MONASH University

Identification of cancer specific endogenous HLA class I peptide repertoire in haematological and paediatric brain malignancy

Kirti Pandey
Bachelor of Zoology (Honours)
Masters of Zoology

A thesis submitted for the degree of Doctor of Philosophy at
Monash University in 2020

Department of Biochemistry and Molecular Biology,
Biomedicine Discovery Institute,
Faculty of Medicine, Nursing and Health Science,
Monash University, Clayton, Victoria, Australia

Copyright notice

I certify that I have secured copyright permissions for third-party content included in this thesis (mentioned below) and have not knowingly added copyright content to my work without the owner's permission.

1) **Figure 1 - The cancer immunity cycle that leads to recognition and elimination of cancer cells.** Figure from Chen *et al* (Chen and Mellman 2013). Copyright permission granted by Elsevier with copyright licence number 4897400429092.

2) **Figure 2 - Anatomic location of site of origin of DIPG tumour.** DIPG arises in the pons brainstem region of the brain in children. Figure from Juratli *et al.* (Juratli et al. 2018). Copyright permission granted by Elsevier with copyright licence number 4897391067327.

ABSTRACT

Over the years an improved understanding of the immune system and the role it plays in cancer development and progression has led to pathbreaking discoveries and advances in the field of cancer immunotherapy. This has also ushered a renewed interest in T cell mediated immunotherapy as it relies on training the immune system to fight cancer by itself. The success of HLA-mediated immunotherapy depends on extensive studies of HLA peptide repertoire or the immunopeptidome of the cancer. By identifying TAAs, CTAs, neoantigens, the HLA-peptide repertoire can help turn the tide in the favour of peptide-based immunotherapy. HLA molecules play a crucial role in immune surveillance making it adept in sampling the tumour microenvironment and representing a snapshot of dysregulated cellular activities. The immunopeptidome can pave way for improved immunotherapy for both haematological and paediatric tumours.

Using a combination of high-resolution mass spectrometry (MS) flow cytometry, and other biochemical techniques the study investigates in-depth the immunopeptidome of AML including cell lines and clinical samples. It provides an insight into not only the HLA class I immunopeptidome at the site of tumour but also on how it is influenced by downregulation of HLA or loss of chromosome 6 in case of HM. Through our MS approach we have identified a plethora of leukaemia specific antigens derived from oncogenes, gonadotrophic proteins along with post translationally modified (PTM) peptides. By combining MS approach with X-ray crystallography, the study elucidates for the first-time crystal structure of HLA A*02:01 with tyrosine phosphorylated and arginine dimethylated peptides. The position of the PTMs within the peptide and their interactions with the HLA A2 residues highlight their immunogenic potential. The study also explores the relationship between soluble HLA (sHLA) and membrane bound HLA (mHLA) in leukaemia patients and provides an alternative to existing dogma in the field that in the event of HLA downregulation maybe the tumour does not contribute to sHLA to the extent as previously thought.

Lastly, the thesis also thoroughly investigates the immunopeptidome of DIPG, a type of rare and aggressive paediatric glioma. Through our optimised MS workflow, we have identified different peptides which have the potential of being translated at a clinical level for T cell mediated immunotherapy, hence fulfilling the unmet need for DIPG specific treatment.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Print Name: Kirti Pandey

Date: 21/09/2020

Publications during enrolment

1. **Kirti Pandey**, Sri H. Ramarathinam, Nicole A. Mifsud, Terry C.C Lim Kam Sian, Rochelle Ayala, Nicola Ternette, Anthony W. Purcell (2019). In-depth mining of the immunopeptidome using complementary peptide enrichment and mass spectrometric acquisition strategies. *Molecular Immunology* 123:7-17 (https://doi.org/10.1016/j.molimm.2020.04.008)

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes one original paper accepted in a peer reviewed journal. The core theme of the thesis is investigating the tumour repertoire of AML and paediatric brain cancer DIPG. Of the published paper and chapters, the ideas, development and writing in the thesis were the principal responsibility of myself, the student, working within the Department of Biochemistry and Molecular Biology under the supervision of Professor Anthony Purcell, Dr Nicole Mifsud and Dr Sri Ramarathinam.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research

In the case of **Chapter 2** my contribution to the work involved the following:

Thesis Chapter	**Publication Title**	**Status**	**Nature and % of student contribution**	**Co-author name(s) Nature and % of Co-author's contribution***	**Co-author(s), Monash student Y/N***
Chapter 2	In-depth mining of the immunopeptidome of an acute myeloid leukaemia cell line using complementary ligand enrichment and data acquisition strategies	Published	60%. Concept and collecting data and writing first draft and revisions	1) Nicole A. Mifsud - 3%, Cell culture 2) Terry C.C. Lim Kam Sian - 3%, Provided cell line, made figures 3) Rochelle Ayala - 5% Technical assistance in experiments 4) Nicola Ternette- 3% Advice on Manuscript 5) Sri H. Ramarathinam - 12% Overall project direction, supervision of data analysis and finalisation of manuscript 6) Anthony W. Purcell –14% Overall project direction, supervision of data analysis and	No No No No No No

				finalisation of manuscript,	

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student name: Kirti Pandey

Student signature: Kirti Pandey **Date: 17/12/20**

I hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author, I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor name: Professor Anthony Purcell

Main Supervisor signature: **Date: 17/12/2020**

ACKNOWLEDGMENTS

The 3.5 years I have spent working on this thesis would have not been possible without the support and guidance of a lot people. I thank you all!

First and foremost, dear respected sir (Tony), I feel words like thankful and grateful cannot embody how I feel about the past 3.5 years, working as a student in your lab. It has been a whirlwind journey and it is all because you. Boss, you are an amazing supervisor and mentor. Thanks for taking a non-mass spec person under your wings and making me realise just how cool it is. You made everything super easy and you always had answers to all my questions (and I did have a lot of them). Thank you, for your kindness, for sharing my excitement when data was good and guidance during not so good times, your unwavering support helped me steer through the convoluted maze that a PhD thesis can be.

Second biggest thanks to Sri, who I felt did another PhD whilst supervising me. I do not know what my thesis would have looked like if I had not hatched a plan of asking Tony to add you as my supervisor. I do not think I can thank you enough for always being so kind and nice (even at times when your student had gone completely bonkers). For being eternally patient and spending your valuable time on discussions involving not only mass specs but even the very minute details that would make a difference to experiments. A massive thanks to Nic for sharing your expertise and vision for my thesis. Your guidance and support since I joined the lab, from day one and throughout the past 3.5 yrs has been invaluable. Lastly, Dr Nadine Dudek for going through my thesis so meticulously, your comments and corrections have been invaluable in shaping the thesis as a finished product.

I think my PhD would have been non-existent if it were not for my parents. Parents oftentimes make the hard choices for their kids (because kids do not want to make it themselves). I still remember, 3.5 yrs back when I was hesitant about moving 15,000 kms away from home my non-religious father asked me to ‘trust god’ and give Monash ‘a try’, hence they kicked me out and send me to Australia because for some reason they believed I would be a ‘good scientist’. So, I guess it is fair to say that even my PhD was born because of you. I would at the same time acknowledge my grandparents (nana and nani) who took care of me when I was a toddler and taught me how to read and write. I know you would have been crazy happy about the thesis.

A massive thanks to all the other members of the Purcell lab. Rochelle who has been lifeline since day one. You taught me so many things and helped me out every step of the way. Thanks for your kind words and care over the years. Annie who has been my partner in crime as she

has never said no to chips, chai or nachos. I think lab life would have been extremely dull if it were not for my fellow PhD students, Terry, Shutao, Shawn. You guys made 3 years super easy with all the get togethers, movie nights and parties. Not just fun stuff but being helpful in the lab, I always knew I could ask any one of you for help.

Most importantly, the families who not only opened their doors but also their hearts and welcomed a stranger into their families. Life would have been very different if it was not for the family lunches and dinners, especially on Christmas day and Easter holidays….I will always be thankful!

Table of Contents

LIST OF FIGURES

CHAPTER 1 - INTRODUCTION

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF TABLES

CHAPTER - 1

CHAPTER - 2

CHAPTER – 3

CHAPTER – 4

CHAPTER – 5

LIST OF ABBREVIATIONS

Abbreviation	Full name
ACN	Acetonitrile
AGC	Automatic gain control
ALC	Average local confidence
ALL	Acute lymphoid leukaemia
Allo-HCT	Allogenic hematopoietic cell transplant
AML	Acute myeloid leukaemia
APC	Antigen presenting cells
β2m	Beta-2-microglobulin
BiTE	Bispecific T cell engager
BM	Bone marrow
C18	Reverse phase column C18
CAR T cell	Chimeric antigen receptor T cell
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CO_2	Carbon dioxide
CSF	Cerebrospinal fluid
CTA	Cancer testis antigen
CV	Column volume
DCs	Dendritic cells
DDA	Data-dependent acquisition
DIPG	Diffuse intrinsic pontine glioma
DMEM	Dulbecco's modified eagle's medium
DNA	Deoxyribonucleic acid
DRiP	Defective ribosomal products
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum

ERAP	Endoplasmic reticulum aminopeptidase
Erp57	Endoplasmic reticulum resident protein
FA	Formic acid
FAB	French-American-British
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
FDR	False discovery rate
FFPE	Formlain fixed paraffin embedded
FLT3	Fms-like tyrosine kinase 3
FSC	Forward scatter
GO	Gene ontology
hr(s)	Hour(s)
HCD	Higher-energy collisional dissociation
HLA	Human leukocyte antigen
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
HSC	Hematopoietic stem cell
IEDB	Immune Epitope Database
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITD	Internal tandem repeats
IU	International unit
LB	Lysogeny broth
LC	Liquid chromatography
LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry
LOH	Loss of heterozygosity
MAGE	Melanoma associated antigen
MAP4K	Mitogen activated protein kinase kinase kinase kinase
MAP3K	Mitogen activated protein kinase kinase kinase

MAP2K	Mitogen activated protein kinase kinase
MAPK	Mitogen activated protein kinase
µF	Microfarad
MHC	Major histocompatibility complex
min	Minute(s)
MLL	Mixed lineage leukaemia
MRM	Multiple monitoring reaction
MS	Mass spectrometry / Multiple sclerosis
MS/MS	Tandem mass spectrometry
MS1	Mass spectrum level 1
MS2	Mass spectrum level 2 (MS/MS)
MW	Molecular weight
m/z	Mas to charge ratio
NaCl	Sodium chloride
NFκB	Nuclear factor kappa B
NK	Natural killer
NPM1	Nucleophosmin 1
PΩ	C-terminal amino acid residue
PAMP	Pathogen-associated molecular patterns
PBG	Peptide binding groove
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PE	Phycoerythrin
PEG	Polyethylene glycol
pHLA	Peptide-HLA
PLC	Peptide loading complex
PD-L1	Programmed cell death ligand - 1
PML-RARA	Promyelocytic leukaemia and retinoic acid receptor alpha
PMSF	Phenylmethanesulfonylfluoride
ppm	Parts per million

PRR	Pattern recognition receptors
PTM	Post-translational modification
RNA	Ribonucleic acid
RP-HPLC	Reverse-phase high performance liquid chromatography
rpm	Revolutions per minute
RPMI	Roswell Park Memorial Institute Medium
RT	Room temperature
RTK	Receptor tyrosine kinase
SDC	Sodium deoxycholate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	Second(s)
SEM	Standard error of the mean
SSC	Side scatter
TAA	Tumour associated antigen
TAE	Tris-acetate-ethylenediaminetetraacetic
TAP	Transporter associated with antigen processing
TBE	Tris/Borate/EDTA
TBS	Tris buffered saline
TBS-T	TBS with Tween-20
TCEP	Tris(2-carboxyethyl) phosphine
TCR	T cell receptor
TFA	Trifluoroacetic acid
TLR	Toll-like receptor
TME	Tumour microenvironment
TNF	Tumour necrosis factor
TOF	Time of flight
TSA	Tumour specific antigen
U	Enzyme unit
v/v	Volume by volume
WHO	World Health Organisation
w/v	Weight by volume

Chapter 1 – Introduction

The immune system offers multi-layered protection from disease

The human immune system can be divided into two major arms: innate and adaptive. 1) The **innate** immune system is the first line of defence against all pathogens. Despite being quick to instigate a response, it is relatively non-specific and not sufficient to combat all infections and diseases. The other arm comprises of the 2) **adaptive** immune system, a highly specific anti-pathogen response that includes the production of antibodies and cytotoxic cells. This helps in not only clearing pathogens but also resolving different diseases. Together, both branches of immunity work to protect us against infectious agents and rogue self-reactive cells.

1.1 Innate Immunity

Innate immunity comprises of four distinct defence barriers, which are segregated based on anatomical, physiological, phagocytic, and inflammatory features. The anatomical and physiological barriers include intact skin, low stomach pH and bacteriolytic lysozyme in tears, saliva, and other secretions. Pathogens that breach the physical barrier encounter the phagocytic cells of the innate immune system including macrophages and dendritic cells (DCs) (Turvey et al. 2010). These cells possess a variety of surface molecules referred to as pathogen recognition receptors (PRRs) that recognise highly conserved structures present only in microbes known as pathogen associated molecular patterns (PAMPs) (Gordon 2002, Medzhitov et al. 2002, Turvey and Broide 2010). When macrophages encounter an inflammatory stimulus, they produce an array of cytokines including interleukins (IL), interferons (IFNs) and tumour necrosis factor alpha (TNF-α) along with chemokines, leukotrienes, prostaglandins, and complement. These soluble effector molecules result in increased vascular permeability, mobilisation, and infiltration of other first responder cells, including neutrophils, natural killer (NK) cells and natural killer T (NKT) cells. These mediators also produce systemic effects such as fever and the production of acute inflammatory response proteins (Arango Duque et al. 2014). Innate immunity thus provides a fast acting, broad layer of immediate protection against pathogenic assault.

1.2 Adaptive immunity

While innate immunity forms a non-specific barrier, adaptive immunity generates a specific response distinguishing itself from innate immunity owing to presence of a) **antigenic specificity**, wherein B cells and T cells can distinguish subtle differences of even a single amino acid between two proteins, b) **diversity**, the immune system can generate a large repertoire of

receptors which can recognise unique structures on the vast array of potential foreign antigens and c) **memory**, once the immune system has recognised and responded to an antigen, it exhibits immunological memory wherein a second encounter with the same antigen induces a rapid and heightened state of immunity, thereby, conferring protection against pathogens after the initial encounter. The adaptive immune system has two branches, humoral and cellular. The **humoral** branch comprises of B cells which secrete antibodies that recognise epitopes expressed or secreted by the pathogens including regions present on soluble proteins, glycoproteins and polysaccharides. The **cellular** branch consists of T cells which recognise peptides presented by human leukocyte antigen (HLA) molecules. These cells orchestrate most of the adaptive immune systems effects including provision of helping B cells to produce high affinity antibodies and the direct elimination of infected or cancerous cells.

1.3 Human leukocyte antigens are important for immunosurveillance

In humans, the HLA genes are located on chromosome 6 in a region of the genome termed the major histocompatibility complex (MHC) which spans approximately 3,600 kilobases (kb). The MHC complex is separated into three sections, two out of those three i.e. HLA class I and class II loci encode a set of membrane glycoproteins that play a crucial part in immunosurveillance. Their function is based on their structure, the origin of the presented peptides and the subtype of T-cells they interact with, HLA molecules are categorised as the following:

HLA class I molecules – comprises of classical HLA-A, -B and -C molecules, which are expressed on all nucleated cells. The major function of the class I gene products is presentation of endogenously derived 8-15 amino acid long peptide antigens to CD8+ T cells. In addition to the classical class I molecules, HLA-G and -F represent non-classical loci, which have variable cell expression profiles. For example, the expression of the class I HLA-G molecules on cytotrophoblasts at the foetal-maternal interface is known to protect the foetus from being recognised as foreign and being rejected by maternal T cells (Hunt et al. 2005). This may occur when paternal antigens begin to appear.

HLA class II molecules- comprises of HLA-DR, -DQ and -DP loci. These glycoproteins are expressed primarily on professional antigen presenting cells (APCs) such as macrophages, DCs and B cells where they present processed antigenic peptides to CD4+ T cells. HLA class II molecules bind peptides that are mostly derived from exogenous, secreted or membrane proteins but can also be derived from cytosolic proteins. These proteins are degraded in the

endo-lysosomal compartments of the cell to a more variable length of 9 to 25 amino acids, compared to the ligands of the HLA class I molecules.

HLA class III genes - encode several secreted proteins including the components of the complement system and molecules involved in inflammation.

HLA class I and II genes are closely linked and therefore are often inherited in a Mendelian fashion from each parent as an HLA haplotype with the alleles co-dominantly expressed in the cells (Bodmer 1987). HLA alleles are highly polymorphic, and it is postulated that this diversity within HLA molecules has evolved under unique selective pressure in different geographic areas. This could be related to the role of the HLA molecules in the presentation of prevalent pathogens in the different areas of the world. HLA polymorphism is not evenly spread throughout the molecule but is clustered in the peptide binding groove (PBG). Amino acid variations in this region alters the peptide binding specificity of HLA allomorphs as these residues govern the characteristic amino acid residue patterns in the bound peptide sequences (Falk et al. 1991). Conserved amino acid residues located at particular positions of the peptides such as P2 and PΩ (the last residue of any peptide) act as anchoring residues in the peptide-binding groove.

The expression of HLA molecules is regulated by various cytokines this included TNF-α and family of IFN proteins i.e. IFN-α, -β and -γ have been shown to increase expression of HLA class I molecules on cells (Végh et al. 1993, Keskinen et al. 1997). IFN-γ induces the formation of a specific transcription factor that binds to the promoter sequence flanking the HLA class I genes (Raval et al. 1998). This leads to transcriptional up-regulation of the genes encoding the class I α chain, β2-microglobulin along with components of the antigen processing pathway including proteasome subunits and the two transporters associated with antigen processing (TAP1 and TAP2) proteins.

This thesis exclusively focuses on HLA class I associated peptides, and these will be discussed in detail below.

1.4 HLA class I molecules present peptides with well-defined binding features

Classical HLA class I molecules (HLA-A, B and C) are structurally heterodimeric and consist of a single 45 kilo Dalton (kDa) transmembrane glycoprotein called the heavy or α chain, which is non-covalently linked to a smaller 12 kDa monomorphic molecule called β-2-microglobulin (β2m) (Cresswell et al. 1973, Parnes et al. 1982). The β2m gene is encoded on chromosome 15 and is essential for stabilisation and cell surface expression of class I molecules (Parnes and Seidman 1982). The heavy chain has three external domains α1, α2 and α3, each containing

approximately 90 amino acids (Bodmer 1987). The α1 and α2 domains interact to form a platform of eight antiparallel β strands bounded by two long α helical regions (Figure 1a). This structure forms a deep groove, known as the PBG, which typically displays peptides of 8–12 amino acids in length, although longer peptides of up to 15 amino acids have been reported to bind as well. The α3 domain adopts an Immunoglobulin (Ig)-like fold and is positioned membrane proximal to α1 and α2, alongside β2m (which also possesses an Ig fold). A C-terminal extension from the α3 domain anchors the complex to the cell membrane. The transmembrane domain has 25 hydrophobic amino acids followed by a short cytoplasmic tail. The α3 domain is highly conserved among HLA class I molecules and contains a sequence that interacts with the CD8 membrane molecule present on certain T cells.

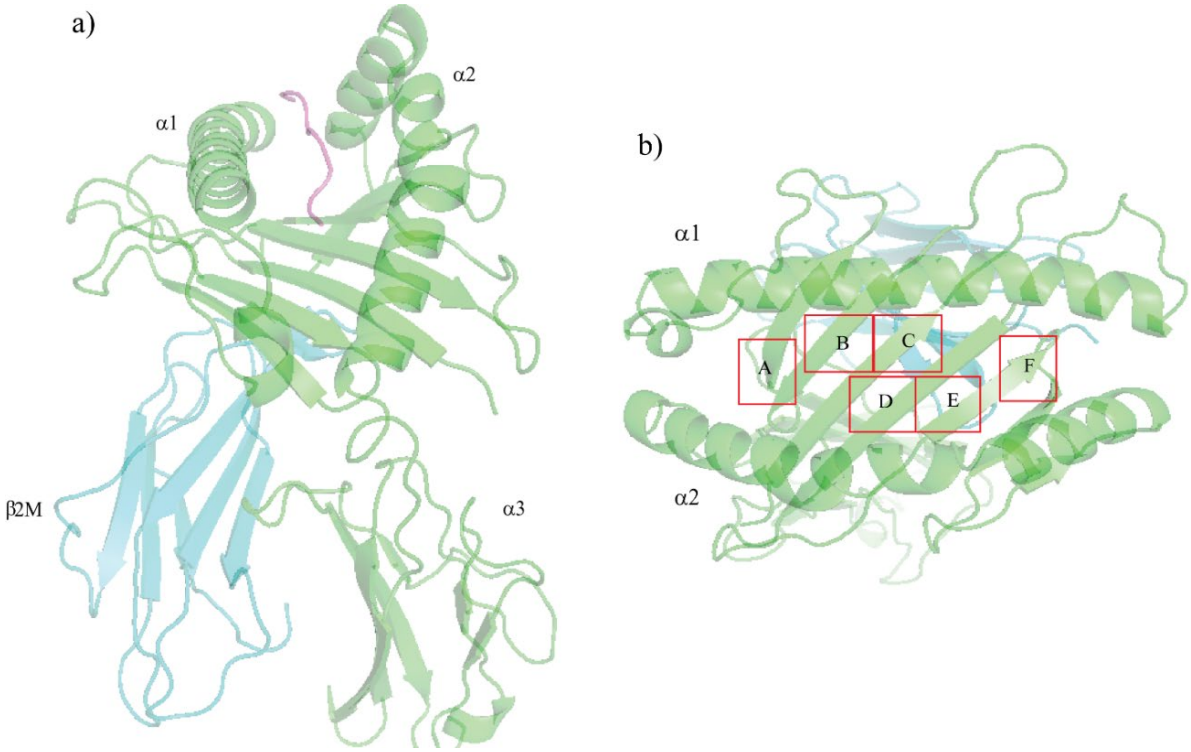


Figure 1 – The structural arrangement of HLA class I molecules. a) A carton representation of HLA A*02:01 in complex with a HIV peptide SLYNTIATL (Protein Data Bank ID: 5NMH) depicting the α1 and α2 domains which form the peptide binding groove whilst the α3 chain of HLA is positioned next to β2M. b) Top view of the peptide binding groove with the peptide antigen removed and highlighting the six pockets (A, B, C, D, E and F) of the HLA A*02:01 molecule.

As mentioned before, most polymorphic residues of HLA class I molecules are localised around the PBG. Changes in amino acid sequence in this region alter the structural shape and electrostatic charge of the groove, which in turn alters the peptide binding properties between

HLA allelic variants or allomorphs (Wieczorek et al. 2017). Key to these allomorph specific interactions are the pockets within the PBG that accommodate the bound peptide. For HLA class I molecules, these pockets are denoted as A, B, C, D, E and F (Figure 1b). The exact topology and electrochemical environment of the anchor pockets favours the binding of specific anchor residues at these sites. As a result, different HLA allotypes will possess different peptide binding preferences. In general, HLA class I molecules present a large and varied repertoire of peptides. Anchors in PBG are often classified as either primary or secondary/auxiliary in nature. **Primary** anchor pockets are highly specific, generally allowing a small subset of similar amino acid side chains to be accommodated. **Secondary** anchor pockets do not place such strict constraints on the peptide but aide in binding affinity. Peptides bind within the PBG using two major non-covalent interactions: a) **conserved contacts**, hydrogen bonds between atoms of the peptide main chain and conserved residues of the HLA heavy chain, and b) **sequence dependent contacts**, between peptide side chains and the polymorphic landscape of the peptide binding groove (Wieczorek et al. 2017). Anchor residues P2 and PΩ reside deep into anchor pocket whilst non-anchor residues of the peptide tend to be more variable as they are often solvent exposed and not intrinsic to binding to the HLA molecule but are known to interact with the T cell receptor (TCR) (Petersen et al. 2009).

1.6 The HLA class I antigen processing and presentation pathway

There are two major proteolytic systems operating within the cell that contribute to the intracellular peptide pool (Rock et al. , Blum 2013). In the cytosol the vast majority of proteolysis is mediated by the proteasome (Maupin-Furlow 2011). The proteasome has a barrel-shaped 20S core structure consisting of four stacked rings of seven subunits each (Blum 2013). The outer rings are composed of α-subunits and the middle two rings of β-subunits, three of which, β1, β2 and β5 (also known as LMP1, LMP2 and MECL1 respectively) constitute the active proteolytic components. The 20S core is capped at each end by a 19S multi-subunit complex that recognises ubiquitin-conjugated proteins marked for degradation (Blum 2013). The process of proteolysis by the 26S proteasome is adenosine 5'-triphosphate (ATP)-dependent and forms the major source for HLA class I peptides. Variants of the β-subunits can be induced by IFN-γ, and are referred to as β1i, β2i and β5i (Gaczynska et al. 1993, Rock and Goldberg 1999). Upon replacement of the constitutive β-subunits with these inducible subunits leads to the formation of the immunoproteasome. The altered composition of the immunoproteasome imparts differences in cleavage specificities (Driscoll et al. 1993, Aki et al. 1994).

Following proteasomal processing, peptides are actively transported into the endoplasmic reticulum (ER) by TAP1 and TAP2 proteins, a member of the ATP binding cassette family (Cresswell et al. 1999, Cresswell 2013). TAP is localised to the ER membrane and its luminal face acts as a scaffold for the assembly of HLA class I molecules (Cresswell et al. 1999). Most proteasomal products are long precursor peptides that are then trimmed at the N-terminal in the cytosol, or later in the ER by aminopeptidases (Rock et al. 2004), referred to as ER aminopeptidase-1 (ERAP1) and ERAP2 in humans. The β2M component of HLA class I molecules are synthesised on polysomes along the rough ER (RER) (Elliott et al. 1990). Assembly of these components into a stable class I HLA molecular complex that can exit the RER requires the presence of a peptide in the binding groove of the HLA class I molecule (Elliott et al. 1990, Blum 2013). The assembly process involves several steps.

HLA class I biosynthesis begins in the ER by trimming of the N-linked glycan by glucosidases I and II (GlsI/ GlsII) to a single terminal glucose residue (G). This permits the interaction of the HLA-I heavy chain with lectin-like chaperones at several stages during folding and assembly. The initial folding events involve the chaperone calnexin, a resident ER membrane protein (Cresswell et al. 1999) (Figure 2). Calnexin associates with the free class I α chain and promotes folding. When β2m binds to the α chain, calnexin is released and the class I molecule associates with the chaperones calreticulin and tapasin (TAP-associated protein) which together with other chaperones and TAP form the peptide loading complex (PLC) (Dong et al. 2009, Panter et al. 2012)(Figure 2). The empty HLA class I heterodimer, which is inherently unstable, binds to calreticulin via the mono glycosylated N linked glycan (Dong et al. 2009, Panter et al. 2012). This stabilises the empty HLA class I molecule and maintains the binding groove in a conformation that favours high affinity peptide loading. Tapasin acts as a bridge and brings the TAP protein into proximity with the class I molecule and allows it to acquire an antigenic or self-peptide (Figure 2). Tapasin is also linked to a second molecule known as endoplasmic reticulum resident protein (Erp57) (Dong et al. 2009). Erp57 is a thiol oxidoreductase that assists in the folding of newly synthesised glycoproteins by mediating disulphide bond formation. The physical association of the α chain/β2m heterodimer with the TAP protein promotes peptide capture by the class I molecule before the peptides are exposed to the luminal environment of the ER (Blum 2013). Peptides that do not bind to class I molecules are rapidly degraded. HLA class I peptides are also sourced from proteins that have undergone post-translational modification like phosphorylation, acetylation and methylation (Trombetta et al. 2005). These modifications can alter the way the peptides interact with HLA

molecules (Zarling et al. 2000). HLA class I molecules with suboptimal peptides are substrates for UDP-Glucose glycoprotein transferase-1 (UGT1) which re-glycosylates the heavy chain glycan, allowing re-entry of the HLA class I molecule into the PLC and exchange for high affinity peptides (Ruddock et al. 2006).

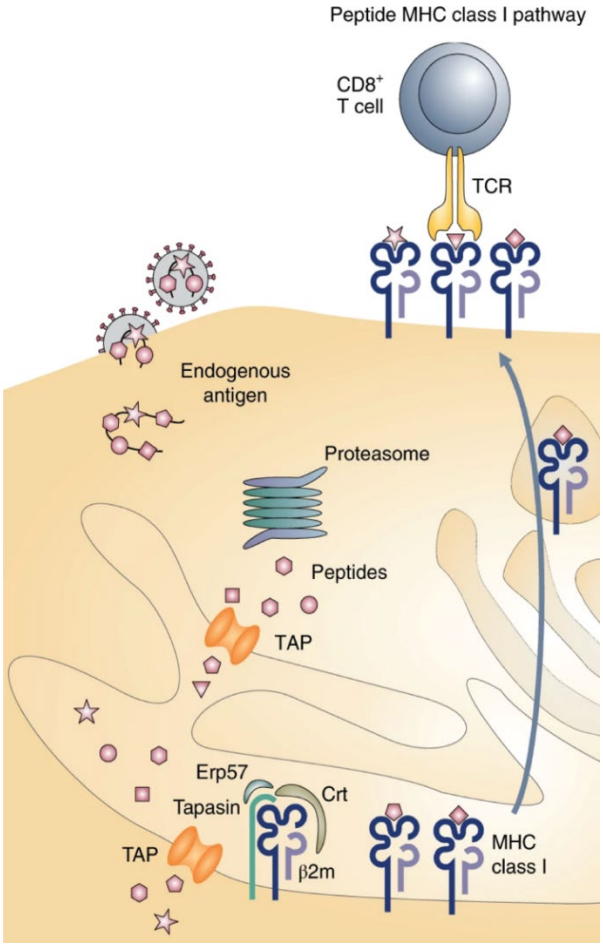


Figure 2 – Assembly and stabilisation of HLA class I molecules. Newly formed class I α chains are stabilised by molecular chaperones in the ER membrane. Subsequent binding to β2m releases these chaperones and allows α chain/β2m heterodimer to bind to calreticulin (Crt) and to tapasin, which is associated with the peptide transporter TAP. This association promotes binding of an antigenic peptide, which stabilises the class I molecule–peptide complex, allowing its release from the ER. Figure adapted from Purcell *et al.* (Purcell et al. 2019).

1.7 T lymphocytes express T cell receptors

Progenitor T lymphocytes arise in the bone marrow (BM) as a result of haematopoiesis. A key process in T cell production is generation of a TCR which is of 2 types, αβ and γδ. This is because the adaptive immune system has to protect the body from a plethora of pathogens and thus a large diversity of αβ TCR capable of recognising any potential pathogen has to be generated. The immune system has developed several mechanisms for TCR generation, this includes **a) germline diversity** which is due to the presence of several variable (V), joining (J) and diversity (D) segments. **b) combinatorial diversity** wherein rearrangement of chromosomal DNA occurs to bring in spaced genes together. To produce a functional αβ TCR all the exons dispersed along the genomic DNA must be joined and transcribed. This process occurs independently for each chain, with genomic DNA rearrangement or recombination. However, gene rearrangement is restricted ensuring that only one fully rearranged TCR chain is produced via the mechanism referred to as allelic exclusion (Abbey et al. 2007). The variable region of the TCR is generated by somatic assembly of components V, D, and J gene segments in a process called V(D)J recombination (Dudley et al. 2005). The TCR α gene is assembled from Vα and Jα regions, whereas the TCR β gene is assembled first by way of the Vβ and Dβ segments, followed by Vβ-Dβ joining to Jβ and finally Vβ-Dβ-Jβ attachment to the Cβ region (Toyonaga et al. 1985, Davis et al. 1988). Transcription of the rearranged genes yields primary transcripts, which are processed to give rise to mRNAs encoding the α and β chains of the membrane bound TCR. The leader sequence is cleaved from the nascent polypeptide chain and the finished αβ TCR is expressed on surface of CD8 T cells.

1.8 Interaction of the αβTCR with the peptide-HLA complex and development in the thymus

All αβ TCRs have a hypervariable loop called complementarity determining regions (CDR), found within the Vα and Vβ domains and constitute the antigen-binding region of the TCR. Each chain has three CDR region (CDR1α, 2α 3α, and CDR1β, 2β, 3β); CDR1 and CDR2 are encoded within V genes and CDR3 is encoded within the DJ genes (Chothia et al. 1988, Jorgensen et al. 1992). The CDR3 is more diverse than the CDR1 or CDR2, due to the incorporation of P and N nucleotides (Davis and Bjorkman 1988, Garcia et al. 1996). The first crystal structure to show this key event was provided by X-ray crystallography in 1996 (Garboczi et al. 1996, Garcia et al. 1996). The structures showed the αβ TCR engaged with the antigenic platform provided by both the HLA molecule and the bound epitope, with the TCR α and β chain above the HLA α2 and α1-helix, respectively. Since then several unique

TCR/peptide/HLA structures have been solved. Another key piece of information gathered from the numerous structures has shown that TCRs assume a strikingly similar docking orientation that is roughly diagonal across the path of the antigenic peptide (Collins et al. 2008). The Vα CDR1 is aligned over the N-terminus of the bound peptide, the Vα and Vβ CDR3 cover the central position of the peptide between the HLA helices and the Vβ CDR1 is over the C-terminus of the peptide. Finally, the Vα and Vβ CDR2 interact primarily with the HLA helices (Bjorkman et al. 1987, Garcia et al. 1996).

Differences in TCR conformation between an unbound and bound state have shown a degree of CDR flexibility to facilitate TCR docking (Kjer-Nielsen et al. 2002, Reiser et al. 2002, Housset et al. 2003). In addition, TCRs in an unliganded state adopt a variety of configurations, with studies suggesting only one configuration is proficient for binding to a specific peptide/HLA (pHLA) complex. Interaction between the TCR and pHLA is critical in determining T cell activation and fate. Several studies by Evavold and Allen showed modifications that affect the ability of the TCR to bind pHLA can alter the ensuing T cell response(Evavold et al. 1991, Evavold et al. 1993). These studies highlighted that by mutating a single TCR-contacting amino acid within a murine haemoglobin peptide the resulting T helper cell response changed from cytokine production and proliferation to cytokine production only.

The T cell repertoire in any individual is generated in the thymus by an extraordinary selection process involving two distinct stages. The first stage is referred to as **positive selection and** permits the survival of only those T cells whose TCRs recognise self-HLA molecules. This process is responsible for generation of a self-HLA-restricted repertoire of T cells. In contrast, the second stage is referred to as **negative selection,** which eliminates T cells that react too strongly with self-HLA. This process is extremely important in generating a primary T cell repertoire that is self-tolerant and would not lead to autoimmunity.

T cells can be broadly categorised in two subtypes, CD8+ and CD4+. CD8+ T cells are activated upon recognition of HLA class I associated peptides. These cells mature into cytolytic cells, though the level of cytotoxicity or cytokine production may vary depending on the stimulus. CD4+ T cells are generally activated upon recognition of HLA class II associated peptides on APCs. More recently CD8+ and CD4+ T cells with regulatory T cell (Treg) properties (CD8+CD25+Foxp3+ and CD4+CD25+Foxp3+) have also been identified. The function of this subset of T cells is immune suppression in order to avoid conflagration of immune response particularly after an acute T helper 1 (Th1) inflammatory responses in order to maintain cellular homeostasis (Yu et al. 2018).

2.CANCER AND IMMUNITY

2.1 The origin of cancer

Cancers are diseases that are caused by genetically deviant cells which proliferate autonomously (Hanahan et al. 2000). The genetic damage consists of mutations (including point mutations, deletions and inversions) and chromosomal rearrangements or losses (Hanahan et al. 2011). Harbouring such oncogenic alterations assist cells to acquire traits such as uncontrolled proliferation and inhibition of cell death thereby causing malignant transformation of the cells. Once malignancy is established, these transformed cells acquire other traits such as reduced growth factor requirements, loss of specialised cell functions, changes in cell motility and the ability to metastasise (Figure 3) (Hanahan and Weinberg 2000). Most initiating mutations affect either/both proto-oncogenes or tumour suppressor genes (TSGs) (Hanahan and Weinberg 2000, Hanahan and Weinberg 2011). Proto-oncogenes code for a variety of growth factors and their receptors (usually tyrosine kinase receptors [RTK] such as FLT, EGFR, VEGFR), enzymes (CDKs, HRAS, NRAS) or transcription factors (MYC, JUN, FOS) that promote cell growth and/or cell division. Mutated versions of proto-oncogenes that promote abnormal cell proliferation are called oncogenes. Another set of genes that play a role in onset of cancer are TSGs (Hanahan and Weinberg 2000, Hanahan and Weinberg 2011), whilst their normal function is suppression of carcinogenesis, their loss facilitates tumour development.

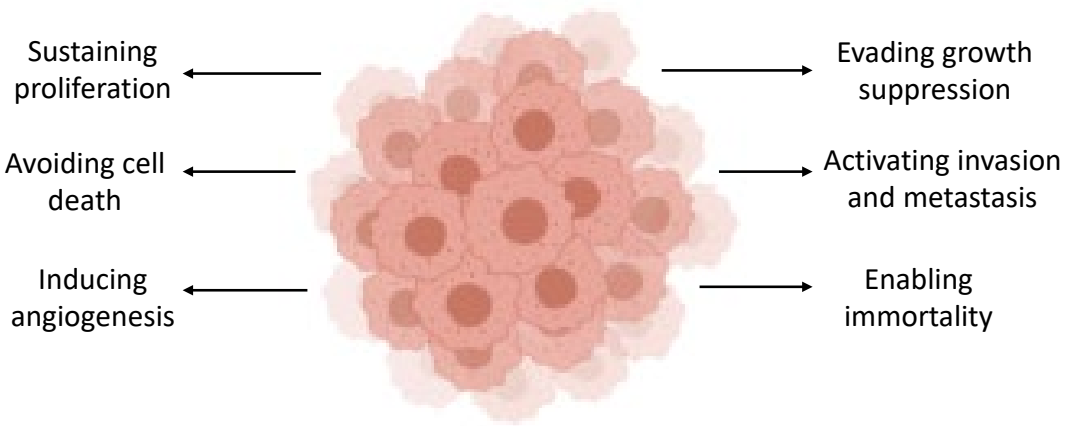


Figure 3 - The six hallmark capabilities acquired by cancers. The six key traits acquired by neoplastic cells eventually leading to cancer. Figure adapted from Hanahan *et al.* (Hanahan and Weinberg 2000).

Two additional features that are involved in cancer pathogenesis are 1) the ability to reprogram cellular metabolism to support neoplastic proliferation, and 2) the ability to evade immune recognition and destruction by T and B lymphocytes, macrophages and NK cells (Rodríguez 2017). Thus, understanding the interplay between the tumour, tumour microenvironment and the immune system is of importance (Rodríguez 2017) as it would guide vaccine design strategies to be adopted for treating patients.

The multistep process in which an apparently normal cell is converted or transformed into a malignant cell is comprised of three stages: initiation, promotion, and progression (Barrett 1993). During the **initiation** phase of carcinogenesis, driver mutations occur that provide a selective growth advantage to a cell and thereby promote cancer development. They influence either/both proto-oncogenes and TSG. Driver mutations in proto-oncogenes are activating or result in new functions, whilst driver mutations in TSGs are inactivating. Mutations in the driver genes result in the dysregulation of biochemical signalling pathways associated with uncontrolled cellular proliferation, survival, and differentiation. This provides the transformed cell with survival and growth advantages over normal cells. The second stage of carcinogenesis is **promotion** which is considered to be a relatively lengthy and reversible process in which actively proliferating preneoplastic cells accumulate (Pitot 1993). Within this period, the process can be altered by chemo preventive agents that affect growth rates. Following initiation and promotion, cells undergo the final process of **progression** wherein the already genetically vulnerable precancerous cells accumulate successive genetic changes including other driver mutations or passenger mutations (that do not provide a selective growth advantage). Phenotypically this leads to dysregulation in downstream signalling pathways giving rise to increasingly malignant subpopulations (Couch 1996). Progression is the phase between a premalignant lesion and the development of invasive cancer (Pitot 1993, Arvelo et al. 2016). As tumour progression advances, cells lose their adherence properties, detach from the tumour mass, and invade neighbouring tissues. The detached cells also enter the circulating blood and lymph and are transported to other organs/tissues away from the site of the primary growth and develop into secondary tumours at these new sites (Pitot 1993). The progress of the neoplastic disease depends on metastatic changes that facilitate; (a) invasion of local normal tissues, (b) entry and transit of neoplastic cells in the blood and lymphatic systems and (c) the subsequent establishment of secondary tumour growth at distal sites (Pitot 1993).

2.2 Cancer immunity is complex and includes subverting immune surveillance by immune editing

The concept of cancer immunosurveillance was first proposed in 1909 by Ehrlich (Muenst et al. 2016) according to which tumours are constantly identified and eradicated by the host immune system even before their clinical manifestations occur. The concept was further refined by Burnet in 1970 when he proposed that genetic changes leading to malignancy are commonplace in somatic cells and that the immune system is responsible for eliminating such malignant cells (Muenst et al. 2016). However, the fact that malignant tumours also develop in patients with a fully functional immune system suggests that immunosurveillance is only a part of the process. Over the years the concept of immunosurveillance has been refined to incorporate a theory known as 'immunoediting'; a dynamic process that involves not only tumour prevention, but also shapes the immunogenicity of developing tumours. Any tumour growth within the body involves three sequential phases**: i) elimination, ii) equilibrium and iii) escape** (Rodríguez 2017).

In the first stage of elimination, the cancer immunity cycle which consists of seven key steps (Figure 4) needs to kick in (Chen et al. 2013). The first step of the cancer immunity cycle involves release of plethora of antigens created as a direct result of oncogenesis. These peptides belong to three broad categories. **1) Tumour-specific antigens** (TSAs) or neoantigens derived from proteins containing driver mutations. **2) tumour-associated antigens** (TAAs) which are proteins with low levels of expression in normal host cells but disproportionately high expression in tumour cells and 3) **cancer testis antigens** (CTA) are proteins absent on normal adult cells, except in reproductive tissues (e.g. testes, foetal ovaries, and trophoblasts) but are selectively expressed by various tumour (Ferguson et al. 2011).

In step 2, APCs such as macrophages and DCs capture these antigens and present them on HLA class I and HLA class II molecules to T cells, resulting in step 3 wherein the priming and activation of effector CD8+ and CD4+ T cell responses occurs. In step 4, the activated effector T cells traffic to and infiltrate the tumour bed (step 5). Step 6 involves TCR specifically recognising and binding to antigen-HLA class I complex present on cancer cells thereby facilitating step 7 which is T cell mediated targeted killing of the cancer cell. Importantly, killing of the cancer cell releases additional TAAs (step 1 again) to increase the breadth and depth of the response in subsequent revolutions of the cycle (Ferguson et al. 2011, Chen and Mellman 2013). These 7 cancer immunity cycle steps result in elimination of the tumour.

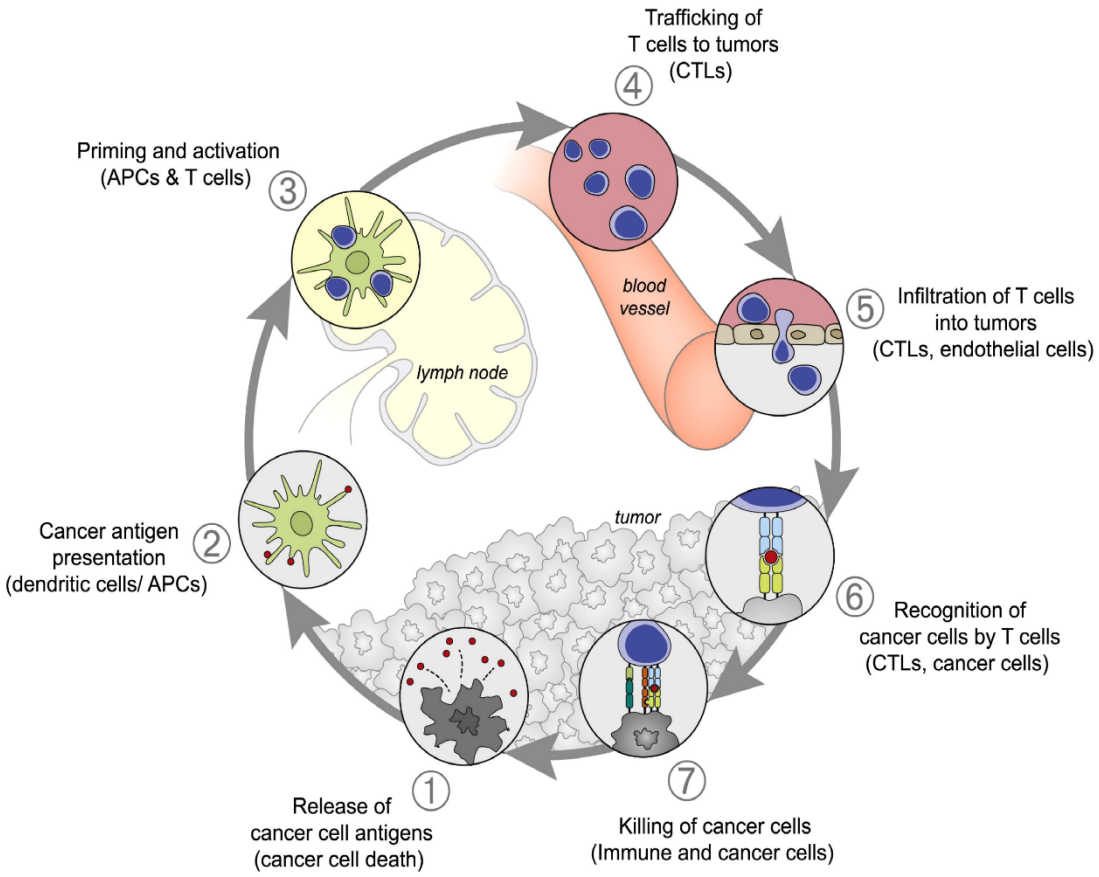


Figure 4 - The cancer immunity cycle that leads to recognition and elimination of cancer cells. Figure from Chen *et al* (Chen and Mellman 2013), copyright permission granted by Elsevier with copyright licence number 4897400429092.

However, some tumour cell variants that survive the elimination phase enter into the second phase which is an equilibrium stage. At this stage the immune system controls tumour outgrowth, but the tumour remains clinically undetectable. For the cancer to progress and spread undetected, tumour cell plasticity results in generation of tumours with low immunogenicity, thereby enabling tumour evasion of immune surveillance (Rodríguez 2017). At the equilibrium stage tumour cells employ a variety of mechanisms to evade immune recognition including downregulation or loss of either HLA or β2M along with mutations influencing other key proteins of antigen processing and presentation machinery such as TAP. This results in decreased TSA, TAA and CTA antigen presentation which dampens the immune response required for elimination of the tumour (Rovatti et al. 2020). Eventually, the third stage of escape is reached. At this stage tumour cells with reduced immunogenicity eventually continue growing; tumour dormancy is broken, and tumours progressively establish an immunosuppressive microenvironment, becoming clinically apparent.

It is known that in cancer patients, the cancer immunity cycle does not perform optimally (Chen and Mellman 2013). Tumour antigens may not be detected, DCs and T cells may treat antigens as self rather than foreign thereby creating Treg responses rather than effector responses, T cells may not properly home to tumours, or may be inhibited from infiltrating the tumour. Factors in the tumour microenvironment might suppress those effector cells that are produced (Chen and Mellman 2013).

2.3 Tumour microenvironment shapes the fate of cancer

The tumour microenvironment (TME) is complex and plays a critical role in tumour origin, maintenance, and its eventual control or escape. The TME is an interplay between different immune cell types, cytokine profiles, and molecular signatures (Rodríguez 2017). At the initial tumour growth stage, tissue damage induces acute Th1 inflammatory responses that favour APC maturation and innate immune cell polarisation, promoting the elimination of developing tumours. APC maturation initiates adaptive immune responses mediated by CD4+ and CD8+ T cells. During inflammation, chemokines control immune cell movement, immune response polarisation and T-cell and DC interactions, while cytokines mediate intercellular communication in the immune system and function as immune regulators. Chemokine and cytokine expression profiles modulate the functional status of the immune system to negatively impact tumour development and progression (Petersson et al. 1998).

In cancer development, TME can be of two kinds pro- or anti-tumour and this is controlled by tumour infiltrating immune cells which modulate the tumour immune response. Anti-tumour microenvironment is inflamed (hot) in nature. Such tumours are characterised by increased presence of T cells particularly CD8+ T cells, cytokines such as IL-2 and other chemokines that mediate high HLA class I and II expression, effector T cell recruitment, macrophage and even plasma B cells. This type of tumour has higher response rates to immunotherapy such as programmed cell death ligand (PDL)-1 therapy (Gajewski et al. 2006). The other category of tumours is pro-tumour or non-inflamed (cold) type. They are characterised by reduced levels of chemokine production, resulting in the poor attraction of CD8+ effector T cells into the tumour mass and poor effector cell trafficking. In addition, high levels of vascular markers also favour tumour growth along with recruitment of tolerogenic DCs, myeloid-derived suppressor cells (MDSCs) and Tregs recruited by the tumour to create an environment with anti-inflammatory properties that favours tumour growth and survival (Chen and Mellman 2013, Rodríguez 2017).

2.4 Modulation of HLA expression in cancer

HLA alleles play a crucial role in antigen presentation which drives immune surveillance and the detection and elimination of malignant cells. Since the process is central to generating an effective anti-tumour response, in several cancers; including pancreatic, ovarian, breast and non-small-cell lung cancer (NSCLC); there is downregulation or complete loss of cell surface HLA expression (Rovatti et al. 2020). HLA-altered tumour cell phenotypes can be classified into two main groups (Rodríguez 2017, Rovatti et al. 2020); 1) **reversible** regulatory or 2) **irreversible** structural defects. Reversible HLA class I regulatory defects may occur at any step of synthesis, assembly, transport and/or molecular surface expression. These events are caused by genetic, epigenetic, transcriptional or post-transcriptional events but HLA expression can be restored either by drugs or cytokines such as IFN-γ.

Reversible downregulation of HLA in context of leukaemia is a direct consequence of the numerous yet commonplace mutations in key proteins such as transcription factors (TFs), chromosome modulators (Rovatti et al. 2020). More recently, downregulation of HLA expression in AML has also been linked to epigenetic silencing due to a highly prevalent Nucleoplasmin 1 (NPM1) and promyelocytic leukaemia and retinoic acid receptor alpha (PML-RARA) oncofusion (Dufva et al. 2020). These mutations cause hypermethylation of the CpG rich promoters such as the class II transactivator (CIITA) promoter of HLA class II genes (Dufva et al. 2020) which leads to reduced expression of HLA class II molecules. There have not been many studies exploring the role of hypermethylation in silencing of HLA class I except for Nie et al (Nie et al. 2001) who reported this mechanism in oesophageal squamous cell carcinoma. Both HLA class I and II expression can be rescued with drugs including decitabine or IFN-γ (Nie et al. 2001, Dufva et al. 2020). Finally, the role of cytokine mediated pathway of HLA silencing in leukaemia by anti-inflammatory cytokines including IL-4, IL-10 and TGF-β is well known (Lee et al. 1997, Tarafdar et al. 2017) thereby rendering HLA molecules less immunogenic and susceptible to T cell recognition. For instance, IL-10 in the tumour microenvironment may result in decreased TAP1 and TAP2 expression and function, resulting in low peptide translocation into the ER that affects HLA-I-mediated antigen presentation, thereby promoting immunosuppression and leading to tumour immune escape (Rabinovich et al. 2007).

In contrast, the irreversible effects are a result of structural defects caused by chromosomal aberrations that disrupt HLA class I heavy chain and β2m gene synthesis and result in complete loss of HLA class I. This frequently occurs due to the coincidence of two molecular events,

either mutation of one β2m gene, or loss of heterozygosity (LOH) of a HLA allele (McGranahan et al. 2017, Rodríguez 2017), which is a frequent mechanism for HLA haplotype loss particularly in solid tumours including NSCLC, cervical cancer (McGranahan et al. 2017). The loss of β2M has been reported in bladder and colon cancer (Maleno et al. 2011).

To conclude, altered HLA I expression on the tumour cell surface is an early and frequent event that promotes carcinogenesis, as HLA-I is critical for the immune recognition of tumour cells and signalling between tumour and immune cells.

