasser, H. C., *et al.* (2023) 'The Immunopeptidome from a Genomic Perspective: Establishing the Noncanonical Landscape of MHC Class I–Associated Peptides', *Cancer Immunology Research*. American Association for Cancer Research, 11(6), p. 747. doi: 10.1158/2326-6066.CIR-22-0621.

Bedran, G., Polasky, D. A., *et al.* (2023) 'Unraveling the glycosylated immunopeptidome with HLA-Glyco', *Nature Communications 2023 14:1*. Nature Publishing Group, 14(1), pp. 1–12. doi: 10.1038/s41467-023-39270-2.

Beninga, J., Rock, K. L. and Goldberg, A. L. (1998) 'Interferon-γ can stimulate post-proteasomal trimming of the N terminus of an antigenic peptide by inducing leucine aminopeptidase', *The Journal of biological chemistry*. United States: Elsevier, 273(30), pp. 18734–18742. doi: 10.1074/jbc.273.30.18734.

Besche, H. C., Peth, A. and Goldberg, A. L. (2009) 'Getting to First Base in Proteasome Assembly', *Cell*. Elsevier B.V., 138(1), pp. 25–28. doi: 10.1016/j.cell.2009.06.035.

Bettoni, F. *et al.* (2009) 'Identification of FAM46D as a novel cancer/testis antigen using EST data and serological analysis', *Genomics*. doi: 10.1016/j.ygeno.2009.06.001.

Blasco, M. A. (2005) ‘Telomeres and human disease: ageing, cancer and beyond.’, *Nature reviews. Genetics*. England, 6(8), pp. 611–622. doi: 10.1038/nrg1656.

Blass, E. and Ott, P. A. (2021) ‘Advances in the development of personalized neoantigen-based therapeutic cancer vaccines’, *Nature Reviews Clinical Oncology 2021 18:4*. Nature Publishing Group, 18(4), pp. 215–229. doi: 10.1038/s41571-020-00460-2.

Bohner, P. *et al.* (2019) ‘Double positive CD4+CD8+ T cells are enriched in urological cancers and favor T helper-2 polarization’, *Frontiers in Immunology*. Frontiers Media S.A., 10(MAR), p. 622. doi: 10.3389/FIMMU.2019.00622/BIBTEX.

Bonaventura, P. *et al.* (2019) ‘Cold tumors: A therapeutic challenge for immunotherapy’, *Frontiers in Immunology*. Frontiers Media S.A., 10(FEB), p. 168. doi: 10.3389/FIMMU.2019.00168/BIBTEX.

Booij, T. H., Cattaneo, C. M. and Hirt, C. K. (2022) ‘Tumor Organoids as a Research Tool: How to Exploit Them’, *Cells*. Multidisciplinary Digital Publishing Institute (MDPI), 11(21). doi: 10.3390/CELLS11213440.

Boulanger, D. S. M. *et al.* (2022) ‘Tapasin-mediated editing of the MHC I immunopeptidome is epitope specific and dependent on peptide off-rate, abundance, and level of tapasin expression’. doi: 10.3389/fimmu.2022.956603.

Bourdetsky, D., Schmelzer, C. E. H. H. and Admon, A. (2014) ‘The nature and extent of contributions by defective ribosome products to the HLA peptidome’, *Proceedings of the National Academy of Sciences*. Proc Natl Acad Sci U S A, 111(16), p. E1591 LP-E1599. doi: 10.1073/pnas.1321902111.

Boyle, L. H. *et al.* (2013) ‘Tapasin-related protein TAPBPR is an additional component of the MHC class I presentation pathway.’, *Proceedings of the National Academy of Sciences of the United States of America*. United States, 110(9), pp. 3465–3470. doi: 10.1073/pnas.1222342110.

Bradner, J. E., Hnisz, D. and Young, R. A. (2017) ‘Transcriptional Addiction in Cancer’, *Cell*. doi: 10.1016/j.cell.2016.12.013.

Brandt, I., Scharpé, S. and Lambeir, A. M. (2007) ‘Suggested functions for prolyl oligopeptidase: A puzzling paradox’, *Clinica Chimica Acta*. Elsevier, 377(1–2), pp. 50–61. doi: 10.1016/J.CCA.2006.09.001.

Brody, J. R. *et al.* (2009) ‘Expression of indoleamine 2,3-dioxygenase in metastatic malignant melanoma recruits regulatory T cells to avoid immune

detection and affects survival', *Cell Cycle*. Taylor and Francis Inc., 8(12), pp. 1930–1934. doi: 10.4161/CC.8.12.8745.

Brossart, P. (2020) 'The role of antigen spreading in the efficacy of immunotherapies', *Clinical Cancer Research*. American Association for Cancer Research Inc., 26(17), pp. 4442–4447. doi: 10.1158/1078-0432.CCR-20-0305/77058/AM/THE-ROLE-OF-ANTIGEN-SPREADING-IN-THE-EFFICACY-OF.

Van Der Bruggen, P. *et al.* (1991) 'A Gene Encoding an Antigen Recognized by Cytolytic T Lymphocytes on a Human Melanoma', *Science*. American Association for the Advancement of Science, 254(5038), pp. 1643–1647. doi: 10.1126/SCIENCE.1840703.

Bulik-Sullivan, B. *et al.* (2018) 'Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification', *Nature Biotechnology 2018 37:1*. Nature Publishing Group, 37(1), pp. 55–63. doi: 10.1038/nbt.4313.

Bullock, A. *et al.* (2022) 'LBA O-9 Botensilimab, a novel innate/adaptive immune activator, plus balstilimab (anti-PD-1) for metastatic heavily pretreated microsatellite stable colorectal cancer', *Annals of Oncology*. Elsevier BV, 33, p. S376. doi: 10.1016/j.annonc.2022.04.453.

Burdon, R. H., van Knippenberg, P. H. P. and Sharpe, P. T. (1988) 'Centrifugation', in *Laboratory Techniques in Biochemistry and Molecular Biology*. Elsevier, pp. 18–69. doi: 10.1016/S0075-7535(08)70628-4.

Burke, F. *et al.* (1997) 'Interferon gamma induces cell cycle arrest and apoptosis in a model of ovarian cancer: Enhancement of effect by batimastat', *European Journal of Cancer*. Pergamon, 33(7), pp. 1114–1121. doi: 10.1016/S0959-8049(97)88065-3.

Burkhart, D. L. and Sage, J. (2008) 'Cellular mechanisms of tumour suppression by the retinoblastoma gene', *Nature Reviews Cancer 2008 8:9*. Nature Publishing Group, 8(9), pp. 671–682. doi: 10.1038/nrc2399.

Burks, J., Reed, R. E. and Desai, S. D. (2015) 'Free ISG15 triggers an antitumor immune response against breast cancer: a new perspective', *Oncotarget*. Impact Journals, LLC, 6(9), p. 7221. doi: 10.18632/ONCOTARGET.3372.

Burnet, M. (1957) 'Cancer: a biological approach', *British medical journal*, 1(5023), pp. 841–847. doi: 10.1136/bmj.1.5023.841.

Burton, P. R. *et al.* (2007) 'Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants.', *Nature genetics*. United States,

39(11), pp. 1329–1337. doi: 10.1038/ng.2007.17.
Butte, M. J. *et al.* (2007) 'Programmed death-1 ligand 1 interacts specifically with the B7-1 costimulatory molecule to inhibit T cell responses.', *Immunity*. United States, 27(1), pp. 111–122. doi: 10.1016/j.immuni.2007.05.016.
Cai, H. *et al.* (2019) 'Mutational landscape of gastric cancer and clinical application of genomic profiling based on target next-generation sequencing', *Journal of Translational Medicine*. BioMed Central Ltd., 17(1), pp. 1–12. doi: 10.1186/S12967-019-1941-0/FIGURES/6.
Caramelo, J. J. *et al.* (2004) 'The endoplasmic reticulum glucosyltransferase recognizes nearly native glycoprotein folding intermediates.', *The Journal of biological chemistry*. United States, 279(44), pp. 46280–46285. doi: 10.1074/jbc.M408404200.
Carcas, L. P. (2014) 'Gastric cancer review', *Journal of Carcinogenesis*. Wolters Kluwer -- Medknow Publications, 13(1), p. 14. doi: 10.4103/1477-3163.146506.
Carmeliet, P. (2005) 'VEGF as a key mediator of angiogenesis in cancer', *Oncology*. Oncology, 69 Suppl 3(SUPPL. 3), pp. 4–10. doi: 10.1159/000088478.
Caron, E. *et al.* (2011) 'The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation', *Molecular systems biology*. Mol Syst Biol, 7. doi: 10.1038/MSB.2011.68.
Carreno, B. M. *et al.* (2015) 'A dendritic cell vaccine increases the breadth and diversity ofmelanoma neoantigen-specific T cells', *Science (New York, N.Y.)*. NIH Public Access, 348(6236), p. 803. doi: 10.1126/SCIENCE.AAA3828.
Cartwright, E. *et al.* (2021) '443P EMERGE: A phase II trial assessing the efficacy of domatinostat plus avelumab in patients with previously treated advanced mismatch repair proficient oesophagogastric and colorectal cancers – phase IIA dose finding', *Annals of Oncology*, 32, pp. S555–S556.
Cattaneo, C. M. *et al.* (2020) 'Tumor organoid–T-cell coculture systems', *Nature Protocols*. doi: 10.1038/s41596-019-0232-9.
Cavazza, T. *et al.* (2021) 'Parental genome unification is highly error-prone in mammalian embryos', *Cell*. Cell Press, 184(11), pp. 2860-2877.e22. doi: 10.1016/J.CELL.2021.04.013.
Cervantes, A. *et al.* (2022) 'Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up', *Annals of Oncology*. Elsevier Ltd, 34(1), pp. 10–32. doi: 10.1016/J.ANNONC.2022.10.003/ATTACHMENT/5A57CC70-CCE4-401B-

A657-C12E03275943/MMC1.PDF.

Chalabi, M. *et al.* (2020) 'Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers', *Nature Medicine*, 26(4), pp. 566–576. doi: 10.1038/s41591-020-0805-8.

Chang, S. C. *et al.* (2005) 'The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0500721102.

Charles A Janeway, J. *et al.* (2001) *Immunobiology: The Immune System in Health and Disease*. 5th edn. Garland Science.

Chen, D. S. and Mellman, I. (2013) 'Oncology meets immunology: The cancer-immunity cycle', *Immunity*. doi: 10.1016/j.immuni.2013.07.012.

Chen, H. *et al.* (2016) 'ERAP1-ERAP2 dimers trim MHC I-bound precursor peptides; implications for understanding peptide editing', *Scientific Reports*, 6(1), p. 28902. doi: 10.1038/srep28902.

Chen, X. *et al.* (2010) 'Quantitative organellar proteomics analysis of rough endoplasmic reticulum from normal and acute pancreatitis rat pancreas', *Journal of Proteome Research*. NIH Public Access, 9(2), p. 885. doi: 10.1021/pr900784c.

Chkhaidze, K. *et al.* (2019) 'Spatially constrained tumour growth affects the patterns of clonal selection and neutral drift in cancer genomic data', *PLOS Computational Biology*. Public Library of Science, 15(7), p. e1007243. doi: 10.1371/JOURNAL.PCBI.1007243.

Chong, C. *et al.* (2017) 'High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferonγ-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome', *Molecular & Cellular Proteomics*. American Society for Biochemistry and Molecular Biology, 17(3), p. 533. doi: 10.1074/mcp.tir117.000383.

Chong, C., Coukos, G. and Bassani-Sternberg, M. (2022) 'Identification of tumor antigens with immunopeptidomics', *Nature biotechnology*. Nat Biotechnol, 40(2), pp. 175–188. doi: 10.1038/S41587-021-01038-8.