2.5 A tale of two cancers

Cancer is a global health burden and based on estimates produced by the International Agency for Research on Cancer, worldwide there were an estimated 18.1 million new cancer cases and 9.6 million cancer deaths in 2018 (Bray et al. 2018). Tumour classification was originally based upon tumour site (Bosman 2019) and histological characteristics. Using this relatively straightforward approach, a globally accepted taxonomy was developed by the world health organisation (WHO) for classification of human tumours (Bosman 2019). With the advent of methods to explore tumour DNA, RNA and the molecular composition of tissue samples, tumour classification became more stratified.

Cancer is a highly dynamic disease, as its patterns vary worldwide, with different cancers prevalent in different populations and sexes, reflecting different risk factors, genetics and environmental factors. For instance, melanoma is one of the most common forms of cancer in Australia whilst liver cancer is most prevalent in south-east Asian countries including China, India and Taiwan. At the same time men are more likely to have incidences of bladder, prostate and liver cancer whilst the most frequent cancer occurring in women are breast, thyroid and cervical malignancy.

This thesis will focus on two cancers, acute myeloid leukaemia (AML), and a subcategory of high-grade astrocytic tumour referred to as diffuse intrinsic pontine glioma (DIPG).

2.5.1 Acute myeloid leukaemia

2.5.1.1 Haematopoiesis and lineage segregation

All mature blood cells originate from hematopoietic stem cells (HSCs) present in the BM and are traditionally categorised into two distinct lineages: lymphoid and myeloid. As multipotent HSC become more differentiated, they lose their ability to self-renew and undergo lineage commitment (Miyamoto et al. 2002). At this stage they are referred to as multipotent progenitor cells (MPP). Further differentiation of MPP cells result in generation of lineage restricted cells

which morphologically, phenotypically and functionally different from one another. These cells are the common lymphoid progenitors (CLP) that can generate all types of lymphoid cells including T, B, and NK cells. Its counterpart, the common myeloid progenitor (CMP) is the precursor of all myeloid lineage cell types including different subsets of granulocytes (neutrophils, eosinophils, and basophils), monocytes, macrophages, erythrocytes, megakaryocytes and mast cells.

2.5.1.2 Classification of acute myeloid leukaemia

AML accounts for ~25% of all adult-onset leukaemia (DiNardo et al. 2016), with the median age of diagnosis at 67 years (Kumar 2011). AML is an HSC-derived malignancy (DiNardo and Cortes 2016) with pathogenesis characterised by presence of a minimum 20% blast cells (i.e. myeloblasts) in the bone marrow or blood with myeloid lineage. The presence of these neoplastic immature precursors prevents normal bone marrow haematopoiesis (De Kouchkovsky et al. 2016) (Papaemmanuil et al. 2016). AML arises as a result of chromosomal abnormalities, which result in formation of chimeric proteins that hinder the normal maturation process of HSC/CMP cells into functional myeloid cells (Estey 2016). This is also mediated by various chemokines produced by malignant blast cells (Youn et al. 2000). The two common AML classifications are the French-American-British (FAB) and the World Health Organization (WHO) systems. The FAB classification represents the first attempt to distinguish between different types of AML (Bennett et al. 1976). Established in 1976, it defines eight subtypes (M0 through M7) based on the morphological and cytochemical characteristics of the leukemic cells. In 1999, the WHO classification was introduced to include modern prognostic factors, such as molecular markers and chromosome translocations (Harris et al. 1999) (Kumar 2011).

Before the WHO classification, only cytogenetic analysis was used to detect chromosomal abnormalities and formed the first prognostication schema in AML (DiNardo and Cortes 2016). It helped in stratifying patients into favourable, intermediate or adverse cytogenetics, which was found to correlate with 5-year overall survival (Grimwade et al. 1998). Cytogenetic abnormalities could be detected in ~50% to 60% of newly diagnosed AML cases (Kumar 2011). There is a high degree of heterogeneity in AML with respect to chromosome abnormalities as it includes deletions, translocations and inversions (Table 1).

Table 1 - Cytogenetic profiling and associated prognostic group in AML

Prognostic risk group	Cytogenetic profile
Favourable	t(8:21)(q22;q22) inv(16)(p13;1q22) t(15;17)(q22;q12)
Intermediate	Cytogenetically Normal (CN-AML) t(9;11)(p22;q23) Cytogenetic abnormalities not included in the favourable or adverse prognostic risk groups
Adverse	inv(3)(q21q26.2) t(6;9)(p23;q34) 11q abnormalities other than t(9;11) − 5 or del(5q) − 7 Complex karyotype

Over the past decade, DNA sequencing technology has revealed recurrent AML gene mutations, undetectable by standard cytogenetic analysis, which additionally contribute to AML pathogenesis (Papaemmanuil et al. 2016) (DiNardo and Cortes 2016). Several studies have reported that ~95% of AML cases harbour at least 1 somatic alteration (Figure 5) (Schlenk et al. 2008, Patel et al. 2012, DiNardo and Cortes 2016). Gene mutations have helped further refine risk stratification based on cytogenetic changes. This was particularly helpful in the case of cytogenetically normal (CN) AML patients who had a normal karyotype and represent the largest subset of AML occurring in ~40%-50% of all patients (Falini et al. 2005, Kumar 2011). These CN patients were categorised in the intermediate-risk group (Grimwade et al. 1998), yet the group is quite heterogeneous, and not all patients in this subset had the same response to treatment. Many of these gene mutations have prognostic implications, independently and/or in the presence of certain co-occurring driver mutations (gene-gene interactions), and targeted therapeutic strategies directed at specific mutations and molecular classes are currently under development.

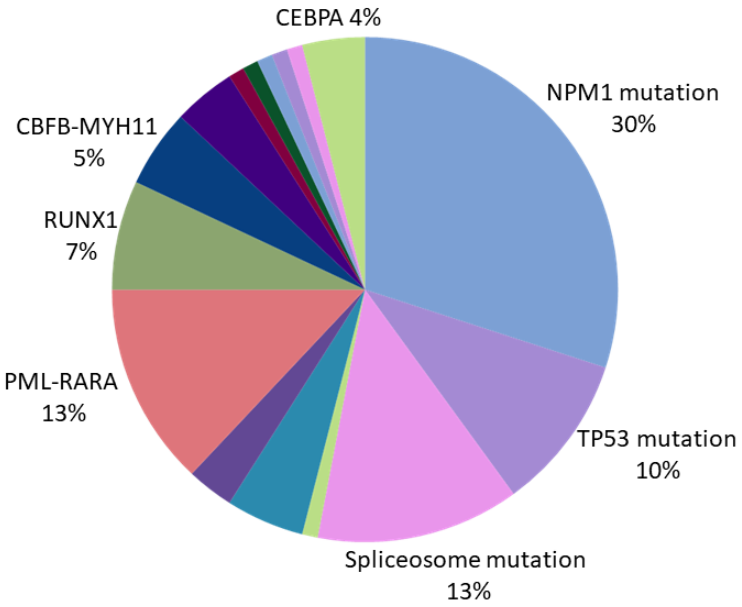


Figure 5 - Distribution of predominant genetic aberrations detected in AML. Pie chart representing distribution of mutations most commonly found in AML. NPM1 mutations are found in ~30% of cases and represent the most common mutations, followed by PML-RARA and spliceosome machinery, followed by RUNX1. The remaining fractions (unlabelled) comprise less frequent mutations. Data from Estey *et al.* (Estey 2018).

These genetic alterations fall into two main groups; class I and class II (Kumar 2011, DiNardo and Cortes 2016, Estey 2016),which offers a conceptual framework for classification. **Class I** alterations comprise of mutations that tend to perturb signalling pathways of well-known tyrosine kinases such as stem cell growth factor receptor Kit (KIT), FMS-like tyrosine kinase 3 (FLT3), PTPN11 and RAS. This results in activation of signal transduction pathways and thereby provide pro-proliferation signals to hematopoietic progenitor cells. **Class II** alterations influence transcription factors thereby resulting in dysregulation of components of the cell cycle machinery and resulting in impaired differentiation. Some genes belonging in the class II include mixed lineage leukaemia (MLL), CCAAT/ enhancer-binding protein α (CEBPα), nucleophosmin1 (NPM1) and RUNX1 (Kumar 2011, DiNardo and Cortes 2016). The best example of such a protein in AML is nucleophosmin (NPM1), a nucleo-cytoplasmic shuttling protein. The mutant NPM1 protein cannot perform its shuttling process efficiently and therefore tends to stay in the cytoplasm for a longer time, where it is susceptible to proteasomal degradation and processing by the HLA class I antigen presentation pathway (van der Lee et al. 2019). More recently, a third category of mutations has emerged, which are found in ~50% of CN-AML (Kumar 2011, DiNardo and Cortes 2016) and influences epigenetic regulators

such as DNMT3A, TET2 and IDH. The three different classes of mutations affect different genes (Table 2) which results in excessive proliferation of abnormal immature leukemic cells known as blasts. The genetic changes in leukemic blasts make them ineffective at generating mature red blood cells, neutrophils, monocytes, and platelets. Hence incorporating gene mutations has helped to further refine risk stratification improving not just prognosis of the disease but also informing treatment options and deciphering the overall survival of patients.

Table 2 - Functional classes and corresponding genes commonly affected in AML

Functional class	Genes affected
Signalling and kinase pathways	FLT3, KIT, KRAS, NRAS and PTPN11
Nucleophosmin	NPM1
Transcription factors	CEBPA, RUNX1 and GATA2
Epigenetic modifiers	DNMT3A, IDH1/2, TET2, MLL, KMT2A and EZH2
Spliceosome complex	SFRB1, SRSF2, U2AF1 and ZRSR2
Tumour suppressors	TP53
Cohesin complex	RAD21, STAG1, STAG2 and SMC3

2.5.1.3 Role of receptor tyrosine kinases in AML

The genes involved in leukemogenesis are related to various cellular functions (Table 2), including ligand-receptor interactions, signal transduction, intracellular localisation, cell cycle control and apoptosis (Hanahan and Weinberg 2011). These enabling signals are conveyed by growth factors that bind to receptors present on cell surface typically containing intracellular tyrosine kinase domains which relay the information downstream. Dysregulation of these signals lead to the emergence of leukemic cells. Particularly in AML, there is disruption in the class III of RTK family, which includes FLT3 (Grafone et al. 2012). Most frequently observed are internal tandem duplications (ITDs) and missense point mutations in the tyrosine kinase domain. These mutations lead to the overexpression or constitutive activation of the RTK. Many studies indicate that patients with FLT3 mutations have a worse prognosis than patients without FLT3 alterations (Estey 2009). In particular, the presence of an FLT3-ITD correlates with an increased risk of relapse and impaired overall survival (Estey 2009). Some other proteins of the RTK family that also get dysregulated are platelet-derived growth factor receptor (PDGFR), macrophage colony-stimulating factor receptor (MCSFR) and KIT, which share the overall same structure. As a result of this there is a dysregulation in downstream signalling in tumour cells which is not found in non-malignant cells. These mutations provide

an advantage to malignant cells, manifested as uncontrolled cell growth, differentiation and evasion of apoptosis (Meyer et al. 2009). Some other mutations acquired by malignant leukemic cell including NRAS and KRAS (DiNardo and Cortes 2016) influencing some other key pathways including Ras/Raf/MEK/ERK, MAP2K/MAPK/JNK, and JAK/STAT signalling (Krueger et al. 2006).

To conclude, research over the years in the field of haematological malignancies at genomic, transcriptomic and protein level has been extensive. This has highlighted the various genetic factors and biological pathways contributing to HM. In contrast, tumours of the central nervous system (CNS) are not well studied, with very little information regarding various key genetic factors such as mutations, chromosomal aberrations and pathways that lead to oncogenesis. The second cancer examined in this thesis involves studying an aggressive paediatric, grade IV astrocytoma known as DIPG which only recently (in 2016) was made into a separate cancer category under WHO classification.

2.5.2 Diffuse intrinsic pontine glioma

Brain tumours are the leading cause of cancer-related mortality and morbidity in children (Fitzmaurice et al. 2017). According to the WHO, primary tumours of the CNS are classified from grades I to IV, based on their morphology, grade of malignancy, proliferative index, response to treatment and survival time (Urbanska et al. 2014, Louis et al. 2016). Grade I brain cancers include non-malignant tumours such as pilocytic astrocytoma. Grade II categorisation is used for relatively non-malignant tumours including central neurocytoma and ependymoma. Grade III brain cancers include low-grade malignancies such as anaplastic ganglioma and meningioma, while grade IV denotes the most malignant brain tumours typically with median survival of 6–12 months (Urbanska et al. 2014, Louis et al. 2016). Diffuse midline gliomas (DMG) encompass a heterogeneous spectrum of disease and includes both adult tumours such as glioblastoma multiforme (GBM) and paediatric gliomas including DIPG. These tumours are mainly categorised as high-grade (III and IV) according to WHO criteria (Louis et al. 2016) and tumour cells share morphological features with normal glial cells.

DIPG is a subtype of high-grade astrocytoma (grade IV category) occurring in the brainstem or the ventral pons of children (Figure 6) (Khuong-Quang et al. 2012). DIPG is spatially and temporally restricted, with diagnosis in infants as young as 3 years of age but with peak incidence between 6 to 9 years of age (Filbin et al. 2018). DIPG diagnosis is based on a combination of clinical and radiologic parameters that include neurological signs, duration of

symptoms and specific neuro-imaging findings (Khuong-Quang et al. 2012, Wang et al. 2019). Radiation is the mainstay of therapy but offers only symptom control, and so far, chemotherapy has shown no benefit (Wang et al. 2019). Surgery is not an option due to the location and the infiltrative nature of the tumour.

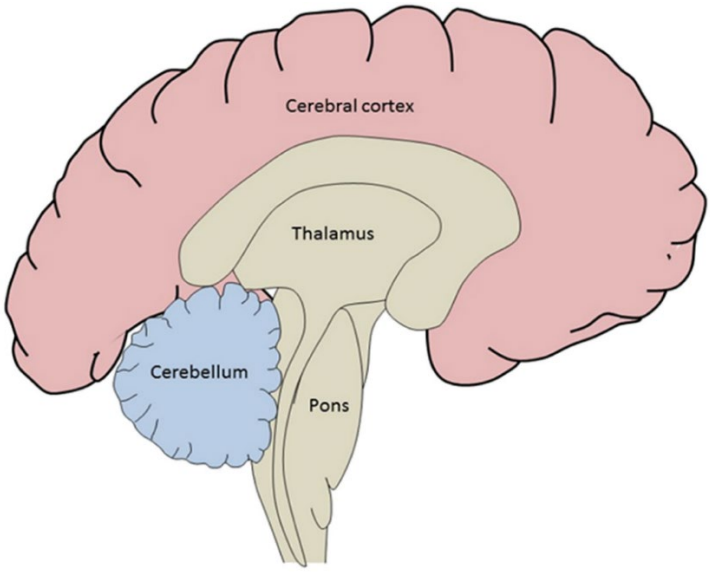


Figure 6 - Anatomic location of site of origin of DIPG tumour. DIPG arises in the pons brainstem region of the brain in children. Figure from Juratli *et al.* (Juratli et al. 2018). Copyright permission granted by Elsevier with copyright licence number 4897391067327.

Previously, DIPGs were classified along with their adult GBM counterparts, but in 2016 they were made a separate category (Louis et al. 2016). Distinction of paediatric DIPGs was driven by mounting evidence that paediatric brain tumours differed from adult GBMs not only in clinical presentation, but also in their unique developmental origins and specific driver mutations (Jones et al. 2014, Wang et al. 2019). For instance, mutations in proteins such as IDH1/2, BRAF, FGFR1 which occur quite frequently in glioma were not found in DIPG (Taylor et al. 2014). Schwartzentruber *et al* (Schwartzentruber et al. 2012) was the first study to highlight driver genetic mutations existing in DIPG using whole exome sequencing (WES). They showed that DIPG molecular pathogenesis was correlated with a missense mutation in histone protein 3.3 (H3.3), wherein a lysine at position 27 changed to methionine (hereby referred to as K27M). The K27M mutation was found in ~36% of cases (n=32/90) compared to only 3.4% (11/318) in adult GBM cases. Other paediatric glioma mutations identified in the same study were in alpha thalassemia/mental retardation syndrome X-linked (ATRX) (35% of

cases, 40/113) and histone chaperone protein, death-associated protein 6 DAXX (6%, 7/124 cases) proteins. This finding was confirmed by Wu *et al.* (Wu et al. 2012), wherein they also reported H3.3K27M mutation occurring in 60% of DIPG cases (n = 30/50) and at a lower percentage 18% (n = 18/50) in histone protein H3.1.

Studies have not only recapitulated these findings but identified other proteins potentially being dysregulated in DIPG including TP53, PDGFRA gain or amplification that was observed exclusively in patients carrying the H3.3K27M mutation (Khuong-Quang et al. 2012). In another study by Buczkowicz *et al* reported numerous previously unreported mutations such as R206H and G328E/V in activin receptor type-1 (ACVR1) protein which is known to constitutively activate the BMP/TGF-β signalling pathway pathways (Buczkowicz et al. 2014, Taylor et al. 2014). An alternate study also highlighted several DIPG prognostic markers, which included chromosomal abnormalities involving homozygous deletion of 9p21 and loss of 10q found in 11% and 47% of cases, respectively (Korshunov et al. 2015). A clear outcome of all the studies was that the survival rate of children with the H3.3K27M mutation was significantly reduced to only 9-12 months when compared to children with wild type histone protein who lived up to 4.5 years (Khuong-Quang et al. 2012). At a cellular level, the H3.3K27M mutation results in loss of function as it hinders methylation or acetylation at K27, which alters epigenetic regulation, transcriptional activation or repression of different downstream genes (Schwartzentruber et al. 2012, Wu et al. 2012).

Most DIPG studies to date are genomics based, and only a few studies have explored the proteomic landscape of DIPG patients. In a study by Saratsis *et al* (Saratsis et al. 2012), proteomic analysis of cerebrospinal fluid (CSF) from DIPG patients (n=10) resulted in identification of 507 unique proteins found exclusively in DIPG patients when compared with GBM (n=1) and healthy controls (n=4). Out of these, 334 proteins were upregulated (fold change 2 or more) in DIPG samples including CypA and DDAH1 which are implicated in glioma genesis identified. The study shed light on proteins which can possibly be used as biomarkers for DIPG. In another study by the same group, (Saratsis et al. 2014) they identified 2,061 proteins from 14 formalin fixed paraffin embedded (FFPE) DIPG samples. Hence this data provides a rich source of information as to how the protein landscape of paediatric brain and its surrounding fluid looks like and can be used for comparison with proteome data from DIPG cell lines and other DIPG samples.

3. Current therapy options for cancer

For decades, cancer treatments have relied heavily on surgery, chemotherapy and radiotherapy. Whilst these therapies offer substantial benefit in terms of tumour eradication, relapse can still occur due to presence of residual malignant cells. Therefore, alternative treatment approaches to eliminate the resistant tumour cells are required. As the concept of cancer immunosurveillance has gained more traction in recent years, cancer immunotherapy is an appealing and attractive strategy. The cornerstone of immunotherapy is that it redirects the body's immune system to induce an anti-tumour response. This is because as mentioned earlier (section 2.2), the immune system's capability for spontaneously recognising and eliminating cancer cells is well known. Immunotherapies can be grouped into i) **active** and ii) **passive** immunotherapies (Zhang et al. 2018), where active immunotherapies rely on activation of the host immune response, whilst passive approaches rely on intrinsic anti-cancer activity of administered drugs. Furthermore, classification of specific or non-specific immunotherapies can be drawn dependent on direct targeting of cancer cells (specific) or a general activation of the immune system (non-specific) (Zhang and Chen 2018).

3.1 Cancer vaccines

The success of vaccines in preventing infections such as measles and influenza even before they occur propagated the idea of a cancer vaccine. These cancer vaccines would rely on stimulating or restoring the ability of immune system to recognise and eliminate the cancer on its own. They would likely to contain at least two components, antigens that can be recognised by the adaptive immune system namely T and B cells and an adjuvant that would activate the adaptive immune response. Vaccines are broadly categorised into i) **preventative** and ii) **therapeutic** vaccines. The goal of a preventative vaccine is stopping cancer from developing. The best example for preventative cancer vaccine is Gardasil (Knutson et al. 2015, Zhang and Chen 2018) which uses purified inactive proteins from human papilloma virus (HPV) strain types 6, 11, 16, and 18 (Hung et al. 2008). This results in production of antibodies thereby preventing HPV infections which is a leading cause of cervical and head and neck cancers (Knutson and Mittendorf 2015). On the other hand, therapeutic cancer vaccines are based on specific stimulation of the immune system using tumour antigens, thereby directly targeting the immune system causing it to attack cancer cells (Knutson and Mittendorf 2015). The central role in therapeutic vaccines would be played by the APC in presenting defined or undefined tumour antigens to T cells. Peptide based vaccines are thus the mainstay of therapeutic vaccination although there are other types of therapeutic vaccine strategies such as whole

cancer cell lysates, DNA/RNA vaccines and viral vectors all under various stages of clinical trials (Zhang and Chen 2018). These methods will not be discussed in this thesis.

3.2 Peptide vaccines

Peptide based vaccines consist of cancer related immunogenic epitopes restricted by HLA class I allomorphs. The immunogenic peptides comprise of any of the following four categories (Table 3), these include 1) TSAs or neoantigens, which arise from proteins that have undergone mutation and occur exclusively in cancer, 2) TAAs and 3) CTAs and 4) post translationally modified (PTM) peptides, an emerging category of peptides of interest which are disproportionately present on tumour cells due to dysregulated oncogenic signalling. Therapeutic cancer vaccines hold the promise of prolongation of survival, efficacious treatment of the cancer along with high safety and tolerability, as evident from several randomised trials (Knutson and Mittendorf 2015).

Table 3 - Overview of different types of cancer antigens found in leukaemia and DIPG

Antigen Type	Description	Example
Tumour specific antigen/Neoantigen	Completely absent from normal host cells. Arise in cancer cells from nonsynonymous somatic mutations	K27M mutation in DIPG. NPM1 and onco-fusion proteins in AML
Tumour associated antigen	Low levels of expression on normal host cells. Disproportionately expressed on tumour cells. Can be selectively expressed by the cell lineage from which the cancer evolved.	RUNX1, NPM1 in HM SOX and chondroitin sulphate family of proteins in brain cancer.
Cancer testis antigen	Absent on normal adult cells, except in reproductive tissues (e.g. testes, foetal ovaries, and trophoblasts). Selectively expressed by various tumour types	MAGE, PRAME family of proteins.
PTM-peptides	Low levels of expression on normal host cells. Disproportionately expressed on tumour cells due to dysregulated signalling.	Varies between cancers

3.2.1 Tumour associated and cancer testis antigen-based peptide vaccines

Peptide based immunotherapy involving TAAs and CTAs received a major boost following 2 large phase III clinical trials. DERMA was a double-blind, randomised, placebo-controlled trial which included 31 countries in which 3,914 patients were screened (Dreno et al. 2018) wherein

melanoma associated antigen A3 (MAGE-A3) based immunotherapy in patients with stage IIIB or IIIC melanoma was assessed. Similarly, the MAGRIT was also a phase III clinical trial (Tyagi et al. 2009) conducted to evaluate anti-MAGE-A3 immune responses in patients who were either stage IB, II, or IIIA non-small-cell lung cancer (NSCLC) and MAGE-A3-positive patients. However, both the treatments failed to show any significant improvement in disease free survival when compared to the placebo groups in the study. Hence the vaccines were not taken to market. Despite this, other ongoing clinical trials using different TAAs and CTAs have shown promise. In a study by Schwartzentruber *et al* (Schwartzentruber et al. 2011) peptide from a well-known CTA protein gp100 (209-217 peptide) was used for treating patients with stage IV or locally advanced stage III cutaneous melanoma. The outcome of the trial was the presence of high levels of circulating T cells that were capable of recognising and killing melanoma cancer cells *in vitro*.

In a similar study by Carreno et al (Carreno et al. 2015) wherein DC based peptide vaccine was administered to three patients in melanoma stage III. The vaccine was a personalised formulation of 7 neoantigenic peptides identified in each patient by WES and predicted to bind to HLA A*02:01. It was found that vaccination resulted not only in augmented T cell responses to the neoantigens in terms of increased frequency of CD8+ T cells but also expansion of the anti-tumour immune response by revealing subdominant neoantigens.

Other studies that demonstrated the efficacy of incorporating multi peptide vaccine approach, was seen in the case of IMA90, a vaccine for renal carcinoma (RCC), currently in process of undertaking phase III clinical trials. IMA901 vaccine comprised of 8 HLA A*02:01restricted peptides (SVASTITGV, VMAGDIYSV, ALADGVQKV, LLGATCMFV, SVFAGVVGV, ALFDGDPHL, YVDPVITSI, STAPPVHNV, LAALPHSCL and FLPSDFFPSV) identified from RCC patients by mass spectrometry (MS). In phase I trials, enrolled patients received IMA901 vaccine along with granulocyte-macrophage colony-stimulating factor (GM-CSF) as an immunomodulator. 20/27 patients showed a vaccine-induced T cell response to at least one peptide and 8/20 responded to multiple peptides. In order to reduce Treg expansion found in some patients in phase I trial, phase II trial involved vaccinating patients with IMA901 along with a single-dose cyclophosphamide (an immunomodulator known to reduce Treg expression). In phase II ~64% of patients responded to the treatment as evident from peptide specific CD8 T cells found in patients' peripheral blood mononuclear cells (PBMCs) along with reduction of Treg cells frequency (Walter et al. 2012). Similar studies and responses have

been reported for different cancers including renal carcinoma (Obara et al. 2018) along with lymphoma (Schuster et al. 2011) myeloid and neuroblastoma (Krishnadas et al. 2015).
More importantly, recent studies where immunopeptidomics has been used to identify TAAs and CTAs in leukaemia patients (Cobbold et al. 2013, Berlin et al. 2015, Kowalewski et al. 2015, Bilich et al. 2019) reported peptide specific CD8+ T cell response in the form of IFN-γ secretion in PBMCs of healthy donors. These studies demonstrate that peptide-based cancer immunotherapy can induce specific immune responses and can have impactful clinical outcomes.

3.2.2 Tumour specific or neoantigen based peptide vaccines

3.2.2.1 In AML

A drawback of relying on TAA and CTA based vaccines is that these peptides are self-antigens and since high-affinity T cells capable of recognising self-antigens are eliminated during T cell development by central and peripheral tolerance mechanisms, oftentimes there is little or no activation of remaining T cells (Hollingsworth et al. 2019). In order to overcome this challenge peptide-based immunotherapy can exploit are mutations occurring in cancer, this includes both driver and passenger mutations as mentioned in section 2.1. Neoantigens can be of two kind, **1) shared** neoantigens arise from the same mutation which might be present across different cancers. In contrast **2) personal** neoantigens arise from mutations which are exclusive to only a particular cancer.

In the case of haematological malignancy, some of the antigens that are known to carry mutations include NPM1, RARA, RUNX1 and FLT3 (Roerden et al. 2019). The unique feature of AML is the presence of different types of chromosomal rearrangements (as listed in Table 1). Hence neoantigens in AML can be classified into 2 categories, 1) simple and 2) complex. Simple neoantigens arise form a single amino acid nonsynonymous mutation occurring in the proteins. Complex neoantigens on the other hand are derived from fused regions of 2 gene products such as PML-RARA and ETV-AML1. There are several studies which have identified and tested the immunogenicity of both these categories of proteins. Some of them are discussed below.

One of the frequently occurring mutations in AML patients occurs in NPM1 protein, wherein a characteristic 4 bp frameshift insertion occurs in exon 12 of the gene. The resulting mutant NPM1 protein is 4 aa longer than the wild type counterpart, and at its C-terminal end there are

11 aa that are translated due to an alternative reading frame. This results in generation of neopeptides. HLA-bound peptides derived from mutant NPM1 have been identified in patients and cell lines (Narayan et al. 2019, van der Lee et al. 2019).
In a study by van der Lee *et al* (van der Lee et al. 2019) 8/12 AML patients were found to have the above mentioned NPM1 mutation. Five neoepitopes from the mutant NPM1 region were found in 7/8 NPM1 positive AML patients by MS. This included two 8-mer peptides (VEEVSLRK and AVEEVSLR), two 9-mer peptides (CLAVEEVSL and AVEEVSLRK), and one 11-mer peptide (CLAVEEVSLRK). Only anti-CLAVEEVSL CD8+ T cells were identified in 4/6 A*02:01 positive healthy donors by tetramer-based enrichment.

The fusion of PML and RARA gene results in addition of a new triplet codon 'GCC' whose translation results in incorporation of an additional Ala residue in the oncofused protein. Gambacorti *et al* (Gambacorti-Passerini et al. 1993) demonstrated that oncofused protein was recognized by CD4 T cells *in vitro* obtained from PBMCs of healthy donors which were HLA DR11 positive.
A similar study by Yotnda *et al* (Yotnda et al. 1998) tested the immunogenicity of peptide RIAECILGM derived from the oncofused protein ETV-AML1. They reported anti-RIAECILGM CD8+ T cell responses in 3 healthy A2 donors *in vitro* and lyses of tumour cells expressing the peptide. Peptide specific CTL cells were also identified in BM of 2 ALL patients. Similar studies have also been done on other oncofusion proteins such as FLT3-ITD, (Graf et al. 2007).
To conclude, these studies highlight the vast potential of neoantigens that can be exploited for immunotherapeutics for AML.

3.2.2.2 In DIPG
With regards to DIPG there is a highly penetrant and specific mutation of H3.3 K27M occurring in ~70% of cases. There have been only 2 studies which have reported immunogenic nature of a potential neoepitope encompassing the K27M mutation. The first study by Ochs et al (Ochs et al. 2017), found that vaccination with longer 27mer peptide K27Mp26–35 induced both CD4+ and CD8+ T cell responses, as measured by IFN-γ secretion, proliferation and granzyme B expression in a HLA-humanised mouse model (expressing HLA A*02:01, HLA-DRA*01:01, and HLA-DRB1*01:01, and lacking mouse MHC class I and II: A2.DR1 mice). This suggested that the 27mer long peptide was being processed by antigen presentation machinery of both HLA class I and II. This resulted in presentation of a peptide encompassing

the H3.3K27M mutation. However, the exact sequence of the peptides was not known. More recently, Chheda *et al.* (Chheda et al. 2018) performed immunopeptidomics to pull down the HLA*02:01 restricted K27M neoantigen followed by followed by high resolution MS to identify its sequence from U-87MG glioblastoma cell line transfected with H3.3K27M protein. They reportedly identified a 10mer peptide RMSAPSTGGV, restricted by HLA A*02:01 encompassing the K27M mutation. The peptide was also found to be capable of eliciting CD8+ T cell response in A*02:01 positive healthy (n =2) and patients (n =2) PBMCs expanded *in vitro*. Although their immunopeptidomics approach was unable to identify the endogenous mutant peptide in glioma cell lines (U87MG) transfected with the mutant histone protein.

A drawback of neoantigens based T cell immunotherapy is that presence of mutations in tumours lie on a spectrum with some cancers having high mutational burden whilst others have low mutational burden. For instance, the DC based vaccine study by Carreno *et al* (Carreno et al. 2015) identified ~18-24 missense mutations in different proteins in each of the 3 patients and hence 7 neoantigenic peptides could be incorporated in the personalised vaccine formulation. However, the same is not true for haematological malignancies and brain tumours wherein there are only a few proteins which are mutated (Berlin et al. 2015, Rosenberg et al. 2015).

3.2.3 Post translationally modified antigen-based vaccines

Genetic aberrations that are acquired by malignant cells also result in alterations in the biochemistry of the cells. For instance, the most frequently observed mutations at the gene level (in the form of gene fusions, transversions and tandem repeats) influence RTK family which are responsible for signal transduction of cytokines and growth factors. RTK dysregulation further drives imbalance in the MAP/Raf/Ras kinase mediated phospho-signalling downstream. Another important PTM in leukaemia is arginine methylation, which is catalysed by a family of nine protein arginine methyltransferases (PRMTs). In normal cells methylation is involved in regulation of transcription and chromatin organisation, RNA processing and DNA damage repair (Wingelhofer et al. 2019). In leukaemia, changes in arginine methylation are critical for maintenance of leukemic stem cell (LSC) potential by sustaining an oncogenic transcriptional program and blocking differentiation and apoptosis (Wingelhofer and Somervaille 2019, Birch et al. 2020). Specifically, in AML, methylation has been reported to promote tumour growth by regulating a feedback loop which ultimately leads

to expression of FLT3 (He et al. 2019) and RUNX1, key transcription factors required for definitive haematopoiesis, myeloid differentiation, and lymphocyte development

Changes in protein expression and or metabolism that accompany cellular transformation can often alter peptides that are displayed by the HLA as these PTMs are retained on the peptides post antigen processing (Zarling et al. 2006). The PTMs that are frequently found in cancer include phosphorylation, methylation and glycosylation (Cobbold et al. 2013, Malaker et al. 2017, Marino et al. 2017). There is a renewed interest in PTM peptides as structural studies via crystallography have shown that some PTMs can have a stabilising role on the pHLA complex (Mohammed et al. 2008, Petersen et al. 2009, Mohammed et al. 2017). It is hypothesised that that phosphorylation can alter the immunogenicity of a peptide by offering a prominent electronegative target for TCR binding (Petersen et al. 2009). Moreover, PTM-peptide specific CTLs have been recognised in healthy donors PBMCs, capable of both IFN-γ release *in vitro* and cytotoxic activity towards tumour cells expressing the PTM peptides (Cobbold et al. 2013, Malaker et al. 2017) suggesting they may play a role in anti-tumour immunity. Similar conclusions can be drawn from study of Jarmalavicius *et al.* (Jarmalavicius et al. 2010) which demonstrated CD8+ T cell reactivity towards a P5 monomethylated peptide SQNPR(m)FYHK restricted by HLA A*11:01 in 9 melanoma patients, in an *ex vivo* IFN-γ measuring ELISPOT assay. In contrast the P5 dimethyl arginine version of the peptide only generated a response in only one melanoma patient. This highlights the potential role of methylation in TCR recognition.

To conclude, therapeutic cancer vaccines hold the promise of efficaciously treating cancer or prolongation of survival with safety and tolerability. However, development of cancer vaccines remains a major challenge, with little knowledge of (i) the optimal tumour antigens to target, (ii) suitable agents to counteract regulatory mechanisms opposing successful immunotherapy This leaves immense scope of improvement which involves not only identifying new peptides as targets but also improve our understanding interplay between mutations, TME, immune cells and cytokines. This can help in not only predicting a patient's response to therapy but help in making an informed decision on the type of therapy prescribed to patients.

3.3 Role of mass spectrometry in investigating the cancer immunopeptidome

3.3.1 Next generation sequencing often fails to identify cancer specific HLA restricted peptides

For some years now, to identify new targets for immunotherapy, scientists have turned to next generation sequencing (NGS). This is because NGS helped to unravel cancer genomics and identify somatic mutations that give rise to potential neoantigens (Zhang et al. 2017). While NGS provides some insight into the mutational landscape of cancer (Zhang et al. 2017), it is still not sufficient to facilitate breakthroughs in peptide-based immunotherapy. This is because DNA and RNA expression does not always translate into HLA-restricted antigen presentation of corresponding gene products (Weinzierl et al. 2007).

This was evident from studies by Yadav *et al.* wherein WES data was used to identify 1,290 and 67 mutations in mouse cell lines MC-38 and TRAMP-C1 cell lines, respectively. Out of those only 170 predicted neoepitopes in MC-38 and 6 predicted neoepitopes in TRAMP-C1 were found to be restricted by different H-2K^{b} and H-2D^{b} based on MHC-binding predictions. Out of 176 predicted neoepitopes evidence for only 7 (only in MC-38 and none in TRAMP-C1) was confirmed by MS. This low hit rate between predicted and peptides that are actually presented by HLA of cells highlights the main problem with excessive reliance on WES/NGS data.

Similar studies have been conducted for leukaemia (Rajasagi et al. 2013, Rajasagi et al. 2014) wherein WES data was used to identify a total of 1838 nonsynonymous mutations from 91 CLL samples. The immunogenicity of 30 and 18 peptides was checked across 2 CLL patients and peptide reactive CD8 T cells capable of secreting IFNγ were only identified only for two and one peptide in patient1 and patient2 respectively by ELISPOT. For DIPG there has been no such study wherein WES/NGS has been combined with prediction algorithms to identify HLA class I restricted peptide antigens.

Hence it can be said that peptide-based immunotherapy has fallen short of its potential to achieve responses both *in vitro* and in the clinical setting of cancer therapy probably due to disequilibrium between identified tumour-associated targets on the input side and functional vaccine candidates on the output side.

3.3.2 Utilising mass spectrometry to identify the cancer immunopeptidome

In order to have a successful peptide-based vaccine new and more targets have to be identified. This provides an immense opportunity to mine HLA-bound peptides of malignant cells to address the unmet need. The challenge presented at this front arises from the very property of HLA that makes it unique amongst all the genes found in humans. As mentioned in section 1.3, HLA has one of the most polymorphic gene cluster responsible for generating its enormous diversity. This results in a two-layered complexity, first is allele-specific peptide restriction wherein HLA A*02:01 peptide repertoire is very distinct to A*03:01 peptide repertoire as the former prefers peptides with small hydrophobic amino acids (Leu/Val/Ileu) whilst the later prefers peptides with basic amino acids (Lys/Arg) . Second, since an individual can express up to 6 different HLA class I molecules, there can be ~10^5-10^6 distinct peptides being presented on the surface of APCs at any given time (Pandey et al. 2020). Hence, this level of complexity can only be resolved by a highly sensitive technique such as MS. The initial framework of utilising MS to identify and characterise the immunopeptidome was laid down by work of Rammensee, Hunt and Engelhard, Sette and others. Over the years, considerable progress has been made in the field which has helped in understanding of immunopeptidome including length distribution of peptides, consensus binding motifs of different HLA and this has in turn lead to significant improvement of prediction algorithms ability to predict binding affinities of peptides to specific HLA alleles. Over the years, MS has facilitated the identification of peptides associated with different cancers including melanoma, breast cancer and leukaemia (Abelin et al., 2017; Bassani-Sternberg et al., 2016; Chikata et al., 2019; Cobbold et al., 2013; Ternette et al., 2018; Zarling et al., 2006).

3.3.3 Previous immunopeptidome studies and bridging the knowledge gap

For haematological malignancy, there have only been three studies aimed to identify therapeutically actionable peptides for AML using immunopeptidomics (Table 4). The studies have interrogated either PBMCs from patients with haematological malignancies or cell lines to identify naturally presented TAAs, CTAs and phosphopeptides (Cobbold et al. 2013, Berlin et al. 2015, Narayan et al. 2019).

Table 4 - Immunopeptidomics studies and peptides identified in haematological malignancy.

Study	Cohort size/Patient number	HLA class I peptides identified	HLA class II peptides identified	Overlapping peptides
Berlin et al.	Class I, n=15 Class II, n=12	Patients - 13,238	Patients - 2,816	NA
Narayan et al.	Class I and II n=13	Patients – 12,406 Cell lines –25,212	Patients – 4,954 Cell lines – 5,204	Patients – 2,809 Cell lines –1,318
Cobbold et al. (Only phosphopeptides)	Class I, n=5	Patients – 95	NA	NA

Also, there are studies that have investigated the immunopeptidome of haematological malignancies other than AML, including Kowalewski *et al.* (Kowalewski et al. 2015) which identified 18,844 HLA class I bound peptides from PBMCs of 30 CLL patients. Whilst a comparative study by Backert *et al.* (Backert et al. 2017) performed a meta-analysis on 83 HM immunopeptidome datasets comprising of 40,361 unique peptides identified across four major HM (19 AML, 16 CML, 35 CLL, and 13 MM/MCL samples). A comparison across the immunopeptidome dataset resulted in identification of 25 peptides common amongst different haematological malignancy included in the study.

A drawback of these studies is the choice of sample, as the site for leukemogenesis is the bone marrow. Hence not incorporating BM samples but relying on PBMCs leaves a wide gap in the field due to anticipated influence of mutations and TME on the endogenous HLA peptide repertoire in the BM. As mentioned previously in section 2.4 tumours have a tendency of downregulating HLA expression which might not be reflected in the peripheral cells or tissues. Hence the peptide repertoire of the cancer from the tumour site i.e. the BM may be completely different to that of the periphery and should influence immunotherapy choice. Hence this thesis will provide new and deeper insights into

1) The level of antigen presentation in AML and ALL bone marrow aspirates
2) Identifying different classes of tumour associated antigens including TAA, CTA and PTM-peptides in AML and ALL.

This will be achieved by investigating the following aims for AML

1) **Aim 1** - Optimisation of mass spectrometry methods for effective identification of pHLA and PTM-peptide repertoire presented on cancer derived cells lines.
2) **Aim 2** - Pan cancer analysis of peptide HLA repertoire presented in haematological malignancy.

3) **Aim 3** - Characterisation of biophysical binary structure interactions between novel PTM-peptides and HLA Class I alleles.

In contrast to AML, there have been only 2 studies which have 1) verified the immunogenicity of the K27M encompassing neoepitope (as mentioned in section 3.1.1) and 2) a study which has used MS to identify HLA A*02:01 restricted 10mer peptide RMSAPSTGGV. Although the MS data was far from convincing as only the synthetic heavy peptides was found whilst the endogenous peptide was not detected.

Hence this thesis is a step towards right direction, particularly for T cell mediated therapy which remains untranslated at the clinical level due to a dearth of DIPG specific peptides. The current study bridges this missing knowledge as it is the first study to report the HLA class I peptide repertoire of DIPG, thereby identifying different types TAAs, CTAs and neoepitopes, via high-resolution MS, that can be used in rational design of DIPG targeted immunotherapeutics. This will be achieved by investigating the following aim for DIPG aspect of the thesis

Aim 4 - Investigating the immunopeptidome of diffuse intrinsic pontine glioma to identify immunotherapy targets.

REFRENCES

Abbey, J. L. and H. C. O'Neill (2007). "Expression of T-cell receptor genes during early T-cell development." Immunol Cell Biol **86**(2): 166-174.

Aki, M., et al. (1994). "Interferon-gamma induces different subunit organizations and functional diversity of proteasomes." J Biochem **115**(2): 257-269.

Arango Duque, G. and A. Descoteaux (2014). "Macrophage Cytokines: Involvement in Immunity and Infectious Diseases." Front Immunol **5**.

Arvelo, F., et al. (2016). "Tumour progression and metastasis." Ecancermedicalscience **10**: 617.

Backert, L., et al. (2017). "A meta-analysis of HLA peptidome composition in different hematological entities: entity-specific dividing lines and "pan-leukemia" antigens." Oncotarget **8**(27): 43915-43924.

Barrett, J. C. (1993). "Mechanisms of Multistep Carcinogenesis and Carcinogen Risk Assessment." Environmental Health Perspectives **100**: 9-20.

Bennett, J. M., et al. (1976). "Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group." Br J Haematol **33**(4): 451-458.

Berlin, C., et al. (2015). "Mapping the HLA ligandome landscape of acute myeloid leukemia: a targeted approach toward peptide-based immunotherapy." Leukemia **29**(3): 647-659.

Bilich, T., et al. (2019). "The HLA ligandome landscape of chronic myeloid leukemia delineates novel T-cell epitopes for immunotherapy." Blood **133**(6): 550-565.

Birch, N. W. and A. Shilatifard (2020). "The role of histone modifications in leukemogenesis." Journal of Biosciences **45**(1): 6.

Bjorkman, P. J., et al. (1987). "Structure of the human class I histocompatibility antigen, HLA-A2." Nature **329**(6139): 506-512.

Blum, J. S. (2013). "Pathways of Antigen Processing." **31**: 443-473.

Bodmer, W. F. (1987). "The HLA system: structure and function." J Clin Pathol **40**(9): 948-958.

Bosman, F. T. (2019). Integrative Molecular Tumor Classification: A Pathologist's View. Encyclopedia of Cancer (Third Edition). P. Boffetta and P. Hainaut. Oxford, Academic Press**:** 279-285.

Bray, F., et al. (2018). "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA: A Cancer Journal for Clinicians **68**(6): 394-424.

Buczkowicz, P., et al. (2014). "Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations." Nat Genet **46**(5): 451-456.

Carreno, B. M., et al. (2015). "Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells." Science **348**(6236): 803-808.

Chen, Daniel S. and I. Mellman (2013). "Oncology Meets Immunology: The Cancer-Immunity Cycle." Immunity **39**(1): 1-10.

Chheda, Z. S., et al. (2018). "Novel and shared neoantigen derived from histone 3 variant H3.3K27M mutation for glioma T cell therapy." J Exp Med **215**(1): 141-157.

Chothia, C., et al. (1988). "The outline structure of the T-cell alpha beta receptor." The EMBO Journal **7**(12): 3745-3755.

Cobbold, M., et al. (2013). "MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia." Sci Transl Med **5**(203): 203ra125.

Collins, E. J. and D. S. Riddle (2008). "TCR-MHC docking orientation: natural selection, or thymic selection?" Immunologic Research **41**(3): 267.

Couch, D. B. (1996). "Carcinogenesis: basic principles." Drug Chem Toxicol **19**(3): 133-148.

Cresswell, P. (2013). "Pathways of Antigen Processing." Annu Rev Immunol.

Cresswell, P., et al. (1999). "The nature of the MHC class I peptide loading complex." Immunol Rev **172**: 21-28.
Cresswell, P., et al. (1973). "Papain-Solubilized HL-A Antigens from Cultured Human Lymphocytes Contain Two Peptide Fragments." Proceedings of the National Academy of Sciences **70**(5): 1603-1607.
Davis, M. M. and P. J. Bjorkman (1988). "T-cell antigen receptor genes and T-cell recognition." Nature **334**: 395.
De Kouchkovsky, I. and M. Abdul-Hay (2016). "'Acute myeloid leukemia: a comprehensive review and 2016 update'." Blood Cancer J **6**(7): e441.
DiNardo, C. D. and J. E. Cortes (2016). "Mutations in AML: prognostic and therapeutic implications." Hematology Am Soc Hematol Educ Program **2016**(1): 348-355.
Dong, G., et al. (2009). "Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer." Immunity **30**(1): 21-32.
Dreno, B., et al. (2018). "MAGE-A3 immunotherapeutic as adjuvant therapy for patients with resected, MAGE-A3-positive, stage III melanoma (DERMA): a double-blind, randomised, placebo-controlled, phase 3 trial." The Lancet Oncology **19**(7): 916-929.
Driscoll, J., et al. (1993). "MHC-linked LMP gene products specifically alter peptidase activities of the proteasome." Nature **365**(6443): 262-264.
Dudley, D. D., et al. (2005). "Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences." Adv Immunol **86**: 43-112.
Dufva, O., et al. (2020). "Immunogenomic Landscape of Hematological Malignancies." Cancer Cell.
Elliott, T., et al. (1990). "Naturally processed peptides." Nature **348**(6298): 195-196.
Estey, E. (2016). "Acute myeloid leukemia: 2016 Update on risk-stratification and management." Am J Hematol **91**(8): 824-846.
Estey, E. H. (2009). "Treatment of acute myeloid leukemia." Haematologica **94**(1): 10-16.
Estey, E. H. (2018). "Acute myeloid leukemia: 2019 update on risk-stratification and management." Am J Hematol **93**(10): 1267-1291.
Evavold, B. D. and P. M. Allen (1991). "Separation of IL-4 production from Th cell proliferation by an altered T cell receptor ligand." Science **252**(5010): 1308-1310.
Evavold, B. D., et al. (1993). "Tickling the TCR: selective T-cell functions stimulated by altered peptide ligands." Immunology Today **14**(12): 602-609.
Falini, B., et al. (2005). "Cytoplasmic Nucleophosmin in Acute Myelogenous Leukemia with a Normal Karyotype." New England Journal of Medicine **352**(3): 254-266.
Falk, K., et al. (1991). "Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules." Nature **351**(6324): 290-296.
Ferguson, T. A., et al. (2011). "Armed response: how dying cells influence T-cell functions." Immunol Rev **241**(1): 77-88.
Filbin, M. G., et al. (2018). "Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq." Science **360**(6386): 331-335.
Fitzmaurice, C., et al. (2017). "Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study." JAMA Oncol **3**(4): 524-548.
Gaczynska, M., et al. (1993). "Gamma-interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes." Nature **365**(6443): 264-267.
Gajewski, T. F., et al. (2006). "Immune resistance orchestrated by the tumor microenvironment." Immunol Rev **213**: 131-145.

Gambacorti-Passerini, C., et al. (1993). "Human CD4 lymphocytes specifically recognize a peptide representing the fusion region of the hybrid protein pml/RAR alpha present in acute promyelocytic leukemia cells." Blood **81**(5): 1369-1375.
Garboczi, D. N., et al. (1996). "Structure of the complex between human T-cell receptor, viral peptide and HLA-A2." Nature **384**(6605): 134-141.
Garcia, K. C., et al. (1996). "An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex." Science **274**(5285): 209-219.
Gordon, S. "Pattern Recognition Receptors." Cell **111**(7): 927-930.
Graf, C., et al. (2007). "A neoepitope generated by an FLT3 internal tandem duplication (FLT3-ITD) is recognized by leukemia-reactive autologous CD8+ T cells." Blood **109**(7): 2985-2988.
Grafone, T., et al. (2012). "An overview on the role of FLT3-tyrosine kinase receptor in acute myeloid leukemia: biology and treatment." Oncol Rev **6**(1): e8.
Grimwade, D., et al. (1998). "The Importance of Diagnostic Cytogenetics on Outcome in AML: Analysis of 1,612 Patients Entered Into the MRC AML 10 Trial." Blood **92**(7): 2322-2333.
Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.
Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.
Harris, N. L., et al. (1999). "World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997." J Clin Oncol **17**(12): 3835-3849.
He, X., et al. (2019). "PRMT1-mediated FLT3 arginine methylation promotes maintenance of FLT3-ITD+ acute myeloid leukemia." Blood **134**(6): 548-560.
Hollingsworth, R. E. and K. Jansen (2019). "Turning the corner on therapeutic cancer vaccines." NPJ Vaccines **4**: 7.
Housset, D. and B. Malissen (2003). "What do TCR–pMHC crystal structures teach us about MHC restriction and alloreactivity?" Trends in Immunology **24**(8): 429-437.
Hung, C. F., et al. (2008). "Therapeutic human papillomavirus vaccines: current clinical trials and future directions." Expert Opin Biol Ther **8**(4): 421-439.
Hunt, J. S., et al. (2005). "HLA-G and immune tolerance in pregnancy." The FASEB Journal **19**(7): 681-693.
Jarmalavicius, S., et al. (2010). "Differential arginine methylation of the G-protein pathway suppressor GPS-2 recognized by tumor-specific T cells in melanoma." Faseb j **24**(3): 937-946.
Jones, C. and S. J. Baker (2014). "Unique genetic and epigenetic mechanisms driving signatures of paediatric diffuse high-grade glioma." Nat Rev Cancer **14**(10).
Jorgensen, J. L., et al. (1992). "Molecular components of T-cell recognition." Annu Rev Immunol **10**: 835-873.
Juratli, T. A., et al. (2018). "Molecular pathogenesis and therapeutic implications in pediatric high-grade gliomas." Pharmacology & Therapeutics **182**: 70-79.
Keskinen, P., et al. (1997). "Regulation of HLA class I and II expression by interferons and influenza A virus in human peripheral blood mononuclear cells." Immunology **91**(3): 421-429.
Khuong-Quang, D. A., et al. (2012). "K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas." Acta Neuropathol **124**(3): 439-447.
Kjer-Nielsen, L., et al. (2002). "The 1.5 Å Crystal Structure of a Highly Selected Antiviral T Cell Receptor Provides Evidence for a Structural Basis of Immunodominance." Structure **10**(11): 1521-1532.
Knutson, K. L. and E. A. Mittendorf (2015). "Cancer vaccines in the new era of cancer immunotherapy." Vaccine **33**(51): 7376.