Cioni, B. *et al.* (2019) 'HLA class II expression on tumor cells and low numbers of tumor-associated macrophages predict clinical outcome in oropharyngeal cancer', *Head & Neck*. John Wiley & Sons, Ltd, 41(2), pp. 463–478. doi: 10.1002/HED.25442.

Clénet, M. L. *et al.* (2017) 'Peripheral human CD4+CD8+ T lymphocytes exhibit

a memory phenotype and enhanced responses to IL-2, IL-7 and IL-15', *Scientific Reports 2017 7:1*. Nature Publishing Group, 7(1), pp. 1–15. doi: 10.1038/s41598-017-11926-2.

Cooper, G. M. (2000) 'The Nuclear Envelope and Traffic between the Nucleus and Cytoplasm', in *The Cell: A Molecular Approach*. 2nd edn. Sinauer Associates.

Corbière, V. *et al.* (2011) 'Antigen spreading contributes to MAGE vaccination-induced regression of melanoma metastases', *Cancer research*. Cancer Res, 71(4), pp. 1253–1262. doi: 10.1158/0008-5472.CAN-10-2693.

Cremolini, C. *et al.* (2015) 'FOLFOXIRI plus bevacizumab versus FOLFIRI plus bevacizumab as first-line treatment of patients with metastatic colorectal cancer: updated overall survival and molecular subgroup analyses of the open-label, phase 3 TRIBE study', *The Lancet. Oncology*. Lancet Oncol, 16(13), pp. 1306–1315. doi: 10.1016/S1470-2045(15)00122-9.

Curto, M. *et al.* (2007) 'Contact-dependent inhibition of EGFR signaling by Nf2/Merlin.', *The Journal of cell biology*. United States, 177(5), pp. 893–903. doi: 10.1083/jcb.200703010.

Der, S. D. *et al.* (1998) 'Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays', *Proceedings of the National Academy of Sciences*, 95(26), pp. 15623–15628. doi: 10.1073/pnas.95.26.15623.

Dhatchinamoorthy, K., Colbert, J. D. and Rock, K. L. (2021) 'Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation', *Frontiers in Immunology*. Frontiers Media S.A., 12, p. 469. doi: 10.3389/FIMMU.2021.636568/BIBTEX.

Dighe, A. S. *et al.* (1994) 'Enhanced in vivo growth and resistance to rejection of tumor cells expressing dominant negative IFNγ receptors', *Immunity*. doi: 10.1016/1074-7613(94)90087-6.

Dijkstra, K. K. *et al.* (2018) 'Generation of Tumor-Reactive T Cells by Co-culture of Peripheral Blood Lymphocytes and Tumor Organoids', *Cell*. doi: 10.1016/j.cell.2018.07.009.

Discontinuation Domatinostat Program - 4SC AG (2022). Available at: https://www.4sc.com/news/4sc-ag-discontinuation-domatinostat-program-2/ (Accessed: 10 November 2022).

Disis, M. L. *et al.* (2004) 'Humoral Epitope-Spreading Following Immunization

with a HER-2/neu Peptide Based Vaccine in Cancer Patients', *Journal of Clinical Immunology*. Springer, 24(5), pp. 571–578. doi: 10.1023/B:JOCI.0000040928.67495.52/METRICS.

Driscoll, J. *et al.* (1993) 'MHC-linked LMP gene products specifically alter peptidase activities of the proteasome.', *Nature*. England, 365(6443), pp. 262–264. doi: 10.1038/365262a0.

Drost, J. and Clevers, H. (2018) 'Organoids in cancer research', *Nature Reviews Cancer*. Nature Publishing Group, 18(7), pp. 407–418. doi: 10.1038/s41568-018-0007-6.

Duan, F. *et al.* (2014) 'Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity', *The Journal of Experimental Medicine*. The Rockefeller University Press, 211(11), p. 2231. doi: 10.1084/JEM.20141308.

Duncan, T. J. *et al.* (2007) 'Loss of IFNγ Receptor Is an Independent Prognostic Factor in Ovarian Cancer', *Clinical Cancer Research*. American Association for Cancer Research, 13(14), pp. 4139–4145. doi: 10.1158/1078-0432.CCR-06-2833.

Dunkley, T. P. J. *et al.* (2004) 'Localization of organelle proteins by isotope tagging (LOPIT)', *Molecular and Cellular Proteomics*. Elsevier, 3(11), pp. 1128–1134. doi: 10.1074/mcp.T400009-MCP200.

Dunn, G. P. *et al.* (2002) 'Cancer immunoediting: from immunosurveillance to tumor escape', *Nature Immunology*, 3(11), pp. 991–998. doi: 10.1038/ni1102-991.

Dunn, G. P., Old, L. J. and Schreiber, R. D. (2004) 'The Three Es of Cancer Immunoediting', *Annual Review of Immunology*. doi: 10.1146/annurev.immunol.22.012703.104803.

Ebert, P. J. R. *et al.* (2016) 'MAP Kinase Inhibition Promotes T Cell and Anti-tumor Activity in Combination with PD-L1 Checkpoint Blockade', *Immunity*. doi: 10.1016/j.immuni.2016.01.024.

Eggermont, A. M. M. *et al.* (2018) 'Adjuvant Pembrolizumab versus Placebo in Resected Stage III Melanoma', *New England Journal of Medicine*, 378(19), pp. 1789–1801. doi: 10.1056/NEJMoa1802357.

van Endert, P. (2008) 'Role of tripeptidyl peptidase II in MHC class I antigen processing – the end of controversies?', *European Journal of Immunology*. John Wiley & Sons, Ltd, 38(3), pp. 609–613. doi: 10.1002/EJI.200838181.

Van Endert, P. (2011) ‘Post-proteasomal and proteasome-independent generation of MHC class I ligands’, *Cellular and Molecular Life Sciences*. Springer, 68(9), pp. 1553–1567. doi: 10.1007/S00018-011-0662-1/METRICS.

Eng, C. *et al.* (2019) ‘Atezolizumab with or without cobimetinib versus regorafenib in previously treated metastatic colorectal cancer (IMblaze370): a multicentre, open-label, phase 3, randomised, controlled trial’, *The Lancet Oncology*. doi: 10.1016/S1470-2045(19)30027-0.

Evans, D. M. *et al.* (2011) ‘Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility.’, *Nature genetics*. United States, 43(8), pp. 761–767. doi: 10.1038/ng.873.

Evnouchidou, I. *et al.* (2008) ‘The internal sequence of the peptide-substrate determines its N-terminus trimming by ERAP1.’, *PloS one*. United States, 3(11), p. e3658. doi: 10.1371/journal.pone.0003658.

Evnouchidou, I. *et al.* (2012) ‘A common single nucleotide polymorphism in endoplasmic reticulum aminopeptidase 2 induces a specificity switch that leads to altered antigen processing.’, *Journal of immunology (Baltimore, Md. : 1950)*. United States, 189(5), pp. 2383–2392. doi: 10.4049/jimmunol.1200918.

Fearon, E. R. and Vogelstein, B. (1990) ‘A genetic model for colorectal tumorigenesis’, *Cell*. Cell, 61(5), pp. 759–767. doi: 10.1016/0092-8674(90)90186-I.

Fehling, H. J. *et al.* (1994) ‘MHC class I expression in mice lacking the proteasome subunit LMP-7’, *Science (New York, N.Y.)*. Science, 265(5176), pp. 1234–1237. doi: 10.1126/SCIENCE.8066463.

Fehrenbacher, L. *et al.* (2016) ‘Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): A multicentre, open-label, phase 2 randomised controlled trial’, *The Lancet*. Lancet Publishing Group, 387(10030), pp. 1837–1846. doi: 10.1016/S0140-6736(16)00587-0.

Ferlay, J. *et al.* (2015) ‘Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012’, *International journal of cancer*. Int J Cancer, 136(5), pp. E359–E386. doi: 10.1002/IJC.29210.

Ferrington, D. A. and Gregerson, D. S. (2012) ‘Immunoproteasomes: structure, function, and antigen presentation.’, *Progress in molecular biology and translational science*. NIH Public Access, 109, p. 75. doi: 10.1016/B978-0-12-397863-9.00003-1.

Ferro, A. *et al.* (2014) ‘Worldwide trends in gastric cancer mortality (1980-2011), with predictions to 2015, and incidence by subtype’, *European journal of cancer (Oxford, England : 1990)*. Eur J Cancer, 50(7), pp. 1330–1344. doi: 10.1016/J.EJCA.2014.01.029.

Ferro, E. S., Gewehr, M. C. F. and Navon, A. (2020) ‘Thimet Oligopeptidase Biochemical and Biological Significances: Past, Present, and Future Directions’, *Biomolecules 2020, Vol. 10, Page 1229*. Multidisciplinary Digital Publishing Institute, 10(9), p. 1229. doi: 10.3390/BIOM10091229.

Fidler, I. J. (2003) ‘The pathogenesis of cancer metastasis: the “seed and soil” hypothesis revisited.’, *Nature reviews. Cancer*. England, 3(6), pp. 453–458. doi: 10.1038/nrc1098.

Fleischmann, G. *et al.* (2015) ‘Mechanistic Basis for Epitope Proofreading in the Peptide-Loading Complex’, *Journal of immunology (Baltimore, Md. : 1950)*. J Immunol, 195(9), pp. 4503–4513. doi: 10.4049/JIMMUNOL.1501515.

Folkins, A. K. *et al.* (2008) ‘A candidate precursor to pelvic serous cancer (p53 signature) and its prevalence in ovaries and fallopian tubes from women with BRCA mutations’, *Gynecologic oncology*. Gynecol Oncol, 109(2), pp. 168–173. doi: 10.1016/J.YGYNO.2008.01.012.

Förster, A., Whitby, F. G. and Hill, C. P. (2003) ‘The pore of activated 20S proteasomes has an ordered 7-fold symmetric conformation’, *The EMBO Journal*. European Molecular Biology Organization, 22(17), p. 4356. doi: 10.1093/EMBOJ/CDG436.

Foster, L. J. *et al.* (2006) ‘A Mammalian Organelle Map by Protein Correlation Profiling’, *Cell*. Elsevier B.V., 125(1), pp. 187–199. doi: 10.1016/j.cell.2006.03.022.

Freeman, G. J. *et al.* (2000) ‘Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation.’, *The Journal of experimental medicine*. United States, 192(7), pp. 1027–1034. doi: 10.1084/jem.192.7.1027.

Frei, M. (2012) ‘Centrifugation Basics Density Gradient Media Cell Viability and Proliferation Organelle Isolation’, *Biofiles*, 6, pp. 6–7.

Früh, K. and Yang, Y. (1999) ‘Antigen presentation by MHC class I and its regulation by interferon γ’, *Current Opinion in Immunology*. Curr Opin Immunol, 11(1), pp. 76–81. doi: 10.1016/S0952-7915(99)80014-4.

Gagné, F. (2014) *Tissue Preparation and Subcellular Fractionation Techniques*,

Biochemical Ecotoxicology: Principles and Methods. Academic Press. doi: 10.1016/B978-0-12-411604-7.00002-7.

Gandhi, L. *et al.* (2018) 'Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer', *New England Journal of Medicine*, 378(22), pp. 2078–2092. doi: 10.1056/NEJMoa1801005.

Gao, J. *et al.* (2016) 'Loss of IFN-γ Pathway Genes in Tumor Cells as a Mechanism of Resistance to Anti-CTLA-4 Therapy', *Cell*. doi: 10.1016/j.cell.2016.08.069.

García-Horsman, J. A., Männistö, P. T. and Venäläinen, J. I. (2007) 'On the role of prolyl oligopeptidase in health and disease', *Neuropeptides*. Churchill Livingstone, 41(1), pp. 1–24. doi: 10.1016/J.NPEP.2006.10.004.

Garralda, E. *et al.* (2021) 'A phase 1 first-in-human study of the anti-LAG-3 antibody MK4280 (favezelimab) plus pembrolizumab in previously treated, advanced microsatellite stable colorectal cancer.', *Journal of Clinical Oncology*, 39(15_suppl), p. 3584. doi: 10.1200/JCO.2021.39.15_suppl.3584.

Geiss-Friedlander, R. *et al.* (2009) 'The cytoplasmic peptidase DPP9 is rate-limiting for degradation of proline-containing peptides', *Journal of Biological Chemistry*. Elsevier, 284(40), pp. 27211–27219. doi: 10.1074/jbc.M109.041871.

Gerlinger, M. *et al.* (2012) 'Intratumor heterogeneity and branched evolution revealed by multiregion sequencing', *The New England journal of medicine*. N Engl J Med, 366(10), pp. 883–892. doi: 10.1056/NEJMOA1113205.

Gerlinger, M. *et al.* (2014) 'Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing', *Nature genetics*. Nat Genet, 46(3), pp. 225–233. doi: 10.1038/NG.2891.

Gerstung, M. *et al.* (2020) 'The evolutionary history of 2,658 cancers', *Nature 2020 578:7793*. Nature Publishing Group, 578(7793), pp. 122–128. doi: 10.1038/s41586-019-1907-7.

Gfeller, D. and Bassani-Sternberg, M. (2018) 'Predicting antigen presentation-What could we learn from a million peptides?', *Frontiers in Immunology*. Frontiers Media S.A., 9(JUL), p. 1. doi: 10.3389/FIMMU.2018.01716/FULL.

Ghanem, E. *et al.* (2010) 'The transporter associated with antigen processing (TAP) is active in a post-ER compartment', *Journal of Cell Science*. doi: 10.1242/jcs.060632.

Girdlestone, J. *et al.* (1993) 'Transcriptional regulation of HLA-A and-B: Differential binding of members of the Rel and IRF families of transcription

factors', *Proc. Natl. Acad. Sci. USA*, 90, pp. 11568–11572.

Glaire, M. A. *et al.* (2022) 'Discordant prognosis of mismatch repair deficiency in colorectal and endometrial cancer reflects variation in antitumour immune response and immune escape', *The Journal of Pathology*. John Wiley & Sons, Ltd, 257(3), pp. 340–351. doi: 10.1002/PATH.5894.

Gomes, A. P. *et al.* (2019) 'Dynamic Incorporation of Histone H3 Variants into Chromatin Is Essential for Acquisition of Aggressive Traits and Metastatic Colonization', *Cancer cell*. Cancer Cell, 36(4), pp. 402-417.e13. doi: 10.1016/J.CCELL.2019.08.006.

Goncalves, G. *et al.* (2021) 'IFNγ Modulates the Immunopeptidome of Triple Negative Breast Cancer Cells by Enhancing and Diversifying Antigen Processing and Presentation', *Frontiers in Immunology*. doi: 10.3389/fimmu.2021.645770.

Gonzalez-Exposito, R. *et al.* (2019) 'CEA expression heterogeneity and plasticity confer resistance to the CEA-targeting bispecific immunotherapy antibody cibisatamab (CEA-TCB) in patient-derived colorectal cancer organoids', *Journal for ImmunoTherapy of Cancer*. doi: 10.1186/s40425-019-0575-3.

Govindarajan, K. R. *et al.* (2003) 'MPID: MHC-Peptide Interaction Database for sequence-structure-function information on peptides binding to MHC molecules', *Bioinformatics (Oxford, England)*. Bioinformatics, 19(2), pp. 309–310. doi: 10.1093/BIOINFORMATICS/19.2.309.

Grasso, C. S. *et al.* (2018) 'Genetic mechanisms of immune evasion in colorectal cancer', *Cancer Discovery*. doi: 10.1158/2159-8290.CD-17-1327.

Grimes, D. R. *et al.* (2020) 'Evidence for hypoxia increasing the tempo of evolution in glioblastoma', *British Journal of Cancer 2020 123:10*. Nature Publishing Group, 123(10), pp. 1562–1569. doi: 10.1038/s41416-020-1021-5.

Grossmann, N. *et al.* (2014) 'Mechanistic determinants of the directionality and energetics of active export by a heterodimeric ABC transporter', *Nature Communications 2014 5:1*. Nature Publishing Group, 5(1), pp. 1–10. doi: 10.1038/ncomms6419.

Guasp, P. *et al.* (2016) 'The Peptidome of Behçet's Disease-Associated HLA-B*51:01 Includes Two Subpeptidomes Differentially Shaped by Endoplasmic Reticulum Aminopeptidase 1.', *Arthritis & rheumatology (Hoboken, N.J.)*. United States, 68(2), pp. 505–515. doi: 10.1002/art.39430.

Hammarström, S. (1999) 'The carcinoembryonic antigen (CEA) family: Structures, suggested functions and expression in normal and malignant tissues',

Seminars in Cancer Biology. doi: 10.1006/scbi.1998.0119.

Hammond, C., Braakman, I. and Helenius, A. (1994) 'Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 91(3), pp. 913–917. doi: 10.1073/pnas.91.3.913.

Hammond, J. C. *et al.* (2012) 'Ampa Receptor Subunit Expression in the Endoplasmic Reticulum in Frontal Cortex of Elderly Patients with Schizophrenia', *PLoS ONE*. Public Library of Science, 7(6), p. 39190. doi: 10.1371/JOURNAL.PONE.0039190.

Hanahan, D. and Weinberg, R. A. (2000) 'The Hallmarks of Cancer', *Cell*. Elsevier, 100(1), pp. 57–70. doi: 10.1016/S0092-8674(00)81683-9.

Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of Cancer: The Next Generation', *Cell*. Elsevier, 144(5), pp. 646–674. doi: 10.1016/j.cell.2011.02.013.

Hanson, A. L. *et al.* (2018) 'Genetic Variants in ERAP1 and ERAP2 Associated With Immune-Mediated Diseases Influence Protein Expression and the Isoform Profile.', *Arthritis & rheumatology (Hoboken, N.J.)*. United States, 70(2), pp. 255–265. doi: 10.1002/art.40369.

Harris, J. L. *et al.* (2001) 'Substrate specificity of the human proteasome', *Chemistry & Biology*. Cell Press, 8(12), pp. 1131–1141. doi: 10.1016/S1074-5521(01)00080-1.

Hassel, J. C. *et al.* (2021) 'Results from the phase Ib of the SENSITIZE trial combining domatinostat with pembrolizumab in advanced melanoma patients refractory to prior checkpoint inhibitor therapy.', *https://doi.org/10.1200/JCO.2021.39.15_suppl.9545*. Wolters Kluwer Health, 39(15_suppl), pp. 9545–9545. doi: 10.1200/JCO.2021.39.15_SUPPL.9545.

Hearn, A., York, I. A. and Rock, K. L. (2009) 'The Specificity of Trimming of MHC Class I-Presented Peptides in the Endoplasmic Reticulum', *The Journal of Immunology*. American Association of Immunologists, 183(9), pp. 5526–5536. doi: 10.4049/jimmunol.0803663.

Heckler, M. and Dougan, S. K. (2018) 'Unmasking Pancreatic Cancer: Epitope Spreading After Single Antigen Chimeric Antigen Receptor T-Cell Therapy in a Human Phase I Trial', *Gastroenterology*. W.B. Saunders, 155(1), pp. 11–14. doi: 10.1053/j.gastro.2018.06.023.

Heiden, M. G. Vander, Cantley, L. C. and Thompson, C. B. (2009) 'Understanding

the Warburg Effect: The Metabolic Requirements of Cell Proliferation', *Science*, 324(5930), pp. 1029–1033. doi: 10.1126/science.1160809.

Held, T. *et al.* (2021) 'Evidence for an involvement of the ubiquitin-like modifier ISG15 in MHC class I antigen presentation', *European Journal of Immunology*. John Wiley & Sons, Ltd, 51(1), pp. 138–150. doi: 10.1002/EJI.202048646.

Hérin, M. *et al.* (1987) 'Production of stable cytolytic T-cell clones directed against autologous human melanoma', *International Journal of Cancer*, 39(3), pp. 390–396. doi: 10.1002/IJC.2910390320.

Hermann, C. *et al.* (2015) 'TAPBPR alters MHC class I peptide presentation by functioning as a peptide exchange catalyst.', *eLife*. England, 4. doi: 10.7554/eLife.09617.

Hilf, N. *et al.* (2018) 'Actively personalized vaccination trial for newly diagnosed glioblastoma', *Nature 2018 565:7738*. Nature Publishing Group, 565(7738), pp. 240–245. doi: 10.1038/s41586-018-0810-y.

Holch, Stintzing and Heinemann (2016) 'Treatment of Metastatic Colorectal Cancer: Standard of Care and Future Perspectives', *Visceral Medicine*, 32, pp. 178–183. doi: 10.1159/000446052.

Hongo, A. *et al.* (2019) 'Upstream Position of Proline Defines Peptide–HLA Class I Repertoire Formation and CD8+ T Cell Responses', *The Journal of Immunology*. American Association of Immunologists, 202(10), pp. 2849–2855. doi: 10.4049/JIMMUNOL.1900029.

Housseau, F. *et al.* (2001) 'N-linked carbohydrates in tyrosinase are required for its recognition by human MHC class II-restricted CD4(+) T cells.', *European journal of immunology*. Germany, 31(9), pp. 2690–2701. doi: 10.1002/1521-4141(200109)31:9<2690::aid-immu2690>3.0.co;2-8.

Howell, W. M. (2014) 'HLA and disease: guilt by association', *International Journal of Immunogenetics*. John Wiley & Sons, Ltd, 41(1), pp. 1–12. doi: 10.1111/IJI.12088.

Huang, B. *et al.* (2022) 'The endoplasmic reticulum chaperone BiP is a closure-accelerating cochaperone of Grp94', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 119(5). doi: 10.1073/PNAS.2118793119/SUPPL_FILE/PNAS.2118793119.SD01.XLSX.

Huber, E. M. *et al.* (2012) 'Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity', *Cell*. Elsevier B.V., 148(4), pp. 727–738. doi: 10.1016/j.cell.2011.12.030.

Hughes, C. S., Postovit, L. M. and Lajoie, G. A. (2010) ‘Matrigel: A complex protein mixture required for optimal growth of cell culture’, *PROTEOMICS*. John Wiley & Sons, Ltd, 10(9), pp. 1886–1890. doi: 10.1002/PMIC.200900758.

Hunt, D. F. *et al.* (1992) ‘Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry’, *Science (New York, N.Y.)*. Science, 255(5049), pp. 1261–1263. doi: 10.1126/SCIENCE.1546328.

Hutchinson, J. P. *et al.* (2021) ‘Common allotypes of ER aminopeptidase 1 have substrate-dependent and highly variable enzymatic properties.’, *The Journal of biological chemistry*. United States, 296, p. 100443. doi: 10.1016/j.jbc.2021.100443.

Ilik, İ. A. *et al.* (2020) ‘SON and SRRM2 are essential for nuclear speckle formation’, *eLife*. Edited by J. P. Staley, K. Struhl, and J. P. Staley. eLife Sciences Publications, Ltd, 9, p. e60579. doi: 10.7554/eLife.60579.

Inderberg-Suso, E. M. *et al.* (2012) ‘Widespread CD4+ T-cell reactivity to novel hTERT epitopes following vaccination of cancer patients with a single hTERT peptide GV1001’, *Oncoimmunology*. Oncoimmunology, 1(5), pp. 670–686. doi: 10.4161/ONCI.20426.

Itzhak, D. N. *et al.* (2016) ‘Global, quantitative and dynamic mapping of protein subcellular localization’, *eLife*. eLife Sciences Publications Ltd, 5(JUN2016). doi: 10.7554/ELIFE.16950.

Jacob, F. *et al.* (2020) ‘A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity’, *Cell*. Cell Press, 180(1), pp. 188-204.e22. doi: 10.1016/J.CELL.2019.11.036.