Korshunov, A., et al. (2015). "Integrated analysis of pediatric glioblastoma reveals a subset of biologically favorable tumors with associated molecular prognostic markers." Acta Neuropathol **129**(5): 669-678.
Kowalewski, D. J., et al. (2015). "HLA ligandome analysis identifies the underlying specificities of spontaneous antileukemia immune responses in chronic lymphocytic leukemia (CLL)." Proceedings of the National Academy of Sciences **112**(2): E166-E175.
Krishnadas, D. K., et al. (2015). "A phase I trial combining decitabine/dendritic cell vaccine targeting MAGE-A1, MAGE-A3 and NY-ESO-1 for children with relapsed or therapy-refractory neuroblastoma and sarcoma." Cancer Immunol Immunother **64**(10): 1251-1260.
Krueger, K. E. and S. Srivastava (2006). "Posttranslational protein modifications: current implications for cancer detection, prevention, and therapeutics." Mol Cell Proteomics **5**(10): 1799-1810.
Kumar, C. C. (2011). "Genetic Abnormalities and Challenges in the Treatment of Acute Myeloid Leukemia." Genes Cancer **2**(2): 95-107.
Lee, Y. J., et al. (1997). "TGF-beta suppresses IFN-gamma induction of class II MHC gene expression by inhibiting class II transactivator messenger RNA expression." The Journal of Immunology **158**(5): 2065-2075.
Louis, D. N., et al. (2016). "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary." Acta Neuropathol **131**(6): 803-820.
Malaker, S. A., et al. (2017). "Identification of Glycopeptides as Posttranslationally Modified Neoantigens in Leukemia." Cancer Immunol Res **5**(5): 376-384.
Maleno, I., et al. (2011). "Frequent loss of heterozygosity in the β2-microglobulin region of chromosome 15 in primary human tumors." Immunogenetics **63**(2): 65-71.
Marino, F., et al. (2017). "Arginine (Di)methylated Human Leukocyte Antigen Class I Peptides Are Favorably Presented by HLA-B*07." Journal of Proteome Research **16**(1): 34-44.
Maupin-Furlow, J. (2011). "Proteasomes and protein conjugation across domains of life." Nat Rev Microbiol **10**(2): 100-111.
McGranahan, N., et al. (2017). "Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution." Cell **171**(6): 1259-1271.e1211.
Medzhitov, R. and C. A. Janeway (2002). "Decoding the Patterns of Self and Nonself by the Innate Immune System." Science **296**(5566): 298-300.
Meyer, V. S., et al. (2009). "Identification of natural MHC class II presented phosphopeptides and tumor-derived MHC class I phospholigands." J Proteome Res **8**(7): 3666-3674.
Miyamoto, T., et al. (2002). "Myeloid or Lymphoid Promiscuity as a Critical Step in Hematopoietic Lineage Commitment." Developmental Cell **3**(1): 137-147.
Mohammed, F., et al. (2008). "Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self." Nat Immunol **9**(11): 1236-1243.
Mohammed, F., et al. (2017). "The antigenic identity of human class I MHC phosphopeptides is critically dependent upon phosphorylation status." Oncotarget **8**(33): 54160-54172.
Muenst, S., et al. (2016). "The immune system and cancer evasion strategies: therapeutic concepts." J Intern Med **279**(6): 541-562.
Narayan, R., et al. (2019). "Acute myeloid leukemia immunopeptidome reveals HLA presentation of mutated nucleophosmin." PLoS One **14**(7): e0219547.
Nie, Y., et al. (2001). "DNA hypermethylation is a mechanism for loss of expression of the HLA class I genes in human esophageal squamous cell carcinomas." Carcinogenesis **22**(10): 1615-1623.
Obara, W., et al. (2018). "Present status and future perspective of peptide-based vaccine therapy for urological cancer." Cancer Science **109**(3): 550-559.

Ochs, K., et al. (2017). "K27M-mutant histone-3 as a novel target for glioma immunotherapy." Oncoimmunology **6**(7): e1328340.
Panter, M. S., et al. (2012). "Dynamics of major histocompatibility complex class I association with the human peptide-loading complex." J Biol Chem **287**(37): 31172-31184.
Papaemmanuil, E., et al. (2016). "Genomic Classification and Prognosis in Acute Myeloid Leukemia." N Engl J Med **374**(23): 2209-2221.
Parnes, J. R. and J. G. Seidman (1982). "Structure of wild-type and mutant mouse β2-microglobulin genes." Cell **29**(2): 661-669.
Patel, J. P., et al. (2012). "Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia." New England Journal of Medicine **366**(12): 1079-1089.
Petersen, J., et al. (2009). "Post-translationally modified T cell epitopes: immune recognition and immunotherapy." Journal of Molecular Medicine **87**(11): 1045.
Petersson, M., et al. (1998). "Constitutive IL-10 production accounts for the high NK sensitivity, low MHC class I expression, and poor transporter associated with antigen processing (TAP)-1/2 function in the prototype NK target YAC-1." J Immunol **161**(5): 2099-2105.
Pitot, H. C. (1993). "The molecular biology of carcinogenesis." Cancer **72**(3 Suppl): 962-970.
Purcell, A. W., et al. (2019). "Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics." Nature Protocols.
Rabinovich, G. A., et al. (2007). "Immunosuppressive strategies that are mediated by tumor cells." Annu Rev Immunol **25**: 267-296.
Rajasagi, M., et al. (2013). "Tumor Neoantigens Are Abundant Across Cancers." Blood **122**(21): 3265-3265.
Rajasagi, M., et al. (2014). "Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia." Blood **124**(3): 453-462.
Raval, A., et al. (1998). "Cytokine regulation of expression of class I MHC antigens." Exp Mol Med **30**(1): 1-13.
Reiser, J. B., et al. (2002). "A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex." Immunity **16**(3): 345-354.
Rock, K. L. and A. L. Goldberg (1999). "Degradation of cell proteins and the generation of MHC class I-presented peptides." Annu Rev Immunol **17**: 739-779.
Rock, K. L., et al. (2004). "Post-proteasomal antigen processing for major histocompatibility complex class I presentation." Nat Immunol **5**(7): 670-677.
Rodríguez, J. A. (2017). "HLA-mediated tumor escape mechanisms that may impair immunotherapy clinical outcomes via T-cell activation." Oncol Lett **14**(4): 4415-4427.
Roerden, M., et al. (2019). "Neoantigens in Hematological Malignancies-Ultimate Targets for Immunotherapy?" Front Immunol **10**: 3004.
Rosenberg, S. A. and N. P. Restifo (2015). "Adoptive cell transfer as personalized immunotherapy for human cancer." Science **348**(6230): 62-68.
Rovatti, P. E., et al. (2020). "Mechanisms of Leukemia Immune Evasion and Their Role in Relapse After Haploidentical Hematopoietic Cell Transplantation." Frontiers in Immunology **11**(147).
Ruddock, L. W. and M. Molinari (2006). "N-glycan processing in ER quality control." Journal of Cell Science **119**(21): 4373-4380.
Saratsis, A. M., et al. (2014). "Comparative multidimensional molecular analyses of pediatric diffuse intrinsic pontine glioma reveals distinct molecular subtypes." Acta Neuropathol **127**(6): 881-895.
Saratsis, A. M., et al. (2012). "Insights into pediatric diffuse intrinsic pontine glioma through proteomic analysis of cerebrospinal fluid." Neuro Oncol **14**(5): 547-560.

Schlenk, R. F., et al. (2008). "Mutations and Treatment Outcome in Cytogenetically Normal Acute Myeloid Leukemia." New England Journal of Medicine **358**(18): 1909-1918.
Schumacher, T., et al. (2014). "A vaccine targeting mutant IDH1 induces antitumour immunity." Nature **512**(7514): 324-327.
Schuster, S. J., et al. (2011). "Vaccination with patient-specific tumor-derived antigen in first remission improves disease-free survival in follicular lymphoma." J Clin Oncol **29**(20): 2787-2794.
Schwartzentruber, D. J., et al. (2011). "gp100 Peptide Vaccine and Interleukin-2 in Patients with Advanced Melanoma." New England Journal of Medicine **364**(22): 2119-2127.
Schwartzentruber, J., et al. (2012). "Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma." Nature **482**(7384): 226-231.
Tarafdar, A., et al. (2017). "CML cells actively evade host immune surveillance through cytokine-mediated downregulation of MHC-II expression." Blood **129**(2): 199-208.
Taylor, K. R., et al. (2014). "Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma." Nat Genet **46**(5): 457-461.
Toyonaga, B., et al. (1985). "Organization and sequences of the diversity, joining, and constant region genes of the human T-cell receptor beta chain." Proc Natl Acad Sci U S A **82**(24): 8624-8628.
Trombetta, E. S. and I. Mellman (2005). "Cell biology of antigen processing in vitro and in vivo." Annu Rev Immunol **23**: 975-1028.
Turvey, S. E. and D. H. Broide (2010). "Chapter 2: Innate Immunity." J Allergy Clin Immunol **125**(2 Suppl 2): S24-32.
Tyagi, P. and B. Mirakhur (2009). "MAGRIT: the largest-ever phase III lung cancer trial aims to establish a novel tumor-specific approach to therapy." Clin Lung Cancer **10**(5): 371-374.
Urbanska, K., et al. (2014). "Glioblastoma multiforme - an overview." Contemp Oncol (Pozn) **18**(5): 307-312.
van der Lee, D. I., et al. (2019). "Mutated nucleophosmin 1 as immunotherapy target in acute myeloid leukemia." J Clin Invest **129**(2): 774-785.
Végh, Z., et al. (1993). "Increased expression of MHC class I molecules on human cells after short time IFN-γ treatment." Molecular Immunology **30**(9): 849-854.
Walter, S., et al. (2012). "Multipeptide immune response to cancer vaccine IMA901 after single-dose cyclophosphamide associates with longer patient survival." Nat Med **18**(8): 1254-1261.
Wang, S. S., et al. (2019). "Towards Immunotherapy for Pediatric Brain Tumors." Trends in Immunology **40**(8): 748-761.
Weinzierl, A. O., et al. (2007). "Distorted Relation between mRNA Copy Number and Corresponding Major Histocompatibility Complex Ligand Density on the Cell Surface." Molecular & Cellular Proteomics **6**(1): 102-113.
Wieczorek, M., et al. (2017). "Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation." Frontiers in Immunology **8**(292).
Wingelhofer, B. and T. C. P. Somervaille (2019). "Emerging Epigenetic Therapeutic Targets in Acute Myeloid Leukemia." Frontiers in Oncology **9**(850).
Wu, G., et al. (2012). "Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas." Nat Genet **44**(3): 251-253.
Yotnda, P., et al. (1998). "Cytotoxic T cell response against the chimeric ETV6-AML1 protein in childhood acute lymphoblastic leukemia." J Clin Invest **102**(2): 455-462.
Youn, B. S., et al. (2000). "Chemokines, chemokine receptors and hematopoiesis." Immunol Rev **177**: 150-174.

Yu, Y., et al. (2018). "Recent advances in CD8(+) regulatory T cell research." Oncol Lett **15**(6): 8187-8194.
Zarling, A. L., et al. (2000). "Phosphorylated Peptides Are Naturally Processed and Presented by Major Histocompatibility Complex Class I Molecules in Vivo." J Exp Med **192**(12): 1755-1762.
Zarling, A. L., et al. (2006). "Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy." Proc Natl Acad Sci U S A **103**(40): 14889-14894.
Zhang, H. and J. Chen (2018). "Current status and future directions of cancer immunotherapy." J Cancer **9**(10): 1773-1781.
Zhang, X., et al. (2017). "Personalized cancer vaccines: Targeting the cancer mutanome." Vaccine **35**(7): 1094-1100.

CHAPTER - 2

Chapter 2 – In-depth mining of the immunopeptidome of acute myeloid leukaemia cell line using complementary ligand enrichment and data acquisition strategies

In-depth mining of the immunopeptidome of an acute myeloid leukemia cell line using complementary ligand enrichment and data acquisition strategies

Molecular Immunology 123 (2020) 77-1

Molecular Immunology

journal homepage: www.elsevier.com/locate/molimm

In-depth mining of the immunopeptidome of an acute myeloid leukemia cell line using complementary ligand enrichment and data acquisition strategies

Kirti Pandey[a], Nicole A. Mifsud[a], Terry C.C. Lim Kam Sian[a], Rochelle Ayala[a], Nicola Ternette[b], Sri H. Ramarathinam[a,*], Anthony W. Purcell[a,*]

a Department of Biochemistry and Molecular Biology and Infection and Immunity Program, Biomedicine Discovery Institute, Monash University, Clayton 3800, VIC, Australia

b The Jenner Institute, Mass Spectrometry Laboratory, University of Oxford, Oxford OX3 7FZ, United Kingdom

Article Info

Keywords:
Cancer
Acute myeloid leukemia
Immunopeptidomics
HLA

ABSTRACT

The identification of T cell epitopes derived from tumour specific antigens remains a significant challenge for the development of peptide-based vaccines and immunotherapies. The use of mass spectrometry-based approaches (immunopeptidomics) can provide powerful new avenues for the identification of such epitopes. In this study we report the use of complementary peptide antigen enrichment methods and a comprehensive mass spectrometric acquisition strategy to provide in-depth immunopeptidome data for the THP-1 cell line, a cell line used widely as a model of human leukemia. To accomplish this, we combined robust experimental workflows that incorporated ultrafiltration or off-line reversed phase chromatography to enrich peptide ligand as well as a multifaceted data acquisition strategy using an Orbitrap Fusion LC-MS instrument. Using the combined datasets from the two ligand enrichment methods we gained significant depth in immunopeptidome coverage by identifying a total of 41,816 HLA class I peptides from THP-1 cells, including a significant number of peptides derived from different oncogenes or over expressed proteins associated with cancer. The physicochemical properties of the HLA-bound peptides dictated their recovery using the two ligand enrichment approaches and their distribution across the different precursor charge states considered in the data acquisition strategy. The data highlight the complementarity of the two enrichment procedures, and in cases where sample is not limiting, suggest that the combination of both approaches will yield the most comprehensive immunopeptidome information.

2.1 INTRODUCTION

Antigen presenting cells (APCs) process both endogenous and exogenous proteins into peptides through two separate intracellular pathways. These peptides are then loaded onto either human leukocyte antigen (HLA) class I or class II molecules, respectively (Aptsiauri et al., 2007). This collective set of peptides presented on the surface of cells is known as the immunopeptidome (Caron et al., 2011). The

immunopeptidome is dynamic and represents a snapshot of the cellular proteome (Blum, 2013). These peptide antigens are displayed to circulating CD8+ or CD4 +T cells in complex with HLA class I and class II molecules, respectively. T cells expressing an antigen-specific T cell receptor (TCR) recognise certain peptide-HLA (pHLA) complexes, leading to activation and induction of effector functions (Aptsiauri et al., 2007). In addition to a large pool of self-peptides, HLA also present peptides associated with various diseases or infections (Chikata et al., 2019; Croft et al., 2019; Purcell et al., 2007; Sahay et al., 2017). Additionally, some peptides may also serve as an immunological indicator of a 'transformed' cell phenotype as they present tumour-associated antigens (TAAs). TAA peptides are derived from cancer-specific antigens whose expression is dysregulated, such as p53, or from proteins that are inappropriately or overexpressed including cancer testis antigens (CTA), or from proteins containing somatic mutations. In these situations, the immune system is capable of recognising these cancer related peptides and eliminating the malignant cells (Freudenmann et al., 2018; Smyth et al., 2006).

One of the challenges in studying the immunopeptidome arises from the highly diverse nature of HLA molecules, which are encoded by the most polymorphic gene cluster in the human genome the major histocompatibility complex (MHC) which to date has 15,966 class I alleles (A, B, C). An individual can possess up to 6 different HLA allomorphs (Cole, 2015). The resulting immunopeptidome therefore contains a complex mixture of peptide ligands displaying different HLA binding characteristics, with estimates of up to 10^5-10^6 distinct peptides presented on the surface of APCs at a given time (Caron et al., 2015; Mason, 1998; Schuster et al., 2018). Over the years considerable progress has been made to identify and characterise immunopeptidomes (Abelin et al., 2017; Freudenmann et al., 2018; Ramarathinam et al., 2018a), which has facilitated a greater understanding of HLA binding characteristics of peptide ligands (binding motifs), as well as improvements in the prediction of peptide binding affinities to specific HLA alleles (Alvarez et al., 2017; Michal Bassani-Sternberg et al., 2017; Garde et al., 2019).

Studies from the groups of Rammensee, Hunt and Engelhard, Sette and others have provided the framework for the identification of cancer-associated neoantigens by mass spectrometry (Barouch et al., 1995; Chicz et al., 1992; Depontieu et al., 2009; Falk et al., 1991; Hunt et al., 1992; Naylor and Tomlinson, 1998; Rammensee, 1995; Zarling et al., 2000). Since these pioneering studies, MS has facilitated the identification of peptides associated not only with different cancers, bacterial and viral infections, as well as other diseases settings of drug hypersensitivity and autoimmunity (Abelin et al., 2017; BassaniSternberg et al., 2016; Chikata et al., 2019; Cobbold et al., 2013; Croft et al., 2019; Dudek et al., 2012; Ebrahimi-Nik et al., 2019; Illing et al., 2012; Loffler et al., 2019; Narayan et al., 2019; Ramarathinam et al., 2018b; Ternette et al., 2018; Zarling et al., 2006). These studies have aided the development of new diagnostic tools and provision of novel targets for immunotherapy. However, further improvements for comprehensive immunopeptidome characterisation are still required, with the greatest hindrance being the bottleneck associated with overall number of peptides identified per sample. In the literature typical yields of 5000–6000 peptides for a sample size of 10^8 cells are reported (Purcell et al., 2019). These numbers can be significantly enhanced with comprehensive fractionation strategies with recent reports of 10–30,000 peptides achieved (Woods et al., 2019). Most published studies investigating pHLA complexes have used a single approach to enrich for the peptides extracted from the cleft of HLA molecules, incorporating either reverse phase high performance liquid chromatography (RP-HPLC) or molecular weight cut-off (MWCO) filters (hereafter referred to as filter) to remove the proteinaceous material from the immunoaffinity-purified pHLA. In this study, we have separated the pHLA eluate derived from a human acute monocytic leukaemia (AML) cell line, known as THP-1, using two different separation techniques (RP-HPLC and filter) to initially separate the eluted peptides. Our data demonstrates surprising complementarity between these two different sample preparation techniques. Moreover, using an optimised data acquisition strategy on an Orbitrap Fusion LC-MS instrument we show significant improvements in immunopeptidome coverage by strategically

analysing singly charged precursor ions in addition to the usually targeted multiply charged ions. 2. Experimental procedures

2.2 MATERIALS AND METHODS

2.2.1. Cell lines and purified antibodies

All the cell lines used in the study i.e. THP-1 parental and THP-1 parental transfected with HLA B*27:05 (Battle et al., 2013; Tsuchiya et al., 1980) along with HMy2.C1R cells (Zemmour et al., 1992) transfected with HLA A*02:01 (hereafter named C1R.A*02:01; see Supplementary Table 1 for details) were maintained in RF10 (RPMI 1640 (Gibco, ThermoFisher, USA) supplemented with 2mM MEM nonessential amino acid solution (Gibco, ThermoFisher, USA), 100mM HEPES (Gibco, ThermoFisher, USA), 2mM L-glutamine (Gibco, ThermoFisher, USA), Penicillin/Streptomycin (Gibco, ThermoFisher, USA), 50 µM 2-mercaptoethanol (Sigma-Aldrich, USA) and 10% heat inactivated foetal calf serum (FCS) (Sigma-Aldrich, USA). Maintenance of C1R.A*02:01 during long-term culture was facilitated by addition of Hygromycin B (0.3 mg/mL; Thermo Fisher Scientific, USA). The HLA class I typing of THP-1 cell line was confirmed as homozygous A*02:01, B*15:11 and C*03:03 (Victorian Transplantation and Immunogenetics Service, Australia). HLA class I expression of cell lines was measured by flow cytometry following surface staining using hybridoma supernatants for both anti-HLA-A2 (BB7.2 [HB-82 ; ATCC] (Parham and Brodsky, 1981); produced in-house) and pan-HLA class I (W6/32 [HB-95; ATCC] (Brodsky and Parham, 1982); produced inhouse) as primary antibodies and goat anti-mouse IgG PE (Southern Biotech, USA) as the secondary antibody. The cell lines were expanded to 1×10^9 in roller bottles, harvested by centrifugation (3724g, 15 mins, 4°C), snap frozen in liquid nitrogen and stored at −80°C until required.

Purified antibody was generated from both HB-82 and HB-95 hybridoma cells cultured in RF5 (same constituents as RF10 but 5% FCS) and expanded in roller bottles at 37°C, 5% CO_2. Secreted monoclonal antibodies, BB7.2 (anti-HLA-A2) and W6/32 (pan-HLA class I), were harvested from spent media and purified using Protein A Sepharose (PAS, CaptivA®, Repligen, USA) using a Profinia purification system (Biorad).

2.2.2. Immunoprecipitation of pHLA complexes

Affinity matrix wherein 10 mg/mL of BB7.2 or W6/32 was crosslinked to 1 mL of PAS resin was prepared as previously described (Ramarathinam et al., 2018b) and used for the immunoaffinity capture of solubilised native pHLA from the THP-1 or C1R-A2 cells. Briefly, frozen cell pellets (1×10^9) were pulverised using a cryogenic mill (Retsch Mixer Mill MM 400), reconstituted in Lysis Buffer [0.5% IGEPAL (Sigma-Aldrich, USA), 50mM Tris pH 8, 150mM NaCl and protease inhibitors (Complete Protease Inhibitor Cocktail Tablet, 1 tablet per 50 mL solution; Roche Molecular Biochemicals, Switzerland)] and incubated for 1 hr at 4°C with rotation. The supernatant was passed through a PAS pre-column (500 µL) to remove non-specific binding material, followed by serial affinity capture of HLA A*02:01 peptides (BB7.2 column) and HLA-B*15:11 and -C*03:03 peptides (W6/32 column). Bound complexes were eluted from the column by using 5 column volumes (CV) of 10% acetic acid (Supplementary Fig. 1).

The eluate was mixed and split into two equal volumes representing peptides isolated from 5×10^8 cells, which were then processed differently. One part of the eluate was fractionated by RP-HPLC on a 4.6 mm internal diameter 100 mm long RP monolithic C18 HPLC column (Chromolith Speed Rod, Merck-Millipore, Germany) using an ÄKTA micro HPLC system (GE Healthcare) running a mobile phase consisting of buffer A (0.1% trifluoroacetic acid [TFA; Thermo Fisher Scientific, USA]) and buffer B (80% acetonitrile [ACN; Thermo Fisher Scientific, USA]/0.1% TFA). A flowrate of 2 mL/min was used to separate the peptides into 1 mL fractions across a changing gradient; 0–15% buffer B over 0.25 min., 15–30% buffer B over 4 mins, 30–40% buffer B over 8 mins, 40–45% buffer B over 10 mins, and 45–

99% buffer B over 2 mins. Peptide-containing fractions were collected, pooled using a concatenation strategy into 10 fractions, vacuum concentrated, and reconstituted in 0.1% formic acid (FA; Thermo Fisher Scientific, USA). The second half of the eluate was passed through a 5 kDa MWCO filter (Amicon), which had been pre-washed with optima water (Thermo Fisher Scientific, USA) and then equilibrated with 10% acetic acid by centrifugation at 16,060g for 30 mins. at room temperature (RT). The sample flow through containing HLA peptides was collected. Finally, the filter was rinsed with buffer A and centrifuged at 16,060g for 30 mins. at RT. The filtered sample was centrifugally evaporated and desalted by reverse phase C_{18} stage tips (Omix, Agilent). The C_{18} tips were washed with 0.1% TFA and peptides were eluted with 40% ACN in 0.1% TFA. Similar processing was conducted for C1R.A*02:01 cells wherein peptides were eluted with BB7.2 antibody.

2.2.3. Analysis of HLA class I bound peptides by LC-MS/MS

All 10 pooled peptide containing fractions following RP-HPLC and the single peptide containing fraction obtained from the 5 kDa MWCO filter were analysed by LC-MS-MS. A mixture of 11 indexed retention time (iRT) peptides were spiked into each sample to aid retention time alignment (Escher et al., 2012). Peptides were loaded onto a PepMap Acclaim 100 C_{18} trap column 5 µm particle size, 100 µm ×2 cm and 100 Å (Thermo Scientific) at 15 µL/min using an Ultimate 3000 RSLC nano-HPLC (Thermo Scientific). The column was equilibrated with 2% ACN and 0.1% FA. Peptides were eluted and separated at a flow rate of 250 µL/min on an in-line analytical column (PepMap RSLC C_{18}, 2 µm particle size, 75 µm × 50 cm and 100 Å, Thermo Scientific) using a 125 mins linear gradient of 2.5–99% of buffer B (80% ACN, 0.1%FA) in buffer A (0.1% FA). The gradient started from 2.5% buffer B and increased to 7.5% buffer B in a minute followed by linear gradient to 37.5% buffer B for 90 mins. followed by increase to 99% buffer B in 10 mins. Peptides were introduced using nano-electrospray ionisation (nano-ESI) method into the Orbitrap Fusion Tribrid MS (Thermo Scientific) at a source temperature of 275°C.

All MS spectra (MS1) profiles were recorded from full ion scan mode 375−1800 *m/z*, in the Orbitrap at 120,000 resolution with automatic gain control (AGC) target of 400,000 and dynamic exclusion of 15 s. The top 12 precursor ions were selected using top speed mode at a cycle time of 2 s. For MS/MS a decision tree was made to aid selecting peptides of charge state 1 and 2–6 separately. For singly charged analytes only ions falling within the range of *m/z* 800–1800 were selected, whereas for +2 to +6 *m/z* s no such parameter was set. The c-trap was loaded with a target of 200,000 ions with an accumulation time of 120 ms and isolation width of 1.2 amu. Normalised collision energy was set to 32 (high energy collisional dissociation [HCD]) and fragments were analysed in the Orbitrap at 30,000 resolution.

The raw data files obtained from the instrument were analysed using Peaks 8.5 software (Bioinformatics solutions) against the human proteome (Uniprot 15/06/2017; 20,182 entries). The following search parameters were used: error tolerance of 10 ppm using monoisotopic mass for precursor ions and 0.02 Da tolerance for fragment ions; enzyme used was set to none with following variable modifications: oxidation at Met, deamidation at Asp and Gln and phosphorylation at Ser, Thr and Tyr. False discovery rate (FDR) was estimated using a decoy fusion method (Zhang et al., 2012) and all datasets were analysed at 5% FDR.

2.2.4. Experiment design and statistical rationale

For the THP-1 cell line the entire experiment was performed in triplicate, wherein each replicate eluate was split into two and the whole procedure performed separately. All the data from triplicate dataset was analysed using GraphPad Prism version 7 (GraphPad Software, USA, www.graphpad.com). The C1R.A*02:01 cell line experiment was performed once (n =1) to corroborate the results from THP-1. The amino acid frequency was graphed as grouped data and the statistically significant P-value was measured by two-way ANOVA corrected for multiple comparisons (Sidak). Area proportional Venn diagrams were prepared using BioVenn (Hulsen et al., 2008). Hydrophobicity of the peptides was

calculated using GRAVY Calculator (accessible at www.gravy-calculator.de) and the percentage of hydrophobic peptides (GRAVY score greater than zero) was graphed as column data and statistically significant P-value was calculated by unpaired T-test with Welch's correction. Error bars for both representations show mean ± SD. Presence and validation of phosphorylated sites was done by PhosphoSitePlus (Hornbeck et al., 2012)

2.3 RESULTS

Three independent biological replicates of THP-1 cells were expanded in vitro and harvested for immunoaffinity purification of their cell surface pHLA. The solubilised pHLA complexes were dissociated in acid and the peptide cargo isolated using two different strategies; monolithic RP-HPLC or 5 kDa MWCO ultrafiltration. All resultant peptide fractions were then run on an Orbitrap Fusion Tribrid LC-MS using identical chromatographic and data acquisition methods. The data after application of a 1% or 5% FDR cut off is provided in the main results files (Supplementary data files 1–4). Only peptides of 7–15 amino acids in length are reported and peptide scans with alternative explanations due to isobaric isoleucine or leucine substitutions were reported separately in Supplementary Data files 5−6. We provide both the more stringent dataset at 1%FDR (Supplementary data files 1 and 3) and the slightly less stringent dataset at 5% FDR (Supplementary data files 2 and 4) since both these cut offs are used in the literature and often many peptides that fail the 1% FDR represent bona fide HLA ligands (Faridi et al., 2018). Peptides from different HPLC and filter replicates were collated together into a master list in Supplementary data file 7 for easy comparison. A summary of data obtained using FDR cut off 1% and 5% is provided in Table 1, for the rest of analyses, peptides obtained at a 5% FDR cut off were used reflecting a common cut-off used for similar immunopeptidome studies (Chikata et al., 2019; Croft et al., 2019; Ebrahimi-Nik et al., 2019; Loffler et al., 2019; Narayan et al., 2019; Ramarathinam et al., 2018a).

Table 1: Summary of results stratified by FDR cut off values. Peptide count at 1% and 5% FDR cut off for 7-15mers post removal of redundancy (peptides with I=L interchange) and modifications. iRT and decoy peptides and redundant peptides with I-L interchange were removed.

	5% FDR redundant count	5% FDR final peptide count	1% FDR redundant count	1% FDR final peptide count
BB72	518	20,316	312	11,821
W632	409	21,520	308	14,319
Total	927	41,816	620	26,201

2.3.1. Combination of datasets from complementary peptide enrichment strategies increases the total number of peptides identified in the THP-1 immunopeptidome

The solubilised pHLA complexes obtained from the THP-1 cell line post-immunoaffinity purification were dissociated in acid and their peptide cargo isolated using two separate methods; RP-HPLC or a 5 kDa MWCO filter. Using the combined datasets and a 5% FDR cut-off, a total of 41,816 unique HLA class I peptides were identified from the THP-1 cell line (n =3), of which 20,316 restricted to HLA A*02:01. The peptide length distribution was as expected, with the majority of peptides being 9mers (Fig. 1a). The HLA binding motif of all 9mer peptides (Fig. 1b) found across the triplicate dataset matched the existing HLAA*02:01 motifs constructed from HLA A*02:01 peptides reported in the immune epitope database (IEDB) (Supplementary Fig. 2a) (Barouch et al., 1995; Parker et al., 1992). Similarly, for pan HLA class I alleles a total of 21,520 peptides were identified and the majority were also 9mers (Fig. 1c). The mixed motif (Fig. 1d) was deconvolved by Gibbs clustering to delineate

peptides bound to either HLA B*15:11 or HLAC*03:03 and motifs for these peptides closely matched those reported for HLA B*15:11 and HLA-C*03:03 peptides in the IEDB database (Supplementary Fig. 2b and 2c). Approximately 75% of the eluted peptides were predicted to bind to one of the HLA allotypes expressed by THP-1 cells using NetMHC pan (4.0 server) with a rank threshold below 2.0. Importantly, we assessed the peptide overlap identified from each of these enrichment procedures (Fig. 2). Surprisingly, there was only 30% and 28% overlap between the two methods for HLA A*02:01 and pan HLA class I alleles, respectively. Thus, peptides identified from the filter fraction contributed an additional 3265 HLA-A2 and 2450 pan class I peptides to the immunopeptidome of THP-1 cells.

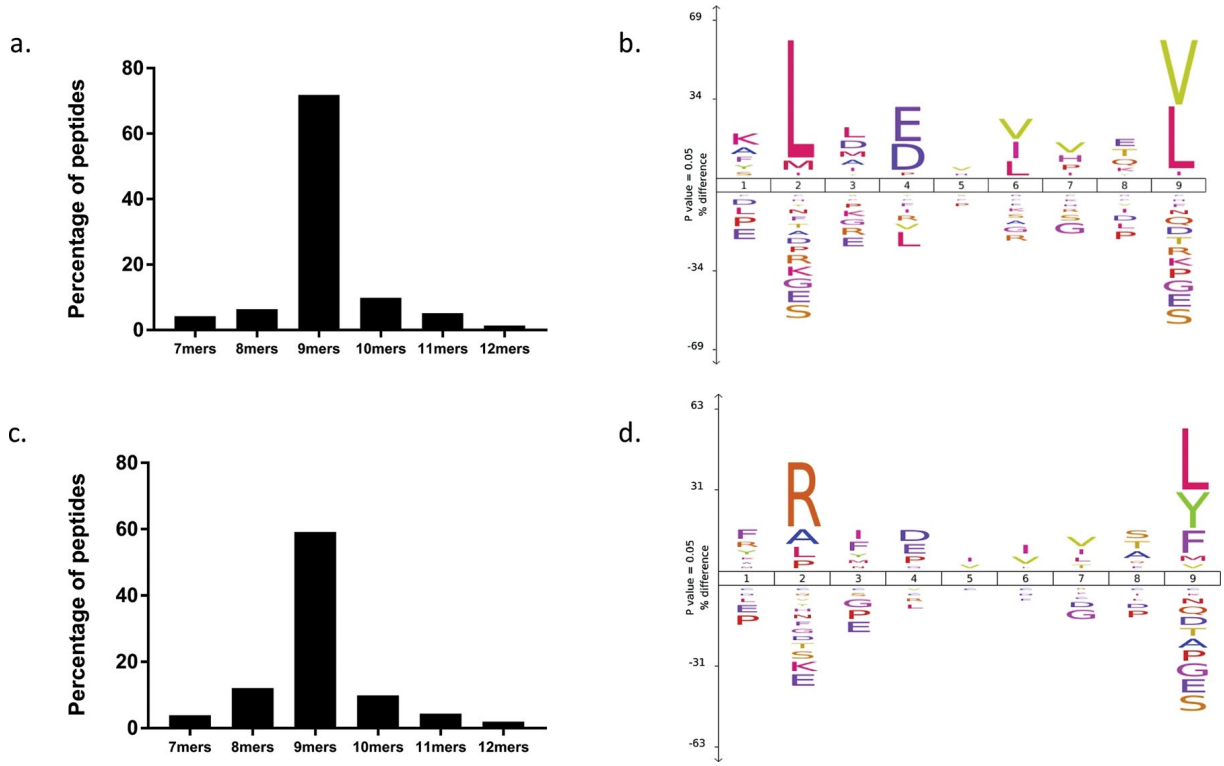


Fig. 1. Peptides isolated from of THP-1 cells follow the canonical peptide length distribution and binding motifs. For peptide length distribution, the percentage of peptides present in higher numbers (i.e. 7 to 12mers) were counted across a triplicate dataset for both (a) HLA A*02:01 and (c) pan HLA class I. The majority of peptides identified for all alleles were 9mers and their binding motif was interpreted using IceLogo (Colaert et al., 2009) for both (b) HLA A*02:01 (n = 12,405) and (d) pan HLA class I (n = 12,459).

2.3.2. Unique singly charged peptides expand the detectable immunopeptidome

Unlike tryptic peptides, HLA bound peptides often contain sequences that tend to form dominant singly charged ions. Therefore, as a part of our study we included LC-MS/MS acquisition of single charged peptides for all HLA class I peptides in addition to double and triple charged peptides. Whilst other studies have investigated singly charged species (Bassani-Sternberg et al., 2016; Ternette et al., 2018) to our knowledge this is the first study to systematically analyse the properties of peptides identified with unique precursor charge states. Interestingly, for both HLA-A2 (Fig. 3) and the remaining pan class I peptides (Supplementary Fig. 3) there was very little overlap in the peptides identified from precursor ions with +1 and +2 charge states or peptides identified from precursor ions with +2 and +3 charge states. This suggests that HLA-bound peptide precursor ions rarely show multiple charge states and highlights the utility of including all relevant peptide precursor ion charge states to obtain comprehensive immunopeptidomes. A total of 6669 and 5424 unique singly charged peptide precursors

were identified restricted to HLA A*02:01 and the remaining class I alleles HLA-B*15:11 and -C*03:03, respectively. The length distribution of peptides present in the different charge states was also evaluated. The +1 charge state mostly comprised of 9mers (~50−52%) followed by shorter 7- and 8-mer peptides (13–14% each) and a very low percentage 10−11mers and near absence of longer peptides. Similar results were observed for the +2 charge state, with majority of peptides (68%) being 9mers along with 10mers (12%). In contrast, there was no clear dominant length for +3 charge state peptides with the frequency of longer peptides on par with the 9−10 mers (Supplementary Fig. 4).

Singly charged peptides were observed to contain more small hydrophobic amino acids including Ala and Ser at P1; Ala, Gly and Ser at P4; and Thr at P8 (Fig. 3b). In contrast, precursor ions that adopted +2 and +3 charge states represented peptides that were enriched in charged amino acids including Lys at P1; Asp and Glu at P4; and Lys and Glu at P8 (Fig. 3c). At the same time peptides presented in different charge states retained their preference for Leu and Val at P2 and P9, which represent key HLA-binding anchor residues for these peptides. An interesting observation was made for the other class I alleles in the study which included HLA B*15:11, C*03:03 and B*27:05 (expressed in transgenic THP-1). Since this dataset contained a mix of peptides analysis was done at two levels. We first performed Gibbs clustering to stratify peptides and assign them as ligands to a putative HLA-allotype and further refined their restriction based on NetMHC 4.0 binding predictions. The majority of +1 charge state peptide precursors were restricted by either HLA B*15:11 or HLA C*03:03. Whereas for peptide precursors with a +2 charge state, there was a marked increase in peptides restricted by B*27:05 along with the other alleles.

2. 3.3. Properties of the unique 9mer peptides identified in RP-HPLC and filter fractions

To ascertain the difference between the unique 9mer peptides identified using either the RP-HPLC or filter fractions, several parameters including amino acid frequency and peptide hydrophobicity were evaluated. The amino acid frequency was calculated for each position from P1 to P9 for all peptides unique to either RP-HPLC or filter enriched fractions. For THP-1-derived HLA A*02:01 peptides the most significant increase was observed for Leu at P2 ($p < 0.0001$) and P9 ($p < 0.0001$) in the filter fraction (Fig. 4). This represents an enrichment of P9 Leu and a decrease in the other anchor residue, Val, at this position in the filter enriched peptidome. Some significant changes were also shown in the frequency of different amino acids at various positions from P3 to P8 (Supplementary Fig. 5), with most changes occurring for hydrophobic residues.

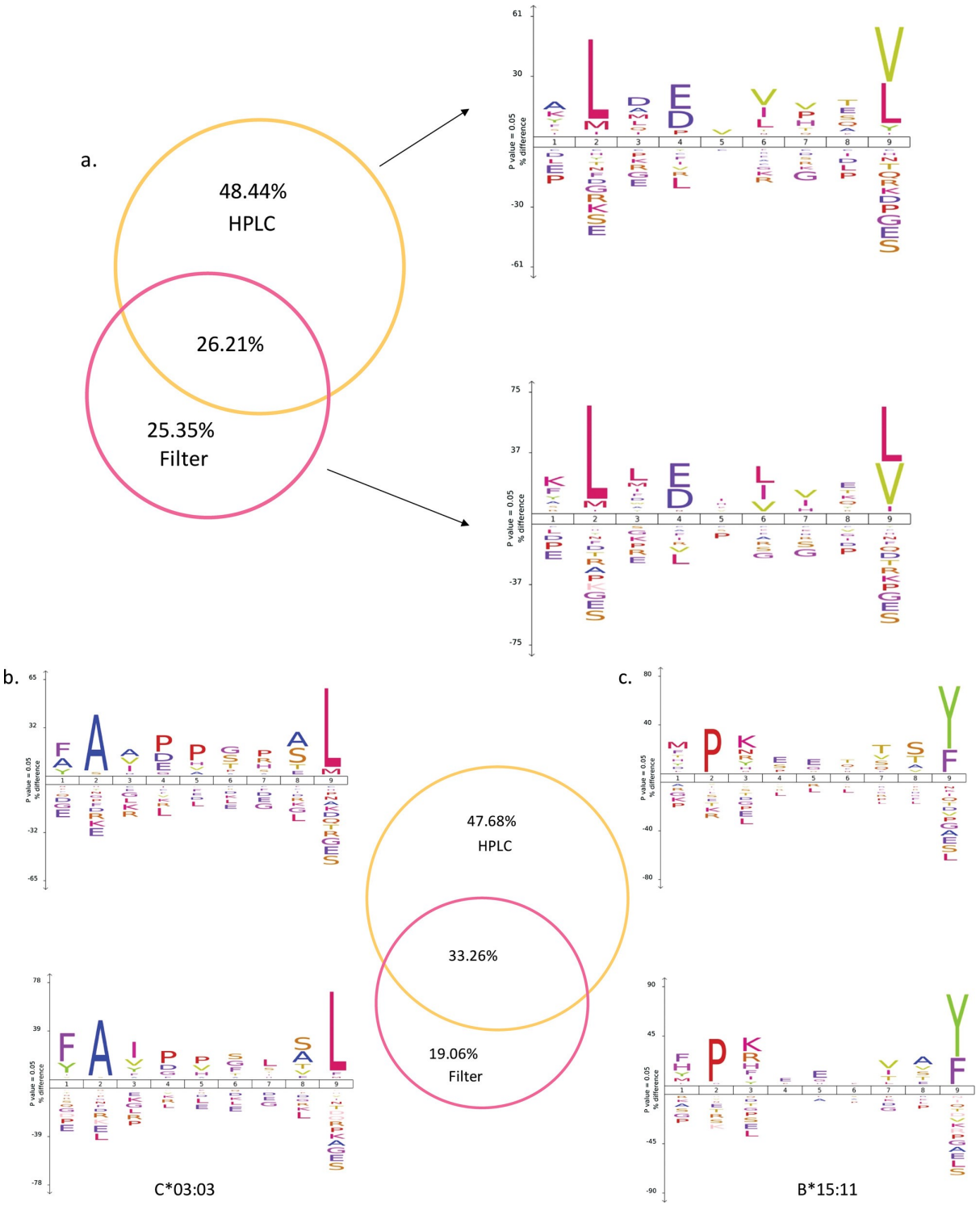


Fig. 2. Orthogonal separation allows in-depth coverage of the immunopeptidome by identifying peptides unique to each method. (a) For HLA A*02:01 9mers were identified in each method showed very little overlap. The peptide motifs unique to each method were plotted using IceLogo and showed marked differences. Similarly, for the pan HLA class I alleles (b) unique peptides were identified, and Gibbs clustering was applied to reveal peptide motifs for alleles (b) C*03:03 and (c) B*15:11.

For pan HLA class I unique 9mers were selected and Gibbs clustering was performed to tease apart the peptides belonging to HLAB*15:11 or C*03:01. For HLA B*15:11, significant changes were observed marked by increase in Pro and Phe at P2 and P9, respectively (Fig. 4b) along with significant increases

in Phe at P1 and charged amino acids including Lys and Arg at P3 (Supplementary Figure 6). Surprisingly, for HLA-C*03:03 no difference in frequency of small amino acids Ala and Leu at anchor residue positions P2 and P9, respectively were observed (Fig. 4c). However, significant changes were found for the percentage of Phe and Tyr at P1 and Phe at P6 (Supplementary Figure 7).

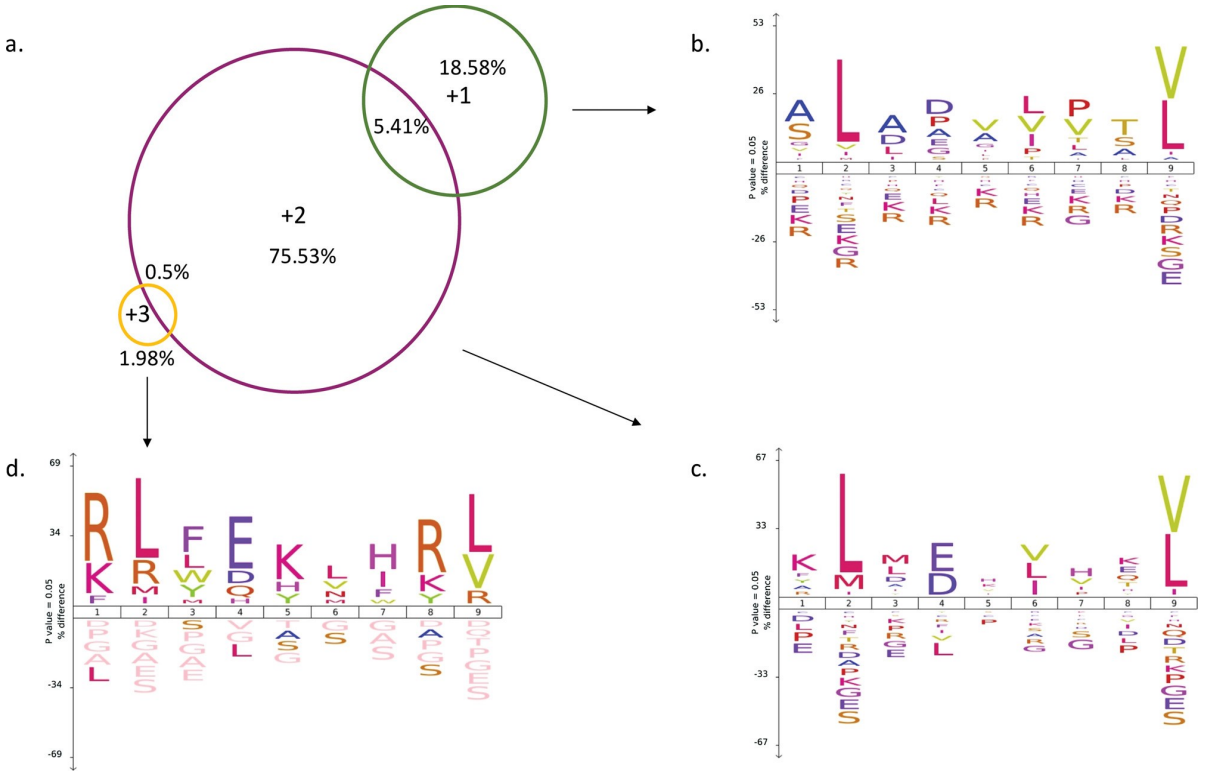


Fig. 3 - Little overlap in peptide identity between precursor ions that adopt +1, +2 and +3 charge states in the HLA–A*02:01 restricted THP-1 immunopeptidome (n =3). (a) The peptides (7 to 15mers) found in different charge states had very little overlap amongst them. The 9mers unique to each charge state were examined and their binding motif was analysed using IceLogo (Colaert et al., 2009) showing the difference in the amino acid composition of (b) +1 charge state (n = 3521) c) +2 charge state (n =9742) and d) +3 charge state (n = 95).

The overall hydrophobicity of the peptides was determined by calculating the average of hydropathy score (GRAVY score (Kyte and Doolittle, 1982)), wherein a score greater than 0 indicates a hydrophobic nature. The percent of hydrophobic peptides was determined for unique peptides found in both RP-HPLC and filter fractions. Surprisingly, unique HLA A*02:01 peptides identified from the filter fraction were more hydrophobic than the peptides only identified in the RPHPLC fractions (Fig. 5).

2.3.4. The majority of HLA-A2 bound phosphopeptides were identified in the RP-HPLC enriched fraction

Phosphorylation is one of the most important post-translational modifications (PTM) studied, as the process itself is dysregulated in some diseases, especially cancer (Cobbold et al., 2013; Zarling et al., 2006). Phosphorylated peptides being different from self-peptides can potentially act as neo-epitopes in cancer (Cobbold et al., 2013; Zarling et al., 2006). A total of 1075 phosphopeptides were identified in the study and they constituted only 2.6% of total peptides (Supplementary Table 2). For both the datasets (A2 and pan class I) the majority of the peptides (52–55%) were phosphorylated at Ser followed by Thr (~34–35%) and remaining at Tyr. With the exception of 8mer peptides (P3, P4 and P5 are possible sites for phosphorylation), the observed phospho-sites for 9mer and 10mers were similar to previous studies

where in P4 was the preferred position of phosphorylation (Supplementary Figure 12). For HLA-A2, a total of 487 phosphopeptides were identified across the THP-1 dataset. Of the HLA-A2 bound phosphopeptides, 340 were identified from the RP-HPLC fractions, whilst the filter fraction contributed the remaining 127 phosphopeptides. The overlap between the peptides was minimal, with only 20 common peptides identified from both enrichment strategies. Likewise, for the pan HLA class I immunopeptidomes a total 588 phosphopeptides were identified, with 375 peptides identified from the RP-HPLC enriched fraction and remaining 175 unique to filter fraction, with an overlap of 38 peptides. The increased identification of phosphopeptides in RP-HPLC fractions is a result of extensive fractionation that helps to identify less abundant PTM peptides versus filter fraction, which is an unfractionated single injection onto the mass spectrometer. The difference can also be explained in terms of hydrophilicity of peptides, with phosphopeptides being more hydrophilic in nature were found in the RP-HPLC fraction.

2.3.5. Transgenic THP-1 and mono-allelic HLA cell line data

In addition to parental THP-1 cells, we also performed a similar investigation using a transgenic THP-1 expressing HLA B*27:05 along with the other HLA expressed by parental THP-1. This experiment was performed to see if the presence of a charged and bulky amino acid Arg would suffer any bias in filter separation. Similar to previous parental THP-1 experiments, the transgenic THP-1 eluate was split and processed by both methods. To identify peptides unique to each method, overlap between peptides were examined and showed that there was a 20.3% overlap for HLA-A2 and 35.3% overlap for pan HLA class I alleles. As shown for the HLA allotypes expressed by the parental THP-1 unique peptides, there was also a difference in motifs for the peptides bound to the introduced HLA B*27:05 allele when enriched using HPLC or filter, particularly at P1, P4 and P8 (Supplementary Figure 8).

A similar study was also performed on the “mono allelic” cell line expressing HLA A*02:01 (C1R.A2). A total of 14,916 peptides (7 to 15mers at 5% FDR) were identified, with 8280 peptides from RP-HPLC and 6636 peptides from filter fractions. Similar to the THP-1 dataset, the majority of peptides from both methods were 9mers (Supplementary Figure 9a and 9b) with binding motifs matching existing data in IEDB (Supplementary Fig. 2a). Unique peptides identified in the filter enriched fraction had a subtle difference in their motif (Supplementary Figure 9c and 9d). This variation could be attributed to an increase in the frequency of hydrophobic residues in the filter fraction, including Leu at P2 and P9 and Val at P9 (Supplementary Figures 10a and 10b). Similar to THP-1, a difference was seen in hydrophobicity with more of such peptides in filter fraction (Supplementary Figure 10c).

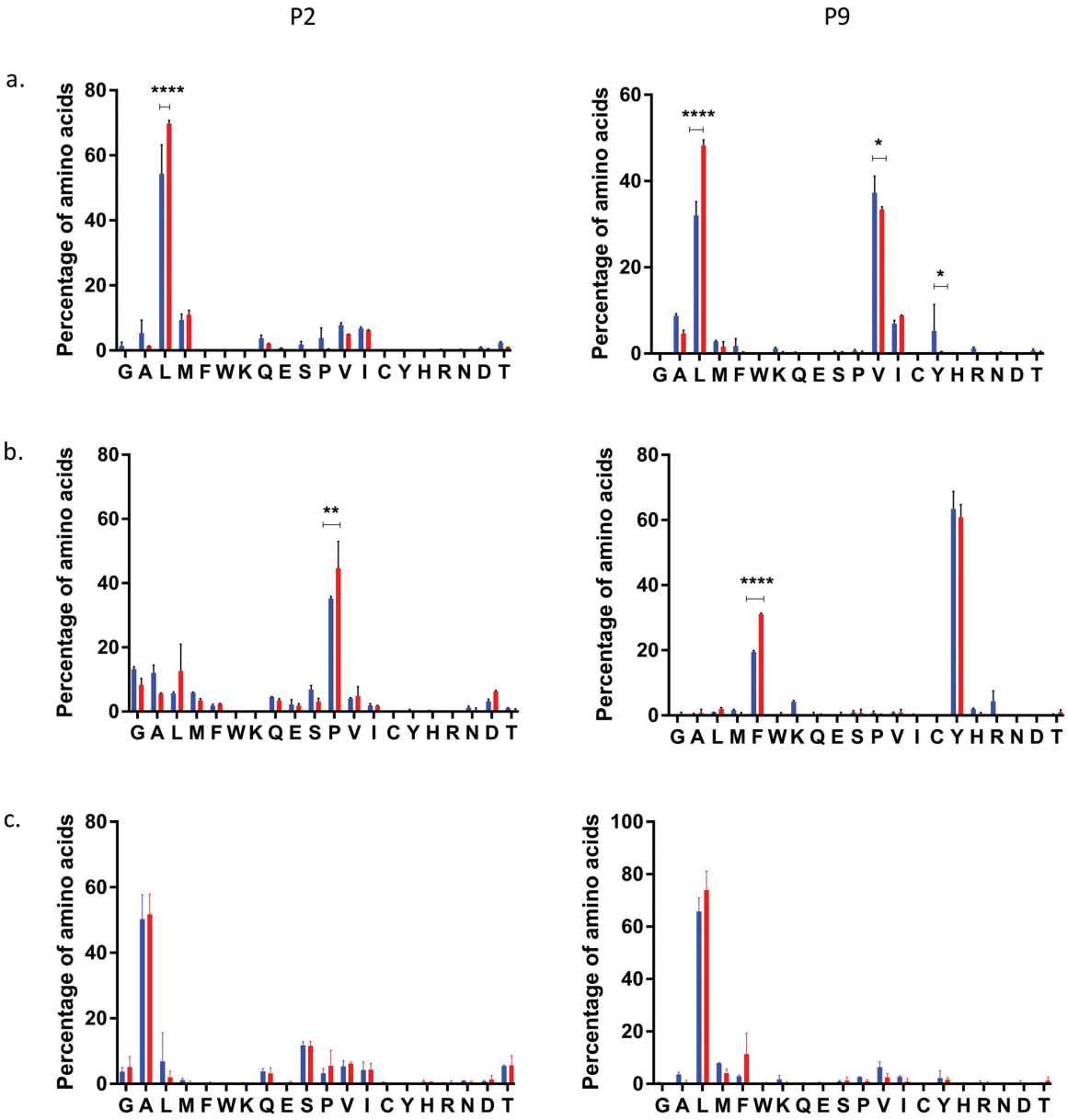


Fig. 4. **Amino acid frequency at anchor residues P2 and P9 for unique 9mers identified** in RP-HPLC (blue) and filter (red) enriched fractions. The amino acid frequency was calculated for 9mers across a duplicate THP-1 dataset. (a) For HLA-A2, a significant increase was observed in frequency of Leu at P2 in the filter fraction and at P9 significant increases were seen in the prevalence of Leu and Tyr in RP-HPLC 9mer peptides. (b) For B*15:11, a significant increase was observed in frequency of Pro at P2 and Phe at P9 and c) no significant difference was seen for amino acid frequencies at the anchor residues for the C*03:03. p-value was measured by two-way Anova corrected for multiple comparisons (Sidak) where **** $p < 0.0001$ and ** $p = 0.0023$, ** $p = 0.0089$, * $p = 0.0213$, error bars represent mean ± SD from a duplicate dataset (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

2.3.6. Mining of the immunopeptidome data for cancer associated antigens

The list of peptides identified in THP-1 cell line was also interrogated for presence of TAAs and CTAs using Tumour T-cell antigen database (TANTIGEN, http://cvc.dfci.harvard.edu/tantigen) (Olsen et al., 2017) and CTdatabase (http://www.cta.lncc.br/) respectively (Almeida et al., 2009). A total of 1687 TAAs were identified across the THP-1 dataset with 826 HLA-A2 and 861 pan HLA class I peptides (Table 2 and Supplementary file 8).

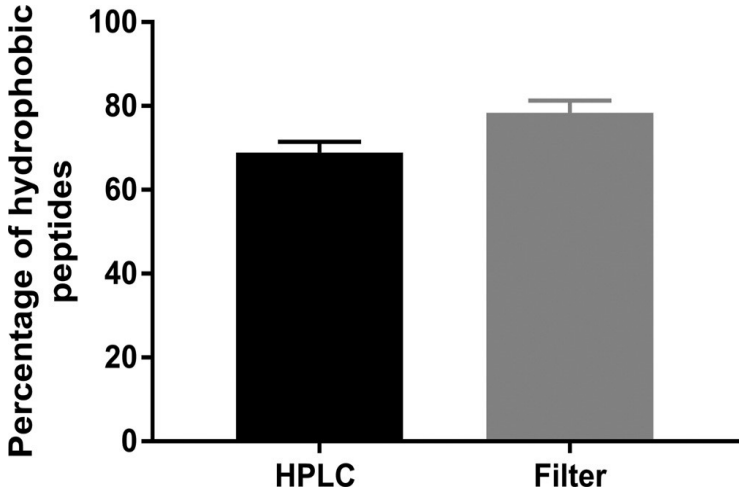


Fig. 5. Presence of more hydrophobic peptides in the filter fraction. The GRAVY score of all the unique 9mer peptides identified in RP-HPLC and filter fractions was calculated using www.gravy-calculator.de and the percentage of hydrophobic peptides (GRAVY score greater than zero) were plotted as column data and statistically significant p-value was calculated by unpaired T-test with Welch's correction. Error bars for both representations show mean ± SD. The RP-HPLC fraction contained less hydrophobic 9mers than the filter fraction.

Table 2 – The number and annotation of HLA bound peptides identified in THP-1 cell line at 5% FDR

HLA Allele (Antibody used)	Total number of peptides	CTA peptides (CTdatabase)	Oncogene associated peptides (TANTIGEN)	Phosphopeptides
A*02:01	20,316	55	826	487
B*15:11 and C*03:03	21,520	85	861	588
Total	41,816	140	1,687	1,075

For nonamers TAAs found in HLA-A2, 77% of were predicted to be strong binders (rank threshold of 0.5). Whereas for pan HLA class I, 25% of B*15:11, 24% of C*03:03 and 29% of B*27:05 were strong binders. A total of 140 peptides were found to be originating form CTA proteins, wherein 55 were restricted by HLA-A2 and 85 peptides restricted by alleles in the pan class I fraction. Nonamers from these CTA were assessed for their binding ability and it was found that 83% of A2 peptides were strong binders (rank threshold of 0.5) whereas for B*15:11, 16 peptides were found to be strong binders, 17 peptides were strong binders for C*03:03 and 12 strong binders for B*27:05. A detailed list of the peptides and their source proteins is provided in Supplementary file 8.

2.4 DISCUSSION

In this study, we determined the HLA class I peptide repertoire from a human AML cell line (THP-1) using two different strategies, namely RP-HPLC and filter enrichment of the acidified immunoprecipitates of pHLA. By orthogonally separating the same sample eluate we successfully expanded the known pHLA immunopeptidome. The peptides were studied both at 1% and 5% FDR and we found a total of 26,201 and 41,816 unique HLA class I peptides at the respective FDR cut-off values

considered (Table 1 and Supplementary files 1–4), thereby making this the largest reported dataset in AML (Narayan et al., 2019) and for the frequently studied THP-1 cell line. To ensure that the peptides identified were bona fide ligands we analysed nonamers (which constituted the majority of the peptides identified) and their motifs matched those previously reported from the HLA allotypes expressed by THP-1. The dataset was also predominated by binders as verified by NetMHC pan 4.0 (Hoof et al., 2009; Jurtz et al., 2017). The current IEDB dataset for HLA-B*15:11 and -C*03:03 allotypes is very limited with only 88 and 1017 nonamers, respectively. Here, we identified a total of 2769 matched to HLA-B*15:11 and 3519 for HLA-C*03:03 providing a considerable enhancement of allotype specific peptides.

In recent years, there have been only a handful of studies which have included singly charged peptides in their data acquisition strategies (Bassani-Sternberg et al., 2015; Ternette et al., 2018) and due to advent of additional fragmentation techniques, including EtHCD, the focus has slowly shifted to identification of peptides with charge states of +3 and above (Frese et al., 2013; Mommen et al., 2014). Although these techniques enhance fragmentation or provide added spectral information, thereby aiding in accurate identification of peptides, they do not consider singly charged peptide ions. In our study, we not only used a decision tree-based design to examine singly charged peptide precursor ions (*m/z* 800–1800 amu), which resulted in the identification of 12,093 unique peptides across HLA-A2 and pan HLA class I datasets. The +1 charge state peptides constituted around 27% and 21% of total peptides in HLA A2 and pan class I across triplicate dataset. Only 4–5% of these peptides had a corresponding +2 charge species and none +3 species. The data therefore suggests that the populations of peptides that yield +1 and +2 charge state precursors are quite different for each allotype suggesting each charge state yields unique peptide identifications. Based on inclusion of a transgenic B*27:05 THP-1 cell line, we also conclude that it is much more beneficial to investigate +1 charge state peptides for alleles which present hydrophobic peptides e.g. HLA A*02:01, B*15:11 and C*03:03. We also suggest that regardless of the alleles expressed by the cell line gathering +1 peptide data is important when exploring immunogenic T cell epitopes as this property is independent of charge state of the peptide. For instance, we found that 5 known T cell epitopes derived from TAAs, including peptide SVASTITGV (Perilipin) and LLDVAPLSL (Heat shock protein 70) which were previously reported as either potential T cell epitopes or known immunogens (Faure et al., 2004; Gloger et al., 2016), that were contained within our THP-1 immunopeptidome and identified only from the singly charged precursor data. Thus, the inclusion of single charged peptides in future studies is imperative for a more comprehensive analysis of the immunopeptidome.

Surprisingly, there was little overlap between peptides identified from both ligand enrichment strategies. To delineate the underlying reason, we first checked if this was due to variability in the data itself. For this we checked difference in biological and technical replicates. We found that there was ~32% overlap between peptides identified in individual replicates, which is consistent with other studies that have reported 23–46% overlap between samples (Newey et al., 2019). For technical replicates we found 52% and ~66% for A2 and pan class I respectively between any two replicates and between all three replicates ~39% and ~53% overlap in A2 and pan class I respectively with few hundred peptides unique to each replicate. This is similar to other studies (Partridge et al., 2018) wherein the reproducibility and overlap of immunopeptidomics data has been reported and reflects the sample complexity. Hence, we conclude that the differences in peptide identifications seen between the two enrichment strategies is more likely due to an inherent bias in these two ligand enrichment techniques. Our study shows that both approaches although complementary, yield unique peptides, and this cannot simply be explained by the degree of fractionation in case of HPLC or multiple runs of the unfractionated MWCO sample as none of them are able to recapitulate the population of peptides unique to the other method. We evaluated several parameters which might contribute to the different subsets of unique peptides identified including differences in positional amino acid frequencies and hydrophobicity. For HLA A*02:01 filter enriched peptides, a significant increase in hydrophobic

residues including Leu, Val and Ile was seen at P2 and P9 along with significant increases in bulky hydrophobic residues Phe at P3 and P5 and Trp at position P3. This contributed to the increase in average hydrophobicity of the peptides identified from the filter enriched fraction. Similar results were seen for the HLA A*02:01-bound peptides isolated from a mono-allelic C1R cell line and for peptides mapped to HLA B*15:11 and C*03:03 alleles.

We also studied the influence of both separation techniques had on identifying PTMs, particularly phosphorylation. Identification of phosphopeptides can be a tedious process due to their relatively low abundance, thereby occasionally requiring an additional enrichment step for detection especially for global phosphoproteomics. Recently a few immunopeptidomics studies have identified phosphopeptides without enrichment (Abelin et al., 2017; Bassani-Sternberg et al., 2016; Demmers et al., 2019). With our method wherein, we have used high resolution at both MS1 and MS2 level and have incorporated longer accumulation times leading to high quality spectra we were able to identify over 1000 phosphopeptides. This included several previously reported pHLA species (Cobbold et al., 2013; Solleder et al., 2020). Moreover, ~23% of the phosphosites were also present in the PhosphoSitePlus database (Hornbeck et al., 2012) thereby supporting the correct identification of some of these phosphopeptides. (Casado et al., 2013; Cobbold et al., 2013). Surprisingly, most phosphopeptides were detected by RP-HPLC enrichment, consistent with the bias towards more hydrophilic peptides in this enrichment strategy. Another reason contributing to this might be that phosphopeptides identified in the THP-1 cell line, irrespective of the HLA allele, have the XSPXX motif which is enriched in RP-HPLC enriched peptides (Supplementary Figure 11) as evident from the increase in the amino acid frequencies of Ser and Pro at P4 and P5 in these fractions for all alleles studied (A*02:01, B*15:11, C*03:03 and B*27:05) (Supplementary Figure 5, 6 and 7).

HLA class I peptide ligands are challenging to analyse due to enormous diversity in the repertoire and the low abundance of most of these species. Herein we present a large immunopeptidomics data set from a highly studied model cell line of human AML by using two widespread and complementary approaches i.e. RP-HPLC and MWCO. Our study suggests that for abundant samples RP-HPLC is the preferred method for ligand enrichment since it facilitates deeper coverage and aids in the identification of less abundant peptides. Moreover, we show that the combination of RP-HPLC and filter ligand enrichment strategies can expand the detected peptide repertoire of the THP-1 cell line as both approaches provide additive information. Both approaches may be useful when more comprehensive peptidome analysis is required that incorporate more targeted analysis such as data independent acquisition (DIA) strategies. One could envision the generation of extensive DIA spectral libraries from more abundant antigen sources which can then be used to interrogate, at much higher level of sensitivity, more precious clinical samples.

Funding

This work was supported by the Australian Research Council (ARC), Australian National Health and Medical Research Council (NHMRC) Projects (1122099 and 1165490) and an Accelerator Program Grant form Cancer Research UK (A21998) to N.T. and A.W.P. A.W.P. is supported by a NHMRC Principal Research Fellowship (1137739).

Computational resources were supported by the R@CMon/Monash Node of the NeCTAR Research Cloud, an initiative of the Australian Government's Super Science Scheme and the Education Investment Fund

Data availability

TheMSproteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2018) partner repository with the dataset identifier PXD015039 as. raw files, mzIdentML files exported by PEAKS Studio 8.5.

Credit authorship contribution statement

Kirti Pandey: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing review & editing, Visualization. Nicole A. Mifsud: Resources, Writing review & editing, Supervision. Terry C.C. Lim Kam Sian: Resources, Writing - review & editing, Visualization. Rochelle Ayala: Resources, Writing - review & editing. Nicola Ternette: Methodology, Resources, Writing - review & editing, Funding acquisition. Sri H. Ramarathinam: Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision. Anthony W. Purcell: Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgments

We thank staff at the Monash Biomedical Proteomics Facility for technical assistance.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.molimm.2020.04.008.

References

Abelin, J.G., Keskin, D.B., Sarkizova, S., Hartigan, C.R., Zhang, W., Sidney, J., Stevens, J., Lane, W., Zhang, G.L., Eisenhaure, T.M., Clauser, K.R., Hacohen, N., Rooney, M.S., Carr, S.A., Wu, C.J., 2017. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. Immunity 46, 315–326.

Almeida, L.G., Sakabe, N.J., deOliveira, A.R., Silva, M.C., Mundstein, A.S., Cohen, T., Chen, Y.T., Chua, R., Gurung, S., Gnjatic, S., Jungbluth, A.A., Caballero, O.L., Bairoch, A., Kiesler, E., White, S.L., Simpson, A.J., Old, L.J., Camargo, A.A., Vasconcelos, A.T., 2009. CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. Nucleic Acids Res. 37, D816–819.

Alvarez, B., Barra, C., Nielsen, M., Andreatta, M., 2018. Computational tools for the identification and interpretation of sequence motifs in Immunopeptidomes. Proteomics 12, 1700252.

Aptsiauri, N., Cabrera, T., Garcia-Lora, A., Lopez-Nevot, M.A., Ruiz-Cabello, F., Garrido, F., 2007. MHC Class I Antigens and Immune Surveillance in Transformed Cells, Int. Rev. Cytol. Academic Press, pp. 139–189.

Barouch, D., Friede, T., Stevanović, S., Tussey, L., Smith, K., Rowland-Jones, S., Braud, V., McMichael, A., Rammensee, H.G., 1995. HLA-A2 subtypes are functionally distinct in peptide binding and presentation. J. Exp. Med. 182, 1847–1856.

Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L.J., Mann, M., 2015. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol. Cell Proteomics 14, 658–673.

Bassani-Sternberg, M., Braunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., Weber, J., Slotta-Huspenina, J., Specht, K., Martignoni, M.E., Werner, A., Hein, R., D, H.B, Peschel, C., Rad, R., Cox, J., Mann, M., Krackhardt, A.M., 2016. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. Nat. Commun. 7, 13404.

Bassani-Sternberg, M., Chong, C., Guillaume, P., Solleder, M., Pak, H., Gannon, P.O., Kandalaft, L.E., Coukos, G., Gfeller, D., 2017. Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity. PLoS Comput Biol 13, e1005725.

Battle, R., Poole, K., Haywood-Small, S., Clark, B., Woodroofe, M.N., 2013. Molecular characterisation of the monocytic cell line THP-1 demonstrates a discrepancy with the documented HLA type. Int. J. Cancer 132, 246–247.

Blum, J.S., 2013. Pathways Antigen Processing 31, 443–473.

Brodsky, F.M., Parham, P., 1982. Monomorphic anti-HLA-A,B,C monoclonal antibodies detecting molecular subunits and combinatorial determinants. J. Immunol. 128, 129–135.

Caron, E., Vincent, K., Fortier, M.H., Laverdure, J.P., Bramoulle, A., Hardy, M.P., Voisin, G., Roux, P.P., Lemieux, S., Thibault, P., Perreault, C., 2011. The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. Mol. Syst. Biol. 7, 533.

Caron, E., Kowalewski, D.J., Chiek Koh, C., Sturm, T., Schuster, H., Aebersold, R., 2015. Analysis of major histocompatibility complex (MHC) immunopeptidomes using mass spectrometry. Mol. Cell Proteomics 14, 3105–3117.

Casado, P., Alcolea, M.P., Iorio, F., Rodríguez-Prados, J.-C., Vanhaesebroeck, B., SaezRodriguez, J., Joel, S., Cutillas, P.R., 2013. Phosphoproteomics data classify hematological cancer cell lines according to tumor type and sensitivity to kinase inhibitors. Genome Biol. 14, R37.

Chicz, R.M., Urban, R.G., Lane, W.S., Gorga, J.C., Stern, L.J., Vignali, D.A., Strominger, J.L., 1992. Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. Nature 358, 764–768.

Chikata, T., Paes, W., Akahoshi, T., Partridge, T., Murakoshi, H., Gatanaga, H., Ternette, N., Oka, S., Borrow, P., Takiguchi, M., 2019. Identification of immunodominant HIV1 epitopes presented by HLA-C*12:02, a protective allele, using an immunopeptidomics approach. J. Virol. 93.

Cobbold, M., De La Pena, H., Norris, A., Polefrone, J.M., Qian, J., English, A.M., Cummings, K.L., Penny, S., Turner, J.E., Cottine, J., Abelin, J.G., Malaker, S.A., Zarling, A.L., Huang, H.W., Goodyear, O., Freeman, S.D., Shabanowitz, J., Pratt, G., Craddock, C., Williams, M.E., Hunt, D.F., Engelhard, V.H., 2013. MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia. Sci. Transl. Med. 5, 203ra125.

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., Gevaert, K., 2009. Improved visualization of protein consensus sequences by iceLogo. Nat. Methods 6, 786–787.

Cole, D.K., 2015. The ultimate mix and match: making sense of HLA alleles and peptide repertoires. Immunol. Cell Biol. 93, 515–516.

Croft, N.P., Smith, S.A., Pickering, J., Sidney, J., Peters, B., Faridi, P., Witney, M.J., Sebastian, P., Flesch, I.E.A., Heading, S.L., Sette, A., La Gruta, N.L., Purcell, A.W., Tscharke, D.C., 2019. Most viral peptides displayed by class I MHC on infected cells are immunogenic. Proc. Natl. Acad. Sci. 116, 3112–3117.

Demmers, L.C., Heck, A.J.R., Wu, W., 2019. Pre-fractionation extends but also creates a Bias in the detectable HLA class iota ligandome. J. Proteome Res. 18, 1634–1643.

Depontieu, F.R., Qian, J., Zarling, A.L., McMiller, T.L., Salay, T.M., Norris, A., English, A.M., Shabanowitz, J., Engelhard, V.H., Hunt, D.F., Topalian, S.L., 2009. Identification of tumor-associated, MHC class II-restricted phosphopeptides as targets for immunotherapy. Proc. Natl. Acad. Sci. U.S.A. 106, 12073–12078.

Dudek, N.L., Tan, C.T., Gorasia, D.G., Croft, N.P., Illing, P.T., Purcell, A.W., 2012. Constitutive and inflammatory immunopeptidome of pancreatic β-Cells. Diabetes 61, 3018–3025.

Ebrahimi-Nik, H., Michaux, J., Corwin, W.L., Keller, G.L., Shcheglova, T., Pak, H., Coukos, G., Baker, B.M., Mandoiu, I.I., Bassani-Sternberg, M., Srivastava, P.K., 2019. Mass spectrometry driven exploration reveals nuances of neoepitope-driven tumor rejection. JCI Insight 5.

Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., MacCoss, M.J., Rinner, O., 2012. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics 12, 1111–1121.

Falk, K., Rotzschke, O., Stevanovic, S., Jung, G., Rammensee, H.G., 1991. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351, 290–296.

Faridi, P., Purcell, A.W., Croft, N.P., 2018. In immunopeptidomics we need a sniper instead of a shotgun. Proteomics.

Faure, O., Graff-Dubois, S., Bretaudeau, L., Derre, L., Gross, D.A., Alves, P.M., Cornet, S., Duffour, M.T., Chouaib, S., Miconnet, I., Gregoire, M., Jotereau, F., Lemonnier, F.A., Abastado, J.P., Kosmatopoulos, K., 2004. Inducible Hsp70 as target of anticancer immunotherapy: identification of HLA-A*0201-restricted epitopes. Int. J. Cancer 108, 863–870.

Frese, C.K., Zhou, H., Taus, T., Altelaar, A.F.M., Mechtler, K., Heck, A.J.R., Mohammed, S., 2013. Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). J. Proteome Res. 12, 1520–1525.

Freudenmann, L.K., Marcu, A., Stevanović, S., 2018. Mapping the tumour human leukocyte antigen (HLA) ligandome by mass spectrometry. Immunology 154, 331–345. Garde, C., Ramarathinam, S.H., Jappe, E.C., Nielsen, M., Kringelum, J.V., Trolle, T., Purcell, A.W., 2019. Improved peptide-MHC

class II interaction prediction through integration of eluted ligand and peptide affinity data. Immunogenetics.

Gloger, A., Ritz, D., Fugmann, T., Neri, D., 2016. Mass spectrometric analysis of the HLA class I peptidome of melanoma cell lines as a promising tool for the identification of putative tumor-associated HLA epitopes. Cancer Immunol. Immunother. 65, 1377–1393.

Hoof, I., Peters, B., Sidney, J., Pedersen, L.E., Sette, A., Lund, O., Buus, S., Nielsen, M., 2009. NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics 61, 1–13.

Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., Sullivan, M., 2012. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. 40, D261–270.

Hulsen, T., de Vlieg, J., Alkema, W., 2008. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics 9, 488.

Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E., Engelhard, V.H., 1992. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science 255, 1261–1263.

Illing, P.T., Vivian, J.P., Dudek, N.L., Kostenko, L., Chen, Z., Bharadwaj, M., Miles, J.J., Kjer-Nielsen, L., Gras, S., Williamson, N.A., Burrows, S.R., Purcell, A.W., Rossjohn, J., McCluskey, J., 2012. Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. Nature 486, 554–558.

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., Nielsen, M., 2017. NetMHCpan4.0: improved Peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. J. Immunol. 199, 3360–3368.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

Loffler, M.W., Mohr, C., Bichmann, L., Freudenmann, L.K., Walzer, M., Schroeder, C.M., Trautwein, N., Hilke, F.J., Zinser, R.S., Muhlenbruch, L., Kowalewski, D.J., Schuster, H., Sturm, M., Matthes, J., Riess, O., Czemmel, S., Nahnsen, S., Konigsrainer, I., Thiel, K., Nadalin, S., Beckert, S., Bosmuller, H., Fend, F., Velic, A., Macek, B., Haen, S.P., Buonaguro, L., Kohlbacher, O., Stevanovic, S., Konigsrainer, A., Rammensee, H.G., 2019. Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. Genome Med. 11, 28.

Mason, D., 1998. A very high level of crossreactivity is an essential feature of the T-cell receptor. Immunol. Today 19, 395–404.

Mommen, G.P.M., Frese, C.K., Meiring, H.D., van Gaans-van den Brink, J., de Jong, A.P.J.M., van Els, C.A.C.M., Heck, A.J.R., 2014. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc. Natl. Acad. Sci. 111, 4507–4512.

Narayan, R., Olsson, N., Wagar, L.E., Medeiros, B.C., Meyer, E., Czerwinski, D., Khodadoust, M.S., Zhang, L., Schultz, L., Davis, M.M., Elias, J.E., Levy, R., 2019. Acute myeloid leukemia immunopeptidome reveals HLA presentation of mutated nucleophosmin. PLoS One 14, e0219547.

Naylor, S., Tomlinson, A.J., 1998. Membrane preconcentration-capillary electrophoresis tandem mass spectrometry (mPC-CE-MS/MS) in the sequence analysis of biologically derived peptides. Talanta 45, 603–612.

Newey, A., Griffiths, B., Michaux, J., Pak, H.S., Stevenson, B.J., Woolston, A., Semiannikova, M., Spain, G., Barber, L.J., Matthews, N., Rao, S., Watkins, D., Chau, I., Coukos, G., Racle, J., Gfeller, D., Starling, N., Cunningham, D., Bassani-Sternberg, M., Gerlinger, M., 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. J. Immunother. Cancer 7, 309.
Olsen, L.R., Tongchusak, S., Lin, H., Reinherz, E.L., Brusic, V., Zhang, G.L., 2017. TANTIGEN: a comprehensive database of tumor T cell antigens. Cancer Immunol. Immunother. 66, 731–735.
Parham, P., Brodsky, F.M., 1981. Partial purification and some properties of BB7.2. A cytotoxic monoclonal antibody with specificity for HLA-A2 and a variant of HLA-A28. Hum. Immunol. 3, 277–299.
Parker, K.C., Bednarek, M.A., Hull, L.K., Utz, U., Cunningham, B., Zweerink, H.J., Biddison, W.E., Coligan, J.E., 1992. Sequence motifs important for peptide binding to the human MHC class I molecule, HLA-A2. J. Immunol. 149, 3580–3587.
Partridge, T., Nicastri, A., Kliszczak, A.E., Yindom, L.-M., Kessler, B.M., Ternette, N., Borrow, P., 2018. Discrimination between human leukocyte antigen class I-Bound and Co-purified HIV-Derived peptides in immunopeptidomics workflows. Front. Immunol. 9.
Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., 2018. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 47, D442–D450.
Purcell, A.W., McCluskey, J., Rossjohn, J., 2007. More than one reason to rethink the use of peptides in vaccine design. Nat. Rev. Drug Discov. 6, 404–414.
Purcell, A.W., Ramarathinam, S.H., Ternette, N., 2019. Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. Nat. Protoc. 14, 1687–1707.
Ramarathinam, S.H., Croft, N.P., Illing, P.T., Faridi, P., Purcell, A.W., 2018a. Employing proteomics in the study of antigen presentation: an update. Expert Rev. Proteomics 15, 637–645.
Ramarathinam, S.H., Gras, S., Alcantara, S., Yeung, A.W.S., Mifsud, N.A., Sonza, S., Illing, P.T., Glaros, E.N., Center, R.J., Thomas, S.R., Kent, S.J., Ternette, N., Purcell, D.F.J., Rossjohn, J., Purcell, A.W., 2018b. Identification of native and posttranslationally modified HLA-B*57:01-Restricted HIV envelope derived epitopes using immunoproteomics. PROTEOMICS 18, 1700253.
Rammensee, H.G., 1995. Chemistry of peptides associated with MHC class I and class II molecules. Curr. Opin. Immunol. 7, 85–96.
Sahay, B., Nguyen, C.Q., Yamamoto, J.K., 2017. Conserved HIV epitopes for an effective HIV vaccine. J. Clin. Cell. Immunol. 8.
Schuster, H., Shao, W., Weiss, T., Pedrioli, P.G.A., Roth, P., Weller, M., Campbell, D.S., Deutsch, E.W., Moritz, R.L., Planz, O., Rammensee, H.-G., Aebersold, R., Caron, E., 2018. A tissue-based draft map of the murine MHC class I immunopeptidome. Sci. Data 5, 180157.
Smyth, M.J., Dunn, G.P., Schreiber, R.D., 2006. Cancer immunosurveillance and immunoediting: the roles of immunity in suppressing tumor development and shaping tumor immunogenicity, advances in immunology. Academic Press, pp. 1–50.
Solleder, M., Guillaume, P., Racle, J., Michaux, J., Pak, H.-S., Müller, M., Coukos, G., Bassani-Sternberg, M., Gfeller, D., 2020. Mass spectrometry based immunopeptidomics leads to robust predictions of phosphorylated HLA class I ligands. Mol. Cell. Proteom. 19, 390–404.
Ternette, N., Olde Nordkamp, M.J.M., Müller, J., Anderson, A.P., Nicastri, A., Hill, A.V.S., Kessler, B.M., Li, D., 2018. Immunopeptidomic profiling of HLA-A2-positive triple negative breast cancer identifies potential immunotherapy target antigens. Proteomics 18, 1700465.
Tsuchiya, S., Yamabe, M., Yamaguchi, Y., Kobayashi, Y., Konno, T., Tada, K., 1980. Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). Int. J. Cancer 26, 171–176.

Woods, K., Faridi, P., Ostrouska, S., Deceneux, C., Wong, S.Q., Chen, W., Aranha, R., Croft, N.P., Duscharla, D., Li, C., Ayala, R., Cebon, J., Purcell, A.W., Schittenhelm, R.B., Behren, A., 2019. The diversity of the immunogenic components of the melanoma immunopeptidome. bioRxiv, 623223.

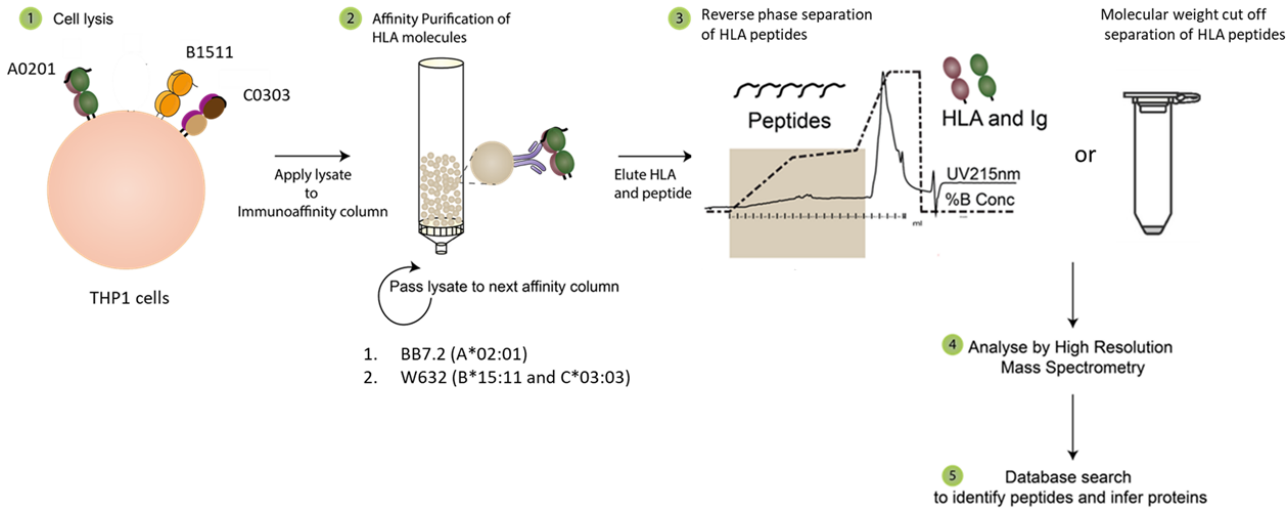
Zarling, A.L., Ficarro, S.B., White, F.M., Shabanowitz, J., Hunt, D.F., Engelhard, V.H., 2000. Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. J. Exp. Med. 192, 1755–1762.

Zarling, A.L., Polefrone, J.M., Evans, A.M., Mikesh, L.M., Shabanowitz, J., Lewis, S.T., Engelhard, V.H., Hunt, D.F., 2006. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc. Natl. Acad. Sci. U.S.A. 103, 14889–14894.

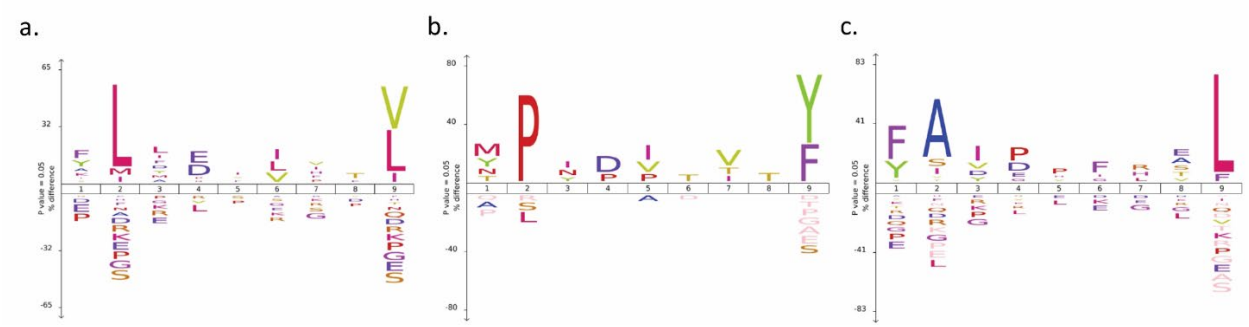
Zemmour, J., Little, A.M., Schendel, D.J., Parham, P., 1992. The HLA-A,B "negative" mutant cell line C1R expresses a novel HLA-B35 allele, which also has a point mutation in the translation initiation codon. J. Immunol. 148, 1941–1948.

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A., Ma, B., 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol. Cell Proteomics 11 M111.010587.

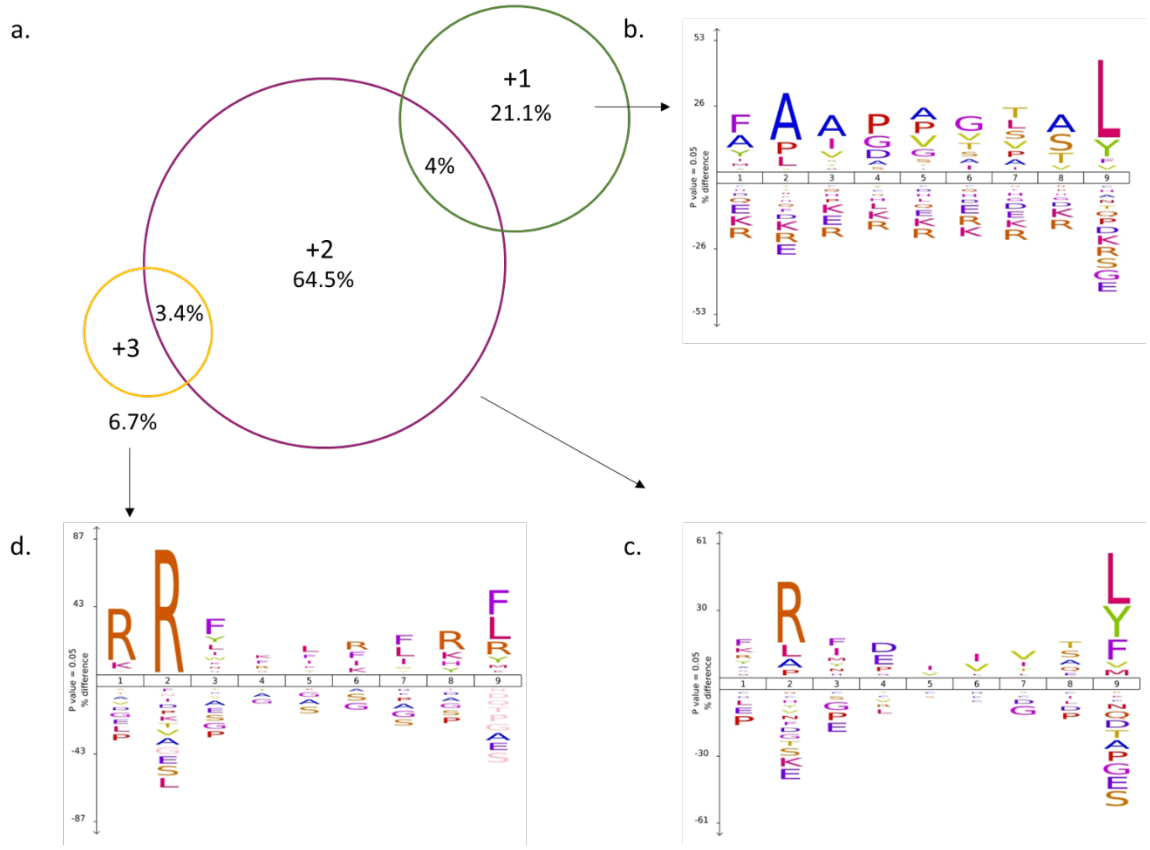
CHAPTER 2 – SUPPLEMENTARY FIGURES



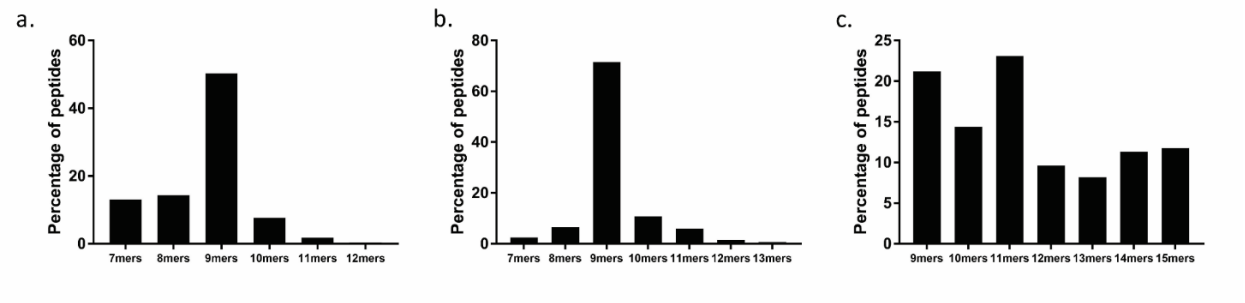
Supplementary Figure 1 – Depicting the methodology adopted for the experiments.



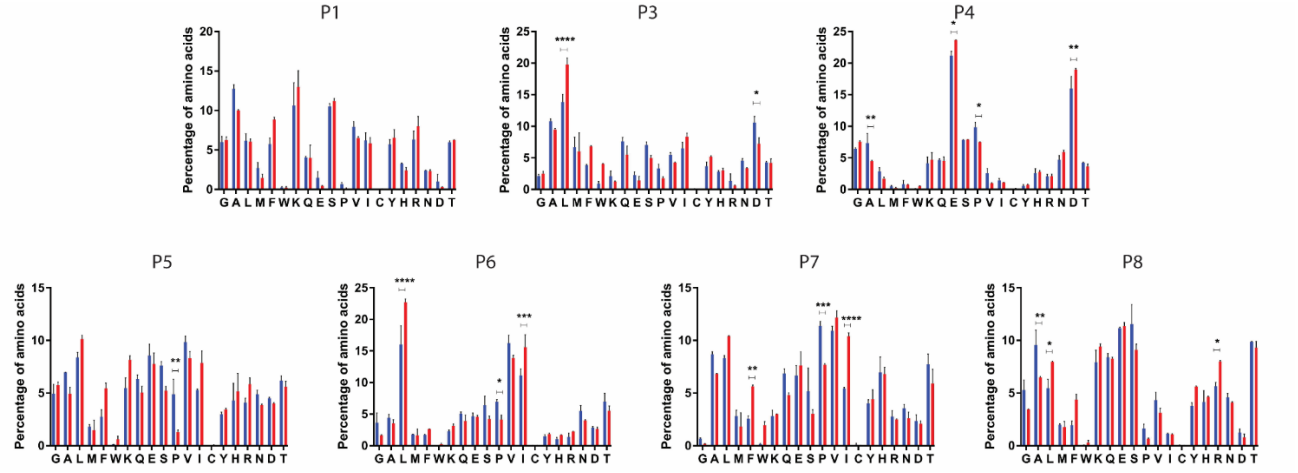
Supplementary Figure 2 – Motif generated using 9mer peptides reported in IEDB for alleles a) A*02:01, **b)** B*15:11 and **c)** C*03:03.



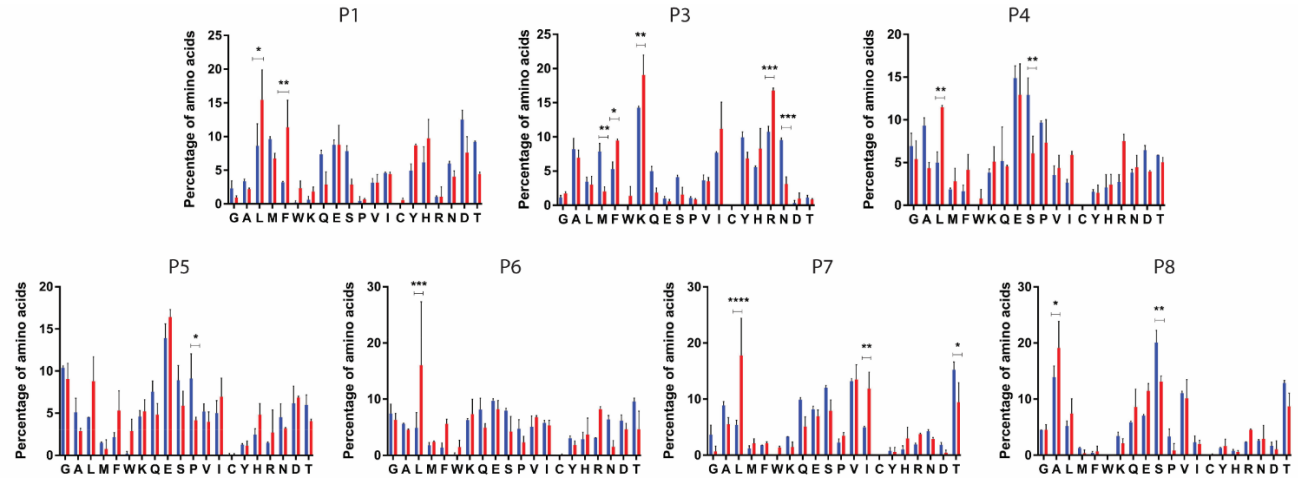
Supplementary Figure 3 - Little overlap between charge state +1, +2 and +3 of pan class I peptides identified in THP-1 immunopeptidome (n=3). a) Peptides (7 to 15mers) found in different charge states had very little overlap amongst them. The 9mers unique to each charge state were examined and their binding motif was analysed using icelogo (Colaert et al. 2009) showing the difference in amino acid composition of **b)** +1 charge state (n=2,749) **c)** +2 charge state (n=9996) and **d)** +3 charge state (n=742).



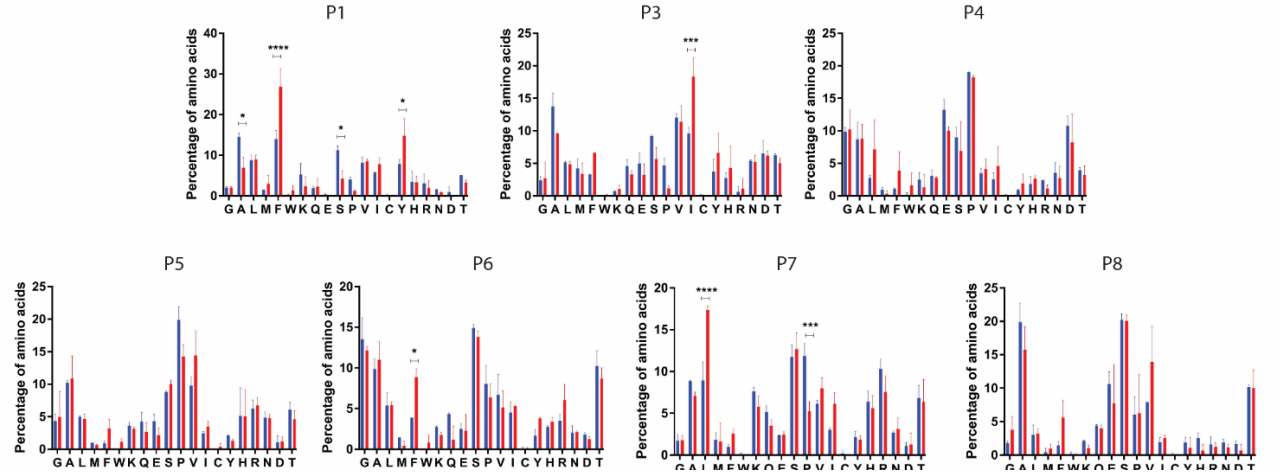
Supplementary Figure 4 – Peptide length distribution in different charge states a) +1, b) +2 and c) +3. In +1 and +2 charge states there is a predominance of 9mer peptides whereas in +3 state the percentage of 9- and 11mers is comparable along with longer length peptides.



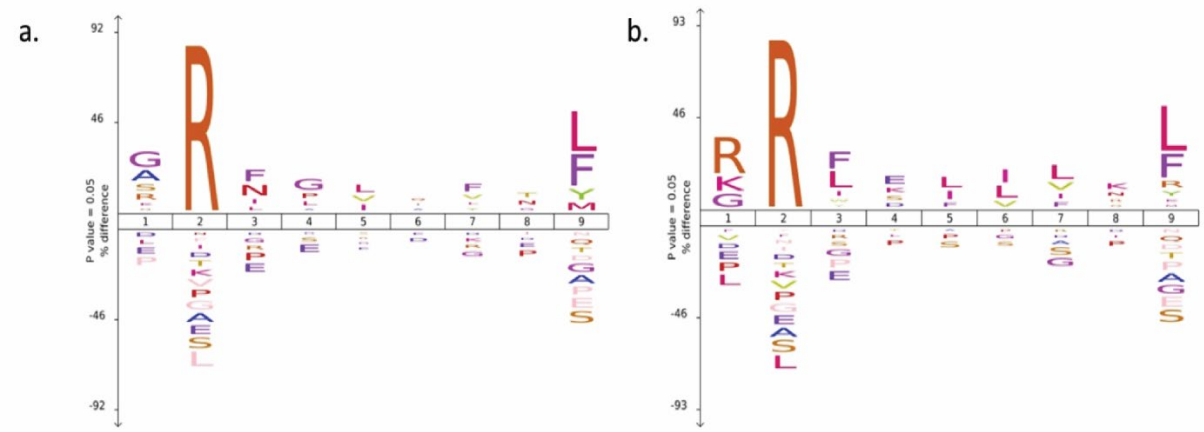
Supplementary Figure 5 – Amino acid frequency at other positions for A*02:01 peptides identified in THP-1 immunopeptidome (n=2). The amino acid frequency was calculated for unique 9mers found in RP-HPLC (blue) and filter (red) fraction and plotted. Significant differences in amino acid frequency was seen in different amino acids at all position except P1. P-value was measured by two-way ANOVA corrected for multiple comparisons (Sidak) where **** p<0.0001, *** p >0.0001, ** p value range 0.006-0.002 and * p value range 0.01-0.05 and error bars represent mean ± SD.



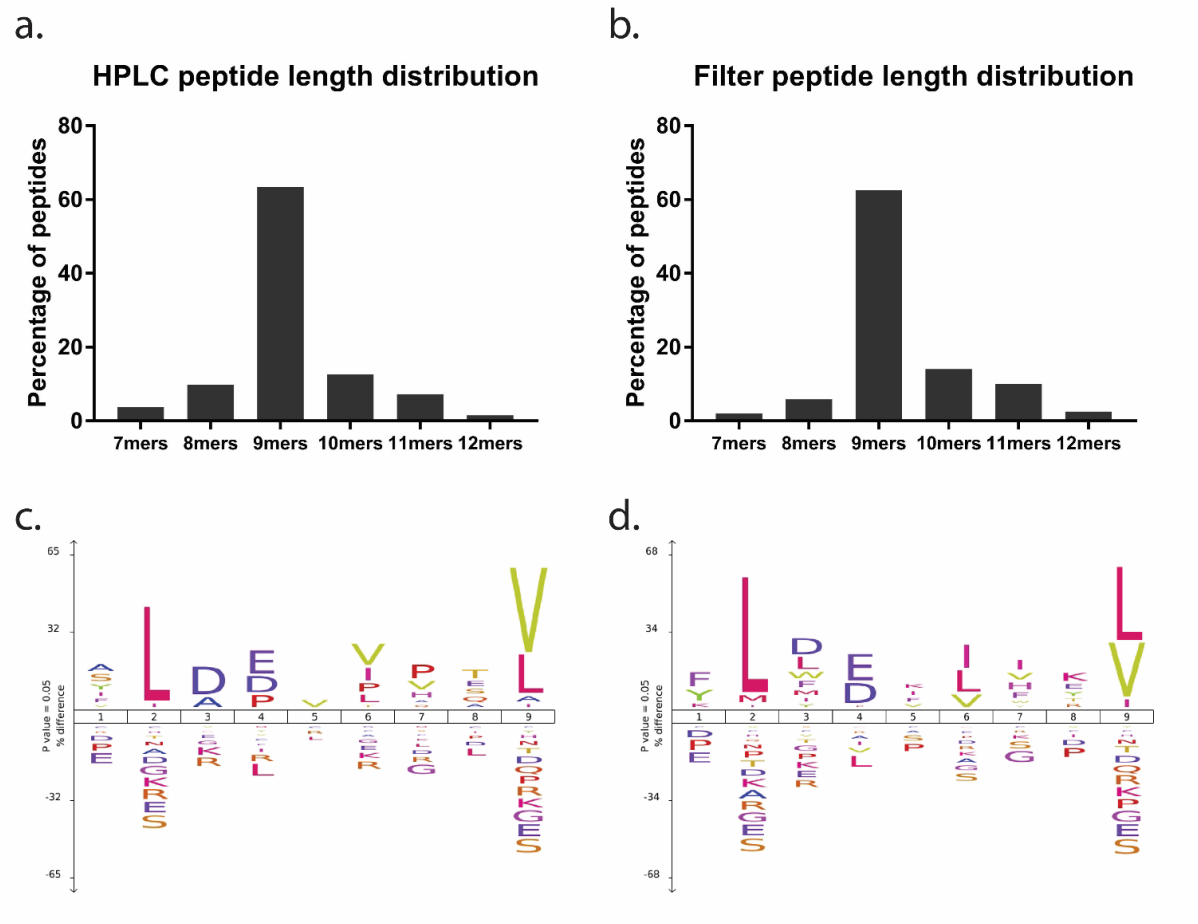
Supplementary Fig 6 - Amino acid frequency at other positions for B*15:11 peptides identified in THP-1 immunopeptidome (n=2). The amino acid frequency was calculated for unique 9mers found in RP-HPLC (blue) and filter (red) fraction and plotted. Significant differences in amino acid frequency was seen in different amino acids at all position except P1. P-value was measured by two-way ANOVA corrected for multiple comparisons (Sidak) where **** p<0.0001, *** p >0.0001, ** p value range 0.006-0.002 and * p value range 0.01-0.05 and error bars represent mean ± SD.



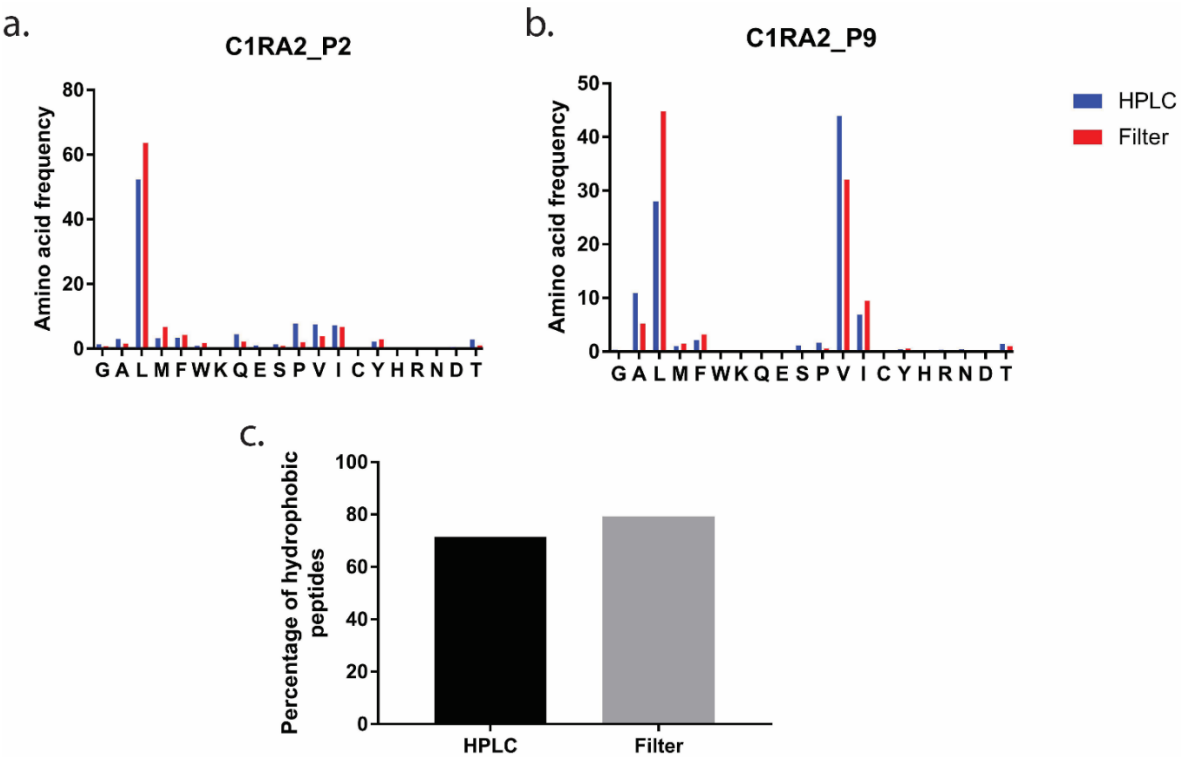
Supplementary Fig 7 - Amino acid frequency at other positions for C*03:03 peptides identified in THP-1 immunopeptidome (n=2). The amino acid frequency was calculated for unique 9mers found in RP-HPLC (blue) and filter (red) fraction and plotted. Significant differences in amino acid frequency was seen in different amino acids at all position except P1. P-value was measured by two-way ANOVA corrected for multiple comparisons (Sidak) where **** p<0.0001, *** p >0.0001, ** p value range 0.006-0.002 and * p value range 0.01-0.05 and error bars represent mean ± SD.