Javitt, A. *et al.* (2019) ‘Pro-inflammatory cytokines alter the immunopeptidome landscape by modulation of HLA-B expression’, *Frontiers in Immunology*. Frontiers Media S.A., 10(FEB), p. 141. doi: 10.3389/FIMMU.2019.00141/BIBTEX.

Javitt, A. *et al.* (2021) ‘The proteasome regulator PSME4 drives immune evasion and abrogates anti-tumor immunity in NSCLC’, *bioRxiv*. doi: 10.1101/2021.10.24.464690.

Jing, X. *et al.* (2019) ‘Role of hypoxia in cancer therapy by regulating the tumor microenvironment’, *Molecular Cancer 2019 18:1*. BioMed Central, 18(1), pp. 1–15. doi: 10.1186/S12943-019-1089-9.

Jørgensen, K. W. *et al.* (2014) ‘NetMHCstab – predicting stability of peptide–MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery’,

Immunology. John Wiley & Sons, Ltd, 141(1), pp. 18–26. doi: 10.1111/IMM.12160.

Jorgovanovic, D. *et al.* (2020) 'Roles of IFN-γ in tumor progression and regression: a review', *Biomarker Research 2020 8:1*. BioMed Central, 8(1), pp. 1–16. doi: 10.1186/S40364-020-00228-X.

Justa-Schuch, D. *et al.* (2016) 'DPP9 is a novel component of the N-end rule pathway targeting the tyrosine kinase Syk', *eLife*. eLife Sciences Publications Ltd, 5(September). doi: 10.7554/ELIFE.16370.

Kadara, H. *et al.* (2017) 'Whole-exome sequencing and immune profiling of early-stage lung adenocarcinoma with fully annotated clinical follow-up', *Annals of Oncology*. Oxford University Press, 28(1), pp. 75–82. doi: 10.1093/annonc/mdw436.

Kantarjian, H. *et al.* (2017) 'Blinatumomab versus Chemotherapy for Advanced Acute Lymphoblastic Leukemia', *New England Journal of Medicine*. Massachusetts Medical Society, 376(9), pp. 836–847. doi: 10.1056/NEJMOA1609783/SUPPL_FILE/NEJMOA1609783_DISCLOSURES.PDF.

Kaplan, R. N. *et al.* (2005) 'VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche.', *Nature*. England, 438(7069), pp. 820–827. doi: 10.1038/nature04186.

Kawahara, M. *et al.* (2009) 'Analysis of the role of tripeptidyl peptidase II in MHC class I antigen presentation in vivo', *Journal of immunology*. United States, 183(10), pp. 6069–6077. doi: 10.4049/jimmunol.0803564.

Keskin, D. B. *et al.* (2019) 'Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial', *Nature*. NIH Public Access, 565(7738), p. 234. doi: 10.1038/S41586-018-0792-9.

Kessler, J. H. *et al.* (2011) 'Antigen processing by nardilysin and thimet oligopeptidase generates cytotoxic T cell epitopes', *Nature immunology*. Nat Immunol, 12(1), pp. 45–53. doi: 10.1038/NI.1974.

Kim, E., Kwak, H. and Ahn, K. (2009) 'Cytosolic Aminopeptidases Influence MHC Class I-Mediated Antigen Presentation in an Allele-Dependent Manner', *The Journal of Immunology*. American Association of Immunologists, 183(11), pp. 7379–7387. doi: 10.4049/JIMMUNOL.0901489.

Kim, J. E. *et al.* (2019) 'Mutation Burden and I Index for Detection of Microsatellite Instability in Colorectal Cancer by Targeted Next-Generation Sequencing',

Journal of Molecular Diagnostics. Elsevier B.V., 21(2), pp. 241–250. doi: 10.1016/j.jmoldx.2018.09.005.

Kim, S. I. *et al.* (2003) 'Regulation of cell-surface major histocompatibility complex class I expression by the endopeptidase EC3.4.24.15 (thimet oligopeptidase)', *Biochem. J*, 375, pp. 111–120.

Kirino, Y. *et al.* (2013) 'Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B*51 and ERAP1.', *Nature genetics*. United States, 45(2), pp. 202–207. doi: 10.1038/ng.2520.

Kisselev, A. F. *et al.* (1999) 'The Sizes of Peptides Generated from Protein by Mammalian 26 and 20 S Proteasomes', *Journal of Biological Chemistry*. Elsevier BV, 274(6), pp. 3363–3371. doi: 10.1074/jbc.274.6.3363.

Kitamura, T. *et al.* (2007) 'SMAD4-deficient intestinal tumors recruit CCR1+ myeloid cells that promote invasion.', *Nature genetics*. United States, 39(4), pp. 467–475. doi: 10.1038/ng1997.

Klein, G. (1966) 'Tumor Antigens', *Annual Review of Microbiology*, 20(1), pp. 223–252. doi: 10.1146/annurev.mi.20.100166.001255.

Knutson, K. L. *et al.* (2016) 'Improved survival of HER2+ breast cancer patients treated with trastuzumab and chemotherapy is associated with host antibody immunity against the HER2 intracellular domain', *Cancer Research*. American Association for Cancer Research Inc., 76(13), pp. 3702–3710. doi: 10.1158/0008-5472.CAN-15-3091/652289/AM/IMPROVED-SURVIVAL-OF-HER2-BREAST-CANCER-PATIENTS.

Komov, L. *et al.* (2018) 'Cell Surface MHC Class I Expression Is Limited by the Availability of Peptide-Receptive "Empty" Molecules Rather than by the Supply of Peptide Ligands', *Proteomics*. doi: 10.1002/pmic.201700248.

Komov, L. *et al.* (2021) 'The effect of interferons on presentation of defective ribosomal products as HLA peptides', *Molecular and Cellular Proteomics*. American Society for Biochemistry and Molecular Biology Inc., 20, p. 100105. doi: 10.1016/J.MCPRO.2021.100105/ATTACHMENT/F6B67BCF-2DEF-476F-9D4D-7BFE61FD7DB7/MMC6.XLSX.

Koopmann, J. O. *et al.* (1996) 'Translocation of long peptides by transporters associated with antigen processing (TAP)', *European Journal of Immunology*. John Wiley & Sons, Ltd, 26(8), pp. 1720–1728. doi: 10.1002/EJI.1830260809.

Kreiter, S. *et al.* (2015) 'Mutant MHC class II epitopes drive therapeutic immune

responses to cancer', *Nature 2015 520:7549*. Nature Publishing Group, 520(7549), pp. 692–696. doi: 10.1038/nature14426.

Lang, S. *et al.* (2017) 'An update on Sec 61 channel functions, mechanisms, and related diseases', *Frontiers in Physiology*. Frontiers Media S.A., 8(NOV), p. 887. doi: 10.3389/FPHYS.2017.00887/BIBTEX.

Larance, M. *et al.* (2013) 'Global subcellular characterization of protein degradation using quantitative proteomics', *Molecular and Cellular Proteomics*. Elsevier, 12(3), pp. 638–650. doi: 10.1074/mcp.M112.024547.

Latchman, Y. E. *et al.* (2004) 'PD-L1-deficient mice show that PD-L1 on T cells, antigen-presenting cells, and host tissues negatively regulates T cells.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 101(29), pp. 10691–10696. doi: 10.1073/pnas.0307252101.

Laumont, C. M. *et al.* (2016) 'Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames', *Nature Communications*. Nature Publishing Group, 7. doi: 10.1038/ncomms10238.

Le, D. T. *et al.* (2015) 'PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.', *The New England journal of medicine*. doi: 10.1056/NEJMoa1500596.

Leboeuf, D. *et al.* (2020) 'Downregulation of the Arg/N-degron Pathway Sensitizes Cancer Cells to Chemotherapy In Vivo', *Molecular Therapy*. Cell Press, 28(4), pp. 1092–1104. doi: 10.1016/J.YMTHE.2020.01.021.

Lee, A. S. (2007) 'GRP78 Induction in Cancer: Therapeutic and Prognostic Implications', *Cancer Research*. American Association for Cancer Research, 67(8), pp. 3496–3499. doi: 10.1158/0008-5472.CAN-07-0325.

Lee, S. Y. *et al.* (2016) 'APEX Fingerprinting Reveals the Subcellular Localization of Proteins of Interest', *Cell Reports*. Elsevier B.V., 15(8), pp. 1837–1847. doi: 10.1016/j.celrep.2016.04.064.

Lehnert, E. and Tampé, R. (2017) 'Structure and dynamics of antigenic peptides in complex with TAP', *Frontiers in Immunology*. Frontiers Research Foundation, 8(JAN), p. 10. doi: 10.3389/FIMMU.2017.00010/BIBTEX.

Leivonen, S. K. *et al.* (2019) 'T-cell inflamed tumor microenvironment predicts favorable prognosis in primary testicular lymphoma', *Haematologica*. Ferrata Storti Foundation, 104(2), p. 338. doi: 10.3324/HAEMATOL.2018.200105.

Leone, P. *et al.* (2013) 'MHC class i antigen processing and presenting

machinery: Organization, function, and defects in tumor cells', *Journal of the National Cancer Institute*. doi: 10.1093/jnci/djt184.

Levine, A. J., Momand, J. and Finlay, C. A. (1991) 'The p53 tumour suppressor gene', *Nature*. Nature, 351(6326), pp. 453–456. doi: 10.1038/351453A0.

Li, H. *et al.* (2008) 'Identification of an N-linked glycosylation in the C4 region of HIV-1 envelope gp120 that is critical for recognition of neighboring CD4 T cell epitopes', *Journal of immunology (Baltimore, Md. : 1950)*. J Immunol, 180(6), pp. 4011–4021. doi: 10.4049/JIMMUNOL.180.6.4011.

Li, Y. *et al.* (2020) 'Patterns of somatic structural variation in human cancer genomes', *Nature*, 578(7793), pp. 112–121. doi: 10.1038/s41586-019-1913-9.

Liberzon, A. *et al.* (2015) 'The Molecular Signatures Database (MSigDB) hallmark gene set collection', *Cell systems*. NIH Public Access, 1(6), p. 417. doi: 10.1016/J.CELS.2015.12.004.

Liepe, J. *et al.* (2016) 'A large fraction of HLA class I ligands are proteasome-generated spliced peptides', *Science (New York, N.Y.)*. Science, 354(6310), pp. 354–358. doi: 10.1126/SCIENCE.AAF4384.

Lim, M. Y., Paulo, J. A. and Gygi, S. P. (2019) 'Evaluating False Transfer Rates from the Match-between-Runs Algorithm with a Two-Proteome Model', *Journal of Proteome Research*. American Chemical Society, 18(11), pp. 4020–4026. doi: 10.1021/ACS.JPROTEOME.9B00492/SUPPL_FILE/PR9B00492_SI_002.PDF.

Lipford, G. B. *et al.* (1995) 'In vivo CTL induction with point-substituted ovalbumin peptides: immunogenicity correlates with peptide-induced MHC class I stability', *Vaccine*. Vaccine, 13(3), pp. 313–320. doi: 10.1016/0264-410X(95)93320-9.

Liu, L. *et al.* (2015) 'The BRAF and MEK inhibitors dabrafenib and trametinib: Effects on immune function and in combination with immunomodulatory antibodies targeting PD-1, PD-L1, and CTLA-4', *Clinical Cancer Research*. doi: 10.1158/1078-0432.CCR-14-2339.

Liu, S. Q. *et al.* (2021) 'A CRISPR Screen Reveals Resistance Mechanisms to CD3-Bispecific Antibody Therapy', *Cancer immunology research*. NIH Public Access, 9(1), p. 34. doi: 10.1158/2326-6066.CIR-20-0080.