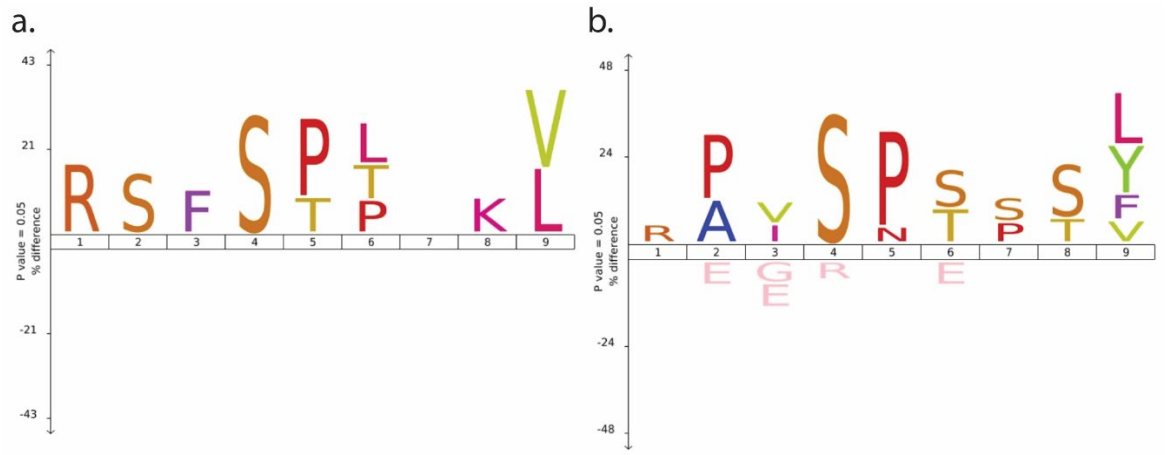
Supplementary Figure 8 – The difference in motif of nonameric HLA B*27:05 peptides unique to a) RP-HPLC and b) filter fractions. Although the amino acids at anchor positions P2 and P9 is same dominated by Arg and Leu and Phe at P2 and P9 respectively, differences in amino acids are seen at P1, P4 and P8 are observed. The binding motif was interpreted using iceLogo (34).



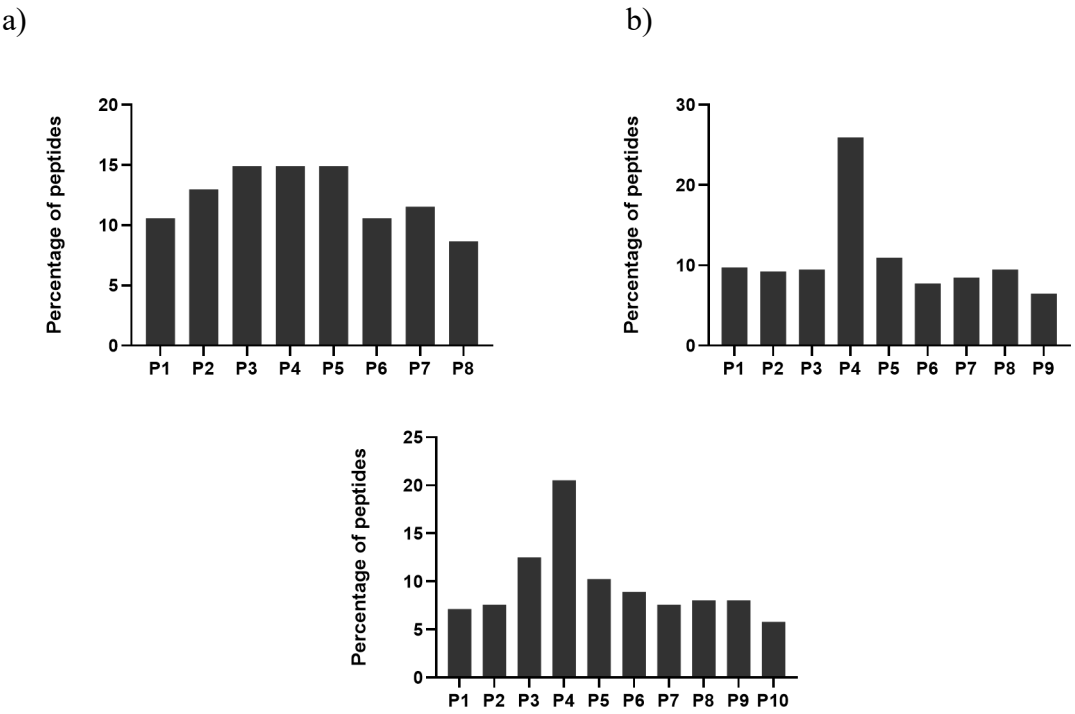
Supplementary Figure 9 - Comparison between the peptide length distribution and binding motifs predicted for A*02:01 peptides in RP-HPLC and filter fractions of C1R.A*02:01 cell line (n = 1)**.** Peptide length distribution was plotted for **a)** RP-HPLC (n=8280) and **b)** filter fractions (n=6636). Since the majority of peptide were 9mers, unique 9mers were selected and their binding motif was interpreted using iceLogo for both **c)** RP-HPLC (n= 2092) and **d)** filter (n=1476) fractions.



Supplementary Figure 10 - Amino acid frequency at P2 and P9 for unique 9mers identified in RP-HPLC and filter from C1R.A*02:01 cell line (n=1)**.** The amino acid frequency was calculated and plotted in Graphpad. **a)** An increase was observed in frequency of Leu (L) at P2 in the filter fraction. **b)** At P9, apart from the increase for L in filter, increases were also shown for the prevalence of Ala (A) and Val (V) in RP-HPLC 9mer peptides. **c) Presence of more hydrophobic peptides in filter.** The GRAVY score of all the unique 9mer peptides identified in RP-HPLC and filter fractions was calculated using www.gravy-calculator.de and the percentage of hydrophobic peptides (GRAVY score more than zero) were plotted. The RP-HPLC fraction contained less hydrophobic 9mers than the filter fraction.



Supplementary Figure 11 - The Proline (P) directed kinase motif (XSPXX) identified for unique 9mer phosphopeptides identified in a) HLA A*02:01 and b) pan HLA class I with Serine (S) dominant at P4 and Proline (P) at P5. The binding motif was interpreted using iceLogo (34).



Supplementary Figure 12: Distribution of phospho-moiety across the length of the peptide in a) 8-mers, b) 9-mers and c) 10-mers.

CHAPTER - 3

Chapter 3: Altered HLA peptide repertoire in haematological malignancy

3.1 INTRODUCTION

Haematological malignancy (HM) is an umbrella term used to refer to neoplasms originating in haematopoietic tissues primarily the bone marrow (BM). They clinically manifest as leukaemia which can be either myeloid or lymphoid in origin. This includes acute myeloid leukaemia (AML), which is characterised by an increase in myeloid blasts cells in the BM (Döhner et al. 2015, De Kouchkovsky and Abdul-Hay 2016, Terwilliger et al. 2017). Adult AML accounts for ~25% of all leukaemia cases in the world (DiNardo and Cortes 2016) with a median age at diagnosis of 67 years (Kumar 2011). AML is one of the most heterogenous types of cancer with cytogenetic abnormalities detected in ~50-60% of newly diagnosed patients. These abnormalities include chromosomal translocations, deletions, and inversions resulting in complex phenotypes that not only have prognostic implications but also influence whether or not the patient responds to certain treatments. Around 30-40% of AML cases which do not have any chromosomal aberrations are considered cytogenetically normal (CN), but they still have an intermediate prognosis.

As per treatment guidelines most AML patients are started on induction therapy (Estey 2009, Morra et al. 2009). Since there is ~100% chance of relapse post one round of induction therapy (Morra et al. 2009), it is followed by consolidation therapy (Morra et al. 2009). At this stage patients are recommended to undergo allogeneic hematopoietic cell transplant (allo-HCT) (Rodríguez 2017). Although allo-HCT continues to be considered an optimal approach there is acute transplant-related mortality (TRM) seen in ~85% of AML patients (Morra et al. 2009). Hence alternative forms of treatment must be explored.

Cancer immunotherapy offers an alternative to cytotoxic chemotherapy and radiotherapy, particularly in B cell leukaemia and lymphoma. Monoclonal antibody mediated immunotherapy such as Brentuximab vedotin (BV) (Vassilakopoulos et al. 2019) which targets CD30 expressing B cells, has resulted in ~5% increased survival rate by lowering combined risk of progression and death (Connors et al. 2018) in refractory B cell lymphoma. Moreover, the success of patient derived and genetically modified T cells which includes bispecific T cell engager (BiTE) antibody targeting CD19 and CD3 used for treating refractory (R) B-cell acute leukaemia (Przepiorka et al. 2015). Another example of modified T cell therapy includes chimeric antigen receptor (CAR) T cell therapies such as Kymriah™ and Yescarta™ (Zheng et al. 2018) which have been used for treatment of refractory or relapse of B cell acute

lymphoblastic leukaemia (ALL) and large B cell lymphoma respectively (Zheng et al. 2018). However the success of the T cell therapies was found to be short-lived as ~30% and 55% of the patients treated with BiTE and Kymriah™ respectively (Ruella et al. 2016, Zheng et al. 2018)relapsed with no prolonged T cell memory established in them (Ruella and Maus 2016).

Much of the above-mentioned immunotherapeutic success remains untranslated in AML. This can be attributed to the complex cytogenetics combined with different mutations that lead to heterogenous cell populations in AML. In contrast, B cell lymphomas are is quite homogenous with a few key mutations such as BCR-ABL translocation and fusion. B cell lymphomas also express several surface antigens such as CD19 and CD30 which can be targeted by immunotherapy. A potential alternative for AML treatment is T cell mediated immunotherapy, involving cytolytic CD8 T cells whose function is to recognise and destroy cancer cells. A critical step in the CD8 T cell response is governed by HLA class I, which presents intra-cellular peptides on the surface of all nucleated cells. This set of peptides is referred to as the immunopeptidome is recognised by T cell receptors. The immunopeptidome of cancer cells oftentimes comprises of cancer specific peptides which are not present on normal cells and include tumour specific mutant peptides or neoantigens and proteins aberrantly expressed on tumour cells such as tumour associated antigens (TAAs) and cancer testis antigens (CTAs). T cell mediated immunotherapy is particularly useful as it trains the patient's own immune system to combat the cancer. However, at the site of leukemogenesis i.e. the BM, the tumour microenvironment (TME) can influence HLA expression by downregulating it epigenetically. Oftentimes chromosomal deletions might result in loss of chromosome 6 or 15 which might lead to loss of HLA expression (Rodríguez 2017, Rovatti et al. 2020). These are one of the most common tumour escape mechanisms, allowing the cancer to proliferate unchecked.
Another facet of the immunopeptidome that can be exploited for both prognosis and diagnostic purposes is soluble HLA (sHLA) identified in plasma samples of patients. sHLA can be a rich source of information as its levels are found to be elevated in both AML and ALL patients (Kantarjian et al. 1992, Albitar et al. 2007). This has been attributed to HLA being rapidly shed from proliferating blasts cells (Albitar et al. 2007). Elevated sHLA has also been found to correlate with both reduced response to treatment and worse survival (Kantarjian et al. 1992, Albitar et al. 2007).

The current study is an attempt to understand the effect cytogenetic aberrations occurring at site of leukemogenesis might have on HLA expression and therefore the endogenous peptide

repertoire of AML and ALL patients. We also wish to identify and expand AML and ALL specific peptide antigens from representative HM cell lines and bone marrow aspirates. Another crucial aspect of the study is to investigate if any mutations occurring in the bone marrow might influence the sHLA-derived immunopeptidome as it is a derivative and therefore a representation of the membrane HLA (mHLA) repertoire of the BM cells. To achieve these aims we have used a combination of immunoprecipitation and high resolution liquid chromatography tandem mass spectrometry (LC-MS/MS) These findings will provide key information for HM-targeted peptide based immunotherapeutics, as well as personalised medicine approaches as it can guide vaccine design to improve survival outcome in patients.

3.2 MATERIALS AND METHODS

3.2.1 Cell lines and HLA expression

All haematological cancer cell lines (THP-1, MV411.1 and HL60) were maintained in RF10 (RPMI 1640 (Gibco, ThermoFisher, USA) supplemented with 2mM MEM nonessential amino acid solution (Gibco, ThermoFisher, USA), 100mM HEPES (Gibco, ThermoFisher, USA), 2mM L-glutamine (Gibco, ThermoFisher, USA), Penicillin/Streptomycin (Gibco, ThermoFisher, USA), 50µM 2-mercaptoethanol (Sigma-Aldrich, USA) and 10% heat inactivated FCS (Sigma-Aldrich, USA)). HLA class I typing of all cell lines was confirmed by genomic analysis at the Victorian Transplantation and Immunogenetics Service (Australia) (Table 1). Cell surface HLA class I expression was measured by flow cytometry following surface staining using hybridoma supernatants for both anti-HLA-A2 (BB7.2 [HB-82; ATCC] (Parham et al. 1981); produced in-house) and pan-HLA class I (W6/32 [HB-95; ATCC] (Brodsky et al. 1982); produced in-house) as primary antibodies and goat anti-mouse IgG PE (Southern Biotech, USA) as the secondary antibody. Cells in the form of events (30,000) were acquired on LSR IIa (BD Biosciences, USA) using the FACSDiva software and analysis of the cell population was performed using FlowJo software (Treestar Inc, USA).

3.2.2 Cell line sample processing

Cell lines were expanded to $1x10^9$ in 800mL of RF10 media in roller bottles (Greiner Bio-One International AG, Austria), washed in PBS, harvested by centrifugation (3724g, 15 mins, 4°C), snap frozen in liquid nitrogen and stored at -80°C until required. Purified antibody was generated from both HB-82 and HB-95 hybridoma cells cultured in RF5 (same constituents as RF10 but 5% FCS) and expanded in roller bottles at 37°C, 5% CO_2. Secreted monoclonal antibodies, BB7.2 (anti-HLA-A2) and W6/32 (pan-HLA class I), were harvested from spent

media and purified using Protein A Sepharose (PAS, CaptivA®, Repligen, USA) using a Profinia purification system (BioRad).

3.2.3 Clinical sample processing

Cryopreserved bone marrow mononuclear cells (BMMC) sample (~25x10^6 per vial) from a healthy individual was commercially sourced (Lonza, Walkersville, USA). Patient BM aspirate (~2 mL, n=3) along with 20 mL of peripheral blood (n=3) were collected in heparinised vacutainer tubes from newly diagnosed leukemic patients recruited at Box Hill Hospital during routine clinical investigations (Eastern Health ethics number LR77/2016 and Monash University ethics number 8088). BMMC samples were pelleted at 2,300g, 20 mins, 4°Ca and cryopreserved at -80°C until required. Plasma was separated from peripheral blood samples following centrifugation for at 2,300g, 10 mins, 22°C also cryopreserved at -80°C until required. The peripheral blood samples were also used to isolate peripheral blood mononuclear cells (PBMCs) using Ficoll-Paque density gradient centrifugation. PBMCs were used to isolate DNA for HLA class I typing, which was performed by the Victorian Transplantation and Immunogenetics Service.

3.2.3 Immunoprecipitation of pHLA complexes

An immunoaffinity matrix was generated wherein 10 mg/mL of BB7.2 or W6/32 was crosslinked to 1 mL of PAS resin as previously described (Ramarathinam et al. 2018) and used for the immunoaffinity capture of solubilised native pHLA complexes. Briefly, frozen cell pellets (1x10^9) and BM pellets (1-2x10^8) were pulverised by cryogenic milling (Retsch Mixer Mill MM 400), reconstituted in Lysis Buffer [0.5% IGEPAL (Sigma-Aldrich, USA), 50mM Tris pH 8, 150mM NaCl and protease inhibitors (Complete Protease Inhibitor Cocktail Tablet, 1 tablet per 50 mL solution; Roche Molecular Biochemicals, Switzerland)] and proteins solubilised by incubation for 1 hr at 4°C with rotation. The supernatant was passed through a PAS pre-column (500µL) to remove non-specific binding material, followed by serial affinity capture of HLA A*02:01 peptides (BB7.2 column) and the remaining class I peptides (W6/32 column) for the THP-1 cell line and healthy BM samples. For all other samples only W6/32 was used to pull down pan class I peptides. Bound peptide-HLA complexes were eluted from the column in 10% acetic acid.

For THP-1, the eluate was mixed and split into two equal volumes and treated as mentioned previously (Pandey et al. 2020). In brief, one part of the eluate was fractionated by RP-HPLC and the second half of the eluate was passed through a 5kDa MWCO filter (Amicon). The sample flow through containing HLA peptides was collected, centrifugally evaporated, and

desalted by reversed phase C18 stage tips (Omix, Agilent). The C18 tips were washed with 0.1% TFA and peptides were eluted with 40% ACN in 0.1% TFA

For the other two cell lines and clinical samples the eluate was fractionated only by RP-HPLC into 48 fractions which were concatenated, vacuum concentrated, into 10 fractions, reconstituted in 0.1% formic acid and acquired on the MS.

3.2.4 Immunoprecipitation of sHLA complexes

To isolate sHLA bound peptides from plasma samples from leukemic patients first affinity matrix wherein 2 mg/mL W6/32 was crosslinked to 200µL of PAS resin was prepared as previously described (Ramarathinam et al. 2018) and used for the immunoaffinity capture of solubilised native sHLA from the patient plasma samples. Samples were thawed completely and mixed with lysis buffer [0.5% IGEPAL (Sigma-Aldrich, USA), 50mM Tris pH 8, 150mM NaCl and protease inhibitors (Complete Protease Inhibitor Cocktail Tablet, 1 tablet per 50 mL solution; Roche Molecular Biochemicals, Switzerland)] and incubated for 1 hr at 4°C with rotation. The samples were centrifuged at 3724g, 10 mins, 4°C and the supernatant was transferred to a fresh tube. The lysate was then passed through a PAS pre-column (500µL) to remove non-specific binding material, followed by affinity capture of pHLA complexes using the W6/32 immunoaffinity column. Bound complexes were eluted from the column in 10% acetic acid. The eluate was passed through a 5kDa MWCO filter (Amicon) which had been pre-washed with MS-grade optima water (Thermo Fisher Scientific, USA) and then equilibrated with 10% acetic acid by centrifugation at 16,060g for 30 mins at room temperature (RT). The sample flow through containing sHLA-bound peptides was collected. Finally, the filter was rinsed with buffer A and centrifuged at 16,060g for 30 mins at RT. The filtered flow through was centrifugally evaporated and desalted by reversed phase C18 stage tips (Omix, Agilent). The C18 tips were washed with 0.1% TFA and peptides were eluted with 40% ACN in 0.1% TFA.

3.2.5 Analysis of HLA class I bound peptides by LC-MS/MS

All 10 pooled peptide containing fractions following RP-HPLC and the single peptide containing fraction obtained from the 5kDa MWCO filter were analysed by LC-MS/MS. A mixture of 11 iRT peptides were spiked into each sample to aid retention time alignment/ (Escher et al. 2012). Peptides were loaded onto a PepMap Acclaim 100 C_{18} trap column (5 µm particle size, 100 µm x 2 cm and 100Å (Thermo Scientific)) at 15µL/min using an Ultimate 3000 RSLC nano-HPLC (Thermo Scientific). The column was equilibrated with 2% ACN and 0.1% FA. Peptides were eluted and separated at a flow rate of 250 nL/min on an in-line

analytical column (PepMap RSLC C_{18}, 2 µm particle size, 75 µm x 50 cm and 100Å, Thermo Scientific) using a 125 mins linear gradient of 2.5- 99% of buffer B (80% ACN, 0.1%FA). The gradient started from 2.5% buffer B and increased to 7.5% buffer B in a minute followed by linear gradient to 37.5% buffer B for 90 mins followed by increase to 99% buffer B in 10 mins. Peptides were introduced using nano-electrospray ionisation (nano-ESI) method into the Orbitrap Fusion Tribrid MS (Thermo Scientific, USA) at a source temperature of 275° C.

All MS spectra (MS1) profiles were recorded from full ion scan mode 375-1800 *m/z*, in the Orbitrap at 120,000 resolution with automatic gain control (AGC) target of 400,000 and dynamic exclusion of 15 secs. The top 12 precursor ions were selected using top speed mode at a cycle time of 2 secs. For MS/MS a decision tree was made to aid selecting peptides of charge state 1 and 2-6 separately. For singly charged analytes only ions falling within the range of *m/z* 800-1800 were selected, whereas for +2 to +6 *m/z* no such parameter was set. The C-trap was loaded with a target of 200,000 ions with an accumulation time of 120 ms and isolation width of 1.2 amu. Normalised collision energy was set to 32 (high energy collisional dissociation [HCD]) and fragments were analysed in the Orbitrap at 30,000 resolution.

3.2.6 Data analysis

The raw data files obtained from the mass spectrometer were analysed using Peaks 8.5 software (Bioinformatics Solutions) against the human proteome (Uniprot 15/06/2017; 20,182 entries). The following search parameters were used: error tolerance of 10 ppm using monoisotopic mass for precursor ions and 0.02 Da tolerance for fragment ions; enzyme used was set to none with following variable modifications: oxidation at Met, deamidation at Asp and Gln and phosphorylation at Ser, Thr and Tyr. The false discovery rate (FDR) was estimated using a decoy fusion method (Zhang et al. 2012) and all datasets were analysed using at 5% FDR cut off. For data filtering first, iRT and decoy peptides were removed along with any peptide present in duplicate. Secondly, only peptides 7-15 amino acid in length were retained and peptide scans with alternative I-L explanations were removed from the dataset. The haematological malignancy elution experiments were performed in duplicate. The data was plotted in GraphPad Prism (Version 8.4). Venn diagrams were prepared using Venny (Oliveros) and binding affinity of the peptides was predicted using NetMHC 4.0(Jurtz et al. 2017).

3.3 RESULTS

3.3.1 Mapping endogenously presented HLA class I peptide repertoire in haematological cell lines and bone marrow samples.

This study investigated the immunopeptidome in haematological malignancy, using THP-1, MV411.1 and HL60 cell lines, as well as clinical samples comprising of BM aspirates from one healthy, 2 AML and 1 ALL patients (details in Table 1). AML is a highly heterogenous form of cancer with ~ 95% of AML cases having at least one mutation. In order to determine if any of them might influence HLA peptide presentation, we catalogued the presence of any aberrations present in the samples (Table 1). Cell lines THP-1 and MV411.1 are reported to contain translocations of chromosome 11q to chromosome 9 and 4, respectively. This leads to the generation of mixed lineage leukaemia (MLL)-AF9 and MLL-AF4 oncofusion proteins in THP-1 and MV411.1, respectively. No cytogenetic aberration has been reported for the HL60 cell line. Cytogenetic profiling of clinical samples (performed at Box Hill Hospital) revealed that AML patient1 was CN, a condition that is quite prevalent in AML (Estey 2018) whilst AML patient2 reportedly had a mutation in NPM1 protein. ALL patient3 was found to have trisomy of 18 and 21, loss of chromosome 6, 2 and 9 as well as monosomy of other chromosomes. Particularly, for chromosome 6, due to lack of WES data it is difficult to comment whether the aberration was clonal or sub clonal in origin.

Prior to undertaking the antigen discovery approach, HLA class I expression was analysed for all three cell lines by FACS using both an allele specific antibody such as BB7.2 (anti-HLA-A2) for THP-1 and W6/32 (pan-HLA class I) for all three cell lines. Based on FACS data, all cell lines were found to have high HLA expression (Figure 1).

Table 1 – Genetic characteristics of cell lines and clinical samples

Cell line/sample	Molecular genetics/Karyotype	HLA class I typing
THP-1	t(9,11)(p22,q23) translocation - MLL-AF9	A*02:01, B*15:11, C*03:03
MV411.1	t(4,11)(q21,q23) translocation – MLL-AF4	A*03:01, A*68:01 B*14:02, B*18:01, C*08:02, C*15:02
HL60	Normal	A*01:01, B*57:01, C*06:02
Healthy BM	Normal	A*02:01, A*74:01, B*07:02, B*45:01 C*07, C*16
AML patient 1	No mutation or cytogenetic aberrations	A*03:01, A*31:01, B*07:02, B*56:01, C*01:02, C*07:02
AML patient 2	NPM1 mutation, Normal cytogenetics, CD13+, CD33+, MPO+, CD77 and CD34-	A*26:01, A*33:03 B*27:04, B*58:01, C*03:02, C*12:02
ALL patient 3	Monosomy of chromosome 3,4,7,12,13, 16 and 17 Trisomy of chromosome 18 and 21 Loss of 2,6 and 9	A*29:02, A*68:01, B*15:17, B*44:03, C*07:01, C*16:01

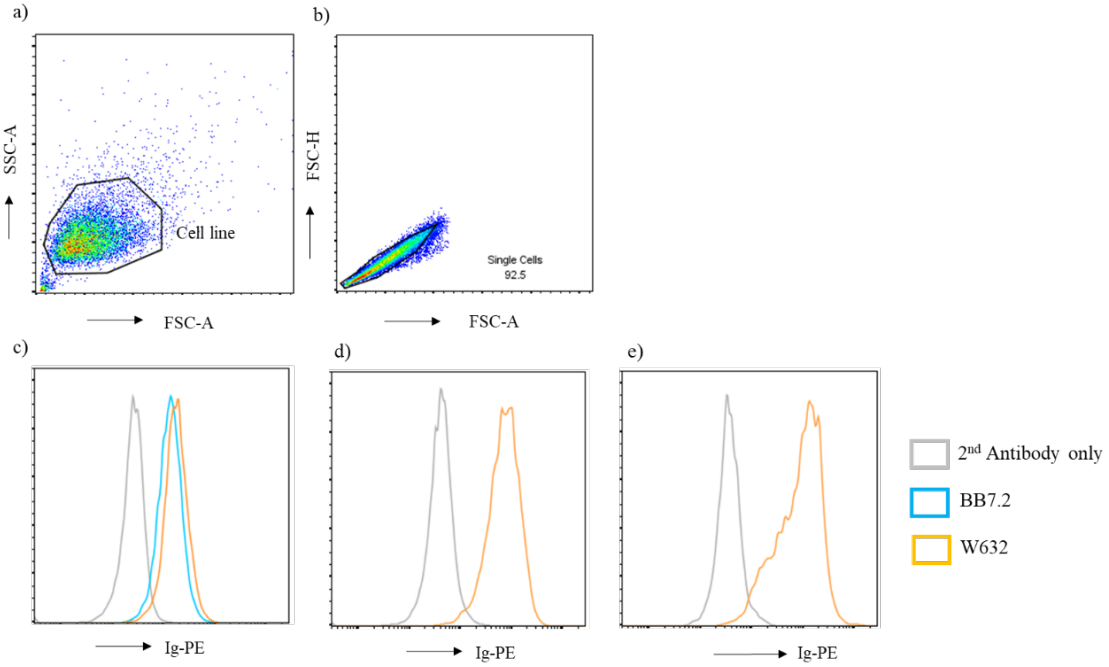


Figure 1 – HLA cell surface expression on cell lines. Cells were stained with BB7.2 (anti-HLA-A2) and/or W6/32 (pan-HLA class I) antibodies followed by secondary antibody (goat anti-mouse IgG PE). a) Total of 20,000 cells were acquired on LSR IIa (BD Biosciences) using the FACSDiva software. b) Analysis of the cell population was performed following removal of doublet cells using FlowJo software (Treestar Inc, USA). The level of HLA cell surface expression was observed using histograms for c) THP-1, d) MV411.1 and e) HL60 cell lines.

HLA expression on the surface was seen as a shift in mean fluorescence intensity compared to control sample of cells stained only with secondary antibody.

The cell lines were cultured and expanded *in vitro* to obtain a pellet of $1x10^9$ cells. For patient material, ~2mL of BM aspirate (for each patient) was centrifuged, yielding ~$1\text{-}2x10^8$ cells across the different samples. To isolate pHLA complexes, cell lysis was performed followed by immunoaffinity purification using HLA specific antibodies such as BB7.2 for capture of HLA A*02:01-peptide complexes (THP-1 and healthy BM sample) followed by/or solely W6/32 for capturing of other HLA class I allotype-peptide complexes. For each cell line two biological replicates were used, whilst the clinical samples were acquired only once by Tribrid Fusion LC–MS/MS (ThermoScientific, USA).

This resulted in generation of largest haematological malignancy data reported, with a total of 81,027 unique 7-15 amino acid long HLA class I peptides identified across cell lines and clinical samples. This included 61,445 unique peptides present only in the cell lines and 16,130 unique peptides identified only in the clinical samples with 3,452 overlapping peptides between both sample sets. The peptides were restricted by 28 different HLA alleles (nine HLA-A, ten HLA-B and nine HLA-C alleles (Table 1). Some of the HLA allotypes were shared between some samples for HLA A*02:01 (n = 2), -A*03:01 (n = 2), -A*68:01 (n = 2), -B*07:02 (n = 2) and -C*07 (n = 3). The overlap between cell line biological replicates ranged from 39%-50% across different cell lines (Supplementary Figures 1-3).

The data also highlighted the potential role of cytogenetic aberrations on the HLA class I immunopeptidome. The chromosome 11 translocation giving rise to MLL oncofusion proteins did not influence HLA class I expression or the size of immunopeptidome repertoire, as the three cell lines had similar number of peptides identified in biological duplicate data despite HL60 being cytogenetically normal (Table 2). Similarly, we observed that the CN patient (i.e. AML patient1) yielded a similar number of peptides as those identified from the healthy BM. In contrast, AML patient2 with mutated NPM1 protein had a significantly reduced number of HLA class I peptides thereby suggesting perturbation of HLA class I expression. This finding has not been previously reported. At the same time HLA class I peptides (n = 95) were identified from ALL patient3 wherein karyotyping revealed loss of chromosome 6. These might have come from non-ALL cells contaminating the bone marrow aspirate.

Table 2 – HLA class I peptides (7-15 mers) identified across the cell lines and clinical samples

Cell line/sample	A*02:01 peptides	Pan class I peptides
THP-1	13,411	11,274
MV411.1		23,345
HL60		20,100
Healthy BM	3,628	7,203
AML patient1		11,642
AML patient2		1,223
ALL patient3		95

3.3.2 An overview of immunopeptidome of clinical samples and cell lines

To characterise the mHLA class I repertoire of clinical samples and cell lines further analysis including peptide length distribution and motif analysis was performed. For peptide length distribution, the percentage of peptides was calculated for each sample and plotted. As anticipated both clinical (Figure 2a) and cell line (Figure 2b) samples had typical class I peptide length distribution, wherein nonamers were the dominant length followed by longer 10 and 11mers. The variation seen in the percentages amongst different samples could be attributed to immunogenetics as each sample had different HLA alleles. For instance, THP-1 expressed high levels of HLA A*02:01, which is known to prefer nonamers (Sarkizova et al. 2020). Similarly, a significantly higher percentage for longer peptides in MV411.1 and HL60 cell lines were predicted to be restricted by HLA allotypes A*68:01 and B*57:01, both known to prefer longer peptides due to C- and N terminal extensions, respectively (Pymm et al. 2017, Guillaume et al. 2018). Peptides identified from ALL patient3 also followed typical HLA class I length distribution with ~54% peptides (n=50) being nonamers.

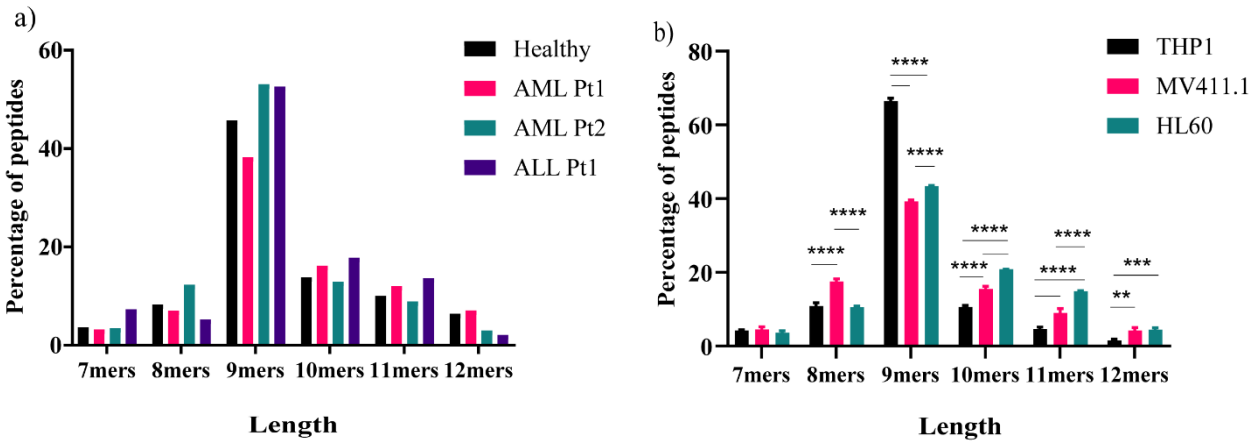
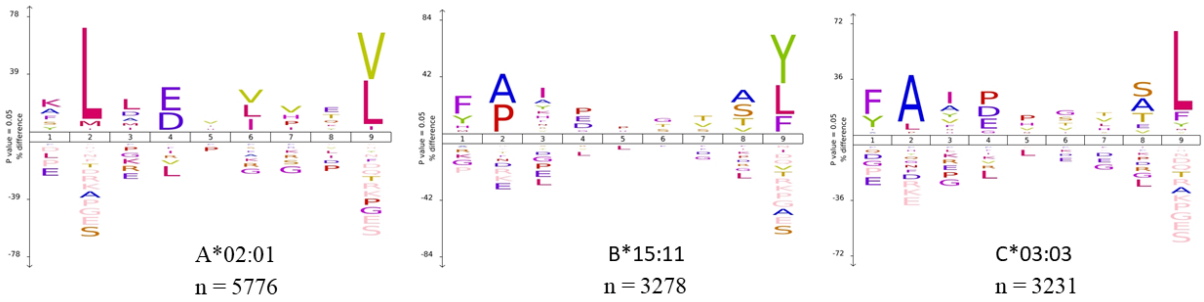


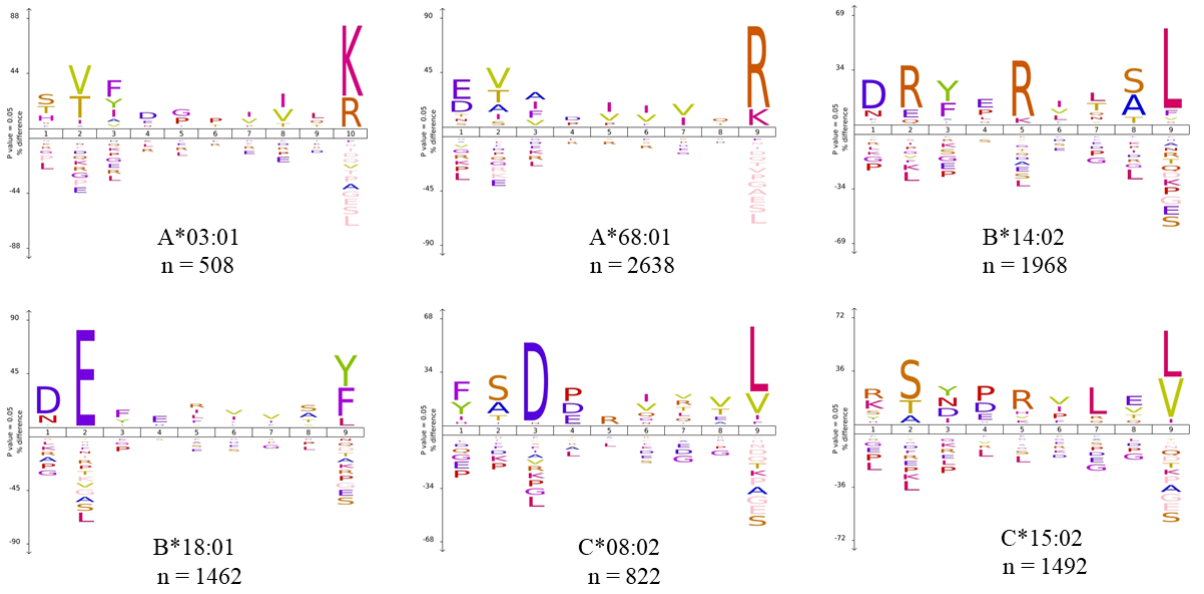
Figure 2 –HLA peptides identified in clinical and cell line samples follow canonical class I length distribution. a) Length distribution of 7- to 12 amino acid long peptides. b) Length distribution of 7- to 12 amino acid long peptides identified across each cell line. THP-1 was dominated by 9mers compared to other two cell lines where 8-, 10- and 11mers were also present in significantly higher percentage. The data was plotted in GraphPad Prism and p-value was measured by two-way Anova corrected for multiple comparisons (Sidak) where **** $p<0.0001$, ***$p= 0.0006$ and ** $p= 0.0010$, error bars represent mean±SD from a duplicate dataset.

In order to determine the potential HLA restriction of the identified peptides, nonamers were selected as they were the dominant length across samples. Allele-specific binding motifs for 9mers were determined by predicting binding affinity using NetMHC Pan 4.0 (Jurtz et al. 2017). Peptides with a rank threshold of less than 0.5 were considered as strong binders (SB), whilst those between 0.5-2.0 were considered as weak binders (WB). Surprisingly, in ALL patient3, 45/50 nonamers were identified as binders and majority of them (~84%) were SB, especially for the A*68:01 and B*15:17 alleles. The HLA-binding predisposition data was then used to plot allotype-specific binding motifs for both the cell lines (Figure 3) and clinical samples (Figure 4) using Icelogo.

a) THP-1



b) MV411.1



c) HL60

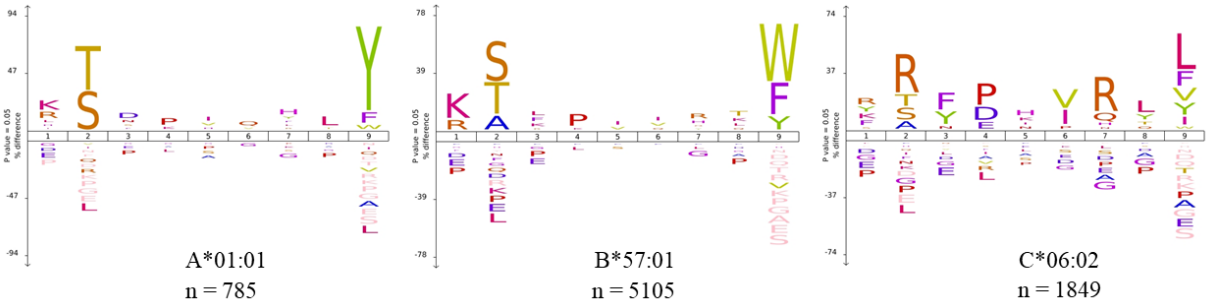
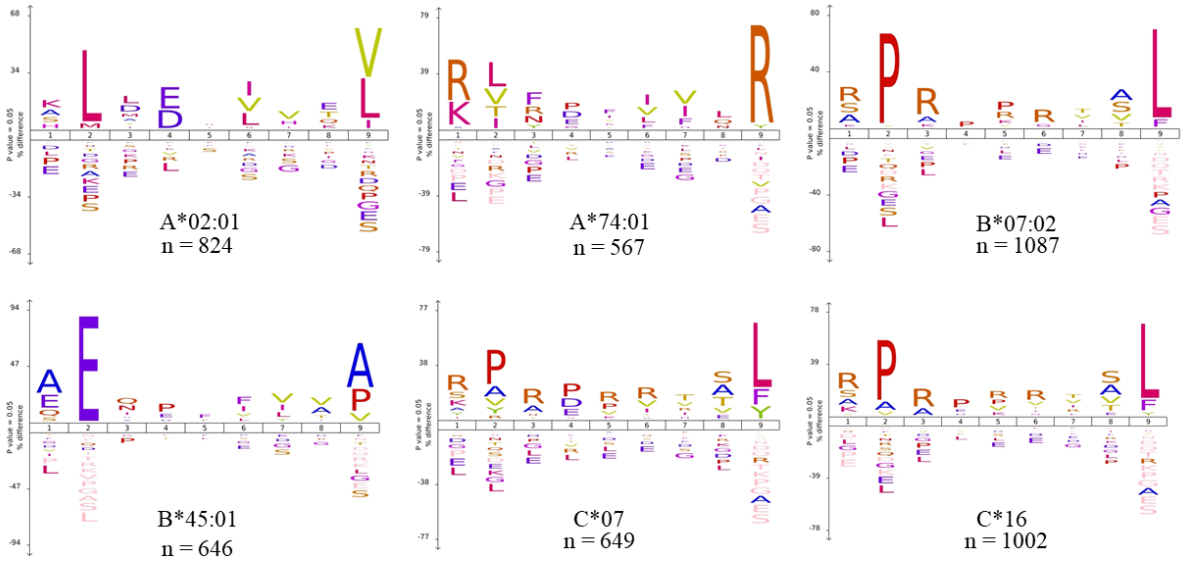
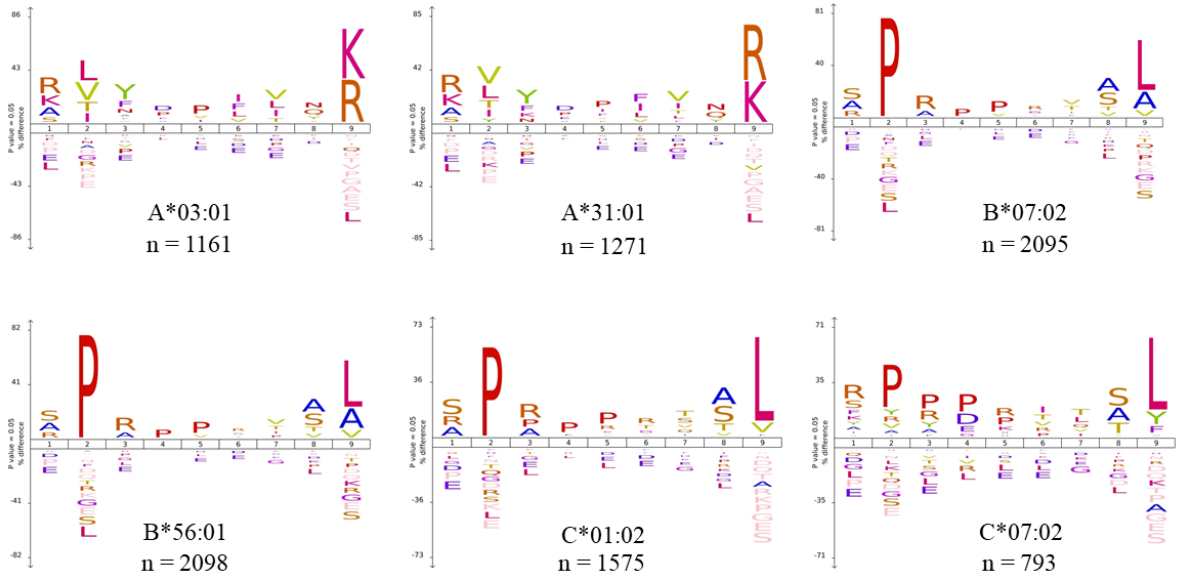


Figure 3 – Nonamers identified from cell lines a) THP-1, b) MV411.1 and c) HL60 have canonical HLA-allotype specific binding motifs. The expected HLA restriction of nonamers was determined based on their predicted HLA-binding affinity using NetMHC Pan 4.0. Icelogo representations of these allotype specific HLA-binding motifs of peptides isolated from the cell lines with n representing number of binders predicted for each allele.

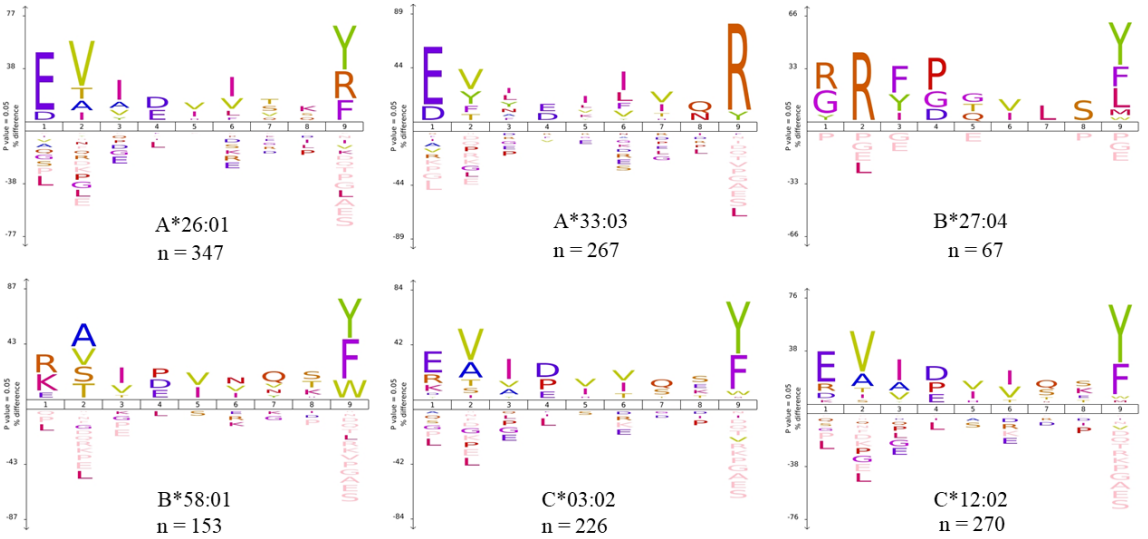
a) Healthy BM



b) AML patient1 BM



c) AML patient2 BM



d) ALL patient3 BM

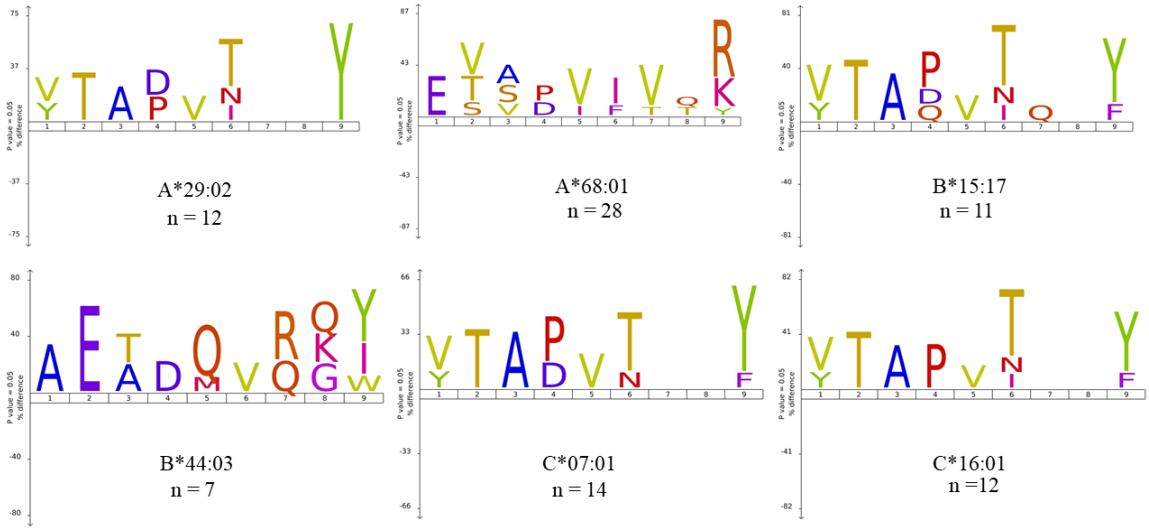


Figure 4 – Nonamers identified from bone marrow samples a) Healthy, b) AML patient 1, c) AML patient2 and d) ALL patient3 have canonical binding motifs. The expected HLA restriction of nonamers was determined based on their predicted HLA-binding affinity using NetMHC Pan 4.0. Icelogo representations of these allotype specific HLA-binding motifs of peptides isolated from the cell lines with n representing number of binders predicted for each allele.

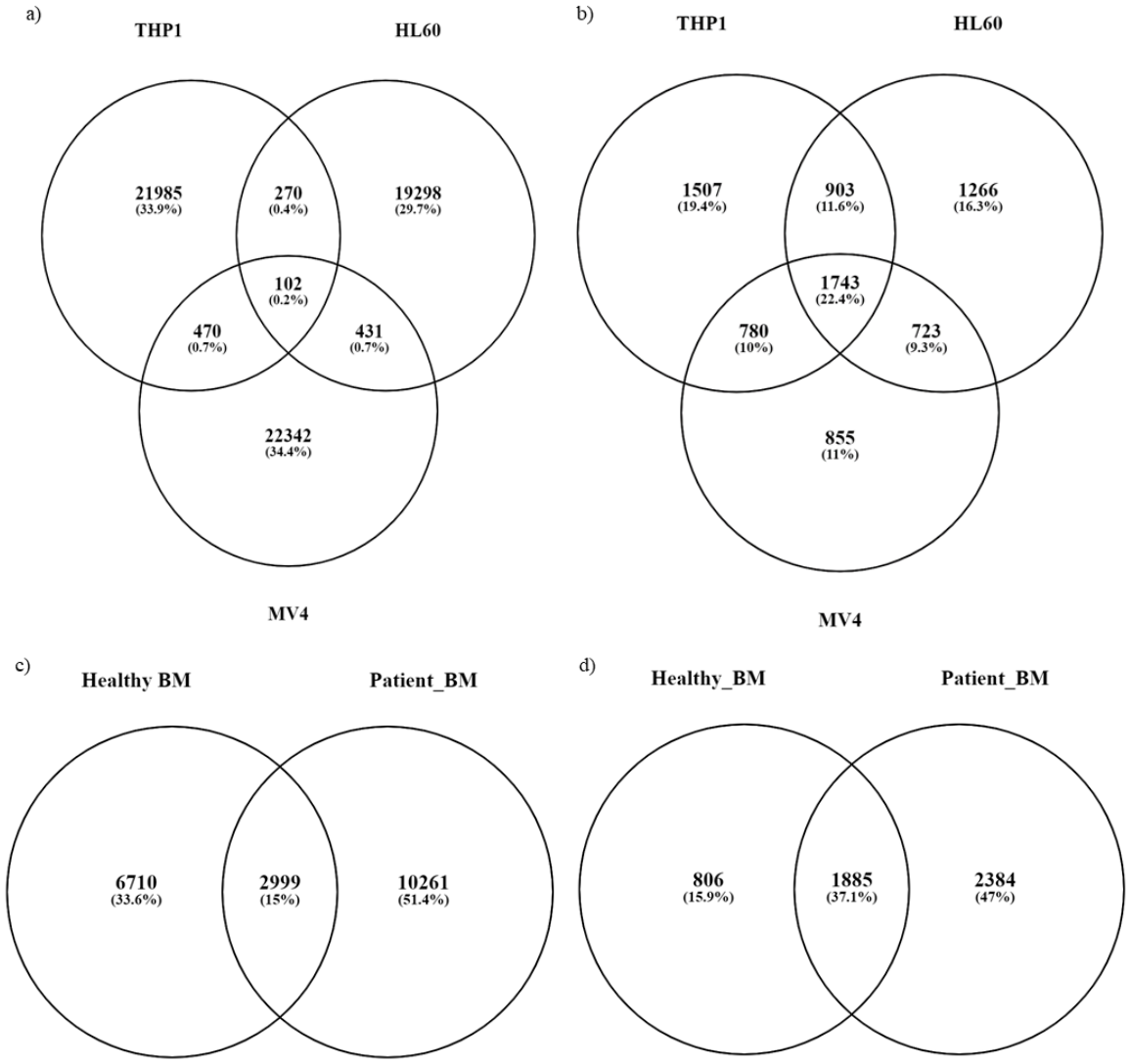
3.3.3 Cell lines recapitulate the immunopeptidome of clinical samples

In order to gain deeper insight if shared HLA class I allotypes were sampling the same or similar peptides in different datasets analysis was performed at 2 levels. First, intra-immunopeptidome comparison between the three cell lines (i.e. THP-1, HL60, MV411.1) revealed only 102 or 0.2% shared peptides (Figure 5a), out of a total 64,898 unique peptides between them. This was expected given the lack of shared HLA allotypes between the cell lines. Examination of the source proteins showed similar result as different allotypes seem to sample different antigens as evident from the small overlap of 1,743 shared proteins out of the total 7,777 unique proteins contributing to the immunopeptidome across the three cell lines (Figure 5b). These shared 1,743 proteins belonged to pathways commonplace in cell lines such as nucleic acid metabolism, mitotic cell cycle, RNA synthesis and mRNA splicing.

Second, inter-immunopeptidome comparison between healthy and patient samples demonstrated that shared alleles presented a much higher proportion of shared peptide antigens. A total of 19,970 unique peptides were identified between the healthy and patient samples, with an overlap of 2,999 peptides between them (Figure 5c), (Supplementary file 1, tab1) representing 1885 common source proteins (Figure 5d). This could be attributed to shared HLA alleles including B*07:02, C*07:02 and A3 supertype alleles sampling the same peptides. To verify this, a sub analysis of 9mers (n = 1155), which constituted the dominant length amongst the shared peptides, was performed using NetMHC4.0. Binding affinity for the 9mers was predicted for all HLA alleles shared between healthy individuals and patients. It was found that out of the 1,155 nonamers, 670 were restricted to HLA B*07:02 followed by 489 restricted by HLA-C*07:02 and 240 by HLA A*74:01. Not surprisingly, HLA alleles that have similar motifs shared nonamers amongst them. For instance, HLA A*74:01 peptides (in healthy BM) were also restricted to other alleles of the HLA-A3 supertype including HLA A*03:01, -A*31:01, -A*33:03 and -A*68:01 (patient HLA alleles) and HLA C*07:02 shared 321/489 peptides with HLA-C*16:02. Based on this data it could be inferred that inter sample homogeneity or heterogeneity at the immunopeptidome level was governed by the presence of similar HLA allotypes between samples.

Another major focus of the study was to determine whether immortalised cell lines could recapitulate the immunopeptidome found in primary tumour samples. This is clinically relevant as oftentimes primary tumour sample is not available or not available in sufficient quantity for immunopeptidomics analysis and hence cell lines might be a good surrogate to understand the disease. To investigate this further, inter-immunopeptidome analysis between cell lines and

clinical samples with similar HLA genotypes was performed. Similar to above results, the THP-1 cell line and healthy BM both of which express HLA A*02:01, shared a total of 984 peptides (Supplementary file 1, tab 2) derived from 810 shared antigens (Figure 5e and 5f). Whilst the MV411.1 cell line and AML patient 1 both of which express HLA allotypes belonging to the HLA-A3 supertype (i.e. HLA A*03:01, -A*68:01 and -A*31:01) and were observed to share 1,433 common peptides (Figure 5g) derived from 2,204 proteins (Figure 5h), Supplementary file 1, tab3). Hence, it would be reasonable to assume that cell lines, to some extent, are capable of recapitulating the immunopeptidome of clinical samples if there are shared or related HLA allotypes expressed.



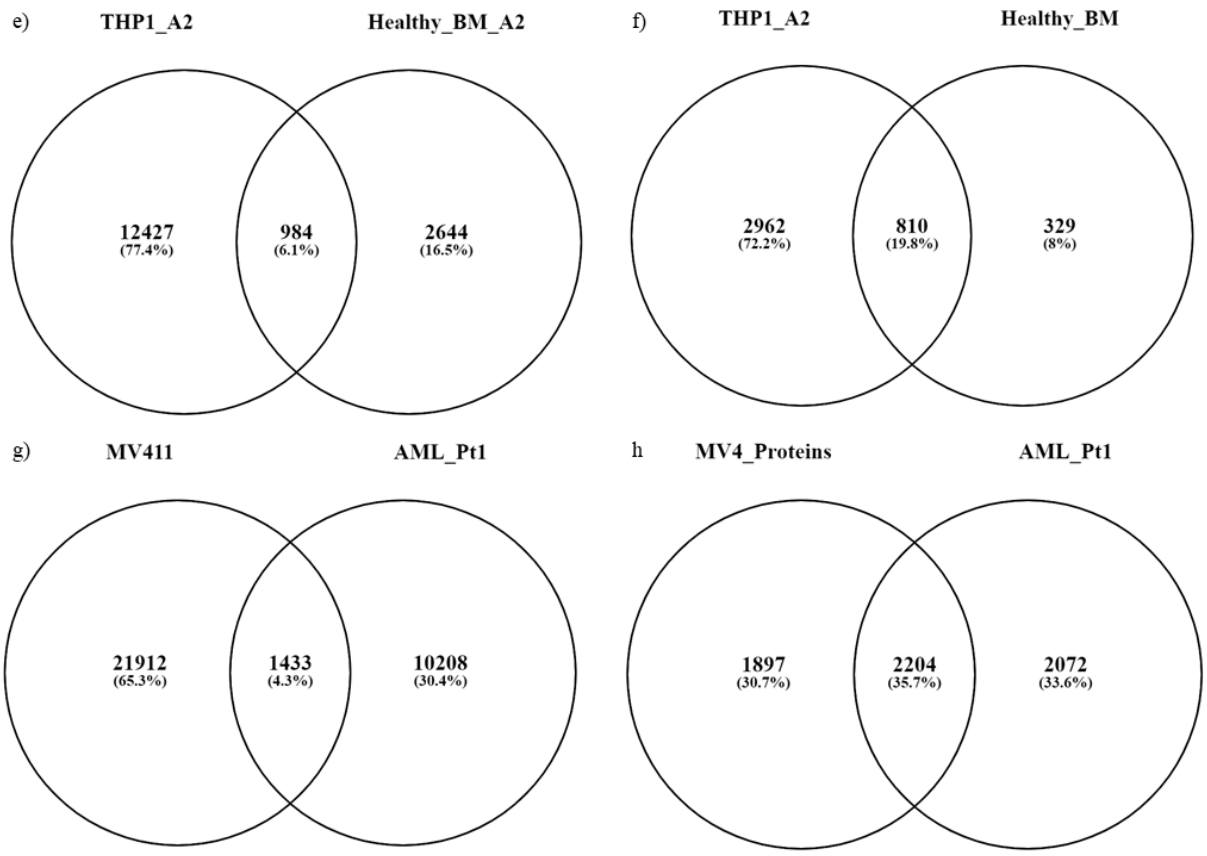


Figure 5 - Shared immunopeptidomes contributed by common source proteins are observed across the different leukemic samples. Venn diagram depicting shared HLA class I immunopeptidome between a) cell lines THP-1, MV411.1 and HL60 c) healthy and patient BM samples e) HLA A2 restricted peptides identified in THP-1 and healthy BM and g) MV411.1 and AML patient1 BM. Venn diagrams depicting the overlap in source proteins between b) cell lines THP-1, MV411.1 and HL60 d) healthy and patient BM samples f) HLA A2 in THP-1 and healthy BM and h) MV411.1 and AML patient1 BM.

3.3.4 Differential membrane and soluble HLA-derived immunopeptidomes observed in HM patients

The current dogma in the field of immunopeptidomics is that sHLA isolated from peripheral blood is representative of the tumour as these molecules are constantly being shed from the tumour site from where they circulate in the periphery. In a previous study by Bassani-Sternberg et al (Bassani-Sternberg et al. 2010) demonstrated an overlap of up to 89% in the immunopeptidomes of sHLA derived from the plasma with the immunopeptidomes of mHLA derived from PBMCs of leukemic patients.

Table 3 – sHLA class I peptides identified in patient BM

Sample	sHLA class I peptides
Healthy BM	NA*
AML patient 1	144
AML patient 2	552
ALL patient 3	2,277

*NA – Not available

In order to investigate if the tumour contributed to sHLA class I repertoire, we performed a direct comparison of the mHLA class I repertoire from the site of leukemogenesis (i.e. BM) and sHLA from peripheral blood. Peripheral blood (20mL) was collected from newly diagnosed leukaemia patients at the same time as collection of BM aspirate. Plasma was separated from whole blood by centrifugation and frozen at -80°C. Immunoaffinity purification was performed to capture sHLA and mHLA peptides from plasma and BM samples, respectively, using immobilised W6/32 antibody. Both mHLA and sHLA samples were acquired on the Tribrid Fusion MS and data analysis was performed as described above.

A total of 2,737 unique sHLA class I peptides were identified across three patient samples (Figure 6a). Surprisingly, maximum number of sHLA peptides (n = 2,277) were identified from ALL patient3 (Table 3), who had a loss of chromosome 6 indicating that the cancer did not contribute to their sHLA immunopeptidome. Majority of sHLA peptides identified in AML patient1 belonged to albumin, hence patient1 data was excluded from any further analysis. Peptide length distribution of the peptides identified in patient2 and patient3 was performed and the peptides were found to have canonical length distribution with ~61% and 64% of peptides identified in AML patient 2 and ALL patient 3, were 9mers respectively (Figure 6b). Second, to ascertain the HLA restriction of the 9mers, NetMHC Pan 4.0 was then used to predict the binding ability of the nonamers for the six HLA class I alleles , with strong and weak binders being used to plot the consensus binding motifs (Figure 7a and 7b).

Another indication that the tumour was not contributing to the sHLA repertoire was the presence of little overlap was found between sHLA and mHLA peptides in patient samples. In AML patient2 there was only 9% overlap between sHLA and mHLA peptides (Supplementary Figure 4a) whilst the overlap was only ~1% for AML patient 3 (Supplementary Figure 4b). The overlap between mHLA of BM and sHLA from plasma did not improve at protein level,

with only 16% and 2.3% commonly shared in Patient2 and Patient3, respectively indicating that mHLA and sHLA were sampling different proteins.

To determine if this was because sHLA was sampling source proteins from different cellular compartment than mHLA, functional annotation based on gene ontology (GO) was performed. It was found that ~79% of source proteins contributing to both sHLA and mHLA came from the cytoplasm followed by membrane (55 -58%) and nucleus (52 - 55%) in both patients (Figure 6c) which is line with previous studies (Scull et al. 2012).

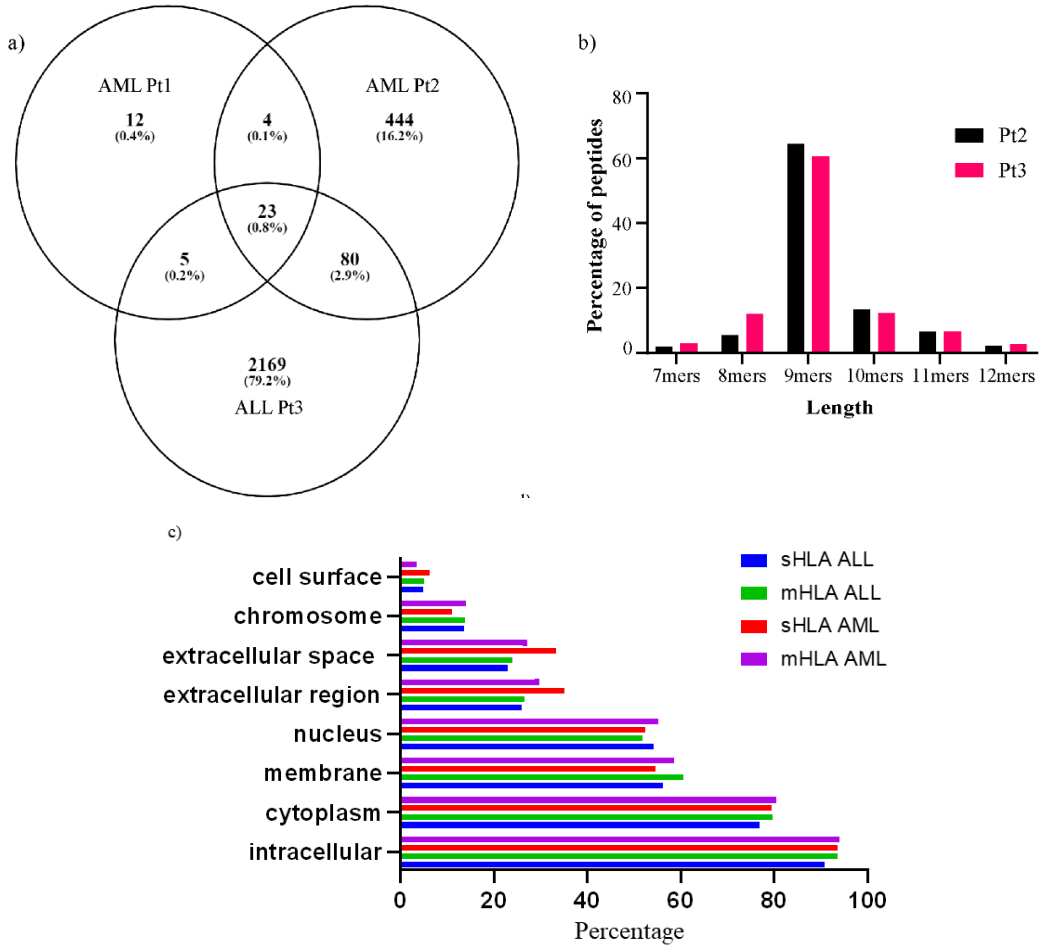
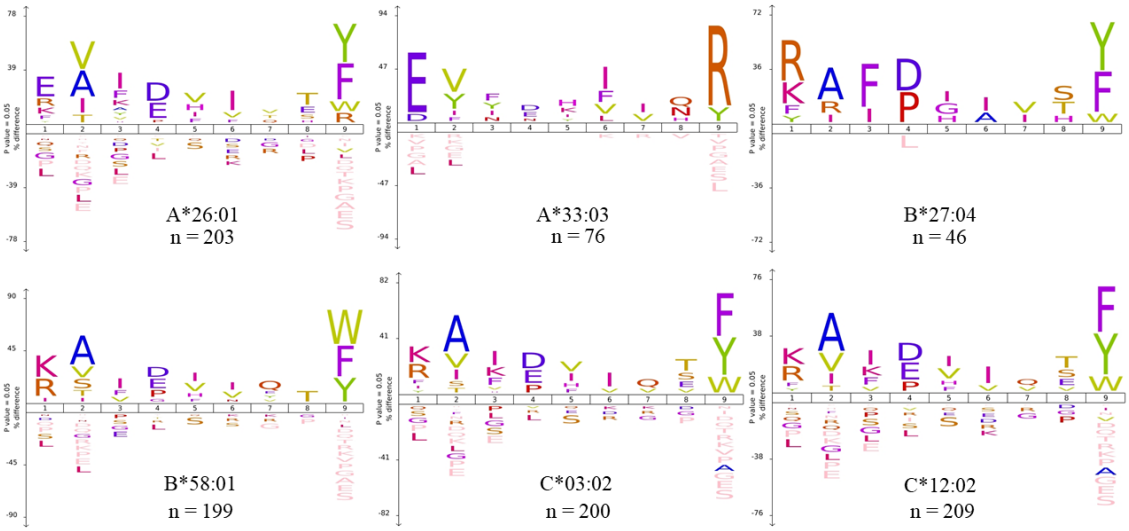


Figure 6 - Comparison of sHLA peptides purified from patient BM. a) Venn diagram to depict number of unique and overlapping sHLA peptides eluted from AML Patient1, AML Patient2 and ALL Patient3. b) Comparison of peptide length distribution identified in Patient2 and Patient3. c) Comparison looking into proportions of proteins associated with different cellular components for mHLA and sHLA in AML Pt2 and ALL Pt3.

a) AML patient2



b) ALL patient3

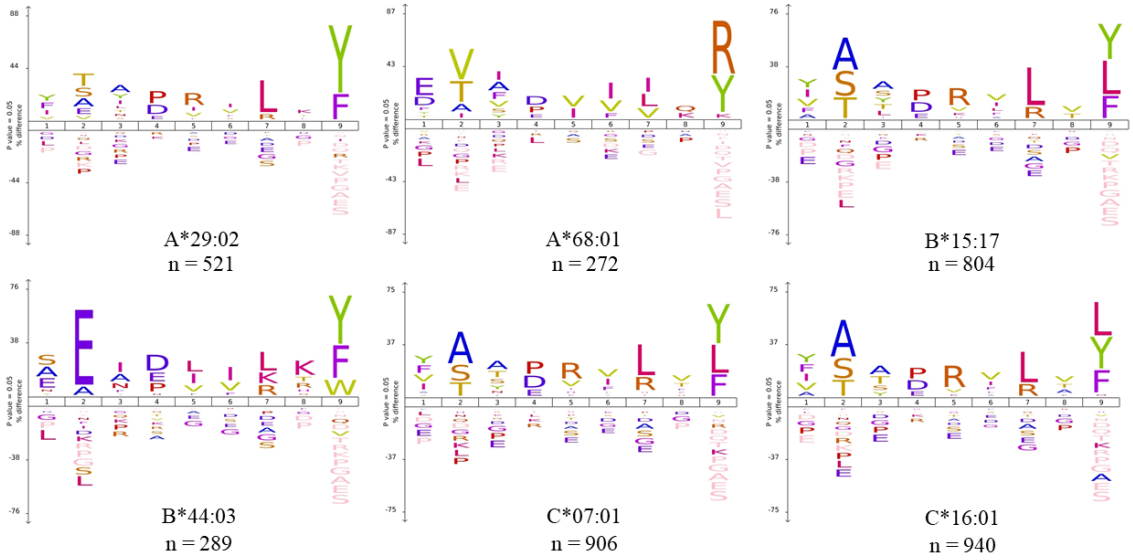


Figure 7 - Predicted nonameric sHLA binders identified from a) AML patient2 and ALL patient3 plasma have canonical binding motifs. 9mers constituted the majority of peptides and were selected for HLA-binding affinity prediction using NetMHC Pan 4.0. Icelogo representation of motifs of peptides predicted to bind to HLA allotypes expressed by each patient with **n** representing number of binders predicted for each allele.

3.3.5 Leukaemia specific antigens identified in clinical samples

To identify cancer-specific peptides, samples were interrogated for the presence of both TAAs and CTAs using both the TANTIGEN (Olsen et al. 2017)and CTdatabase (Almeida et al. 2009)

publicly available databases, respectively. A total of 3,874 TAAs and CTAs were identified across the HM samples (Table 4).

Table 4 – TAA and CTA peptides identified across haematological cell lines and patient samples.

Cell line/sample	TANTIGEN peptides	CTA peptides
THP-1	945	83
MV411.1	1038	70
HL60	628	68
Healthy BM	464	16
AML patient 1	490	30
AML patient 2	39	0
ALL patient 3	3	0

A total of 3,294 unique TAAs were identified across the HM samples with very little overlap between samples (Figure 8a). Not surprisingly, due to large number of peptides identified in the three leukemic cell lines, they contributed to majority of TAA (73%) identified across samples with 98 peptides overlapping with healthy BM and 41 peptides overlapping with patient BM samples (Figure 8a). Out of the 3,294 unique peptides, only 21 were previously reported. 16/21 peptides were restricted by HLA-A2, 3/21 by HLA-B8 and 2/21 peptides by HLA A*24:02 and A*68:01 (Table 5). Several peptides of interest were identified such as VLQELNVTV which was a part of randomised phase II trial in chronic myeloid leukaemia (CML) patients (QuintásCardama et al. 2008). It was found that the vaccination induced specific CD8 T cell response (evident from HLA-A2 peptide-tetramer staining) which improved overall immunological response in CML patients. Another peptide, FLDPRPLTV was found to produce cytotoxic T lymphocyte (CTL) response leading to cell lysis (as measured through Cr release assays) in healthy and multiple myeloma patients in a study by Maecker et al (Maecker et al. 2005).

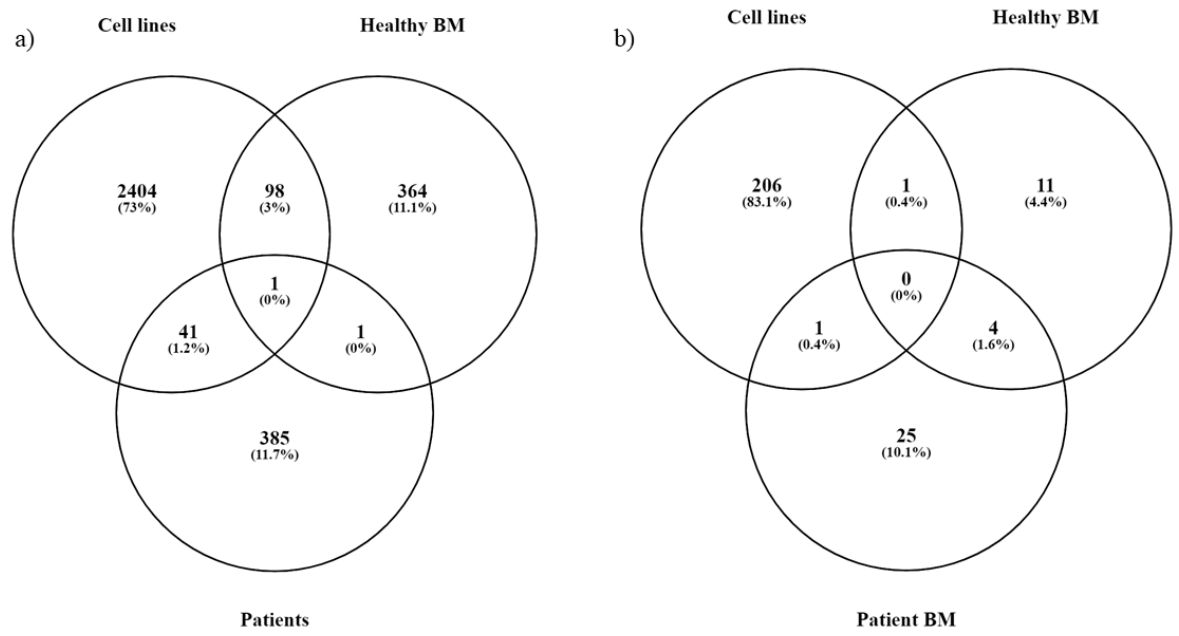


Figure 8 – Venn diagram depicting number of unique and overlapping peptides a) TAAs and b) CTAs identified in haematological malignancy.

Table 5 – Previously cancer-specific HLA class I peptides identified in cell lines and clinical samples

A2 restricted peptides

Peptide	**Cell line/Clinical sample**	**Source Protein**	**Study identified in**
SVASTITGV	THP-1	PLIN2	
FLDPRPLTV	THP-1	CYP1B1	Cytotoxic CTL response seen in healthy and in patients with MM (Maecker et al. 2005)
VLQELNVTV	THP-1 Healthy BM	PRTN3	Randomised phase II trial in CML (QuintásCardama et al. 2008) CTL response in MM (Alatrash et al. 2018)
CLYGNVEKV	THP-1	HNRNP	
KIMDQVQQA	THP-1	APC	Peptide identified in ovarian cancer cell line, anti-peptide CTL response detected as Cr release in healthy A2 donors (Ramakrishna et al. 2003)
RLQEDPPAGV	THP-1	UBE2A	Ovarian cancer (Ramakrishna et al. 2003)
KLDVGNAEV	THP-1 MV411.1 Healthy BM	BCAP31	Ovarian cancer (Ramakrishna et al. 2003)
FLYDDNQRV	THP-1 Healthy BM	TOP2A	Ovarian cancer (Ramakrishna et al. 2003)

ILDDIGHGV	THP-1	ARGBPIA	Ovarian cancer (Ramakrishna et al. 2003)
LLDRFLATV	THP-1 Healthy BM	CCNI	Ovarian cancer (Ramakrishna et al. 2003)
RLYPWGVVEV	THP-1	SEPT2	Ovarian cancer (Ramakrishna et al. 2003)
KLQELNYNL	THP-1 Healthy BM	STAT1	Ovarian cancer (Ramakrishna et al. 2003)
FMFGQKLNV	THP-1	HNRNPL	Identified in patient BLCL
LLDVAPLSL	THP-1	HS71L	CTL response resulting in lysis seen in healthy donors (Faure et al. 2004)
SLLQHLIGL *	THP-1	PRAME	High-avidity allo-HLA restricted T cells identified against the peptide from PBMCs collected from an AML patient suffering from acute GVHD after single HLA-A2 mismatched SCT followed by DLI (Amir et al. 2011)

* Identified in the CTdatabase, all other peptides were identified in the TANTIGEN database

B*08 restricted peptides

Peptide	**Cell line/Clinical sample**	**Source Protein**	**Study identified in**
VGLIRNLAL	MV411.1	CTNNB1	
APLLRWVL	AML Patient1	HMOX1	
ELLIRKLPF ^	AML Patient3	H3.3	

^ Identified in sHLA

Other alleles restricted peptide

Peptide	**Cell line/Clinical sample**	**Source Protein**	**Study identified in**
AFLRHAAL (A*24:02)	MV411.1	NOB1	CTL precursor detected across different cancers (Maeda et al. 2002)
ETIPLTAEKL ^ (A*68:01)	AML Patient3	CCND1	Identified in renal carcinoma study

^ Identified in sHLA

To identify peptides specifically of interest in AML, a curated list of proteins previously reported to be specific to AML was compiled (Klar et al. 2014, Berlin et al. 2015, Backert et al. 2017). The list comprised of 26 proteins including Runt related transcription factor (RUNX1), NPM1, MPO, PRAME, MYC and RARA. In cell lines a total of 198 peptides (Supplementary file 2) were identified from 21 proteins on the list whilst in patients 81 peptides (Supplementary file 3) were identified from the 11 proteins. Even at TAA level, different alleles were presenting different peptides from same source proteins as evident from non-

overlap amongst the identified peptides. Most shared peptides (10/18) came from MPO protein followed by 4/18 peptides were from PRTN3 protein (Figure 9).

Other peptides of clinical interest included 126 unique peptides identified from well documented oncogenic proteins including TP53, NRAS, KRAS, probable E3 ubiquitin-protein ligase (HERC1), ETS Variant Transcription Factor (ETV6) present in both patient BM and healthy BM (Supplementary file 4, tab1 and tab2) and cell lines (Supplementary file 5). Some peptides from proteins such as WT and Lyc were identified only in cells lines.

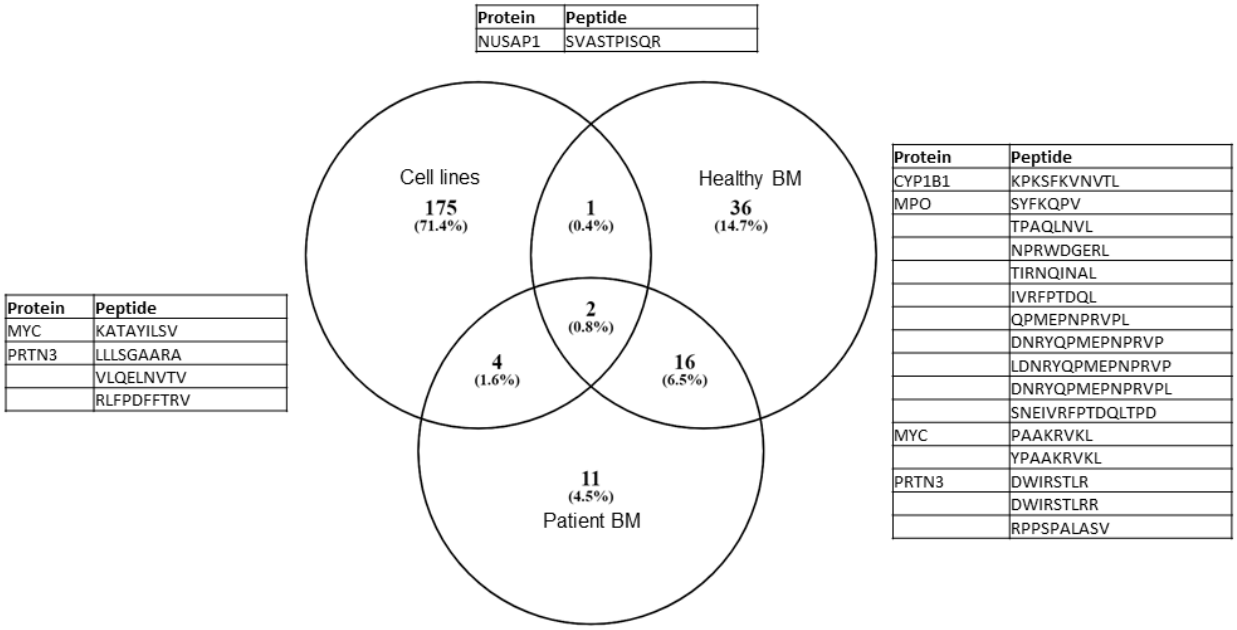


Figure 9 – Unique and shared AML specific peptides identified across cell lines and clinical samples. Peptides in boxes represent shared peptides between each sample. There were 16 shared peptides between healthy and patient BM with maximum peptides (n = 10) coming from MPO protein.

The dataset was also interrogated for CTA peptides, with a total of 248 unique peptides identified across the HM dataset (Table 4). Similar to TAA data, most CTA peptides (~83%) were unique to the leukemic cell lines (Figure 8b) with only 25 peptides found in healthy (Supplementary file 6) and 11 unique peptides in patient BM (Supplementary file 7). Only one of the identified CTA HLA A*02:01-restricted peptide (SLLQHLIGL, Table 5) has been previously reported in an AML patient suffering from acute graft versus host disease (GVHD) after mismatched SCT followed by donor lymphocyte infusion (DLI). Peptide- specific high avidity allo-HLA restricted T cells were identified in the patient PBMCs (Amir et al. 2011).

Some other peptides of interest came from well-known antigens such as MAGE, XAGE and PRAME proteins.

Table 6 – TAA and CTA peptides identified in sHLA of patient samples.

Sample	TANTIGEN peptides	CTA peptides
AML patient 2	22	0
ALL patient 3	97	17

To investigate whether chromosomal loss in the setting of HM influences the presentation of TAAs and CTAs, the sHLA dataset was also mined for presence of both TAAs and CTAs. No TAA or CTA was identified in AML patient1 sHLA peptide repertoire due to low numbers of peptides identified in it. From AML patient2 and ALL patient3, a combined total of 109 unique TAAs (Table 6) were identified (Supplementary file 8 and file 9 respectively), whilst CTAs were identified in ALL patient 3 only (n =17) (Supplementary file 9). Surprisingly in ALL patient 3, two previously reported peptides ETIPLTAEKL [CCND1 source protein] (Weinschenk et al. 2002) restricted to HLA A*68:01 and a HLA B*08:01-restricted peptide ELLIRKLPF [Histone 3.3 source protein]) (Demine et al. 2005) were identified (Table 5). Other peptides of interest were derived from oncogenic proteins such as MYC, NPM1 and DNAJC2.

3.4 DISCUSSION

In this study, we have utilised an optimised experimental workflow as described previously in Pandey *et al* (Pandey et al. 2020) to identify peptides from immunoprecipitated HLA proteins by high-performance MS in both leukemic cell lines and BM samples. To our knowledge, this is first study to identify mHLA class I peptides from patient BM, thereby providing significant information for the field. Our in-depth immunopeptidomics approach resulted in a large dataset comprising of 81,027 mHLA and 2,737 sHLA peptides from a total of 28 HLA allotypes and based on frequency distribution they together cover ~95% of the Caucasian population. Here, we used three established myeloid leukemic cell lines (THP-1, MV411.1 and HL60) as an effective model for AML and performed a comparison with patient BM to determine whether cell lines represent an experimental surrogate in lieu of clinical samples.

The first major outcome was a first-hand insight into how the presence of well documented cytogenic mutations might influence HLA expression and immunopeptidome. This was evident from significantly reduced number of peptides identified in AML patient2 with mutated

NPM1 protein which is known to downregulate HLA class II expression (Dufva et al. 2020). The mutation only influenced HLA expression and not the peptide repertoire itself, as the peptides identified still retained all the characteristics typical of class I peptides including length distribution, and presence of P2 and P9 anchor residues for their respective alleles. Similarly, reduced HLA expression observed in All patient3, due to loss of chromosome 6, is quite uncommon in newly diagnosed leukemic patients (Campoli et al. 2008, Rovatti et al. 2020). Although there are various studies over the years that have reported it in different solid tumours such as lung, cervical cancer, melanoma and adenocarcinomas (Campoli and Ferrone 2008, Shukla et al. 2015, McGranahan et al. 2017). From the limited endogenous peptide repertoire found in ALL patient 3 it could be inferred that although loss of HLA expression is directly associated with reduced antigen presentation at the site of tumour origin, it does not influence the characteristics of class I peptides such as length distribution or presence of anchor residues in 9mer. This study therefore highlights the need to investigate the immunopeptidome at the site of origin as immunotherapy may also include drugs which can target HLA expression in the patients which might aid in recovery and survival of patients.

The second major outcome was a comparative analysis of the immunopeptidome of well-known leukemic cell lines and patient BM. This is a crucial finding as tumour samples from patients are either unavailable or may be too limiting for immunopeptidomics analysis. It was found that samples which shared HLA allotype or supertype shared common peptides derived from a range of antigens. This has been previously reported in immunopeptidomics studies including those in AML patients (Narayan et al. 2019) and pan haematological malignancy immunopeptidome studies (Backert et al. 2017).

Our study also highlighted that cell lines might be able to recapitulate the immunopeptidome of clinical samples to some extent with respect to shared proteins as well. This was evident from overlap observed at source protein level between patient BM and cell lines (~19%) compared either to lower overlaps between healthy BM and cell lines (6.4%) or between patient and healthy BM (2.6%) (Supplementary Figure 5). However, the cell lines did not fully represent a healthy BM or the malignant transformation occurring at the onset of disease as proteins unique to both the conditions (662 in patient BM and 268 in healthy BM) were absent in the cell lines. This observation implies that the cell lines are phenotypically different and probably have reached a state of non-differentiation having been in culture for decades in some circumstances. Interestingly, this also suggests that the immunopeptidome directly reflects

tumour/lineage-specific biology and hence is a useful tool for not only vaccine design but also understanding basic tumour biology and microenvironment.

The third major outcome presents an alternative to existing dogma that sHLA peptides are constantly shed from the tumour site and then circulated throughout the body. The limited number of mHLA peptides identified in ALL patient3 BM was consistent with the cytogenetic report of loss of chromosome 6, however, the identification of 2657 bona fide and unique sHLA class I peptides in the plasma (obtained from peripheral blood at the same time as BM aspirate) is indicative that cancer does not always contribute to sHLA peptide repertoire. This observation is further substantiated by non-overlap observed between mHLA and sHLA peptides identified in different patients. Interestingly, at a source protein level sHLA provides a very clear indication of dysregulation occurring at cellular level associated with oncogenesis. Pathway analysis based on enrichment network visualisation in AML patient2 revealed "gene silencing" and "chromatin remodelling" influencing "MHC class II biosynthesis" via CIITA pathway (Supplementary Figure 6). As mentioned previously, this method of gene silencing is how mutated NPM1 influences downregulation of class II expression. This shows that even if sHLA is not derived from the tumour itself, it can be of immense diagnostic and prognostic value. TAA and CTAs were also identified in sHLA dataset, constituting ~5% of the total sHLA repertoire identified in the study.

Previous studies have relied on gene expression approaches to identify neoepitopes (Rajasagi et al. 2013, Rajasagi et al. 2014). This is not always a good strategy particularly for leukaemia since 1) they have low mutational burden (Rosenberg and Restifo 2015) and 2) RNA expression does not always translate into HLA-restricted presentation of corresponding gene products (Weinzierl et al. 2007). Recent studies have demonstrated that tumour-associated peptides identified by direct analysis of primary tumour-derived HLA ligands are capable of eliciting specific immune responses that are associated with improved clinical outcomes (Schwartzentruber et al. 2011, Berlin et al. 2015, Neidert et al. 2018). Excitingly, our study, has not only generated the largest dataset for HM but has expanded the list of potential peptides that can be tested as potential targets for T cell mediated immunotherapy. We identified a total of 3,874 peptides including the 245 peptides from AML specific proteins including RUNX1, NPM1, RARA, MPO and PRTN3. When compared to published AML immunopeptidomes (derived from PBMCs and cell lines) (Berlin et al. 2015) (Narayan et al. 2019) we identified only two peptides (ILWETVPSM, RESEYKQVY) found in leukemic cell lines that were

shared with Berlin *et al* (Berlin et al. 2015), whilst 23 peptides (derived from proteins including MCL1, and Fibronectin type III domain containing protein) were found overlapping with Narayan *et al* (Narayan et al. 2019) (Supplementary file 10), with the majority of these peptides derived from MV411.1. Some other peptide of interest belonged to well-known, universal oncogenic proteins including KRAS, NRAS and WT, these peptides can also be incorporated either in AML related immunotherapy or might have a wider potential in other cancers.

To conclude, the current study therefore provides a deeper insight into the know-how of HLA class I dysregulation in leukaemia and therefore the peptide repertoire presented in such clinical samples. The extensive amount of data unveiled by our experimental approach aids in the expansion of the current list of peptides to be considered for peptide-based immunotherapy. This can also help in making informed decision with regards to immunotherapy, particularly in favour combined immunotherapeutic which might help in attaining desired clinical outcome of remission and overall survival of patients.

REFRENCES

Alatrash, G., et al. (2018). "Targeting the Leukemia Antigen PR1 with Immunotherapy for the Treatment of Multiple Myeloma." Clinical Cancer Research **24**(14): 3386-3396.

Albitar, M., et al. (2007). "Levels of soluble HLA-I and β2M in patients with acute myeloid leukemia and advanced myelodysplastic syndrome: association with clinical behavior and outcome of induction therapy." Leukemia **21**(3): 480-488.

Almeida, L. G., et al. (2009). "CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens." Nucleic Acids Res **37**(Database issue): D816-819.

Amir, A. L., et al. (2011). "PRAME-specific Allo-HLA-restricted T cells with potent antitumor reactivity useful for therapeutic T-cell receptor gene transfer." Clin Cancer Res **17**(17): 5615-5625.

Backert, L., et al. (2017). "A meta-analysis of HLA peptidome composition in different hematological entities: entity-specific dividing lines and "pan-leukemia" antigens." Oncotarget **8**(27): 43915-43924.

Bassani-Sternberg, M., et al. (2010). "Soluble plasma HLA peptidome as a potential source for cancer biomarkers." Proc Natl Acad Sci U S A **107**(44): 18769-18776.

Berlin, C., et al. (2015). "Mapping the HLA ligandome landscape of acute myeloid leukemia: a targeted approach toward peptide-based immunotherapy." Leukemia **29**(3): 647-659.

Brodsky, F. M. and P. Parham (1982). "Monomorphic anti-HLA-A,B,C monoclonal antibodies detecting molecular subunits and combinatorial determinants." J Immunol **128**(1): 129-135.

Campoli, M. and S. Ferrone (2008). "HLA antigen changes in malignant cells: epigenetic mechanisms and biologic significance." Oncogene **27**(45): 5869-5885.

Connors, J. M., et al. (2018). "Brentuximab Vedotin with Chemotherapy for Stage III or IV Hodgkin's Lymphoma." N Engl J Med **378**(4): 331-344.

De Kouchkovsky, I. and M. Abdul-Hay (2016). "'Acute myeloid leukemia: a comprehensive review and 2016 update'." Blood Cancer J **6**(7): e441.

Demine, R. and P. Walden (2005). "Testing the Role of gp96 as Peptide Chaperone in Antigen Processing." Journal of Biological Chemistry **280**(18): 17573-17578.

DiNardo, C. D. and J. E. Cortes (2016). "Mutations in AML: prognostic and therapeutic implications." Hematology Am Soc Hematol Educ Program **2016**(1): 348-355.

Döhner, H., et al. (2015). "Acute Myeloid Leukemia." New England Journal of Medicine **373**(12): 1136-1152.

Dufva, O., et al. (2020). "Immunogenomic Landscape of Hematological Malignancies." Cancer Cell.

Escher, C., et al. (2012). "Using iRT, a normalized retention time for more targeted measurement of peptides." Proteomics **12**(8): 1111-1121.

Estey, E. H. (2009). "Treatment of acute myeloid leukemia." Haematologica **94**(1): 10-16.

Faure, O., et al. (2004). "Inducible Hsp70 as target of anticancer immunotherapy: Identification of HLA-A*0201-restricted epitopes." Int J Cancer **108**(6): 863-870.

Gobin, S. J. P., et al. (1997). "Site α Is Crucial for Two Routes of IFNγ-Induced MHC Class I Transactivation: The ISRE-Mediated Route and a Novel Pathway Involving CIITA." Immunity **6**(5): 601-611.

Guillaume, P., et al. (2018). "The C-terminal extension landscape of naturally presented HLA-I ligands." Proceedings of the National Academy of Sciences **115**(20): 5083-5088.

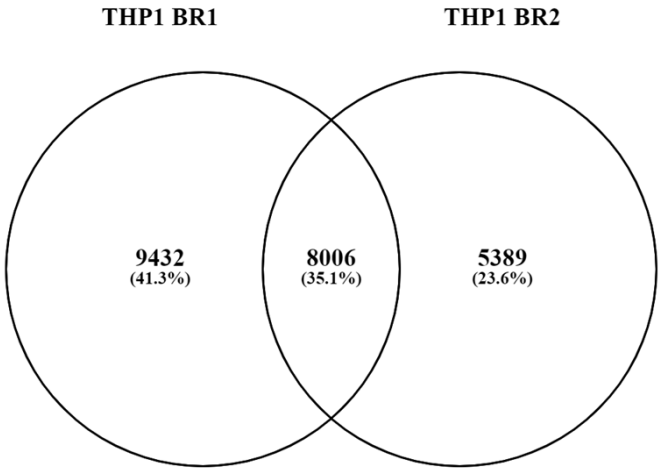
Jurtz, V., et al. (2017). "NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data." J Immunol **199**(9): 3360-3368.

Kantarjian, H. M., et al. (1992). "Prognostic significance of elevated serum beta 2-microglobulin levels in adult acute lymphocytic leukemia." Am J Med **93**(6): 599-604.

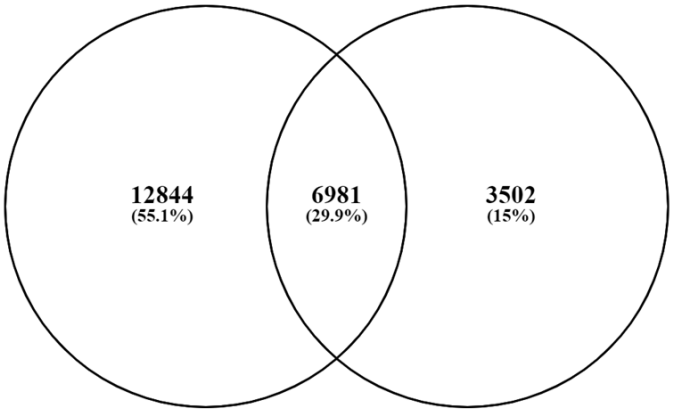
Klar, R., et al. (2014). "Therapeutic targeting of naturally presented myeloperoxidase-derived HLA peptide ligands on myeloid leukemia cells by TCR-transgenic T cells." Leukemia **28**(12): 2355-2366.
Kumar, C. C. (2011). "Genetic Abnormalities and Challenges in the Treatment of Acute Myeloid Leukemia." Genes Cancer **2**(2): 95-107.
Maecker, B., et al. (2005). "Identification of a new HLA-A*0201-restricted cryptic epitope from CYP1B1." International Journal of Cancer **115**(2): 333-336.
Maeda, Y., et al. (2002). "Detection of peptide-specific CTL-precursors in peripheral blood lymphocytes of cancer patients." Br J Cancer **87**(7): 796-804.
Martin, B. K., et al. (1997). "Induction of MHC class I expression by the MHC class II transactivator CIITA." Immunity **6**(5): 591-600.
McGranahan, N., et al. (2017). "Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution." Cell **171**(6): 1259-1271.e1211.
Morra, E., et al. (2009). "Clinical management of primary non-acute promyelocytic leukemia acute myeloid leukemia: Practice Guidelines by the Italian Society of Hematology, the Italian Society of Experimental Hematology, and the Italian Group for Bone Marrow Transplantation." Haematologica **94**(1): 102-112.
Narayan, R., et al. (2019). "Acute myeloid leukemia immunopeptidome reveals HLA presentation of mutated nucleophosmin." PLoS One **14**(7): e0219547.
Neidert, M. C., et al. (2018). "The natural HLA ligandome of glioblastoma stem-like cells: antigen discovery for T cell-based immunotherapy." Acta Neuropathologica **135**(6): 923-938.
Oliveros, J. "VENNY. An interactive tool for comparing lists with Venn diagrams." BioinfoGP, CNB-CSIC.
Olsen, L. R., et al. (2017). "TANTIGEN: a comprehensive database of tumor T cell antigens." Cancer Immunol Immunother **66**(6): 731-735.
Pandey, K., et al. (2020). "In-depth mining of the immunopeptidome of an acute myeloid leukemia cell line using complementary ligand enrichment and data acquisition strategies." Mol Immunol **123**: 7-17.
Parham, P. and F. M. Brodsky (1981). "Partial purification and some properties of BB7.2. A cytotoxic monoclonal antibody with specificity for HLA-A2 and a variant of HLA-A28." Hum Immunol **3**(4): 277-299.
Przepiorka, D., et al. (2015). "FDA Approval: Blinatumomab." Clinical Cancer Research **21**(18): 4035-4039.
Pymm, P., et al. (2017). "MHC-I peptides get out of the groove and enable a novel mechanism of HIV-1 escape." Nat Struct Mol Biol **24**(4): 387-394.
QuintásCardama, A., et al. (2008). "Randomized Phase II Study of Proteinase 3-Derived PR1 Peptide Vaccine and GM-CSF with or without PEG-Interferon ALFA-2B to Eradicate Minimal Residual Disease in Chronic Myeloid Leukemia." Blood **112**(11): 3219-3219.
Rajasagi, M., et al. (2013). "Tumor Neoantigens Are Abundant Across Cancers." Blood **122**(21): 3265-3265.
Rajasagi, M., et al. (2014). "Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia." Blood **124**(3): 453-462.
Ramakrishna, V., et al. (2003). "Naturally occurring peptides associated with HLA-A2 in ovarian cancer cell lines identified by mass spectrometry are targets of HLA-A2-restricted cytotoxic T cells." Int Immunol **15**(6): 751-763.
Ramarathinam, S. H., et al. (2018). "Employing proteomics in the study of antigen presentation: an update." Expert Review of Proteomics **15**(8): 637-645.
Rodríguez, J. A. (2017). "HLA-mediated tumor escape mechanisms that may impair immunotherapy clinical outcomes via T-cell activation." Oncol Lett **14**(4): 4415-4427.

Rosenberg, S. A. and N. P. Restifo (2015). "Adoptive cell transfer as personalized immunotherapy for human cancer." Science **348**(6230): 62-68.
Rovatti, P. E., et al. (2020). "Mechanisms of Leukemia Immune Evasion and Their Role in Relapse After Haploidentical Hematopoietic Cell Transplantation." Frontiers in Immunology **11**(147).
Ruella, M. and M. V. Maus (2016). "Catch me if you can: Leukemia Escape after CD19-Directed T Cell Immunotherapies." Computational and Structural Biotechnology Journal **14**: 357-362.
Sarkizova, S., et al. (2020). "A large peptidome dataset improves HLA class I epitope prediction across most of the human population." Nat Biotechnol **38**(2): 199-209.
Schwartzentruber, D. J., et al. (2011). "gp100 Peptide Vaccine and Interleukin-2 in Patients with Advanced Melanoma." New England Journal of Medicine **364**(22): 2119-2127.
Scull, K., et al. (2012). "Secreted HLA recapitulates the immunopeptidome and allows in-depth coverage of HLA A*02:01 ligands." Molecular immunology **51**: 136-142.
Shukla, S. A., et al. (2015). "Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes." Nature Biotechnology **33**(11): 1152-1158.
Terwilliger, T. and M. Abdul-Hay (2017). "Acute lymphoblastic leukemia: a comprehensive review and 2017 update." Blood Cancer J **7**(6): e577.
Vassilakopoulos, T. P., et al. (2019). "Immunotherapy in Hodgkin Lymphoma: Present Status and Future Strategies." Cancers (Basel) **11**(8).
Weinschenk, T., et al. (2002). "Integrated functional genomics approach for the design of patient-individual antitumor vaccines." Cancer Res **62**(20): 5818-5827.
Weinzierl, A. O., et al. (2007). "Distorted Relation between mRNA Copy Number and Corresponding Major Histocompatibility Complex Ligand Density on the Cell Surface." Molecular & Cellular Proteomics **6**(1): 102-113.
Zhang, J., et al. (2012). "PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification." Mol Cell Proteomics **11**(4): M111.010587.
Zheng, P.-P., et al. (2018). "Approved CAR T cell therapies: ice bucket challenges on glaring safety risks and long-term impacts." Drug Discovery Today **23**(6): 1175-1182.