Liu, Y. T. and Sun, Z. J. (2021) 'Turning cold tumors into hot tumors by improving T-cell infiltration', *Theranostics*. Ivyspring International Publisher, 11(11), p. 5365. doi: 10.7150/THNO.58390.

Loffler, M. W. *et al.* (2018) 'Mapping the HLA ligandome of colorectal cancer reveals an imprint of malignant cell transformation', *Cancer Research*. doi:

10.1158/0008-5472.CAN-17-1745.

Löffler, M. W. *et al.* (2019) 'Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma', *Genome Medicine*. doi: 10.1186/s13073-019-0636-8.

von Loga, K. *et al.* (2020) 'Extreme intratumour heterogeneity and driver evolution in mismatch repair deficient gastro-oesophageal cancer', *Nature Communications*. Nature Publishing Group, 11(1). doi: 10.1038/S41467-019-13915-7.

López de Castro, J. A. (2018) 'How ERAP1 and ERAP2 Shape the Peptidomes of Disease-Associated MHC-I Proteins', *Frontiers in Immunology*. Frontiers Media SA, 9, p. 2463. doi: 10.3389/FIMMU.2018.02463.

Lote, H. *et al.* (2022) 'Advances in immunotherapy for MMR proficient colorectal cancer', *Cancer Treatment Reviews*. Elsevier, 111, p. 102480. doi: 10.1016/J.CTRV.2022.102480.

Luksza, M. *et al.* (2017) 'A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy', *Nature 2017 551:7681*. Nature Publishing Group, 551(7681), pp. 517–520. doi: 10.1038/nature24473.

Lysell, J. *et al.* (2013) 'Genetic Association with ERAP1 in Psoriasis Is Confined to Disease Onset after Puberty and Not Dependent on HLA-C*06', *Journal of Investigative Dermatology*. Elsevier, 133(2), pp. 411–417. doi: 10.1038/JID.2012.280.

Lyssiotis, C. A. and Kimmelman, A. C. (2017) 'Metabolic Interactions in the Tumor Microenvironment', *Trends in cell biology*. Trends Cell Biol, 27(11), pp. 863–875. doi: 10.1016/J.TCB.2017.06.003.

Ma, W. *et al.* (1997) 'Interferon-gamma rapidly increases peptide transporter (TAP) subunit expression and peptide transport capacity in endothelial cells', *The Journal of biological chemistry*. J Biol Chem, 272(26), pp. 16585–16590. doi: 10.1074/JBC.272.26.16585.

Madeleine, M. M. *et al.* (2008) 'Comprehensive Analysis of HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 Loci and Squamous Cell Cervical Cancer Risk', *Cancer Research*. American Association for Cancer Research, 68(9), pp. 3532–3539. doi: 10.1158/0008-5472.CAN-07-6471.

Maeda, M. *et al.* (2019) 'COPII proteins exhibit distinct subdomains within each ER exit site for executing their functions', *Scientific Reports 2019 9:1*. Nature Publishing Group, 9(1), pp. 1–7. doi: 10.1038/s41598-019-43813-3.

Manguso, R. T. *et al.* (2017) 'In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target', *Nature*. Nature, 547(7664), pp. 413–418. doi: 10.1038/nature23270.

Männistö, P. T. and García-Horsman, J. A. (2017) 'Mechanism of action of Prolyl oligopeptidase (PREP) in degenerative brain diseases: Has peptidase activity only a modulatory role on the interactions of PREP with proteins?', *Frontiers in Aging Neuroscience*. Frontiers Media S.A., 9(FEB), p. 27. doi: 10.3389/FNAGI.2017.00027/BIBTEX.

Mansukhani, S. *et al.* (2018) 'Iconic: Peri-operative immuno-chemotherapy in operable oesophageal and gastric cancer.', *https://doi.org/10.1200/JCO.2018.36.15_suppl.TPS4139*. American Society of Clinical Oncology, 36(15_suppl), pp. TPS4139–TPS4139. doi: 10.1200/JCO.2018.36.15_SUPPL.TPS4139.

Marijt, K. A. *et al.* (2018) 'Identification of non-mutated neoantigens presented by TAP-deficient tumors', *Journal of Experimental Medicine*. doi: 10.1084/jem.20180577.

Martincorena, I. *et al.* (2018) 'Somatic mutant clones colonize the human esophagus with age', *Science*. American Association for the Advancement of Science, 362(6417), pp. 911–917. doi: 10.1126/SCIENCE.AAU3879/SUPPL_FILE/AAU3879_TABLES4.XLSX.

Martinez-Fonts, K. *et al.* (2020) 'The proteasome 19S cap and its ubiquitin receptors provide a versatile recognition platform for substrates', *Nature Communications 2020 11:1*. Nature Publishing Group, 11(1), pp. 1–16. doi: 10.1038/s41467-019-13906-8.

Marzluff, W. F. J. (1978) 'Transcription of RNA in isolated nuclei.', *Methods in cell biology*. United States, 19, pp. 317–332. doi: 10.1016/s0091-679x(08)60032-1.

Matsushita, H. *et al.* (2012) 'Cancer Exome Analysis Reveals a T Cell Dependent Mechanism of Cancer Immunoediting', *Nature*. NIH Public Access, 482(7385), p. 400. doi: 10.1038/NATURE10755.

Matyskiela, M. E., Lander, G. C. and Martin, A. (2013) 'Conformational switching of the 26S proteasome enables substrate degradation', *Nature Structural & Molecular Biology 2013 20:7*. Nature Publishing Group, 20(7), pp. 781–788. doi: 10.1038/nsmb.2616.

Mavridis, G. *et al.* (2021) 'The ERAP1 active site cannot productively access the N-terminus of antigenic peptide precursors stably bound onto MHC class I.',

Scientific reports. England, 11(1), p. 16475. doi: 10.1038/s41598-021-95786-x.

McCarthy, M. K. and Weinberg, J. B. (2015) 'The immunoproteasome and viral infection: a complex regulator of inflammation', *Frontiers in Microbiology*. Frontiers Media SA, 6(JAN). doi: 10.3389/FMICB.2015.00021.

McGranahan, N. *et al.* (2017) 'Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution', *Cell*. doi: 10.1016/j.cell.2017.10.001.

McShan, A. C. *et al.* (2018) 'Peptide exchange on MHC-I by TAPBPR is driven by a negative allostery release cycle', *Nature Chemical Biology*, 14(8), pp. 811–820. doi: 10.1038/s41589-018-0096-2.

Megger, D. A. *et al.* (2017) 'Deciphering of the Human Interferon-Regulated Proteome by Mass Spectrometry-Based Quantitative Analysis Reveals Extent and Dynamics of Protein Induction and Repression', *Frontiers in immunology*. Front Immunol, 8. doi: 10.3389/FIMMU.2017.01139.

Michelsen, U. and von Hagen, J. (2009) 'Chapter 19 Isolation of Subcellular Organelles and Structures', in *Methods in Enzymology*. Academic Press, pp. 305–328. doi: 10.1016/S0076-6879(09)63019-6.

Mishto, M. *et al.* (2014) 'Proteasome isoforms exhibit only quantitative differences in cleavage and epitope generation', *European journal of immunology*. Eur J Immunol, 44(12), pp. 3508–3521. doi: 10.1002/EJI.201444902.

Moore, L. *et al.* (2020) 'The mutational landscape of normal human endometrial epithelium', *Nature 2020 580:7805*. Nature Publishing Group, 580(7805), pp. 640–646. doi: 10.1038/s41586-020-2214-z.

Morozov, G. I. *et al.* (2016) 'Interaction of TAPBPR, a tapasin homolog, with MHC-I molecules promotes peptide editing.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 113(8), pp. E1006-15. doi: 10.1073/pnas.1519894113.

Müller, M. *et al.* (2017) '"Hotspots" of antigen presentation revealed by human leukocyte antigen ligandomics for neoantigen prioritization', *Frontiers in Immunology*. doi: 10.3389/fimmu.2017.01367.

Munn, D. H. and Mellor, A. L. (2016) 'IDO in the Tumor Microenvironment: Inflammation, Counter-Regulation, and Tolerance', *Trends in Immunology*. doi: 10.1016/j.it.2016.01.002.

Naranbhai, V. *et al.* (2022) 'HLA-A*03 and response to immune checkpoint blockade in cancer: an epidemiological study', *The Lancet. Oncology*. NIH Public Access, 23(1), p. 172. doi: 10.1016/S1470-2045(21)00582-9.

Nathan, P. *et al.* (2021) 'Overall Survival Benefit with Tebentafusp in Metastatic Uveal Melanoma', *New England Journal of Medicine*, 385(13), pp. 1196–1206. doi: 10.1056/NEJMoa2103485.

Neerincx, A. *et al.* (2017) 'TAPBPR bridges UDP-glucose:glycoprotein glucosyltransferase 1 onto MHC class I to provide quality control in the antigen presentation pathway.', *eLife*. England, 6. doi: 10.7554/eLife.23049.

Newey, A. *et al.* (2019) 'Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment', *Journal for ImmunoTherapy of Cancer*, 7(1), p. 309. doi: 10.1186/s40425-019-0769-8.

Newey, A. *et al.* (2022) 'Multifactorial remodeling of the cancer immunopeptidome by interferon gamma', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2022.03.23.485466. doi: 10.1101/2022.03.23.485466.

Nguyen, A. T., Szeto, C. and Gras, S. (2021) 'The pockets guide to HLA class I molecules', *Biochemical Society Transactions*. Portland Press, 49(5), pp. 2319–2331. doi: 10.1042/BST20210410.

Nguyen, T. T. *et al.* (2011) 'Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.2021.

Nicastri, A. *et al.* (2020) 'The Choice of HLA-Associated Peptide Enrichment and Purification Strategy Affects Peptide Yields and Creates a Bias in Detected Sequence Repertoire', *Proteomics*. Proteomics, 20(12). doi: 10.1002/PMIC.201900401.

Nicholls, S. *et al.* (2009) 'Secondary anchor polymorphism in the HA-1 minor histocompatibility antigen critically affects MHC stability and TCR recognition', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(10), pp. 3889–3894. doi: 10.1073/PNAS.0900411106/SUPPL_FILE/0900411106SI.PDF.

Nielsen, M. *et al.* (2007) 'NetMHCpan, a Method for Quantitative Predictions of Peptide Binding to Any HLA-A and -B Locus Protein of Known Sequence', *PLOS ONE*. Public Library of Science, 2(8), p. e796. doi: 10.1371/JOURNAL.PONE.0000796.

Nowell, P. C. (1976) 'The clonal evolution of tumor cell populations', *Science (New York, N.Y.)*. Science, 194(4260), pp. 23–28. doi: 10.1126/SCIENCE.959840.

Old, L. J. and Boyse, E. A. (1964) ‘Immunology of Experimental Tumors’, *Annual Review of Medicine*, 15(1), pp. 167–186. doi: 10.1146/annurev.me.15.020164.001123.

Oldham, M. L., Grigorieff, N. and Chen, J. (2016) ‘Structure of the transporter associated with antigen processing trapped by herpes simplex virus’, *eLife*. eLife Sciences Publications Ltd, 5(DECEMBER2016). doi: 10.7554/ELIFE.21829.

Olsson, N. *et al.* (2021) ‘An Integrated Genomic, Proteomic, and Immunopeptidomic Approach to Discover Treatment-Induced Neoantigens’, *Frontiers in Immunology*. Frontiers, 0, p. 1129. doi: 10.3389/FIMMU.2021.662443.

Olvera, A. *et al.* (2021) ‘Does Antigen Glycosylation Impact the HIV-Specific T Cell Immunity?’, *Frontiers in Immunology*. Frontiers Media S.A., 11. doi: 10.3389/FIMMU.2020.573928.

Ombrello, M. J., Kastner, D. L. and Remmers, E. F. (2015) ‘Endoplasmic reticulum-associated amino-peptidase 1 and rheumatic disease: genetics’, *Current opinion in rheumatology*. NIH Public Access, 27(4), p. 349. doi: 10.1097/BOR.0000000000000189.

Ott, P. A. *et al.* (2017) ‘An immunogenic personal neoantigen vaccine for patients with melanoma.’, *Nature*. Nature Publishing Group, 547(7662), pp. 217–221. doi: 10.1038/nature22991.

Ott, P. A. *et al.* (2020) ‘A Phase Ib Trial of Personalized Neoantigen Therapy Plus Anti-PD-1 in Patients with Advanced Melanoma, Non-small Cell Lung Cancer, or Bladder Cancer’, *Cell*. Cell, 183(2), pp. 347-362.e24. doi: 10.1016/J.CELL.2020.08.053.

Pagliuca, S. *et al.* (2022) ‘Individual HLA heterogeneity and its implications for cellular immune evasion in cancer and beyond’, *Frontiers in Immunology*. Frontiers Media S.A., 13, p. 5104. doi: 10.3389/FIMMU.2022.944872/BIBTEX.

Papakyriakou, A. *et al.* (2018) ‘The partial dissociation of MHC class I-bound peptides exposes their N terminus to trimming by endoplasmic reticulum aminopeptidase 1.’, *The Journal of biological chemistry*. United States, 293(20), pp. 7538–7548. doi: 10.1074/jbc.RA117.000313.

Parel, Y. and Chizzolini, C. (2004) ‘CD4+ CD8+ double positive (DP) T cells in health and disease’, *Autoimmunity Reviews*. Elsevier, 3(3), pp. 215–220. doi: 10.1016/J.AUTREV.2003.09.001.

Patel, S., Sanjana, N., Kishton, R. et al. *et al.* (2017) ‘Identification of essential

genes for cancer immunotherapy', *Nature*. Nature, 548(7669), pp. 537–542. doi: 10.1038/nature23477.

Pearson, H. *et al.* (2016) 'MHC class I-associated peptides derive from selective regions of the human genome', *The Journal of clinical investigation*. J Clin Invest, 126(12), pp. 4690–4701. doi: 10.1172/JCI88590.

Perez, M. *et al.* (2019) 'Analysis of Secondary Structure Biases in Naturally Presented HLA-I Ligands', *Frontiers in Immunology*. Frontiers Media S.A., 10, p. 2731. doi: 10.3389/fimmu.2019.02731.

Peth, A., Uchiki, T. and Goldberg, A. L. (2010) 'ATP-Dependent Steps in the Binding of Ubiquitin Conjugates to the 26S Proteasome that Commit to Degradation', *Molecular Cell*. Cell Press, 40(4), pp. 671–681. doi: 10.1016/J.MOLCEL.2010.11.002.

Pfammatter, S. *et al.* (2020) 'Extending the Comprehensiveness of Immunopeptidome Analyses Using Isobaric Peptide Labeling', *Analytical Chemistry*. American Chemical Society, 92(13), pp. 9194–9204. doi: 10.1021/ACS.ANALCHEM.0C01545/SUPPL_FILE/AC0C01545_SI_004.XLSX.

Polom, K. *et al.* (2018) 'Meta-analysis of microsatellite instability in relation to clinicopathological characteristics and overall survival in gastric cancer', *British Journal of Surgery*. Oxford Academic, 105(3), pp. 159–167. doi: 10.1002/BJS.10663.

Praveen, P. V. K. *et al.* (2010) 'Tapasin edits peptides on MHC class I molecules by accelerating peptide exchange', *European Journal of Immunology*, 40(1), pp. 214–224. doi: https://doi.org/10.1002/eji.200939342.

Pryor, P. R. (2015) *Subcellular Fractionation: A Laboratory Manual*. Cold Spring Harbor Laboratory Press.

Purcell, A. W. (2000) 'The peptide-loading complex and ligand selection during the assembly of HLA class I molecules', *Molecular Immunology*. doi: 10.1016/S0161-5890(00)00075-4.

Purcell, A. W. *et al.* (2001) 'Quantitative and Qualitative Influences of Tapasin on the Class I Peptide Repertoire', *The Journal of Immunology*. American Association of Immunologists, 166(2), pp. 1016–1027. doi: 10.4049/jimmunol.166.2.1016.

Purcell, A. W., Ramarathinam, S. H. and Ternette, N. (2019) 'Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics', *Nature Protocols 2019 14:6*. Nature Publishing Group, 14(6), pp. 1687–1707. doi:

10.1038/s41596-019-0133-y.

Racle, J. *et al.* (2019) 'Deep motif deconvolution of HLA-II peptidomes for robust class II epitope predictions', *bioRxiv*. doi: 10.1101/539338.

Radcliffe, C. M. *et al.* (2002) 'Identification of specific glycoforms of major histocompatibility complex class I heavy chains suggests that class I peptide loading is an adaptation of the quality control pathway involving calreticulin and ERp57', *The Journal of biological chemistry*. J Biol Chem, 277(48), pp. 46415–46423. doi: 10.1074/JBC.M202466200.

Ramarathinam, S. H. *et al.* (2020) 'A peptide-signal amplification strategy for the detection and validation of neoepitope presentation on cancer biopsies', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.06.12.145276. doi: 10.1101/2020.06.12.145276.

Rammensee, H. G., Friede, T. and Stevanoviíc, S. (1995) 'MHC ligands and peptide motifs: first listing.', *Immunogenetics*. United States, 41(4), pp. 178–228. doi: 10.1007/BF00172063.

Rasmussen, M. *et al.* (2016) 'Pan-Specific Prediction of Peptide–MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity', *The Journal of Immunology*. American Association of Immunologists, 197(4), pp. 1517–1524. doi: 10.4049/JIMMUNOL.1600582.

Reche, P. A., Glutting, J. P. and Reinherz, E. L. (2002) 'Prediction of MHC class I binding peptides using profile motifs', *Human Immunology*. Elsevier, 63(9), pp. 701–709. doi: 10.1016/S0198-8859(02)00432-9.

Reeves, E. *et al.* (2013) 'Naturally occurring ERAP1 haplotypes encode functionally distinct alleles with fine substrate specificity.', *Journal of immunology (Baltimore, Md.: 1950)*. United States, 191(1), pp. 35–43. doi: 10.4049/jimmunol.1300598.

Reeves, E. *et al.* (2014) 'Functionally distinct ERAP1 allotype combinations distinguish individuals with ankylosing spondylitis', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(49), pp. 17594–17599. doi: 10.1073/PNAS.1408882111/SUPPL_FILE/PNAS.201408882SI.PDF.

Reits, E. *et al.* (2003) 'Peptide diffusion, protection, and degradation in nuclear and cytoplasmic compartments before antigen presentation by MHC class I.', *Immunity*. United States, 18(1), pp. 97–108. doi: 10.1016/s1074-7613(02)00511-3.

Reits, E. *et al.* (2004) 'A Major Role for TPPII in Trimming Proteasomal Degradation Products for MHC Class I Antigen Presentation', *Immunity*. Cell Press, 20(4), pp. 495–506. doi: 10.1016/S1074-7613(04)00074-3.

Reits, E. A. J. *et al.* (2000) 'The major substrates for TAP in vivo are derived from newly synthesized proteins', *Nature 2000 404:6779*. Nature Publishing Group, 404(6779), pp. 774–778. doi: 10.1038/35008103.

Reynisson, B. *et al.* (2020) 'NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data', *Nucleic acids research*. doi: 10.1093/nar/gkaa379.

Van Rhee, F. *et al.* (2005) 'NY-ESO-1 is highly expressed in poor-prognosis multiple myeloma and induces spontaneous humoral and cellular immune responses', *Blood*. doi: 10.1182/blood-2004-09-3707.

Ribas, A. *et al.* (2003) 'Determinant spreading and tumor responses after peptide-based cancer immunotherapy', *Trends in Immunology*. Elsevier Ltd, 24(2), pp. 58–61. doi: 10.1016/S1471-4906(02)00029-7.

Rickwood, D., Ford, T. and Graham, J. (1982) 'Nycodenz: A new nonionic iodinated gradient medium', *Analytical Biochemistry*. Academic Press, 123(1), pp. 23–31. doi: 10.1016/0003-2697(82)90618-2.

Ridge, D. (1978) *Practical Aspects of Rate-Zonal Centrifugation*, *Centrifugal Separations in Molecular and Cell Biology*. Butterworth-Heinemann. doi: 10.1016/B978-0-408-70803-6.50007-8.

Ritter, C. *et al.* (2017) 'Epigenetic priming restores the HLA class-I antigen processing machinery expression in Merkel cell carcinoma', *Scientific Reports 2017 7:1*. Nature Publishing Group, 7(1), pp. 1–11. doi: 10.1038/s41598-017-02608-0.

Robert, C. *et al.* (2015) ' Nivolumab in Previously Untreated Melanoma without BRAF Mutation ', *New England Journal of Medicine*. Massachusetts Medical Society, 372(4), pp. 320–330. doi: 10.1056/NEJMOA1412082/SUPPL_FILE/NEJMOA1412082_DISCLOSURES.PDF.

Robinson, J. *et al.* (2015) 'The IPD and IMGT/HLA database: allele variant databases', *Nucleic Acids Research*. Oxford University Press, 43(Database issue), p. D423. doi: 10.1093/NAR/GKU1161.

Roche, P. A. and Furuta, K. (2015) 'The ins and outs of MHC class II-mediated

antigen processing and presentation', *Nature Reviews Immunology 2015 15:4*. Nature Publishing Group, 15(4), pp. 203–216. doi: 10.1038/nri3818.

Rock, K. L. *et al.* (1994) 'Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules', *Cell*. Elsevier, 78(5), pp. 761–771. doi: 10.1016/S0092-8674(94)90462-6.

Rock, K. L., Farfán-Arribas, D. J. and Shen, L. (2010) 'Proteases in MHC class I presentation and cross-presentation', *Journal of immunology (Baltimore, Md. : 1950)*. NIH Public Access, 184(1), p. 9. doi: 10.4049/JIMMUNOL.0903399.

Rosenthal, R. *et al.* (2019) 'Neoantigen-directed immune escape in lung cancer evolution', *Nature 2019 567:7749*. Nature Publishing Group, 567(7749), pp. 479–485. doi: 10.1038/s41586-019-1032-7.

Roumeliotis, T. I. *et al.* (2017) 'Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells', *Cell Reports*. Elsevier, 20(9), p. 2201. doi: 10.1016/J.CELREP.2017.08.010.

Rousselle, T. V. *et al.* (2022) 'An optimized protocol for single nuclei isolation from clinical biopsies for RNA-seq', *Scientific Reports 2022 12:1*. Nature Publishing Group, 12(1), pp. 1–12. doi: 10.1038/s41598-022-14099-9.

Roux, K. J. *et al.* (2018) 'BioID: A Screen for Protein-Protein Interactions', *Current Protocols in Protein Science*. John Wiley & Sons, Ltd, 91(1), pp. 19.23.1-19.23.15. doi: 10.1002/CPPS.51.

Rubin, D. M. *et al.* (1998) 'Active site mutants in the six regulatory particle ATPases reveal multiple roles for ATP in the proteasome.', *The EMBO Journal*. European Molecular Biology Organization, 17(17), p. 4909. doi: 10.1093/EMBOJ/17.17.4909.

Sahin, U. *et al.* (2017) 'Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer', *Nature*. doi: 10.1038/nature23003.

Salk, J. J., Fox, E. J. and Loeb, L. A. (2010) 'Mutational heterogeneity in human cancers: origin and consequences.', *Annual review of pathology*. United States, 5, pp. 51–75. doi: 10.1146/annurev-pathol-121808-102113.

Saper, M. A., Bjorkman, P. J. and Wiley, D. C. (1991) 'Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution', *Journal of Molecular Biology*. Academic Press, 219(2), pp. 277–319. doi: 10.1016/0022-2836(91)90567-P.