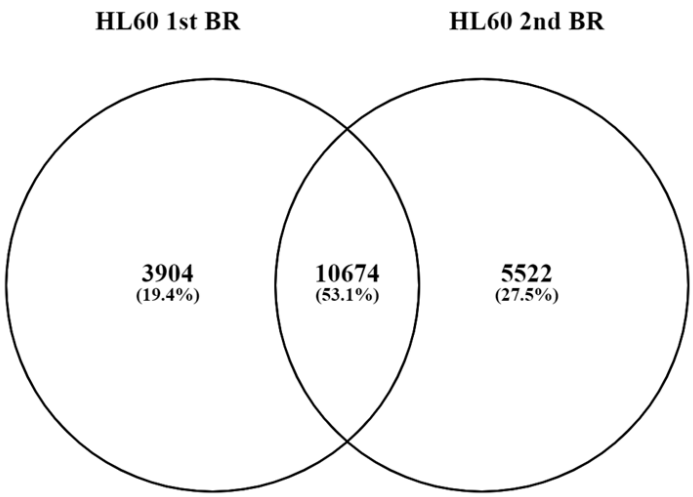
Chapter 3 - Supplementary figures



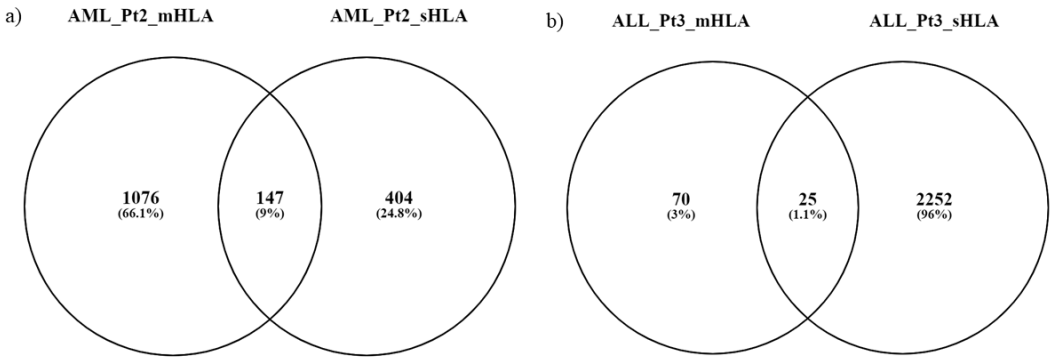
Supplementary Figure 1 – Venn diagram depicting unique and overlapping peptides between AML cell line THP1 biological replicates (BR). An overlap of only 35% was found with 9,432 unique peptides in 1st BR and 5389 unique peptides in 2nd BR. Venn diagram made with help of Venny 2.1. (https://bioinfogp.cnb.csic.es/tools/venny/index.html)



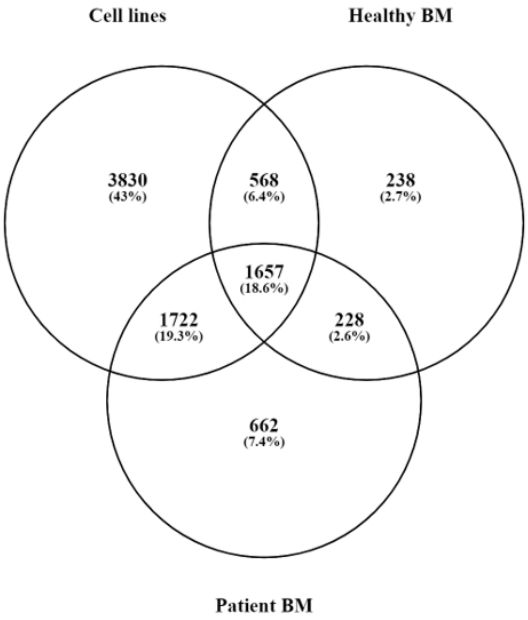
Supplementary Figure 2 - Venn diagram depicting unique and overlapping peptides between AML cell line MV411.1 biological replicates (BR). An overlap of only ~30% was found with 12,844 unique peptides in 1st BR and 3502 unique peptides in 2nd BR. Venn diagram made with help of Venny 2.1, (https://bioinfogp.cnb.csic.es/tools/venny/index.html)



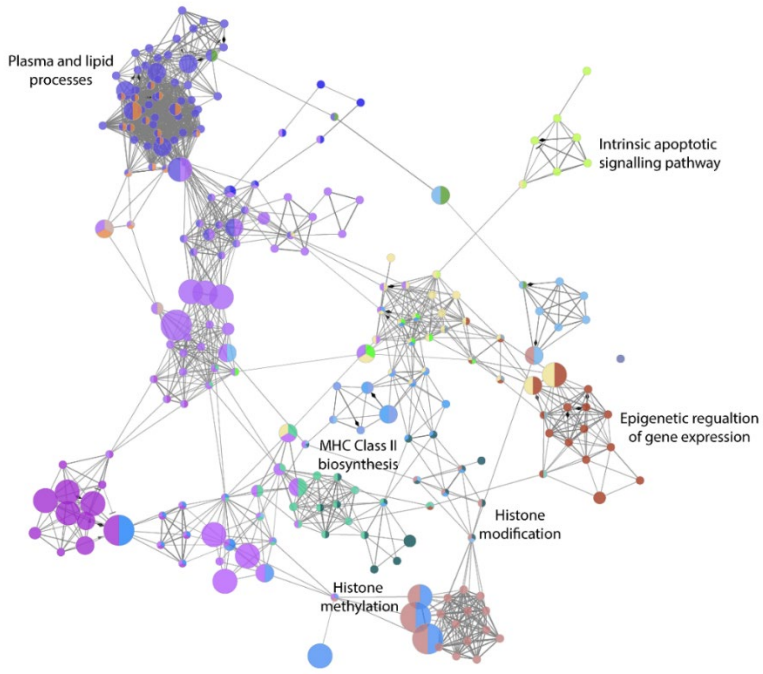
Supplementary Figure 3 - Venn diagram depicting unique and overlapping peptides between AML cell line HL60 biological replicates (BR). An overlap of only 53% was found with 3,904 unique peptides in 1st BR and 5,522 unique peptides in 2nd BR. This was the highest overlap found in biological replictae dataset amongst the three cell lines. Venn diagram made with help of Venny 2.1. (https://bioinfogp.cnb.csic.es/tools/venny/index.html)



Supplementary Figure 4 - Venn diagram depicting unique and overlapping peptides between mHLA and sHLA identified in a) AML patient 2 wherein an overlap of 9% was found between the two dataset and b) ALL patient 3 wherein an overlap of 1.1% was found between the two dataset**.** Venn diagram made with help of Venny 2.1 (https://bioinfogp.cnb.csic.es/tools/venny/index.html).



Supplementary Figure 5 – Venn diagram depicting unique and overlapping source proteins identified across HM samples. High overlap is seen between cell lines and patient source proteins whilst very little overlap is seen between cell lines and healthy and between patients and healthy BM. Venn diagram made by Venny (https://bioinfogp.cnb.csic.es/tools/venny/index.html).



Supplementary Figure 6 - Network analysis of sHLA class I source proteins in AML patient2 depict downregulation of MHC Class II biosynthesis via CIITA pathway. The figure was made using ClueGO (v2.1) plugin for Cytoscape (v3.7).

CHAPTER - 4

Chapter 4: Oncogenic signatures within the PTM-immunopeptidome of haematological malignancies

4.1 INTRODUCTION

There exists a plethora of post translationally modifications (PTMs) each of which can diversify and/or modulate the role and function of proteins. Recent studies have addressed the prevalence of PTMs within the proteome (Mann et al. 2003, Olsen et al. 2013). Signalling dysregulation in cancers is known to alter the PTM footprint in the cells which might be linked to characteristic signatures of malignancy. This is because the acquired capabilities of proliferation, differentiation and survival in cancer cells arise through mutations leading to aberrant signalling (Hanahan and Weinberg 2000). The two most dysregulated signalling pathways in cancer are phosphorylation and methylation. Particularly in leukaemia, aberrant phosphorylation is a dominant mechanism involved in leukemogenesis (Chase et al. 2006, Zhu et al. 2015). This is because the most frequently observed mutations at the gene level (in the form of gene fusions, transversions and tandem repeats) influence kinases belonging to the receptor tyrosine kinase (RTK) family which further leads to dysregulation in Raf/Ras/MAPK kinase or the JAK/STAT pathway downstream. Another important PTM in leukaemia is arginine methylation, which is catalysed by a family of nine protein arginine methyltransferases (PRMTs). In leukaemia arginine methylation seems critical for maintenance of leukemic stem cell (LSC) potential by sustaining an oncogenic transcriptional program and blocking differentiation and apoptosis (Wingelhofer and Somervaille 2019, Birch and Shilatifard 2020).

These cancer specific PTMs can also be represented in the HLA bound peptides (Mei et al. 2020) hence investigating the immunopeptidome can provide insight into oncogenesis. The most prevalent modifications on HLA bound peptides include oxidation of Met (up to ~83% of all PTM-peptides) followed by deamidation of Asn or Glu (up to ~7% of all PTM-peptides) (Mei et al. 2020). Rare but important modifications such as phosphorylation at Ser/Thr/Tyr account for less than 3% of modified peptides or all peptides in the total immunopeptidome(Abelin et al. 2017) whilst methylation at Arg/Lys constitute even lower, less than 1% of the total immunopeptidome (Marino et al. 2017). Up until recently, identification of phosphopeptides required some form of enrichment (Abelin et al. 2015) but as mentioned in Chapter 2 our in-depth immunopeptidomics approach combined with high-resolution MS demonstrates the potential of identifying phosphopeptides without enrichment.

Crystallographic studies (Mohammed et al. 2008, Petersen et al. 2009, Mohammed et al. 2017) have revealed an interesting facet of PTM-peptides, that PTMs such as phosphorylation can stabilise the pHLA complex by interacting with residues in the peptide binding groove. In addition to this, studies investigating the immunogenicity of PTM peptides suggested their potential role in anti-tumour immunity. In a study by Cobbold *et al* (Cobbold et al. 2013) they identified 95 phosphopeptides from leukemic patients. They successfully demonstrated that phosphopeptides specific CD8 T cells identified in healthy donors and expanded *in vitro,* elicited not only strong IFN-γ response on stimulation, but were also capable of recognising and lysing tumour cells expressing the phosphopeptides. Similar CD8 T cell response were found upon vaccinating humanised HLA A2 expressing mouse with 2 phosphopeptides KLIS(Ph)VETDI and RQLS(Ph)SGVSEI (Lin et al. 2019).

Similarly, the impact of methylation on T cell recognition has rarely been explored, with only a single study by Jarmalavicius *et al.* (Jarmalavicius et al. 2010) wherein they identified a monomethylated peptide SQNPR(m)FYHK derived from the tumour suppressor gene, G-protein pathway suppressor 2 (GPS2). The peptide restricted by HLA A*11:01, was found to induce CD8+ T cell mediated INF-γ response (as measured by ELISPOT assay) directly *ex vivo* in PBMCs of nine melanoma patients. At the same time the P5 dimethyl arginine version of the peptide SQNPR(dm)FYHK generated a response only in one patient, thereby highlighting that TCRs can distinguish between dimethyl and monomethyl Arg in this peptide. Together these studies highlight the potential of PTM-peptides for novel immunotherapeutics.

With regards to HLA bound phosphopeptides, all studies to date have investigated only phosphoSer (pSer) or phosphoThr (pThr) containing peptides (Zarling et al. 2000, Zarling et al. 2006, Depontieu et al. 2009, Zarling et al. 2014) (Cobbold et al. 2013, Mommen et al. 2014). This almost certainly reflects the abundance of these two phospho amino acids compared to phosphoTyr (pTyr). In contrast, cell surface presentation of methylated peptides by HLA class I alleles is one of the least explored domains of immunopeptidomics. Currently, there are 3 studies that have identified mono- and/or dimethylated Arg(dmArg) containing peptides presented by different HLA alleles including HLA A*11:01, -B*39:10 and -B*07:02 allotypes (Yague et al. 2000, Jarmalavicius et al. 2010, Marino et al. 2017). In the current study, the use of an optimised immunopeptidomics workflow combined with high resolution MS as described previously (Pandey et al. 2020), has resulted in the generation of one of the largest datasets for

HLA-phosphopeptide ligands identified from HM. These were derived from well characterised cell lines and healthy and patient-derived bone marrow aspirates. Prior to the current study there has been no investigation of the impact of phosphorylation mediated signalling and its associated oncogenic signature within the immunopeptidome. Furthermore, there is no reported structural data for pTyr and asymmetric dmArg peptide antigens restricted by HLA A*02:01 and the influence these PTMs might have on the stability and conformation of the HLA peptide complex.

4.2 MATERIALS AND METHODS

4.2.1 Tumour samples and cell lines

Cryopreserved bone marrow mononuclear cells (BMMC) sample (~25x10^6per vial) from healthy individual was commercially sourced (Lonza, Walkersville, USA). Patient BM aspirate (~2mL each patient) and PBMC (20mL) were collected from three newly diagnosed leukemic patients (n = 3) in heparinised vacutainer tubes recruited at Box Hill Hospital during routine clinical investigations (Eastern Health ethics number LR77/2016 and Monash University ethics number 8088). Two patients were diagnosed with AML whilst one patient was diagnosed with ALL. Cells were pelleted at 2,300g, 20mins, 4°C and cryopreserved at -80°C until required. Patient cytogenetic details and HLA typing is mentioned in Table 1, chapter 3.

All the cell lines used in the study including THP-1, MV411.1 and HL60 were cultured in 800mL of RF10 (RPMI (Gibco, ThermoFisher, USA) supplemented with 2mM MEM nonessential amino acid solution (Gibco, ThermoFisher, USA), 100mM HEPES (Gibco, ThermoFisher, USA), 2mM L-glutamine (Gibco, ThermoFisher, USA), 100 U/mL/100 µg/mL penicillin/streptomycin (Gibco, ThermoFisher, USA), 50µM 2-mercaptoethanol (Sigma-Aldrich, USA) and 10% heat inactivated FCS (Sigma-Aldrich, USA)) in roller bottles (Greiner Bio-One International AG, Austria). All cell lines were maintained and cultured at 37°C with 5% CO_2. HLA class I expression on each of the cell lines was measured by flow cytometry following surface staining using neat hybridoma supernatants for anti-HLA-A2 (BB7.2 [HB-82 ; ATCC], Parham and Brodsky 1981) and pan-HLA class I (W6/32 [HB-95; ATCC], (Brodsky and Parham 1982) as primary antibodies (all antibodies were produced in-house), with a goat anti-mouse IgG PE (Southern Biotech, USA) as the secondary antibody.

4.2.2 Purification of HLA-peptide complexes

For each cell line ~1x10^9 cells were harvested by centrifugation (1200g, 20 mins, 4°C), washed twice in chilled phosphate-buffered saline and frozen by submersion in liquid nitrogen. Pelleted cells were stored at −80°C until time of use.

Based on HLA typing of the samples, immunoaffinity columns were generated by crosslinking 10mg of HLA specific antibodies including BB7.2 (for THP1 an healthy BM) and pan class I antibody W632 (for all samples) to 1 mL of PAS resin (CaptivA®, Repligen, USA) as previously described (Pandey et al. 2020).

To isolate pHLA complexes from the samples, first, frozen cell pellets (1x10^9) and BM pellets (1-2x10^8) were pulverised by cryogenic milling (Retsch Mixer Mill MM 400), and solubilised in Lysis Buffer [0.5% IGEPAL (Sigma-Aldrich, USA), 50mM Tris pH 8, 150mM NaCl and protease inhibitors (Complete Protease Inhibitor Cocktail Tablet, 1 tablet per 50 mL solution; Roche Molecular Biochemicals, Switzerland)] for 1 hr at 4°C with rotation. The supernatant was passed through a PAS pre-column (500µL) to remove non-specific binding material, followed by serial affinity capture wherein allele specific peptides were pulled down first, either by BB7.2 followed by pulling down of the remaining class I peptides by W6/32 (for all samples). Bound pHLA complexes were eluted from the column by using 5 CV of 10% acetic acid.

Only THP1 eluate was simultaneously processed by RP-HPLC and MWCO (as mentioned previously in chapter 2 whilst all the other samples were fractionated by RP-HPLC into 48 fractions which were concatenated, vacuum concentrated, into 10 fractions, reconstituted in 0.1% formic acid and were acquired on high resolution MS Tribrid Fusion (Thermo Scientific, USA).

4.2.3 Mass spectrometry data acquisition

In brief, all MS spectra (MS1) profiles were recorded from full ion scan mode 375-1800 *m/z*, in the Orbitrap at 120,000 resolution with automatic gain control (AGC) target of 400,000 and dynamic exclusion of 15 secs. The top 12 precursor ions were selected using top speed mode at a cycle time of 2 secs. For MS/MS a decision tree was made to aid selecting peptides of charge state 1 and 2-6 separately. For singly charged analytes only ions falling within the range of *m/z* 800-1800 were selected, whereas for +2 to +6 *m/z* s no such parameter was set. The c-trap was loaded with a target of 200,000 ions with an accumulation time of 120 ms and isolation width of 1.2 amu. Normalised collision energy was set to 32 (high energy collisional dissociation (HCD) and fragments were analysed in the Orbitrap at 30,000 resolution.

The LC-MS/MS data was analysed using Peaks 8.5 software (Bioinformatics solutions) against the human proteome (Uniprot 15/06/2017; 20,182 entries). The following search parameters were used: error tolerance of 10 ppm using monoisotopic mass for precursor ions and 0.02 Da tolerance for fragment ions; enzyme used was set to none with following variable modifications: oxidation at Met, deamidation at Asp and Gln, phosphorylation at Ser, Thr and Tyr. False discovery rate (FDR) was estimated using a decoy fusion method (Zhang et al. 2012) and all datasets were analysed at 5% FDR. Files were exported as peptide.csv (Appendix for details). Mono- and dimethylation at Lys and Arg was searched in second round with help of Peaks PTM and files were exported separately as Peaks PTM.csv. Biological replicate data is merged into single file, as mentioned in the appendix, data in chapter 4 in Peaks PTM methylation files.

For motif analysis Icelogo (Colaert et al. 2009) stand-alone software was used. Amino acid enrichment in the experimental dataset over prevalence in the human proteome was determined using the static reference method with Homo sapiens SwissProt as reference set. The motif was depicted as percentage (difference in frequency for an amino acid at a certain location in the experimental set and the reference set) within p-value of 0.05. The difference in percentage is depicted as measure of the height of a letter in the amino acid stack.

4.2.4 Recombinant HLA-peptide complex generation

Inclusion-body of the HLA heavy (α) chain and β2M proteins was prepared as per the method of Garboczi, et al (Garboczi et al. 1996). In brief, the recombinant expression plasmids were transformed into BL21 (DE3) *E. coli*. A positive transformant was selected and grown at 37°C for 16 hrs in LB broth supplemented with kanamycin for both HLA-A2 and β2m. The inoculate was diluted 1/100 into a litre of LB medium and grown to an OD600 of 0.6. The culture was induced with 1mM IPTG and grown for a further 4 hrs. Bacteria were pelleted at 5,000g (SLC-3000 angle rotor, Thermo Fischer Scientific) for 10 mins at 4°C and stored at -80°C until further use.

The bacterial pellet was thawed and then lysed by addition of 22.5ml of lysis buffer containing 50mM Tris pH 8.0, 1% (v/v) Triton X-100, 1% (w/v) sodium deoxycholate, 100mM NaCl, 10mM DTT, 1 mg DNase I, 5mM $MgCl_2$ to extract the inclusion bodies as described previously (Clements et al. 2002). After 40 mins of continuous rocking at room temperature, the samples were homogenised for 30 secs using a Polytron homogeniser, after which 1mM EDTA was added. Inclusion bodies were isolated by centrifugation at 30,000g, 15 mins, 4°C in Sorvall GSA rotor. The pellets were resuspended in 150ml of wash buffer containing 50 mM Tris pH

8.0, 0.5% (v/v) Triton X-100, 100mM NaCl, 1mM EDTA, 1mM DTT, 0.2mM PMSF and 1µg/ml pepstatin, homogenised and centrifuged. This wash step was repeated four times. A final wash in a buffer containing 50mM Tris pH 8.0, 1mM EDTA, 1mM DTT, 0.2mM PMSF and 1µg/ml pepstatin A was then performed. The inclusion bodies were resuspended in guanidinium tris buffer. Inclusion body protein present in the supernatant was quantified by comparing the aliquots and protein standards on a reducing SDS–PAGE visualised by using Coomassie blue-stained, before freezing at -80°C.

To make secondary pHLA complexes HLA A*02:01 and β2M proteins were refolded in the presence of the 4 peptides, KLDYITLY, KLDY(Ph)ITYL, ILAEIARIL and ILAEIAR(dm)IL and purified as described previously by Clements *et al* (Clements et al. 2002). Briefly, 18 mg of HLA A*02:01 was refolded by rapid dilution in a solution containing 100mM Tris-HCl, pH 8.0 (Sigma-Aldrich, USA), 400mM L-arginine-HCl, 5mM reduced glutathione (Sigma-Aldrich, USA), and 0.5mM oxidized glutathione (Sigma-Aldrich, USA) in the presence of 6 mg of β2m and 2 mg of the peptides (Synpeptides, China) for 6 hrs. The refolded HLA-peptide complexes were dialysed using membrane with a cut-off of 12-14 kDa into 10 mM Tris, pH 8.0 overnight. Dialysed protein was loaded by gravity onto a diethyl-aminoethyl (DEAE) column that was first pre-washed with 5 CV of 1M NaCl and then equilibrated with 2 CV of 10mM Tris, pH 8.0. The protein was eluted with ~20mL of elution buffer i.e. 10mM Tris and 150mM NaCl and then concentrated to less than 4mL using 30 kDa cut off concentrator. The concentrated pHLA complex was purified by anion exchange using a HiTrap Q Fast Flow column (GE Healthcare, USA) on AKTA system in 10mM Tris pH 8.0 buffer with a 1M NaCl gradient from 0 to 500mM over 45 mins. The fractions with refolded pHLA complex was visualised on SDS-PAGE gel, pooled and concentrated to 500µL in volume. Final purification was performed by size-exclusion chromatography using HiLoad 10/30 Superdex 75 pg (GE Healthcare, USA) columns on an AKTA Purifier (GE Healthcare, USA) FPLC chromatography systems using isocratic gradient of 150mM Tris, pH 8.0.

4.2.5 X-ray crystallography

All four pHLA complexes were concentrated to ~5mg/mL and sent to the Monash Crystalmation facility wherein broad crystallisation screens using commercially available kits were set up with 200 nL of protein per condition. Proteins were crystallised by the sitting drop method. The A*02:01-KLDY(Ph)ITYL (from here on referred as KLDY(Ph)) complex protein

crystallised in condition B4 of the Shotgun1 (SG1) screen HT96 (Molecular Dimensions, USA) which comprised of 30 % w/v PEG 5000 MME, 0.2 M Ammonium sulfate and 0.1 M MES (pH 6.5). The A*02:01-KLDY native protein did not crystallise in any of the broad screens. Similar broad screens were set up for both A*02:01 ILAEIARIL native (from here on referred as ILAE(native)) and A*02:01 ILAEIAR(dm)IL (from here on referred as ILAE(dm) protein). Both proteins crystallised condition C7 of SG1 HT96 screen (Molecular Dimensions, USA) comprising of 25 % w/v PEG 3350 and 0.1 M Sodium HEPES (pH 7.5).

Manual crystallisation using "hand trays" were set up for ILAE(native) and ILAE(dm) proteins recapitulating the C7 condition with 1µL of protein and 1µL of mother liquor. Proteins were crystallised by the hanging-drop vapour-diffusion method at 20°C after seeding with another pHLA complex (Conditions mentioned in Table 3).

Prior to data collection, crystals were equilibrated in reservoir solution with 10% glycerol added as a cryoprotectant and then flash-cooled in a stream of liquid nitrogen at -180°C. Data sets were collected at the MX2 beamline (Australian Synchrotron, Victoria). The data were recorded on a on a Dectris EIGER 16M detector and were integrated and scaled using XDS (Kabsch 2010) and scaling with Aimless (part of CCP4 programme suite) (Collaborative 1994, Evans 2006). Details of the data processing statistics are summarised in Table 3. Phases for the structures were determined by molecular replacement as implemented in PHASER (McCoy et al. 2007) with HLA A*02:01 used as the search model (Protein Data Bank accession number: 5NMH, HLA A*02:01 restricted HIV peptide SLYNTIATL). Refinement of the models proceeded with iterative rounds of manual building in COOT (Emsley et al. 2004), refinement in PHENIX (Adams et al. 2010) and validation with MOLPROBITY (Adams et al. 2010). Refinement statistics are summarised in Table 3. Crystal processing, data acquisition on Synchrotron, data analysis and refinement was performed by Dr Jan Peterson (Rossjohn Lab, Monash University). Final PDB files were visualised in PyMol Molecular Graphics System, Version 2.0 Schrödinger, LLC.

4.2.6 Thermal stability assay

Thermal stability of the pHLA complex was assessed by differential scanning fluorimetry (DSF) (Niesen et al. 2007) wherein 0.5mg/ml of pHLA complex (native and modified) in 10mM Tris and 150mM NaCl, pH 8.0 in a reaction volume of 15 µL in triplicate (n = 3) was used. Protein unfolding was monitored by the addition of 10X fluorescent dye SYPRO® Orange (5000X, Sigma-Aldrich, USA). Refolded complexes were heated from 25 to 90°C at a heating rate of 1°C/min in the Real Time Detection system (Rotor-Gene® Q, QIAGEN) and

fluorescence intensity was measured using an excitation wavelength of 530 nm and emission at 555 nm. This increase in intensity of fluorescence ratio (530/555nm) vs temperature depicts a point wherein half of the pHLA complex unfolds also referred to as melting temperature (Tm) of the protein. Data was plotted using GraphPad Prism software (version 8.1).

4.3 RESULTS

HLA class I peptides were purified from the AML cell lines (n = 3) and clinical samples (n = 4) and identified by high-resolution MS. THP1 eluate was processed differently, wherein the pHLA complexes were separated using 2 complementary peptide enrichment strategies namely, RP-HPLC and 5kDa MWCO filter as described before (Pandey et al. 2020). The pHLA complexes of MV411.1 and HL60 and all clinical samples were separated using only RP-HPLC. A total of 61,445 unique HLA class I peptides were identified from the cell lines. In addition, clinical samples comprising of bone marrow aspirates yielded a total of 16,130 unique peptides with 3,452 overlapping peptides observed between the cell lines and clinical samples. The samples were searched for different post-translational modifications. Majority of modified peptides were oxidised at Met, with deamidation the second most common modification in all samples. This is consistent with other immunopeptidomics studies (Abelin et al. 2017, Mei et al. 2020).

Since altered phosphorylation is considered as a hallmark of cancer, the samples were investigated for peptides phosphorylated at Ser, Thr and Tyr. A total of 1,773 unique phosphopeptides were identified across all HM samples, thereby making this the biggest phosphopeptide immunopeptidome dataset reported in general and specially for haematological malignancy. Phosphopeptides constituted up to 0.7-1.5% of total peptides identified in different samples except for THP-1 wherein their percentage was much higher at 4.7%. This could be attributed to difference sample processing of the cell line mentioned above. The study also investigated the rarer modification of methylation occurring at Arg/Lys. A total of 320 peptides were identified which constituted only 0.4% of total peptides identified in the HM samples.

4.3.1 Serine/Threonine phosphopeptides dominate HLA class I restricted phosphopeptides in haematological malignancy

To delineate the complex data, a more comprehensive analysis of the phosphopeptides identified across cell lines and clinical samples was performed. Out of the 1,773 unique peptides, 1,585 (~89%) were identified from cell lines, with smaller percentage identified in clinical samples (Supplementary Figure 1a). This was expected as more peptides were identified in cell lines (Table 1), due to larger sample size and presence of biological replicates in each cell line (n = 2). Surprisingly, although the number of peptides identified in each cell line was similar, ~67% of phosphopeptides identified in cell lines came from THP-1 followed by MV411.1 (18%) and HL60 (13%). (Supplementary Figure 1b). This demonstrated that complementary peptide enrichment also helped in expanding phosphopeptide repertoire in THP1 sample.

Table 1 – Ser and Thr phosphopeptides constitute a major share of peptides in haematological malignancy dataset.

Cell line	Total number of peptides identified	pSer	pThr	pTyr
THP-1	22,827	602 (56%)	377 (35%)	99 (9%)
MV411.1	23,345	235 (67%)	104 (30%)	12 (3%)
HL60	20,100	135 (65%)	66 (32%)	8 (3%)
Healthy BM	9,710	50 (58%)	35 (40%)	1 (1%)
AML_Pt1	11,642	69 (66%)	29 (28%)	7 (6%)
AML_Pt2	1,223	5 (56%)	3 (33%)	1 (11%)

Out of the 1,733 phosphopeptides identified, pSer and pThr constituted ~96% of the detectable phosphopeptides within the immunopeptidome whilst pTyr accounted for the remaining 4%. Within the pSer/pThr dataset a preference for canonical nonamers was observed, with ~ 40% of the peptides being 9mers followed by preference for longer peptides, with ~30% being 10mers (Figure 1a). A sub analysis of unique 9mer pSer peptides (n = 352) identified across the samples revealed P4 as the preferred site for phosphorylation, found in ~39% (n =135) of the peptides. This was also evident from enriched Ser seen at P4 in the pSer peptide motif (Figure 1b). The other preferred site for Ser phosphorylation in 9mers was P5 which was found in ~10% of peptides.

In contrast, no clear preference for a P4 residue was observed in unique 9mers pThr peptides (n = 195) identified across different samples. This was substantiated when the motif for the peptides was plotted, with pThr being enriched at all positions except P9 (Figure 1c). On further

analysis, it was found that in cell lines only 15% of pThr peptides were phosphorylated at P4, with P1, P2 and P5 equally preferred. In clinical samples such as healthy BM, phosphorylation was only found at P3 and P6 position, although 6/9 (66%) peptides identified in patients were phosphorylated at P4. This deviation might be driven by different class I alleles present in the dataset and allelic peptide binding preferences.

Despite the slight variation, both Ser/Thr phosphopeptide nonamers followed previously reported K/RXXSPX kinase substrate motif (Mohammed et al. 2008, Cobbold et al. 2013), wherein there is a basic amino acid Arg or Lys at P1 and Pro following Ser/Thr (Figure 1b and 1c).

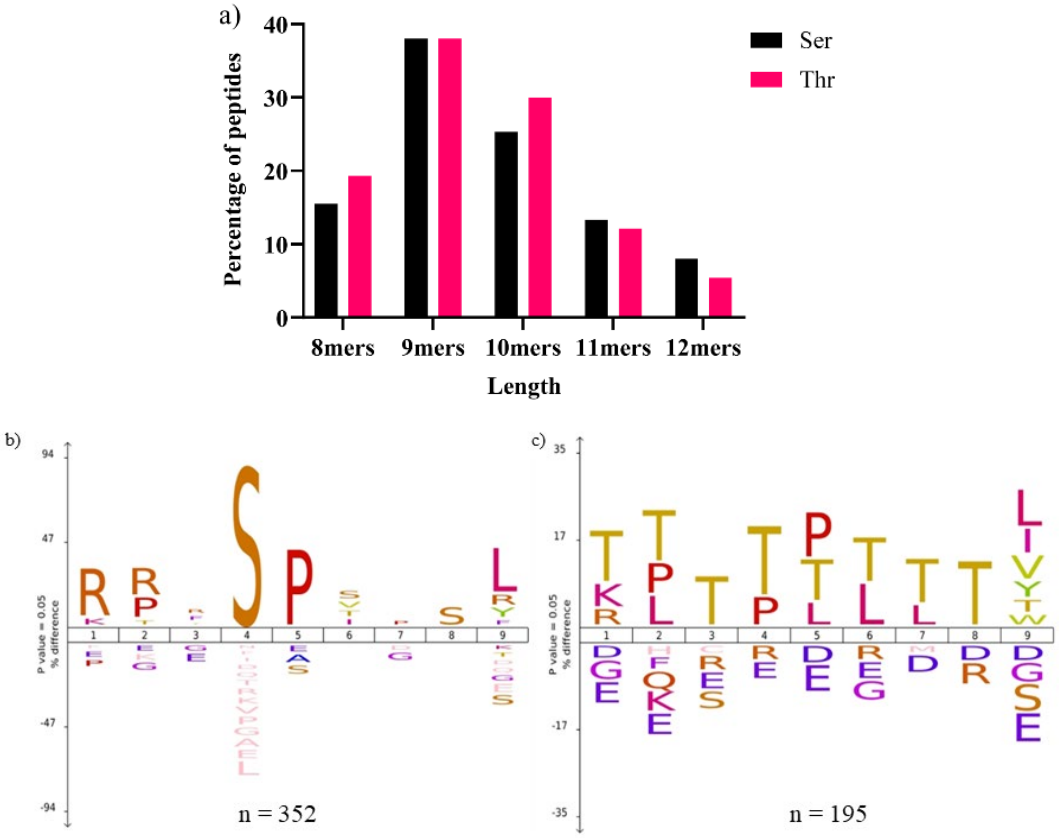


Figure 1 - Nonamers dominate Ser and Thr phosphopeptides identified across haematological malignancy as evident from a) length distribution of class I restricted unique Ser/Thr phosphopeptides identified across cell lines and bone marrow samples. IceLogo depiction of motif of nonameric b) Serine and c) Threonine phosphorylated peptides show preference for basic amino acid at P1 and Pro at P5.

In order to investigate if the site of modification represented in the immunopeptidomics data coincided with sites mapped to the source antigens in broader phosphoproteomics datasets the HM immunopeptidomes were matched to the PhosphoSitePlus database (Hornbeck et al.

2012). It was found that ~32% of the phosphorylation sites within the immunopeptidomics data were also present in PhosphoSitePlus (Supplementary file 1). This included ~42% of phosphopeptides found in BM samples of healthy and AML patients. This indicates that immunopeptidome provides an adept representation of phosphorylation events occurring at whole proteome level.

A total of 89 phosphopeptides were derived from oncogenes and TSG (Supplementary file 2) out of which 61 were also found to be present in PhosphoSitePlus (Supplementary file 3). In cell line derived data, the peptides belonged to well-known oncogenic proteins such as MYC, HERC, NPM, RUNX1, NPM and ATR. In patient sample phosphopeptides originated from different oncogenes, GRK2, MAP3K11, SKI, IRS2 and HIVEP2 proteins. The activation of some of these proteins such as NPM1, MYC, RUNX1 and MAP3K11proteins is associated with onset of leukemic malignancies, thereby making these peptides of particular interest as immunotherapeutic targets.

4.3.2 Tyr phosphopeptides identified in haematological malignancy

To date, there are no studies which have reported HLA restricted pTyr containing peptides. Thus, I provide the first evidence for their existence in the immunopeptidome, with a total of 123 pTyr peptides identified across the HM dataset (Supplementary file 4). In order to understand the nature of pTyr peptides within the immunopeptidome, an in-depth analysis was performed which revealed several key features. Firstly, the peptide length distribution for the pTyr peptides was observed to be skewed towards octamers, with ~43% (n = 53) out of total 123 peptides being 8mers. This was significantly higher when compared to the percentage of 8mers present in either pSer/pThr peptides (15-18%) (Figure 2a) or unmodified 8mers identified across the HM samples (highest percentage ~18% identified in MV411.1, Figure 2b, chapter 3).

Secondly, whilst the preferred site of phosphorylation for pSer peptides was at position P4 (Figure 1b), it was found that 8mer pTyr peptides preferred position P2, P3 and P5. However, this could be attributed to low number of pTyr peptides identified. Thirdly, these octameric peptides contained a L/K/XYXXL/V/P kinase substrate motif, wherein the P1 was either a small hydrophobic amino acid or a positively charged amino acid such as Lys (Figure 2b). The binding motif is similar to a Tyr kinase motif I/V/L/SXYpXXL/I, identified in prostate cancer (Drake et al. 2012) wherein anti-pTyr antibodies (clone 4G10) was used to pull down pTyr peptides from samples treated with trypsin.

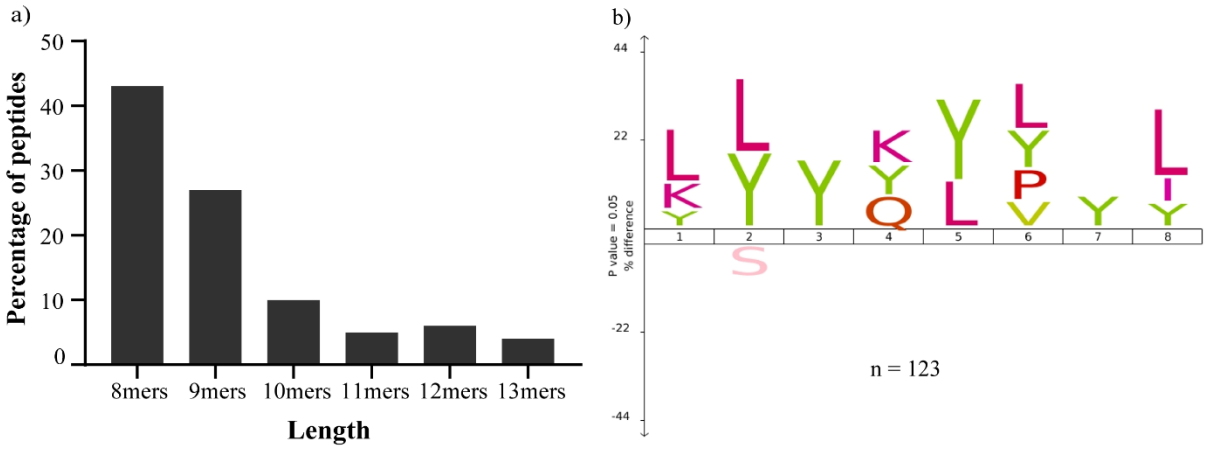


Figure 2 – A preference for octamers in Tyr phosphorylated immunopeptides as evident from a) Length distribution of all Tyr phosphorylated peptides identified in HM dataset, with octamers constituting the major peptide subset. b) Motif of the octameric Tyr phosphorylated peptides as interpreted using IceLogo, which shows predominance of Tyr at P3.

4.3.3 Phospho-immunopeptidome contains signatures associated with oncogenic signalling and malignant transformation

In order to determine if the phosphopeptides identified in the samples depicted the cellular dysregulation occurring in HM within the cell line or patient samples, pathway analysis based on the identified phosphopeptides was performed using Metascape (Zhou et al. 2019). Enrichment network visualisation of 409 unique source proteins identified in THP1 revealed RTK driven signalling directly influencing leukocyte and myeloid differentiation and hemopoiesis (Figure 3a). Some other key pathways whose proteins were found to phosphorylated included histone modification, DNA repair and spliceosome machinery. In order to gain insight into the signalling cascade driving the malignant transformation in THP1, phosphosite analysis was performed which involved identifying any previously reported phosphosite was present in the data and the kinase responsible for it using PhosphoSitePlus database (Hornbeck et al. 2012). RTK mediated signalling was evident in THP1 due to the presence of adaptor proteins SOS and IRS2 (Belov et al. 2012) which are necessary for downstream signal transduction. The signal from adaptor proteins is transduced by Ras/Raf proteins with the help of RhoGTPases and GEF. Several phosphopeptides belonging to both

the protein subtypes were identified in the cell lines. Presence of MAP3K/MAP2K signalling was evident from phosphorylated peptides originating from protein MAP3K1 along with MAP2K1, MAP2K4 and MAPK5 were found (Figure 3b). MAP2K1/5 kinases downstream signalling is known to activate ERK whilst MAP2K4 mediated signalling activates JNK. Some other proteins that participate in the Ras/Raf/MAP3K signalling such as MLK, TAO1 and FAK were also present. The ERK/JNK pathway results in activating genes such as MEF2C (site T293) and CREB-binding protein (CBP). Signalling along this pathway is known to promote both leukocyte and myeloid differentiation. Phosphosite analysis also revealed phosphorylation at a key site in another AML specific protein, RUNX1 (S249) which is known to be mediated by HIP kinase. Several chromatin modifying proteins such as and KMT and EP300 were also found phosphorylated. The presence of RUNX1 with the co-activators p300 and CBP highlights the histone remodelling (Mikhail et al. 2006) which leads to altered gene expression (Wang et al. 2020). To conclude, histone modification together with dysregulation of spliceosome machinery and cell cycle and DNA repair drive oncogenesis in the THP-1 cell line.

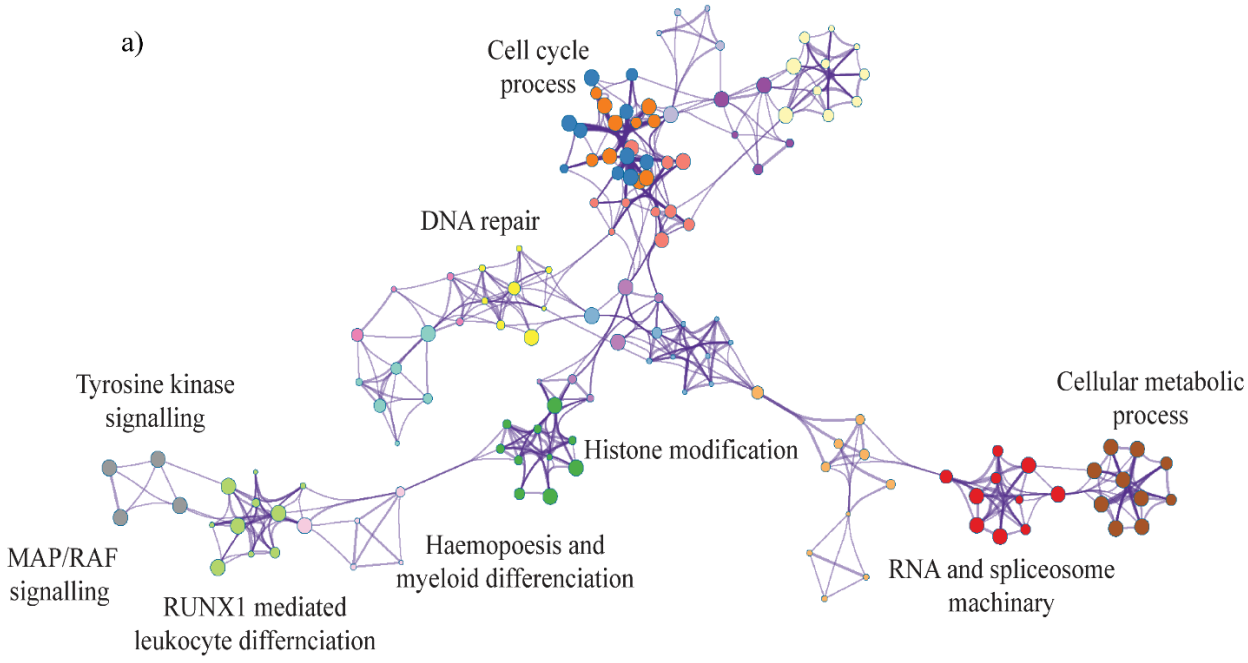


Figure 3a – Phosphorylation controls key pathways involved in leukemogenesis in THP1. Enrichment network visualisation of phosphopeptide source proteins identified in THP1 by Metascape (Zhou et al. 2019) depicting key pathways controlled by phosphorylation. Colour code represents the identities of gene lists.

RTK signalling in the MV411.1 cell line was evident due to the presence of Colony-stimulating factor 1 receptor (CSF1R) which belongs to the family of Fms like tyrosine kinase (FLT) receptors (Figure 3b).

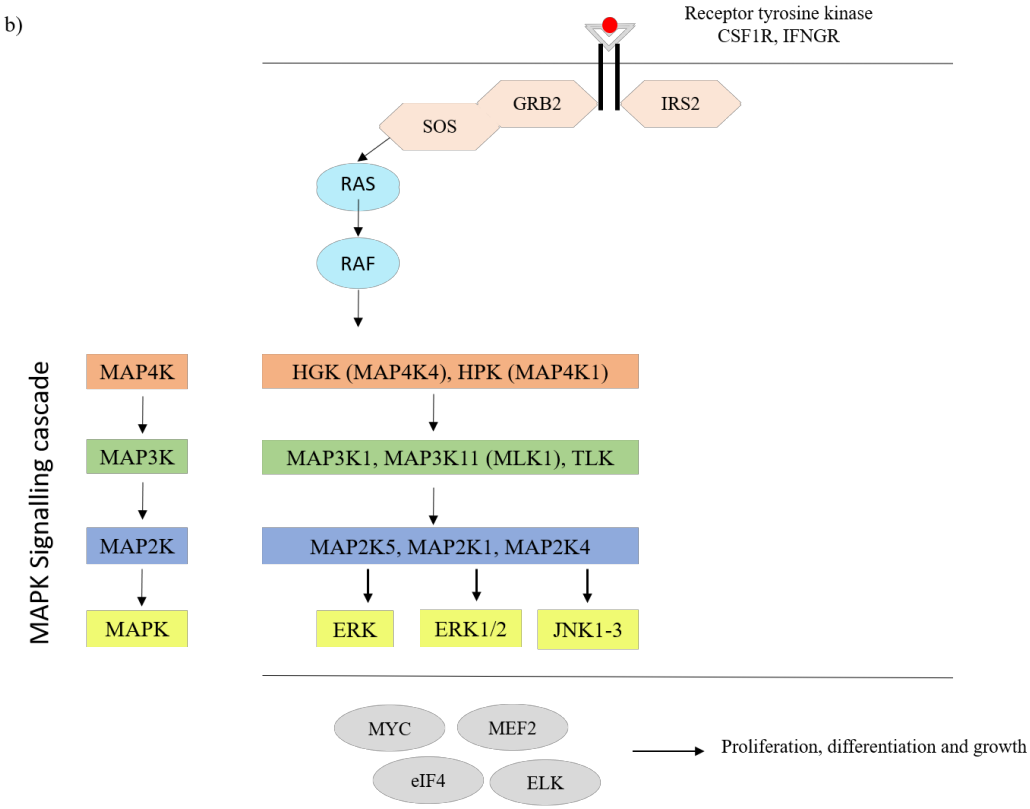


Figure 3b – Predominance of RTK mediated MAPK signalling in haematological malignancy revealed by the phospho immunopeptidome. Potential MAPK signalling cascade identified in cell lines THP-1, MV411.1 and AML patient1.

CSF1R is known to activate multiple signalling via the MAPK pathway which comprises of MAP4K/MAP3K/MAP2K signalling cascades eventually activating either ERK/JNK/p38 proteins. Evidence for the MAPK pathway was confirmed by detection of HGK (a MAP4K) along with MLK3 (a MAP3K). Evidence for downstream signalling for ERK pathway was evident from previously reported and well characterised phosphorylation sites identified in proteins ELK1 (S383), p90RSK (S221) and SMAD2 (T8). In contrast, in the HL60 cell line there was evidence for type 1 IFN-γ mediated signalling through IFN-γR and STAT2 of the JAK/STAT pathway along with mTOR signalling but no evidence for MAPK signalling as neither the adaptor proteins such as IRS2, SOS nor any proteins of the MAPK signalling cascade were detected.

In the AML patient who was cytogenetically normal, phosphosite analysis revealed an interplay between RhoGTPases mediated signalling involving G protein coupled receptor2 (GRK2), Raf, ARHGEF (Figure 3b). There was evidence of both ERK and mTOR signalling as proteins found downstream the two pathways i.e. p90RSK and p70RSK respectively were identified. Similar to that observed for the THP-1 cell line, enrichment network visualisation of phosphopeptide source proteins identified in the patient revealed pathways of chromatin modification and spliceosome machinery were affected by phosphorylation which might be driving oncogenesis (Figure 3c).

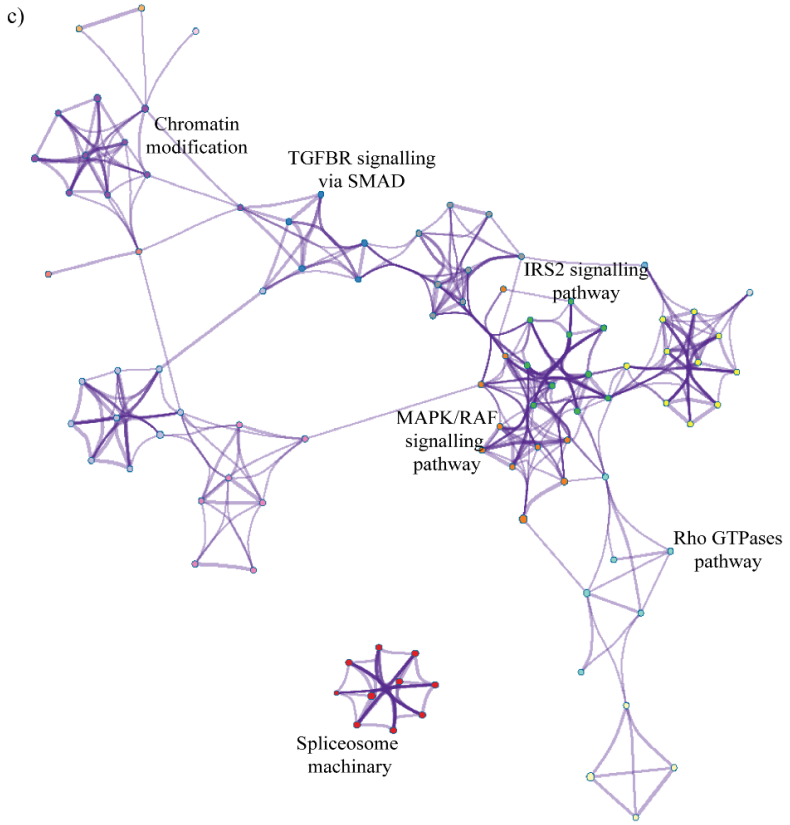


Figure 3c – Phosphorylation driven signalling controls key pathways involved in leukemogenesis in AML patient1. Enrichment network visualisation of phosphopeptide source proteins by Metascape (Zhou et al. 2019) depict key pathways such as chromatin modification, TGFR signalling and spliceosome machinery being controlled by phosphorylation. Colour code represents the identities of gene lists.

Interestingly, pathway analysis of proteins contributing to pTyr peptides revealed targets such as chemokine receptor (CCR), GRK and HIPK. An HLA bound peptide Y(Ph)LQSRYYRA corresponded to Y361 in HIPK protein was identified and phosphorylation at this site by either Src or HIPK itself is known to abrogate HIPK mediated TGFβ signalling, resulting in excessive

proliferation (Shang et al. 2013). Tyrosine phosphorylation seems to occur on proteins responsible for maintaining cellular homoeostasis. These include CBFA2T2, a transcriptional corepressor, NOT1, which is involved in mRNA degradation and miRNA-mediated repression or JKAMP protein which is involved in degradation of misfolded ER luminal proteins. Some other proteins which have a similar regulatory role include SCF and MXRA8 protein which are known to regulate mTOR and p38 kinase pathway. These pathways were found to be enriched in unmodified source proteins of the immunopeptidome across the HM samples.

4.3.4 The dimethylated peptides identified in the haematological malignancy

The HM samples were also interrogated for the presence of methylated peptides but since they are less abundant, only 320 peptides were identified across cell lines and clinical samples (Table 2) although in patient samples, peptides were identified only in AML patient1. Almost equal numbers of 162 monomethylated and 158 dimethylated peptides were identified (Supplementary file 5), wherein Arg methylation was dominant, constituting 52% and 68% in mono- and di-methyl peptides, respectively. In patients, the number of dimethylated peptides identified was twice of that identified in healthy sample, whilst the number of monomethylated peptides was same (Table 2). This increased occurrence of dimethylated in patients is indicative that dimethylation has oncogenic relevance. There was very little overlap amongst the monomethylated (Figure 4a) and dimethylated (Figure 4b) peptides identified across samples. The length distribution analysis of both mono- and dimethylated peptides revealed that 9mer was the dominant length (Figure 4c), with P1 and P4 the preferred site for monomethylation (Figure 4d). For dimethylation there was a clear preference for P3 followed by P4, P5 and P1 (Figure 4d). The motif of the 41 dimethylated nonamers was interpreted using Icelogo and was found to have Pro-Arg-Pro(PRP), Gly-Arg-Gly (GRG) and Pro-Arg-Gly (PRG) motif (Figure 4e), a conventional motif observed for dmArg peptides (Marino et al. 2017). The presence of the 3 substrate motifs is indicative of PRMTs activity as most PRMTs are known to methylate GRG motifs, except PRMT4 and PRMT5, which can also methylate proline-rich motifs. Pathway analysis revealed that the proteins were part of processes involved in RNA processing, gene silencing and translation.

Table 2 – Repertoire of methylated peptides identified across haematological malignancy.