Saric, T. *et al.* (2001) 'Major Histocompatibility Complex Class I-presented

Antigenic Peptides Are Degraded in Cytosolic Extracts Primarily by Thimet Oligopeptidase*', *Journal of Biological Chemistry*. Elsevier, 276(39), pp. 36474–36481. doi: 10.1074/JBC.M105517200.

Saric, T. *et al.* (2002) 'An IFN-γ–induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I–presented peptides', *Nature Immunology 2002 3:12*. Nature Publishing Group, 3(12), pp. 1169–1176. doi: 10.1038/ni859.

Sarkizova, S. *et al.* (2020) 'A large peptidome dataset improves HLA class I epitope prediction across most of the human population', *Nature Biotechnology*. NIH Public Access, 38(2), pp. 199–209. doi: 10.1038/s41587-019-0322-9.

Sasako, M. *et al.* (2011) 'Five-year outcomes of a randomized phase III trial comparing adjuvant chemotherapy with S-1 versus surgery alone in stage II or III gastric cancer', *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. J Clin Oncol, 29(33), pp. 4387–4393. doi: 10.1200/JCO.2011.36.5908.

Sato, T. *et al.* (2009) 'Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche', *Nature*. doi: 10.1038/nature07935.

Sato, T. *et al.* (2011) 'Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium', *Gastroenterology*. doi: 10.1053/j.gastro.2011.07.050.

Saveanu, L. *et al.* (2005) 'Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum', *Nature Immunology*. Nature Publishing Group, 6(7), pp. 689–697. doi: 10.1038/ni1208.

Schmidt, H. *et al.* (1990) 'Differential regulation of HLA class I genes by interferon.', *Immunogenetics*. United States, 31(4), pp. 245–252. doi: 10.1007/BF00204896.

Schnalzger, T. E. *et al.* (2019) '3D model for CAR-mediated cytotoxicity using patient-derived colorectal cancer organoids', *The EMBO journal*. EMBO J, 38(12). doi: 10.15252/EMBJ.2018100928.

Schomburg, L. *et al.* (2000) 'Molecular characterization of a puromycin-insensitive leucyl-specific aminopeptidase, PILS-AP.', *European journal of biochemistry*. England, 267(11), pp. 3198–3207. doi: 10.1046/j.1432-1327.2000.01348.x.

Schuster, H. *et al.* (2017) 'The immunopeptidomic landscape of ovarian carcinomas', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1707658114.

Schwarz, K. *et al.* (2000) 'The proteasome regulator PA28alpha/beta can enhance antigen presentation without affecting 20S proteasome subunit composition.', *European journal of immunology*. Germany, 30(12), pp. 3672–3679. doi: 10.1002/1521-4141(200012)30:12<3672::AID-IMMU3672>3.0.CO;2-B.

Sconocchia, G. *et al.* (2015) 'HLA Class II Antigen Expression in Colorectal Carcinoma Tumors as a Favorable Prognostic Marker', *Neoplasia*. doi: 10.1593/neo.131568.

Seifert, U. *et al.* (2003) 'An essential role for tripeptidyl peptidase in the generation of an MHC class I epitope', *Nature Immunology 2003 4:4*. Nature Publishing Group, 4(4), pp. 375–379. doi: 10.1038/ni905.

Seliger, B. *et al.* (1997) 'IFN-gamma-mediated coordinated transcriptional regulation of the human TAP-1 and LMP-2 genes in human renal cell carcinoma.', *Clinical cancer research : an official journal of the American Association for Cancer Research*. United States, 3(4), pp. 573–578.

Sexton, R. E. *et al.* (2020) 'Gastric Cancer: A Comprehensive Review of Current and Future Treatment Strategies', *Cancer Metastasis Review*, 39(4), pp. 1179–1203. doi: 10.1007/s10555-020-09925-3.

Shiina, T. *et al.* (2009) 'The HLA genomic loci map: expression, interaction, diversity and disease', *Journal of Human Genetics 2009 54:1*. Nature Publishing Group, 54(1), pp. 15–39. doi: 10.1038/jhg.2008.5.

Shimbara, N. *et al.* (1998) 'Contribution of Proline Residue for Efficient Production of MHC Class I Ligands by Proteasomes *', *Journal of Biological Chemistry*. Elsevier, 273(36), pp. 23062–23071. doi: 10.1074/JBC.273.36.23062.

Shukla, S. A. *et al.* (2015) 'Comprehensive analysis of cancer-associated somatic mutations in class i HLA genes', *Nature Biotechnology*. doi: 10.1038/nbt.3344.

Sidney, J. *et al.* (2013) 'Measurement of MHC/Peptide Interactions by Gel Filtration or Monoclonal Antibody Capture', *Current protocols in immunology / edited by John E. Coligan ... [et al.]*. NIH Public Access, 0 18(SUPPL.100), p. Unit. doi: 10.1002/0471142735.IM1803S100.

Siegel, R. L. *et al.* (2020) 'Colorectal cancer statistics, 2020', *CA: A Cancer Journal for Clinicians*. American Cancer Society, 70(3), pp. 145–164. doi: 10.3322/CAAC.21601.

Siegel, R. L. *et al.* (2021) 'Cancer Statistics, 2021', *CA: A Cancer Journal for Clinicians*. American Cancer Society, 71(1), pp. 7–33. doi: 10.3322/CAAC.21654.

Sigma-Aldrich (2006) *Endoplasmic Reticulum Isolation Kit (ER0100) - Technical Bulletin*.

Skoulidis, F. *et al.* (2015) 'Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities', *Cancer Discovery*. American Association for Cancer Research Inc., 5(8), pp. 861–878. doi: 10.1158/2159-8290.CD-14-1236/42538/AM/CO-OCCURRING-GENOMIC-ALTERATIONS-DEFINE-MAJOR.

Smith, D. M. *et al.* (2005) 'ATP binding to PAN or the 26S ATPases causes association with the 20S proteasome, gate opening, and translocation of unfolded proteins', *Molecular cell*. Mol Cell, 20(5), pp. 687–698. doi: 10.1016/J.MOLCEL.2005.10.019.

Souri, Z. *et al.* (2020) 'HDAC Inhibition Increases HLA Class I Expression in Uveal Melanoma', *Cancers*. Multidisciplinary Digital Publishing Institute (MDPI), 12(12), pp. 1–14. doi: 10.3390/CANCERS12123690.

Spierings, E. *et al.* (2009) 'Steric Hindrance and Fast Dissociation Explain the Lack of Immunogenicity of the Minor Histocompatibility HA-1Arg Null Allele', *The Journal of Immunology*. American Association of Immunologists, 182(8), pp. 4809–4816. doi: 10.4049/JIMMUNOL.0803911.

Stamogiannos, A. *et al.* (2015) 'Effects of polymorphic variation on the mechanism of Endoplasmic Reticulum Aminopeptidase 1.', *Molecular immunology*. England, 67(2 Pt B), pp. 426–435. doi: 10.1016/j.molimm.2015.07.010.

Steimle, V. *et al.* (1994) 'Regulation of MHC Class II Expression by Interferon-γ Mediated by the Transactivator Gene CIITA', *Science*. American Association for the Advancement of Science , 265(5168), pp. 106–109. doi: 10.1126/SCIENCE.8016643.

Stoltze, L. *et al.* (2000) 'Two new proteases in the MHC class I processing pathway', *Nature Immunology 2000 1:5*. Nature Publishing Group, 1(5), pp. 413–418. doi: 10.1038/80852.

Storrie, B. and Madden, E. A. (1990) *Isolation of subcellular organelles*, *Methods in enzymology*. Methods Enzymol. doi: 10.1016/0076-6879(90)82018-W.

Strange, A. *et al.* (2010) 'A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1.', *Nature genetics*. United States, 42(11), pp. 985–990. doi: 10.1038/ng.694.

Suski, J. M. *et al.* (2014) ‘Isolation of plasma membrane-associated membranes from rat liver’, *Nature protocols*. Nat Protoc, 9(2), pp. 312–322. doi: 10.1038/NPROT.2014.016.

Tamang, D. L. *et al.* (2006) ‘Induction of granzyme B and T cell cytotoxic capacity by IL-2 or IL-15 without antigens: Multiclonal responses that are extremely lytic if triggered and short-lived after cytokine withdrawal.’, *Cytokine*. NIH Public Access, 36(3–4), p. 148. doi: 10.1016/J.CYTO.2006.11.008.

Tauriello, D. V. F. *et al.* (2018) ‘TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis’, *Nature*. doi: 10.1038/nature25492.

Teixeira, V. H. *et al.* (2019) ‘Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions’, *Nature Medicine*, 25(3), pp. 517–525. doi: 10.1038/s41591-018-0323-0.

Teku, G. N. and Vihinen, M. (2018) ‘Pan-cancer analysis of neoepitopes’, *Scientific Reports 2018 8:1*. Nature Publishing Group, 8(1), pp. 1–10. doi: 10.1038/s41598-018-30724-y.

Thermo Fisher Scientific (2017) ‘Is trypsin digesting your cell surface proteins?’ doi: 10.1634/stemcells.2006-0260/full.

Thomas, C. and Tampé, R. (2017) ‘Proofreading of Peptide-MHC Complexes through Dynamic Multivalent Interactions’, *Front. Immunol*, 8, p. 65. doi: 10.3389/fimmu.2017.00065.

Thul, P. J. *et al.* (2017a) ‘A subcellular map of the human proteome’, *Science*. American Association for the Advancement of Science, 356(6340). doi: 10.1126/SCIENCE.AAL3321/SUPPL_FILE/AAL3321_THUL_SM_TABLE_S9.XLSX.

Thul, P. J. *et al.* (2017b) *The human subcellular proteome - The Human Protein Atlas*, *The Human Protein Atlas*. Available at: https://www.proteinatlas.org/humanproteome/subcellular (Accessed: 27 February 2023).

Tomko, R. J. and Hochstrasser, M. (2013) ‘Molecular Architecture and Assembly of the Eukaryotic Proteasome’, *https://doi.org/10.1146/annurev-biochem-060410-150257*. Annual Reviews , 82, pp. 415–445. doi: 10.1146/ANNUREV-BIOCHEM-060410-150257.

Torrejon, D. Y. *et al.* (2020) ‘Overcoming genetically based resistance mechanisms to PD-1 blockade’, *Cancer Discovery*. American Association for Cancer Research Inc., 10(8), pp. 1140–1157. doi: 10.1158/2159-8290.CD-19-

1409/333445/AM/OVERCOMING-GENETICALLY-BASED-RESISTANCE-MECHANISMS.

Towne, C. F. *et al.* (2005) 'Leucine Aminopeptidase Is Not Essential for Trimming Peptides in the Cytosol or Generating Epitopes for MHC Class I Antigen Presentation', *The Journal of Immunology*. American Association of Immunologists, 175(10), pp. 6605–6614. doi: 10.4049/JIMMUNOL.175.10.6605.

Towne, C. F. *et al.* (2008) 'Puromycin-sensitive aminopeptidase limits MHC class I presentation in dendritic cells but does not affect CD8 T cell responses during viral infections', *Journal of immunology (Baltimore, Md.: 1950)*. J Immunol, 180(3), pp. 1704–1712. doi: 10.4049/JIMMUNOL.180.3.1704.

Tran, E. *et al.* (2015) 'Immunogenicity of somatic mutations in human gastrointestinal cancers', *Science*. doi: 10.1126/science.aad1253.

Tsuji, A. *et al.* (2021) 'The randomized phase II study of FOLFOXIRI plus cetuximab versus FOLFOXIRI plus bevacizumab as the first-line treatment in metastatic colorectal cancer with RAS wild-type tumors: The DEEPER trial (JACCRO CC-13).', *https://doi.org/10.1200/JCO.2021.39.15_suppl.3501*. Wolters Kluwer Health, 39(15_suppl), pp. 3501–3501. doi: 10.1200/JCO.2021.39.15_SUPPL.3501.

Turajlic, S. *et al.* (2017) 'Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis', *The Lancet Oncology*. Elsevier, 18(8), pp. 1009–1021. doi: 10.1016/S1470-2045(17)30516-8.

Uchihara, Y. *et al.* (2022) 'DNA damage promotes HLA class I presentation by stimulating a pioneer round of translation-associated antigen production', *Molecular Cell*. Cell Press, 82(14), pp. 2557-2570.e7. doi: 10.1016/j.molcel.2022.04.030.

Uebel, S. and Tampé, R. (1999) 'Specificity of the proteasome and the TAP transporter', *Current Opinion in Immunology*. Elsevier Current Trends, 11(2), pp. 203–208. doi: 10.1016/S0952-7915(99)80034-X.

Uhlén, M. *et al.* (2019) 'The human secretome', *Science Signaling*, 12(609), p. eaaz0274. doi: 10.1126/scisignal.aaz0274.

Umar, S. *et al.* (2009) 'Functional cross-talk between beta-catenin and NFkappaB signaling pathways in colonic crypts of mice in response to progastrin.', *The Journal of biological chemistry*. United States, 284(33), pp. 22274–22284. doi: 10.1074/jbc.M109.020941.

Urban, J. L. *et al.* (1982) ‘Immunoselection of tumor cell variants by mice suppressed with ultraviolet radiation’, *Journal of Experimental Medicine*. The Rockefeller University Press, 156(4), pp. 1025–1041. doi: 10.1084/JEM.156.4.1025.

Vajdic, C. M. and Van Leeuwen, M. T. (2009) ‘Cancer incidence and risk factors after solid organ transplantation’, *International Journal of Cancer*. John Wiley & Sons, Ltd, 125(8), pp. 1747–1754. doi: 10.1002/IJC.24439.

Vanhoof, G. *et al.* (1995) ‘Proline motifs in peptides and their biological processing’, *The FASEB Journal*. John Wiley & Sons, Ltd, 9(9), pp. 736–744. doi: 10.1096/FASEBJ.9.9.7601338.

Verdegaal, E. M. E. *et al.* (2016) ‘Neoantigen landscape dynamics during human melanoma–T cell interactions’, *Nature 2016 536:7614*. Nature Publishing Group, 536(7614), pp. 91–95. doi: 10.1038/nature18945.

Verschoor, Y. L. *et al.* (2022) ‘Neoadjuvant nivolumab, ipilimumab, and celecoxib in MMR-proficient and MMR-deficient colon cancers: Final clinical analysis of the NICHE study.’, *Journal of Clinical Oncology*. American Society of Clinical Oncology, 40, pp. 3511–3511. doi: 10.1200/JCO.2022.40.16_SUPPL.3511.

Vlachogiannis, G. *et al.* (2018) ‘Patient-derived organoids model treatment response of metastatic gastrointestinal cancers’, *Science*. doi: 10.1126/science.aao2774.

Voeltz, G. K., Rolls, M. M. and Rapoport, T. A. (2002) ‘Structural organization of the endoplasmic reticulum’, *EMBO Reports*. European Molecular Biology Organization, 3(10), p. 944. doi: 10.1093/EMBO-REPORTS/KVF202.

Vogelstein, B. *et al.* (2013) ‘Cancer Genome Landscapes’, *Science (New York, N.Y.)*. NIH Public Access, 339(6127), p. 1546. doi: 10.1126/SCIENCE.1235122.

Wang, M. *et al.* (2017) ‘Role of tumor microenvironment in tumorigenesis’, *Journal of Cancer*, 8(5), pp. 761–773. doi: 10.7150/jca.17648.

Wang, Q. *et al.* (2019) ‘Direct Detection and Quantification of Neoantigens’, *Cancer Immunology Research*, 7(11), pp. 1748–1754. doi: 10.1158/2326-6066.CIR-19-0107.

Wang, S. *et al.* (2021) ‘The state of the art of bispecific antibodies for treating human malignancies’, *EMBO Molecular Medicine*. John Wiley & Sons, Ltd, 13(9), p. e14291. doi: 10.15252/EMMM.202114291.

Warburg, O., Wind, F. and Negelein, E. (1927) ‘The metabolism of tumors in the body’, *Journal of General Physiology*. The Rockefeller University Press, 8(6), pp.

519–530. doi: 10.1085/jgp.8.6.519.
Watson, P. *et al.* (2006) 'Sec16 defines endoplasmic reticulum exit sites and is required for secretory cargo export in mammalian cells', *Traffic*, 7(12), pp. 1678–1687. doi: 10.1111/J.1600-0854.2006.00493.X.
Weast, R. C. (1986) *CRC handbook of chemistry and physics.* 67th edn. CRC Press Inc.
Weber, J. *et al.* (2017) 'Adjuvant Nivolumab versus Ipilimumab in Resected Stage III or IV Melanoma', *New England Journal of Medicine*. New England Journal of Medicine (NEJM/MMS), 377(19), pp. 1824–1835. doi: 10.1056/NEJMOA1709030/SUPPL_FILE/NEJMOA1709030_DISCLOSURES.PDF.
Weeber, F. *et al.* (2015) 'Preserved genetic diversity in organoids cultured from biopsies of human colorectal cancer metastases', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1516689112.
Weis, S. *et al.* (2004) 'Endothelial barrier disruption by VEGF-mediated Src activity potentiates tumor cell extravasation and metastasis', *The Journal of Cell Biology*. The Rockefeller University Press, 167(2), p. 223. doi: 10.1083/JCB.200408130.
Wensink, G. E. *et al.* (2021) 'Patient-derived organoids as a predictive biomarker for treatment response in cancer patients', *npj Precision Oncology*, 5(1), p. 30. doi: 10.1038/s41698-021-00168-1.
Van De Wetering, M. *et al.* (2015) 'Prospective derivation of a living organoid biobank of colorectal cancer patients', *Cell*. doi: 10.1016/j.cell.2015.03.053.
Whiteside, T. L. (2008) 'The tumor microenvironment and its role in promoting tumor growth', *Oncogene*. NIH Public Access, 27(45), p. 5904. doi: 10.1038/ONC.2008.271.
Wieczorek, M. *et al.* (2017) *Major histocompatibility complex (MHC) class I and MHC class II proteins: Conformational plasticity in antigen presentation*, *Frontiers in Immunology*. Frontiers Media SA. doi: 10.3389/fimmu.2017.00292.
Williams, A. P. *et al.* (2002) 'Optimization of the MHC Class I Peptide Cargo Is Dependent on Tapasin', *Immunity*. Cell Press, 16(4), pp. 509–520. doi: 10.1016/S1074-7613(02)00304-7.
Williamson, C. D. *et al.* (2015) 'Isolation of Endoplasmic Reticulum, Mitochondria, and Mitochondria-Associated Membrane and Detergent Resistant Membrane Fractions from Transfected Cells and from Human Cytomegalovirus-Infected

Primary Fibroblasts', *Current Protocols in Cell Biology*. doi: 10.1002/0471143030.cb0327s68.

Wiśniewski, A. *et al.* (2018) 'The association of ERAP1 and ERAP2 single nucleotide polymorphisms and their haplotypes with psoriasis vulgaris is dependent on the presence or absence of the HLA-C*06:02 allele and age at disease onset.', *Human immunology*. United States, 79(2), pp. 109–116. doi: 10.1016/j.humimm.2017.11.010.

Witsch, E., Sela, M. and Yarden, Y. (2010) 'Roles for Growth Factors in Cancer Progression', *Physiology*. American Physiological Society, 25(2), pp. 85–101. doi: 10.1152/PHYSIOL.00045.2009/ASSET/IMAGES/LARGE/PHY0021000080004.JPEG.

Wolf-Levy, H. *et al.* (2018) 'Revealing the Cellular Degradome by Mass Spectrometry Analysis of Proteasome-Cleaved Peptides', *Nature Biotechnology*, 36(11). doi: 10.1038/nbt.4279.

Woolston, A. *et al.* (2019) 'Genomic and Transcriptomic Determinants of Therapy Resistance and Immune Landscape Evolution during Anti-EGFR Treatment in Colorectal Cancer', *Cancer Cell*, 36(1). doi: 10.1016/j.ccell.2019.05.013.

Xie, Y. H., Chen, Y. X. and Fang, J. Y. (2020) 'Comprehensive review of targeted therapy for colorectal cancer', *Signal Transduction and Targeted Therapy 2020 5:1*. Nature Publishing Group, 5(1), pp. 1–30. doi: 10.1038/s41392-020-0116-z.

Yewdell, J. W. (2011) 'DRiPs solidify: progress in understanding endogenous MHC class I antigen processing', *Trends in Immunology*. Elsevier, 32(11), pp. 548–558. doi: 10.1016/J.IT.2011.08.001.

Yewdell, J. W., Antón, L. C. and Bennink, J. R. (1996) 'Defective ribosomal products (DRiPs): a major source of antigenic peptides for MHC class I molecules?', *The Journal of Immunology*. American Association of Immunologists, 157(5), pp. 1823–1826. doi: 10.4049/JIMMUNOL.157.5.1823.

Yewdell, J. W. and Nicchitta, C. V. (2006) 'The DRiP hypothesis decennial: support, controversy, refinement and extension.', *Trends in immunology*. England: Elsevier, 27(8), pp. 368–373. doi: 10.1016/j.it.2006.06.008.

Yu, F., Haynes, S. E. and Nesvizhskii, A. I. (2021) 'IonQuant Enables Accurate and Sensitive Label-Free Quantification With FDR-Controlled Match-Between-Runs', *Molecular & Cellular Proteomics*. Elsevier, 20, p. 100077. doi: 10.1016/J.MCPRO.2021.100077.

Yu, K. *et al.* (2019) ‘Comprehensive transcriptomic analysis of cell lines as models of primary tumors across 22 tumor types’, *Nature Communications*, 10. doi: 10.1038/s41467-019-11415-2.

Zaidi, M. R. and Merlino, G. (2011) ‘The two faces of interferon-γ in cancer’, *Clinical cancer research : an official journal of the American Association for Cancer Research*. Clin Cancer Res, 17(19), pp. 6118–6124. doi: 10.1158/1078-0432.CCR-11-0482.

Zamora, A. E., Crawford, J. C. and Thomas, P. G. (2018) ‘Hitting the Target: How T Cells Detect and Eliminate Tumors.’, *Journal of immunology (Baltimore, Md. : 1950)*, 200(2), pp. 392–399. doi: 10.4049/jimmunol.1701413.

Zhai, L. *et al.* (2018) ‘IDO1 in cancer: a Gemini of immune checkpoints’, *Cellular & Molecular Immunology*, 15(5), pp. 447–457. doi: 10.1038/cmi.2017.143.

Zhang, F. *et al.* (2000) ‘Characterization of ABCB9, an ATP Binding Cassette Protein Associated with Lysosomes *’, *Journal of Biological Chemistry*. Elsevier, 275(30), pp. 23287–23294. doi: 10.1074/JBC.M001819200.

Zhang, J. *et al.* (2023) ‘Single amino acid-based PROTACs trigger degradation of the oncogenic kinase BCR-ABL in chronic myeloid leukemia (CML)’. doi: 10.1016/j.jbc.2023.104994.

Zhao, J. *et al.* (2019) ‘Immune and genomic correlates of response to anti-PD-1 immunotherapy in glioblastoma’, *Nature Medicine 2019 25:3*. Nature Publishing Group, 25(3), pp. 462–469. doi: 10.1038/s41591-019-0349-y.

Zhou, F. (2009) ‘Molecular mechanisms of IFN-γ to up-regulate MHC class i antigen processing and presentation’, *International Reviews of Immunology*. doi: 10.1080/08830180902978120.