Cell lines/clinical samples	Monomethylated	Dimethylated
Cell lines	98	78
Healthy BM	32	29
AML Patient1	37	63

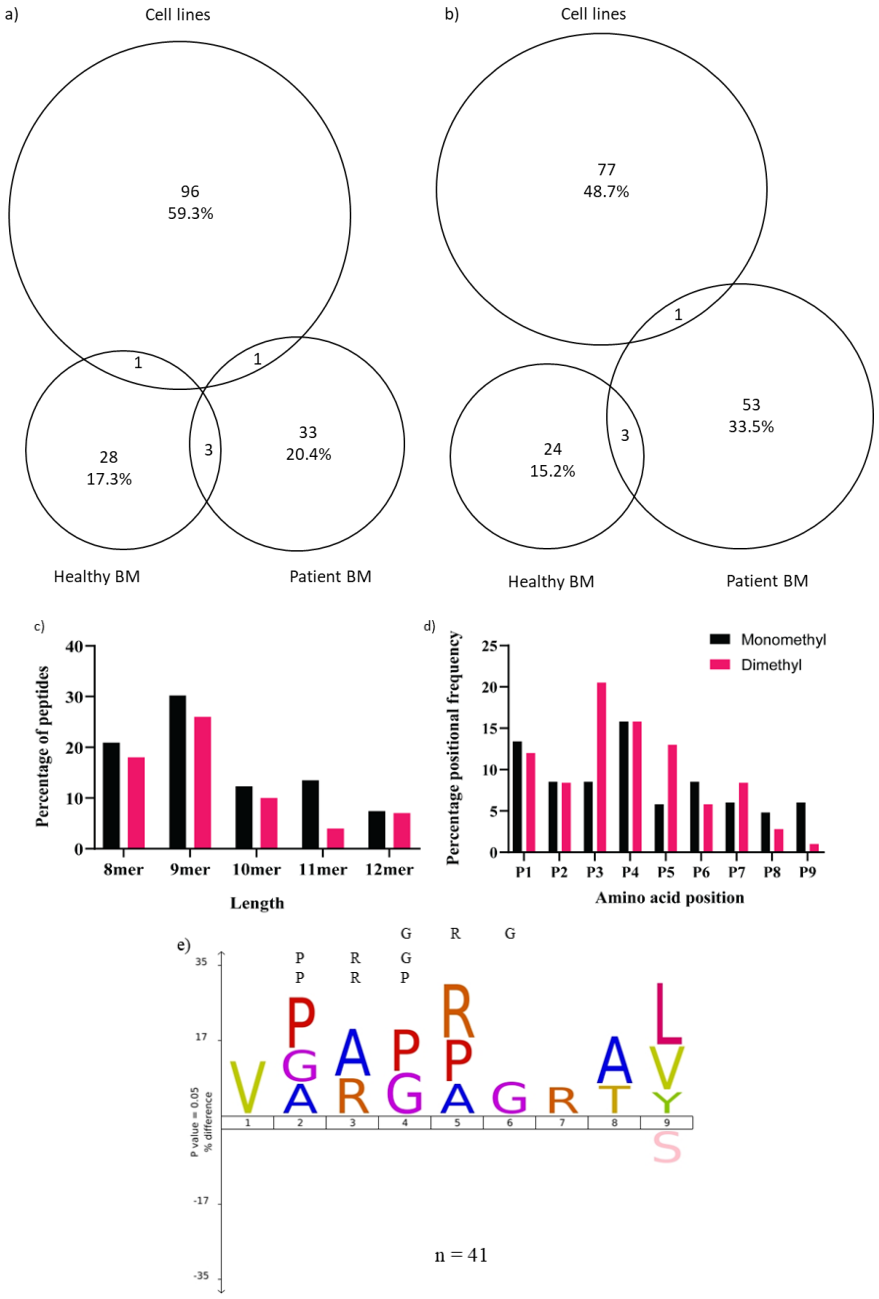


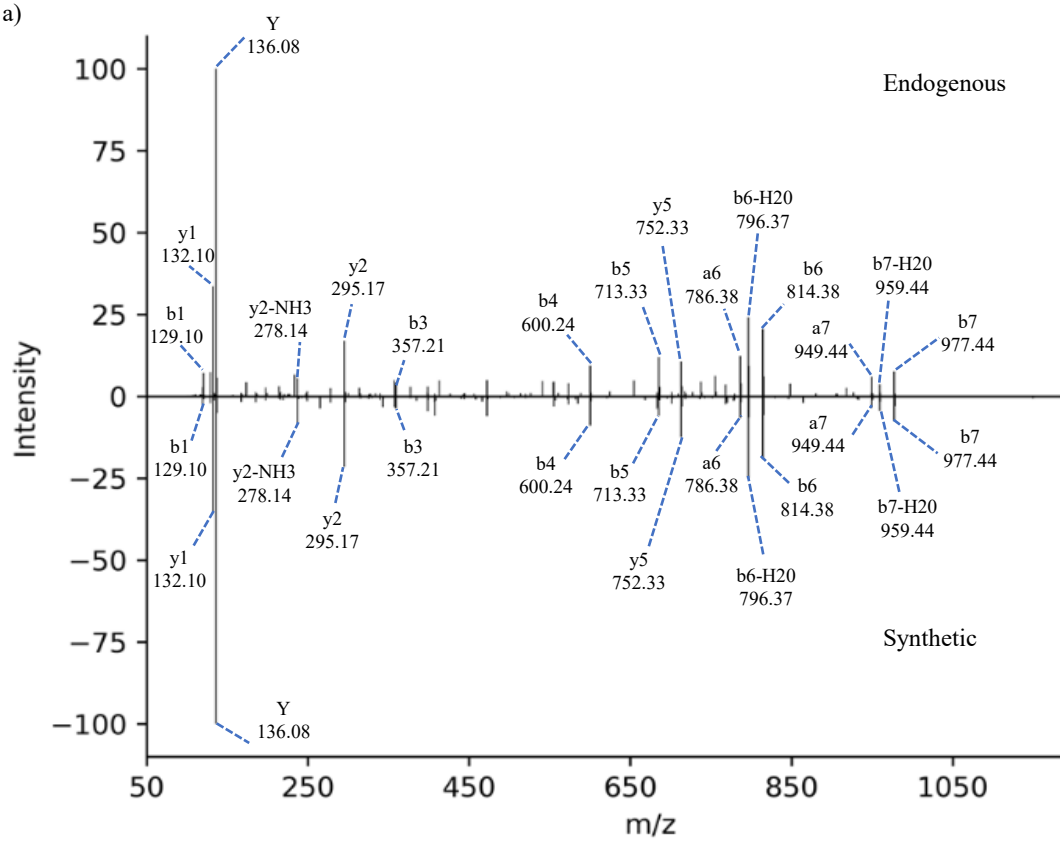
Figure 4 – Properties of mono- and dimethylated peptides identified in haematological malignancy. Very little overlap is seen amongst a) monomethylated and b) dimethylated peptides across cell lines, healthy and patient BM samples. c) Length distribution analysis reveal 9mers as the dominant length for both mono and dimethyl peptides. d) Positional analysis reveals dimethylation prefers P3 whilst monomethylation preferably occurs at P1 and P4. e) Enrichment of PRG/GRG/PRP motif in 9mer dimethylated peptides restricted to different HLA class I was interpreted by IceLogo.

4.3.5 HLA A*02:01 restricted PTM peptides selected for structural work

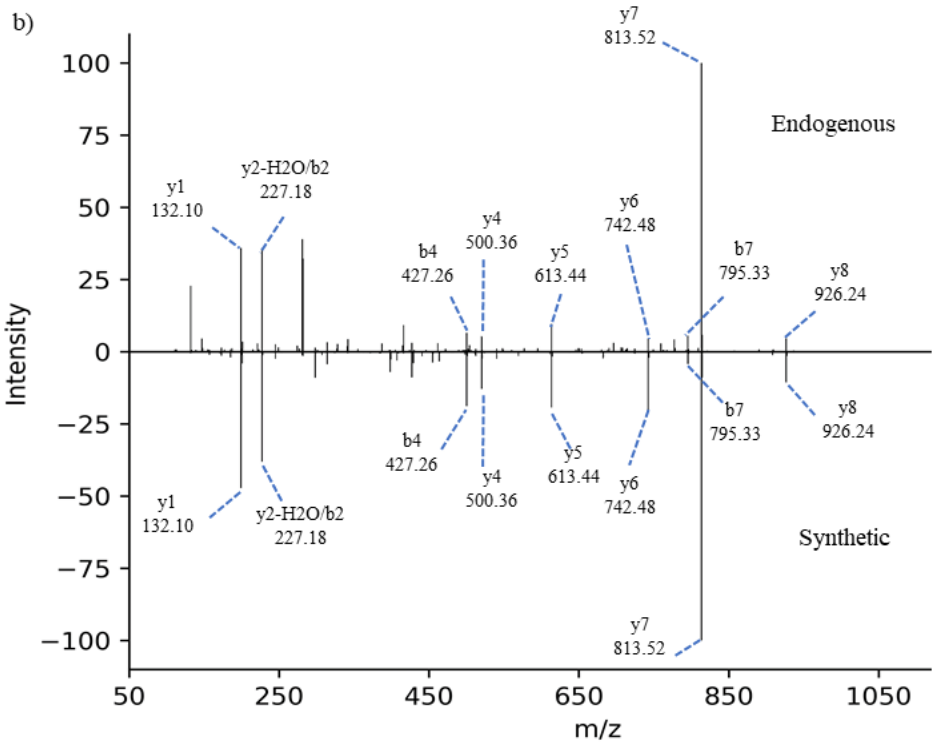
Out of 123 pTyr peptides identified, six HLA A*02:01 restricted peptides were selected and validated by comparing their LC-MS/MS characteristics to their synthetic counterparts. The synthetic and endogenous peptides which matched retention time, MS/MS fragmentation pattern wherein same series of b and y ions were generated were considered genuine. Out of the 6 peptides, KLDY(Ph)ITYL (Figure 5a) along with the remaining 3 peptides matched all the above criteria (Supplementary Figure 2a) and was taken forward with structural studies. This is because KLDY(Ph)ITYL is derived from Splicing factor A3 protein (SF3A) protein, which is known to be dysregulated in ~5-15% of leukaemia cases and is associated with poor disease prognosis (Ogawa 2014, Hou et al. 2016, Visconte et al. 2019) thereby making it clinically relevant. Interestingly, it was identified across the HM dataset, i.e. not only in THP-1 cell line but also in BM aspirate of AML patient1 and AML patient2. It was an 8mer and matched the kinase motif described above.

Similarly, out of the 158 dmArg peptides, five were selected. Two peptides (ILAEIAR(dm)IL, AALGR(dm)GSAL) were restricted to HLA A*02:01 whilst the other three peptides (DGR(dm)VFQVEY, GGFR(dm)GNSGGDY and MPNR(dm)GNYNQNF) were identified in the pan class I fraction of THP1. Validation of the 5 peptides was performed using synthetic peptides. Based on their similar retention time and MS/MS fragmentation pattern, nonamer, ILAEIAR(dm)IL (Figure 5b) along with remaining 4 peptides (Supplementary Figure 2b) were found to be genuine. Peptide ILAEIAR(dm)IL was taken forward for structural studies as it belongs to CPIN1 family of proteins, a CTA which has the highest expression reported in testis (Uhlén et al. 2015). The protein is also associated with several other cancers and has been identified as an unfavourable prognostic marker in head and neck cancer according to TCGA data present in human protein atlas (Uhlén et al. 2015). Hence, one may speculate that these peptides make excellent candidates for T-cell based immune therapy.

a)



Amino Acid	b	b-H20	b-NH3	b (2+)	y	y-NH3
K	101.11	129.10	112.08	65.05	980.44	963.41
L	86.10	242.19	225.16	121.59	867.35	850.33
D	88.04	357.21	340.19	179.11	752.33	735.30
Y(Ph)	216.04	600.24	583.22	300.62	509.30	492.27
I	86.10	713.33	696.30	357.16	396.21	379.19
T	74.06	814.38	797.35	407.69	295.17	278.14
Y	136.08	977.44	960.41	489.22	132.10	115.07
L	86.10					



Amino Acid	Immonium	b	b-H20	b-NH3	b (2+)	y	y-H2O	y-NH3	y(2+)
I	86.10	114.09	96.08	97.07	57.55				
L	86.10	227.18	209.17	210.15	114.09	926.64	908.39	909.43	463.80
A	44.05	298.22	280.20	281.19	149.61	813.52	795.33	796.49	407.26
E	102.06	427.26	409.25	410.23	214.13	742.48	724.28	725.46	371.74
I	86.10	540.33	522.33	523.31	500.67	613.44	595.22	596.41	307.22
A	44.05	611.38	593.37	594.35	429.19	500.36	482.34	483.33	250.68
R(dm)	157.13	795.33	777.32	778.48	398.18	429.32	411.31	412.29	215.14
I	86.10	908.39	890.41	891.47	454.80	245.19	227.18	228.16	123.09
L	86.10					132.10	114.09	115.07	66.55

Figure 5 – Similar MS/MS fragmentation spectra and sequence ladder of endogenous and synthetic peptide a) KLDY(Ph) and sequence ladder of b and y ions identified and b) ILAEIAR(dm)IL sequence ladder of b and y ions.

4.3.6 Phosphorylation stabilises the pHLA complex

To determine if the PTMs included in the study namely phosphorylation of Tyr and dimethylation of Arg, influenced the stability of the pHLA complex, thermal stability assays were performed. Recombinant HLA-A2 (expressed in E. *coli*) was refolded with β2M along with the native or modified peptide. The refolded pHLA complexes were purified by anion exchange and size exclusion chromatography. Both native and modified pHLA complexes refolded well and yielded similar concentration of 5mg/mL and their stability determined by

DSF in triplicate (n = 3). To test the stability of protein complexes, the temperature was gradually increased, and the Tm was measured as an increase in fluorescence intensity.

For the phosphopeptide, it was evident that phosphorylation of P4 Tyr resulted in increased thermostability of the complex, represented by a differential increase of 15°C between the Tm of native (43°C) and the PTM peptide (57°C) (Figure 6a). In contrast, dimethylation at P7 Arg did not contribute to pHLA stability as the Tm of native (53°C) and modified (55°C) peptides remain essentially unaffected (Figure 6b). This suggested that phosphorylation increases the stability of the HLA-A2 complex, whereas dimethylation had no such effect.

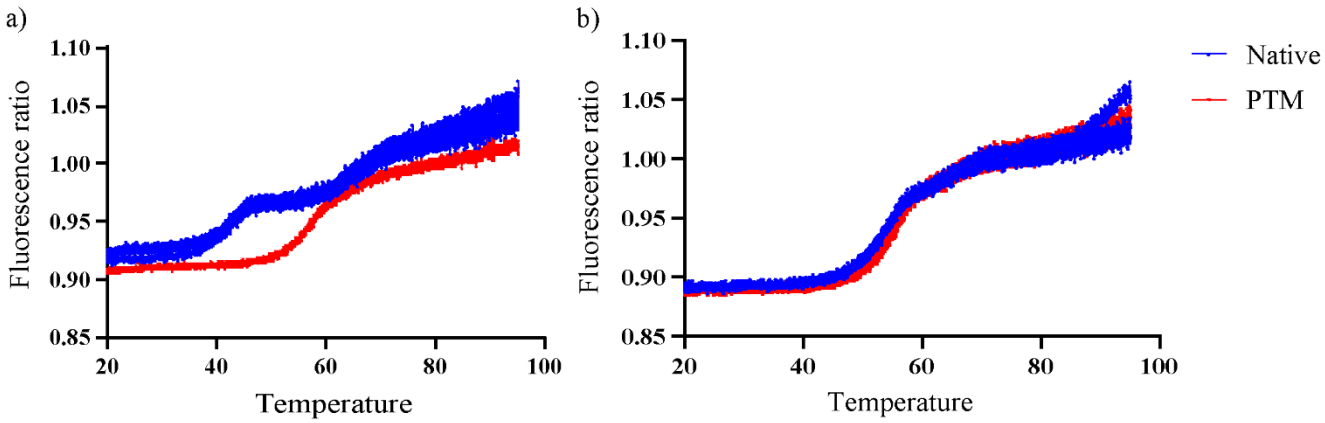


Figure 6 - Phosphorylation of Tyr residue enhances the stability of pHLA complex whilst dimethylation at Arg has no effect. Thermal denaturation assay for native and PTM pHLA complexes of a) HLA A*02:01-KLDYITYL and -KLDY(Ph)ITYL and b) HLA A*02:01-ILAEIARIL and ILAEIAR(dm)IL was performed by DSF. An increase in melting temperature indicates that more energy was required to denature the pHLA complex indicative of stability of the pHLA complex.

Table 3 – Crystallisation conditions for the HLA A*02:01 restricted native and PTM peptides

	SF3A protein Peptide - KLDY(Ph)ITYL	**CPIN1 protein** Peptide - ILAEIARIL	**CPIN1 protein** Peptide - ILAEIAR(dm)IL
Broad screen conditions	B4 of SG1 HT96 screen	C7 of SG1 HT96 screen	C7 of SG1 HT96 screen
Crystallisation conditions	30 % w/v PEG 5000 MME, 0.2 M$(NH_4)_2SO_4$ and 0.1 M MES (pH 6.5)	25% PEG3350, NaHEPES (pH7.5)	25% PEG3350, NaHEPES (pH7.5)
Resolution range	31.32 - 1.75 (1.813 - 1.75)	36.44 - 1.542 (1.598 - 1.542)	45.64 - 1.72 (1.782 - 1.72)
Space group	C 2 221	P 21 21 21	P 21 21 21
Unit cell	83.5104 94.7101 107.39 90 90 90	60.429 80.386 111.025 90 90 90	60.022 79.257 111.639 90 90 90
Total reflections	86078 (8521)	160004 (15329)	114052 (10853)
Unique reflections	43187 (4273)	80052 (7703)	57079 (5460)
Multiplicity	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)
Completeness %	99.71 (99.53)	99.69 (97.19)	99.58 (96.01)
Mean I/sigma(I)	9.86 (1.91)	11.23 (1.36)	10.12 (1.16)
Wilson B-factor	14.88	20.62	24.15
R-merge	0.06194 (0.446)	0.02518 (0.3656)	0.03109 (0.4472)
R-meas	0.0876 (0.6308)	0.03561 (0.517)	0.04397 (0.6324)
R-pim	0.06194 (0.446)	0.02518 (0.3656)	0.03109 (0.4472)
CC1/2	0.996 (0.625)	0.999 (0.721)	0.999 (0.634)
CC*	0.999 (0.877)	1 (0.915)	1 (0.881)
Reflections used in refinement	43101 (4254)	80018 (7672)	57049 (5433)
Reflections used for R-free	1993 (196)	2480 (239)	1764 (165)
R-work	0.1778 (0.2522)	0.1604 (0.2429)	0.1587 (0.2505)
R-free	0.2144 (0.2905)	0.1885 (0.2613)	0.1907 (0.2552)
CC(work)	0.962 (0.836)	0.972 (0.846)	0.973 (0.822)
CC(free)	0.950 (0.828)	0.970 (0.830)	0.965 (0.829)
Number of non-hydrogen atoms	3651	3824	3719
macromolecules	3270	3257	3238
Ligands	24	50	64
Solvent	357	517	417
Protein residues	384	387	388
RMS(bonds)	0.007	0.016	0.012
RMS(angles)	0.92	1.46	1.14
Ramachandran Angles			
Favoured	98.4%	98.15%	98.93%
Allowed	1.60%	1.85%	1.07%
Outliers	0	0	0
Rotamer outliers (%)	0.86	1.45	1.46
Clash score	3.57	4.94	3.71
Average B-factor	19.52	29.1	32.02
macromolecules	18.68	27	30.42
Ligands	35.06	47.75	49.17
Solvent	26.13	40.55	41.82
Number of TLS groups	14	17	18

4.3.7 Novel structure of HLA A*02:01 in complex with a Tyr-phosphorylated peptide

In order to understand the molecular basis for pTyr peptide binding to HLA-A2 structural studies were undertaken. HLA-A2 heavy chain and β2M was expressed and refolded in the presence of both native (KLDYITYL) and PTM (KLDY(Ph)ITYL) peptide. The pHLA complexes were purified and concentrated to 5mg/mL each. For crystallographic studies, crystal trays were set up for both native and modified pHLA complex with different crystallisation conditions using commercially available broad screen kits by the sitting drop method (Fazio et al. 2014).

Only KLDY(Ph) HLA complex was subsequently crystallised (Supplementary Figure 3a) whilst the native pHLA complex failed to crystallise. As per the thermal stability data, this might be attributed to less stable nature of pHLA complex of native peptide. The crystal resolved to a resolution of 1.8 Å, with clear electron density at the phosphorylation site at P4 (Supplementary Figure 3b). Data collection and structural resolution were performed by Dr. Jan Peterson of the Rossjohn Laboratory, Dept. of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University. Final PDB files of the pHLA was visualised on Pymol (version 2.3.2).

The solved HLA A*02:01-KLDY(Ph)ITYL is the first structure of HLA class I restricted pTyr peptide to be reported. The pHLA comprised of heavy α chain constituting of 3 domains namely α1, α2 and α3 along with non-covalently linked β2M (Figure 7a). The α1 and α2 domain form the peptide binding groove whilst the α3 domain was positioned alongside β2M. The peptide binding groove comprised of 6 pockets from A to F. The P2 and PΩ residues of a peptide act as anchor residues, that tether the peptide in the binding grooves whilst residues P1, P4 and P7 are solvent exposed and hence are available to interact with TCR. Residues in peptide (from P1 to PΩ) and HLA A*02:01 interact with one another across different pockets via multiple interaction such as hydrogen bonds, van der Waal (vdw) and salt bridges.

The 8mer peptide KLDY(Ph)ITYL bound to HLA-A2 in an extended conformation (Figure 7b) with the main anchor residues P2-Leu and P8-Leu residing completely in the peptide binding groove pockets B and F respectively (Figure 7b). In addition, P3-Asp was also pointed downward into the peptide binding groove whilst P1-Lys, P4-Tyr and P7-Tyr were solvent exposed and therefore represent potential TCR contact sites (Figure 7b). Hence the peptide

follows a binding pattern typical to most HLA-A2 restricted peptides. The N-terminal P1-Lys is strongly tethered within the cleft, with the residue forming hydrogen bonds with the side chains of Tyr7, Tyr159 and Tyr171 (Figure 7c), and stacking interactions with Met5, Glu63, Trp167, Thr163 and Lys66. The main chain of P2-Leu, a hydrophobic anchor residue forms a hydrogen bond with Tyr7, Glu63 and Lys66 respectively and stacking interactions with other residues of hydrophobic B pocket including Tyr99, Tyr159, Met45, Val67 and Phe9 of the A2 heavy chain. The P3-Asp main chain interacts with Tyr99, via hydrogen bond (Figure 7c). Peptide-HLA interactions are mentioned in Table 4.

The P4-pTyr moiety was projecting upwards thereby forming salt bridges with Lys66 and Arg65. This outward projection of pTyr allowed its O3P atom to interact with the P1-Lys residue directly by hydrogen bond (Supplementary Figure 4a). This type of interaction has previously been reported by Mohammed *et al.* (Mohammed et al. 2008) for peptide RTYS(p)GPMNKV (PDB ID 3BH9) (Supplementary Figure 4b) and may provide enhanced stability to pHLA complex. The phosphate group was prominently surface-exposed and may act as a candidate TCR binding site (Figure 7c).

Another interesting interaction observed for the KLDY(Ph) peptide is at P7, which forms hydrogen bonds with residues of both α1 and α2 chains with Trp147, Gln72 (Figure 7d) respectively. This bridging interaction might provide some level of stabilisation for the pHLA complex. The anchor residue P8-Leu is tethered to F pocket by two hydrogen bonds to Asp77 and Lys146 and main residues also forming salt bridges with Lys146, Leu84 and Asp77.

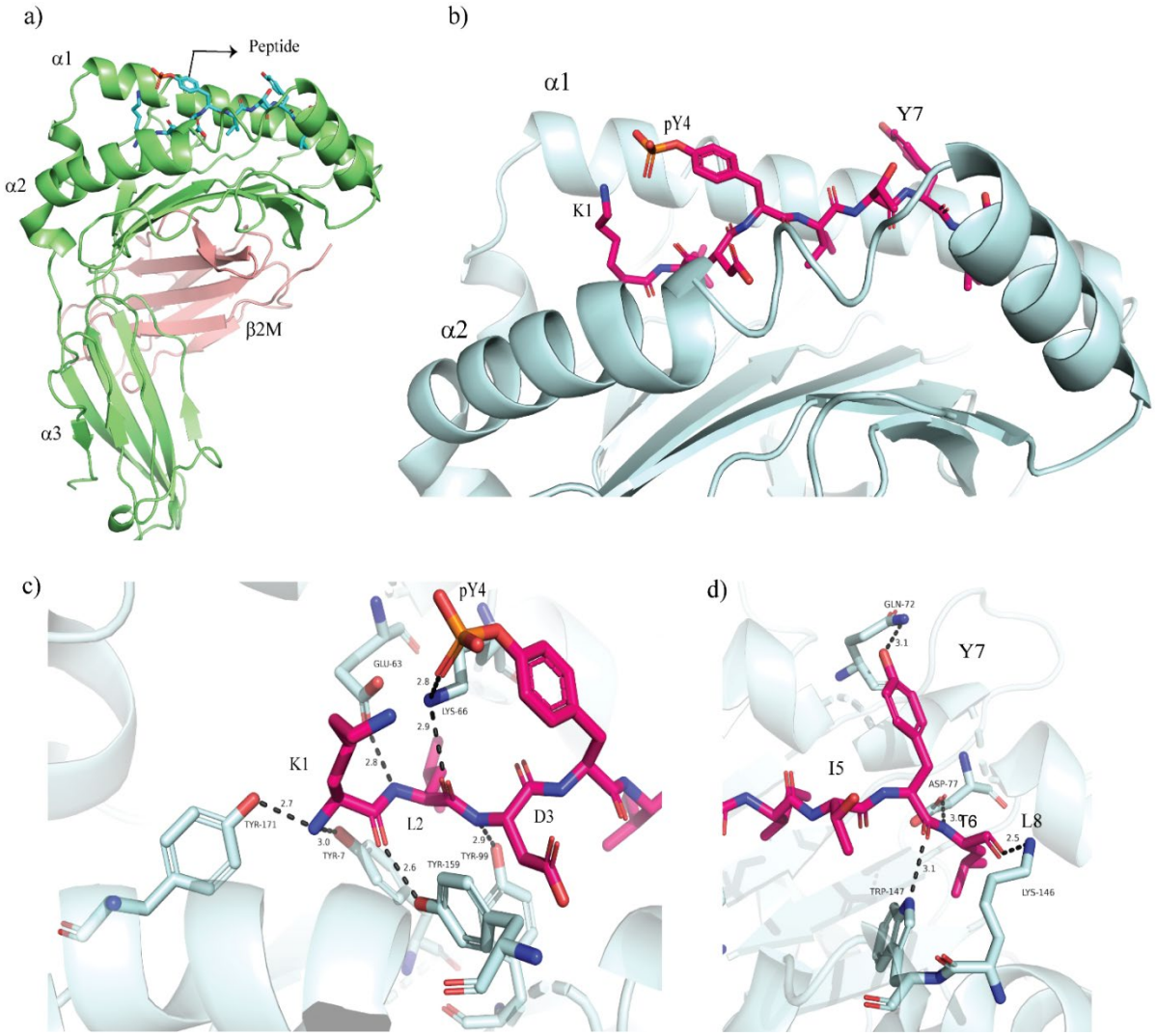


Figure 7 - Structural analysis of a Tyr phosphorylated peptide KLDY(Ph)ITYL presented by HLA A*02:01. a) Complete crystal structure of HLA A*02:01 with KLDY(Ph)ITYL peptide resolved to 1.8 A. b) A cartoon representation of the peptide binding groove with the phosphopeptide sitting in it. Partial cut away view of key hydrogen bond interactions occurring c) from residues P1 to P4 and d) P5 to P8 residues of peptide KLDY(Ph)ITYL peptide.

Table 4 - Table highlighting interactions between KLDY(Ph) peptide with HLA A*02:01

Peptide	Atom	HLA-A2	Type of interaction
Lys[1]	O	Tyr159 OH Met5, Glu63, Trp167, Lys66, Thr163	H- bond vdw
	N	Tyr7 OH, Tyr171 OH	H- bond
Leu[2]	N	Glu63 O ε1, Tyr7 OH	H- bond
	O	Lys66 N ζ Tyr99, Tyr159, Met45, Val67, Phe9	H- bond vdw
Asp[3]	N	Tyr99 OH His70 Nε2, Tyr159 OH	H- bond vdw
pTyr[4]	O3P	Lys66 N ζ	Salt bridge
	O2P	Arg65 NH2	Salt bridge
Ileu[5]		Thr73, Arg97, His114, His70, Arg97	vdw
Thr[6]		Thr73, Trp147, Val152	vdw
Tyr[7]	O	Trp147 Nε1	H- bond
	OH	Gln72 Nε2 Thr73, Val76, Asp77	H- bond vdw
Leu[8]	O	Lys146 Nζ, Tyr84, Leu81, Tyr123, Tyr116, Trp147, Thr143 Asp77 Oδ1	H- bond vdw
	N		H- bond

4.3.8 Novel structure of HLA A*02:01 in complex with a dimethylated peptide and its native counterpart

To determine how dimethylation influences peptide interactions with HLA-A2, the HLA heavy chain and β2M was expressed and refolded in the presence of the both native (ILAEIARIL) and PTM peptide (ILAEIAR(dm)IL). Using broad screens, a few conditions produced small crystals by sitting drop method (Supplementary Figure 5a). Hence manual hand trays were set up for two proteins with a few shortlisted conditions and proteins were crystallised by the hanging drop method. Both pHLA complexes were crystallised in the same conditions (Table 3), and their structures were solved to same planarity P212121 and refined to a resolution of 1.8 Å with clear electron density at the modified site (Supplementary Figure 5b).

The solved HLA A*02:01-ILAEIAR(dm)IL is the first structure of HLA class I restricted dmArg peptide to be reported. Similar to the pTyr structure, the HLA A*02:01 complex for both the peptides comprised of heavy α chain with α1 and α2 domain form the peptide binding

groove formed whilst the α3 domain was positioned alongside β2M (Figure 8a). The peptide binding groove comprised of 6 pockets from A to F. The P2 and P9 residues of a peptides act as anchor residues, that tether the peptide in the binding grooves whilst residues P1, P4, P5 and P7 are solvent exposed and hence are available to interact with TCR. Residues in peptide (from P1 to P9) and HLA A*02:01 interact with one another across different pockets via multiple interaction such as hydrogen bonds, van der Waal (vdw) and salt bridges.

In order to identify the effect dmArg might have on the pHLA complex, the modified peptide was compared against the native pHLA complex. Superimposition of the two revealed an overall similar conformation with a root mean square deviation (r.m.s.d.) of 0.20 Å (Figure 8a). Both the native and modified peptide adopted an extended conformation with a central bulge at residues P4 and P5 (Figure 8b). The P1-Ileu, P4-Glu, P5-Ileu and P7-Arg/dmArg solvent exposed and therefore represent potential TCR contact sites (Figure 8c and 8d) whilst P2-Leu and P9-Leu residing completely in the peptide binding groove B and F pocket respectively. In addition, P3-Ala and P5-Ileu are also pointed downward into the peptide binding groove thereby leaving residues.

Both the peptides followed a binding pattern typical to most A2 restricted peptides. The N-terminal end of P1-Leu of both the peptides was strongly tethered within the cleft with main residues forming hydrogen bonds and with residues Tyr7, Tyr171 and Tyr159 (Figure 8e) and stacking interactions with Met5, Glu63, Tyr59. Thr163, Trp167 and Lys66. The main chain of P2-Leu, a hydrophobic anchor residue forms a hydrogen bond with Glu63 and Lys66 (Figure 8e) and salt bridges with residues of hydrophobic B pocket such as Tyr7, Tyr159, Tyr99, Val67, Met45, Phe9. The P3-Ala interacted with Tyr99 in both the peptides (Figure 8e) and side chain of P4-Glu forming salt bridges with Arg65 and stacking interactions with Lys66.
The P7-Arg in both instances points upwards and interacts with Trp147, Thr73, Val152, Ala150 but unlike the pTyr moiety that directly formed salt bridges with HLA residues, the dmArg group does not interact with the HLA. The only distinguishing interaction between the 2 peptides is P9-Leu which forms hydrogen bond with Thr143 (Figure 8f) in modified peptide, whilst the P9 of the native peptide interacts with Lys146 via hydrogen bond (Table 5a and 5b).

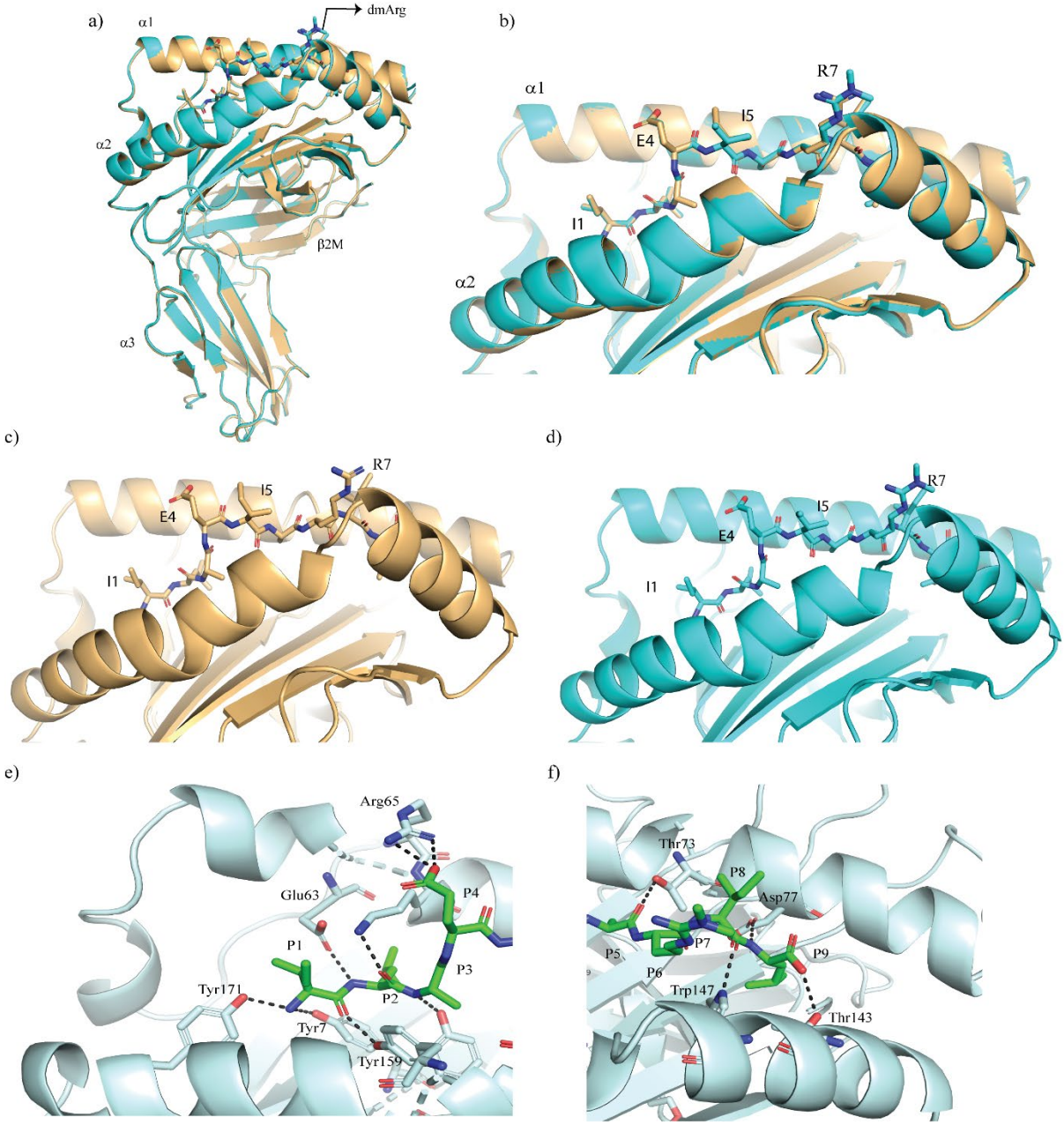


Figure 8 Structures of native ILAEIARIL and dimethylated ILAEIAR(dm)IL peptide presented by HLA A*02:01. a) Superimposed image of native and modified peptide b) The native peptide sitting in the peptide binding groove of HLA A*02:01 c) The modified peptide sitting in the peptide binding groove of HLA A*02:01 d) Network of direct polar contacts occurring between residues P1-P4 and e) P5-P9 of the peptide and HLA A*02:01 allele. The peptide interacts with residues on both the α1 and α2 of the HLA.

Table 5a – Table highlighting interactions between native peptide (ILAEIARIL) with HLA A*02:01

Peptide	Atom	HLA-A2	Type of interaction
Ile[1]	N	Tyr7 OH, Tyr171 OH	H- bond
	O	Tyr 159 OH	H- bond
		Met5, Glu63, Tyr59, Thr163, Trp167, Lys66	vdw
Leu[2]	N	Glu63 O ε1	H- bond
	O	Lys66 N ζ	H- bond
		Tyr7, Tyr159, Tyr99, Val67, Met45, Phe9	vdw
Ala[3]	N	Tyr99 OH	H- bond
		Tyr159, Tyr99, His70, Lys66	vdw
Glu[4]	Oε1	Arg65 NH2	Salt bridge
	Oε2	Arg65 NH1, Arg65 NH2	Salt bridge
		Lys66	vdw
Ileu[5]		Gln155	vdw
Ala[6]	O	Thr73 Oγ1	H- bond
Arg[7]	O	Trp147, Thr73, Val152, Ala150	vdw
Ileu[8]	O	Trp147 Nε1	H- bond
		Thr73, Asp77, Val76	vdw
Leu[9]	N	Asp 77 Oδ1	H- bond
	OXT	Lys 146 Nζ	H- bond
		Trp147, Leu81, Tyr123, Thr143, Tyr84	vdw

Table 5b - Table highlighting interactions between modified ILAEIAR(dm)IL peptide and residues in HLA A*02:01

Peptide	Atom	HLA-A2	Type of interaction
Ile[1]	N	Tyr7 OH, Tyr171 OH	H- bond
	O	Tyr 159 OH Trp167, Met5, Glu63, Thr163, Tyr59, Lys66	H- bond vdw
Leu[2]	N	Glu63 O ε1	H- bond
	O	Lys66 N ζ Tyr7, Tyr159, Tyr99, Val67, Met45, Phe9	H- bond vdw
Ala[3]	N	Tyr99 OH His70, Lys66, Tyr159	H- bond vdw
Glu[4]	Oε1	Arg65 NH2	Salt bridge
	Oε2	Arg65 NH1, Arg65 NH2 Lys66	Salt bridge vdw
Ileu[5]		Gln155	vdw
Ala[6]	O	Thr73 Oγ1	H- bond
Arg[7](dm)		Thr73, Trp147, Val152, Ala150	vdw
Ileu[8]	O	Trp147 Nε1 Thr73, Asp77, Val76	H- bond vdw
Leu[9]	N	Asp 77 Oδ1	H- bond
	OXT	Thr 143 Oγ1 Tyr123, Trp147, Tyr116, Leu81, Tyr84	H- bond vdw

4.4 DISCUSSION

PTMs, such as phosphorylation and methylation are enzymatically added to a wide variety of proteins and are considered a mechanism of protein diversification (Zhao et al. 2009) controlling solubility, activation, degradation or protein–protein interactions (Virág et al. 2020). Phosphorylation controls key processes of cell proliferation and signalling (Manning et al. 2002) and is the dominant mechanism involved in oncogenic signalling. Advancement in MS has made it possible for sensitive and specific detection of low abundance PTM peptides. Optimisation of MS methods and in depth immunopeptidomics can help in identification of these low abundance PTM peptides presented by HLA class I alleles. This is crucial previous studies have shown that PTM antigens are a source of neoantigens and can be exploited as attractive targets for cancer immunotherapy including vaccination. The immunopeptidome can also provide insight into the systemic dysregulation occurring in leukaemia which can also be utilised to make decisions regarding drug therapies targeting dysregulated signalling pathways.

The study highlights that our optimised workflow of peptide elution followed by RP-HPLC provides in-depth coverage of the immunopeptide not only for cell lines wherein the sample is in abundance but also for precious clinical material wherein the sample size is often limitation. With our approach we identified largest number of phosphopeptides ever reported in an immunopeptidomics study and first study to report on tyrosine phosphorylated HLA restricted peptides. Compared to previous studies of Cobbold *et al.* (Cobbold et al. 2013) and Mommen et al (Mommen et al. 2014) wherein they identified 95 and 60 peptides respectively. All peptides identified in the studies were either pSer or pThr. Another key finding of the study is that the phospho immunopeptidome reflects the oncogenic signature in cell lines and patient samples, which suggests that immune surveillance for HM should guide immunotherapy.

In the present study majority of the phosphopeptides identified were Ser and Thr followed by Tyr. The data was unique in several ways, first, there was a distinct length preference amongst phosphopeptides wherein pSer and pThr peptides preferred 9mers and 10mers phosphorylated at P4 and P5 whilst pTyr peptides preferred smaller 8mers phosphorylated at P3. Second, for pTyr peptides, although the motif was different from the pSer or pThr peptides, the preference for Lys at P1 and Pro after pTyr was evident in octamers. The bias became more evident when compared to motif of unmodified nonamers in the immunopeptidomics dataset.

The present study for the first time elucidates novel structures for PTM peptides including Tyr phosphorylated restricted by HLA-A2. The KLDY(Ph) pHLA complex crystallised and adopted a conventional extended conformation in the peptide binding groove forming hydrogen, van der Waal and salt bridge interactions with both the α1 and α2 chains of the HLA. The pTyr residue because of its large size sticks out and two of its oxygen residues (O3P and O2P) interact with residues of HLA. This is in line with previously reported interactions of pSer moiety (Petersen et al. 2009). Hence, similar to previously reported HLA A*02:01 restricted pSer peptides, the pTyr moiety can also influence pHLA-TCR engagement as it introduces an electronegative charge around the peptide binding groove region of HLA-A2. On the other hand the native KLDY peptide was predicted to be a strong binder (peptides with rank threshold of less than 0.5) for HLA A*02:01 by NetMHC 4.0 and pHLA complex refolded similarly to the modified KLDY(Ph) pHLA complex but crystallisation of the native pHLA complex was not achieved. This could be attributed to mainly two factors, a) the appropriate crystallisation condition was not identified or b) the pHLA complex was not stable enough to form crystals. The latter being more probable cause as substantiated by thermal stability assays

wherein a difference of 15 ºC in Tm was observed between the native and modified pHLA complexes.

Mixed lineage leukaemia (MLL) which is found in 10% of AML cases involves reciprocal translocation of chromosome 11q with different chromosomes such as chromosome 9 and 4. This results in fusion of MLL or Lys (K)-specific methyl transferase 2A (KMT2A) gene and leads to the expression of fusion genes termed as MLL-AF9 or MLL-AF4 respectively. THP-1 and MV411.1 reportedly possess MLL-AF9 and MLL-AF4, gene fusions, respectively. In THP-1 there was evidence of dysregulation in MAPK signalling cascade. This is line with previous studies on THP-1 cell line (van Alphen et al. 2020). In contrast, in MV411.1 cell line there was an indication of RTK mediated signalling as evident from presence of CSF1R and its downstream adaptor protein STAT3. Together the RTK/STAT pathway drive the ERK/JNK pathway that leads to cell proliferation and differentiation. Similarity in Ras/Raf/MAPK signalling pathway between THP-1 and MV411.1 also highlights that despite possessing different chromosomal aberrations similar downstream signalling might contribute to oncogenesis (Prange et al. 2017). The study thus highlights that HLA class I is capable of capturing transient events such as phosphorylation occurring inside the cell. This data can be used to determine small molecule inhibitor therapy targeting the above-mentioned pathways and receptors. At the same time, the drawback of this study although is that HLA class I presents peptides derived from proteins present in abundance and might not always accurately depict what is occurring in the cell.

To date, there is no study investigating structural aspects of presentation of methylated peptides. The study for the first time unveils structural insights of an HLA-A2-restricted nonameric peptide ILAEIAR(dm)IL, which contains an asymmetric dimethyl group on Arg at P7. Analysis of the structural data conclusively showed that the conformation of the modified peptide closely matched that of the unmodified counterpart. Unlike the phosphate moiety which played a role in stabilising the pHLA complex, the dmArg does not seem to perform any such function as evident from the thermostability data reporting similar Tm of both the native and modified peptides. The dmArg moiety is solvent accessible but does not interact either with residues of the HLA or residues of the peptide itself. One might speculate that the methyl group might play a role in TCR recognition. This is substantiated by Jarmalavicius *et al* (Jarmalavicius et al. 2010) wherein only the monomethylated peptides was recognised by

CD8+ T cells in all nine patients as evident from IFN-γ release whilst the dimethylated peptide elicited CD8+ T cell response in one patient. Thereby suggesting that methylation has a major influence on antigenicity of the peptides in cancer as it may influence TCR recognition of the peptides.

In conclusion, the study shows the not only diverse repertoire of PTM peptides presented by different alleles of HLA class I but also provides a deeper insight into factors influencing their restriction such as motifs and length. The study also showcases how the PTM-immunopeptidome can be representative of malignant transformation occurring within cells at the time of oncogenesis. Most importantly, the study successfully provides structural insight for the first time into the presentation of two different yet relevant PTMs by pan Caucasian HLA-A2 and the stabilising role a PTM can have on HLA. Together, this clinically relevant information can have much wider implications for cancer treatment. This includes utilising the comprehensive inventory of PTM peptides for multi peptide vaccines which can be either therapeutic or prevention in nature. For patients with suppressed HLA, therapy can include small molecule immunotherapy aiming to increase HLA expression or targeting dysregulated pathways. Hence data from this study provides a scientific background to construct rational therapies which will help to improve vaccination and cell therapy approaches in haematological malignancy.

REFRENCES

Abelin, J. G., et al. (2017). "Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction." Immunity **46**(2): 315-326.

Abelin, J. G., et al. (2015). "Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry." Nat Protoc **10**(9): 1308-1318.

Adams, P. D., et al. (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallographica Section D **66**(2): 213-221.

Bassani-Sternberg, M., et al. (2016). "Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry." Nat Commun **7**: 13404.

Belov, A. A. and M. Mohammadi (2012). "Grb2, a double-edged sword of receptor tyrosine kinase signaling." Sci Signal **5**(249): pe49.

Birch, N. W. and A. Shilatifard (2020). "The role of histone modifications in leukemogenesis." Journal of Biosciences **45**(1): 6.

Brodsky, F. M. and P. Parham (1982). "Monomorphic anti-HLA-A,B,C monoclonal antibodies detecting molecular subunits and combinatorial determinants." J Immunol **128**(1): 129-135.

Chase, A. and N. C. Cross (2006). "Signal transduction therapy in haematological malignancies: identification and targeting of tyrosine kinases." Clin Sci (Lond) **111**(4): 233-249.

Clements, C. S., et al. (2002). "The production, purification and crystallization of a soluble heterodimeric form of a highly selected T-cell receptor in its unliganded and liganded state." Acta Crystallographica Section D **58**(12): 2131-2134.

Cobbold, M., et al. (2013). "MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia." Sci Transl Med **5**(203): 203ra125.

Colaert, N., et al. (2009). "Improved visualization of protein consensus sequences by iceLogo." Nat Methods **6**(11): 786-787.

Cole, D. K. (2015). "The ultimate mix and match: making sense of HLA alleles and peptide repertoires." Immunol Cell Biol **93**(6): 515-516.

Collaborative (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallographica Section D **50**(5): 760-763.

Depontieu, F. R., et al. (2009). "Identification of tumor-associated, MHC class II-restricted phosphopeptides as targets for immunotherapy." Proc Natl Acad Sci U S A **106**(29): 12073-12078.

Drake, J. M., et al. (2012). "Oncogene-specific activation of tyrosine kinase networks during prostate cancer progression." Proceedings of the National Academy of Sciences **109**(5): 1643-1648.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallographica Section D **60**(12 Part 1): 2126-2132.

Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallographica Section D **62**(1): 72-82.

Fazio, V. J., et al. (2014). "A drunken search in crystallization space." Acta Crystallogr F Struct Biol Commun **70**(Pt 10): 1303-1311.

Garboczi, D. N., et al. (1996). "Structure of the complex between human T-cell receptor, viral peptide and HLA-A2." Nature **384**(6605): 134-141.

Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.

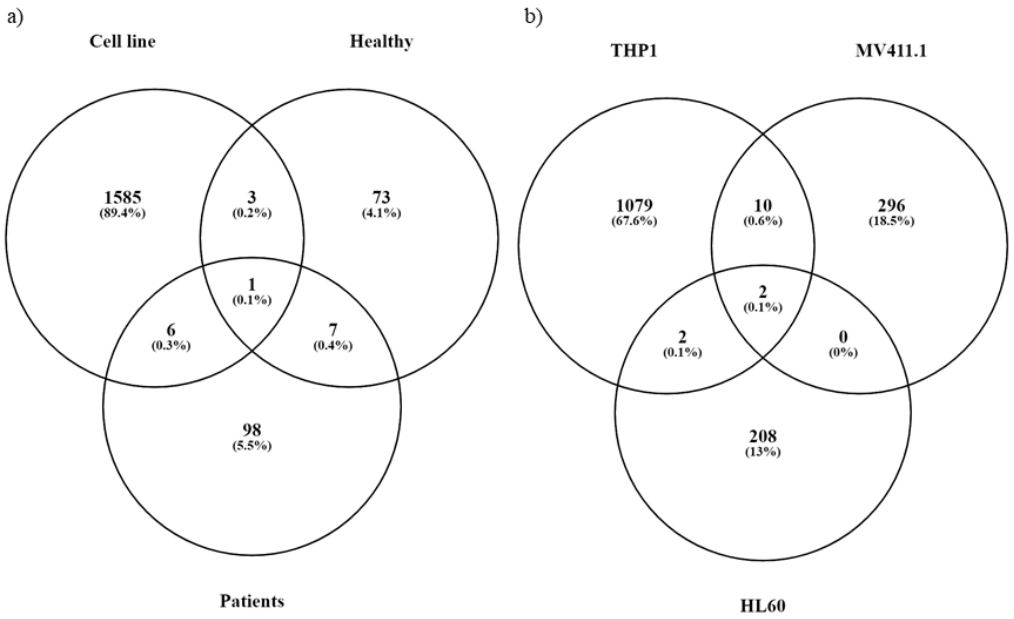
He, X., et al. (2019). "PRMT1-mediated FLT3 arginine methylation promotes maintenance of FLT3-ITD+ acute myeloid leukemia." Blood **134**(6): 548-560.

Hornbeck, P. V., et al. (2012). "PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse." Nucleic Acids Res **40**(Database issue): D261-270.

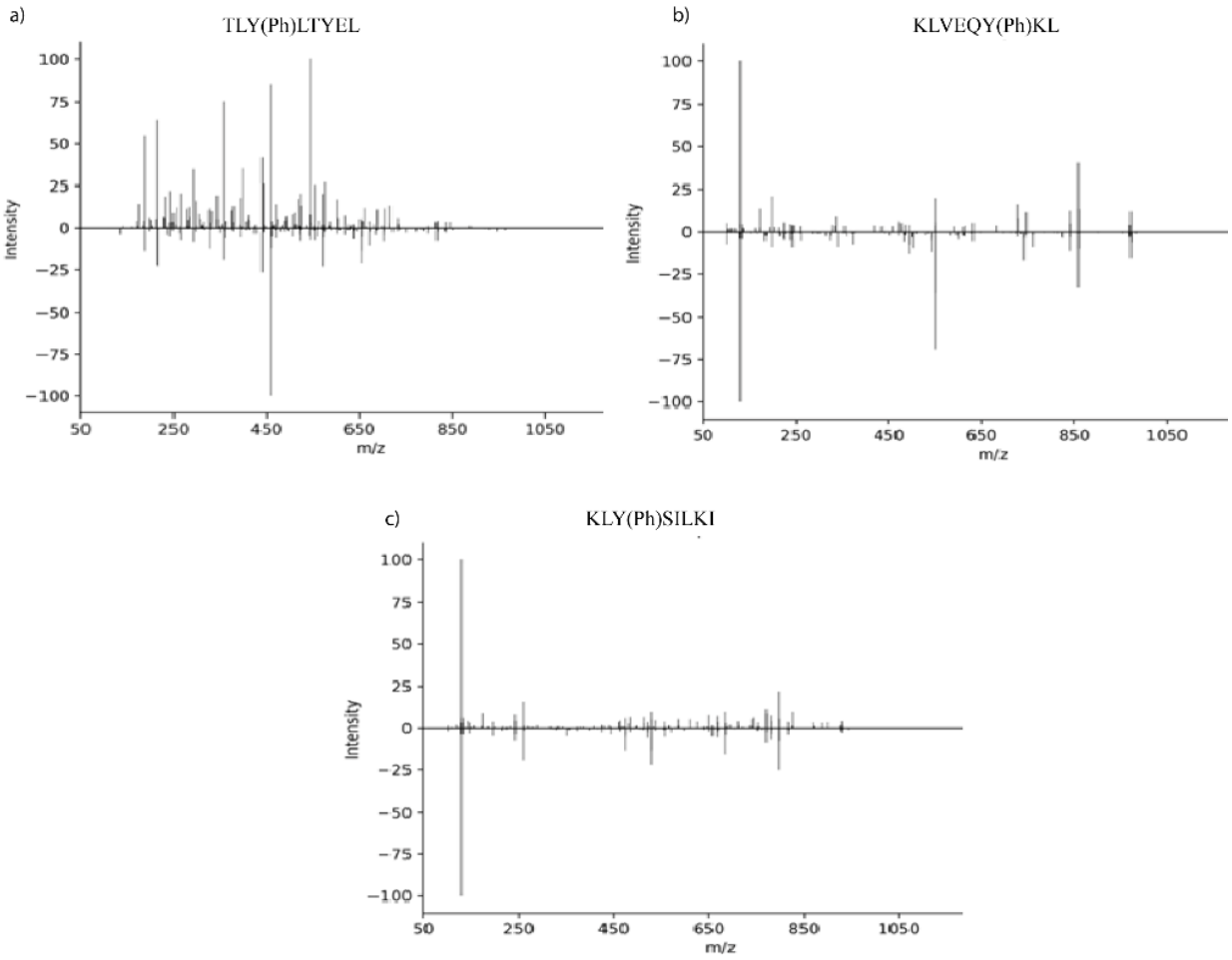
Hou, H. A., et al. (2016). "Splicing factor mutations predict poor prognosis in patients with de novo acute myeloid leukemia." Oncotarget **7**(8): 9084-9101.
Jarmalavicius, S., et al. (2010). "Differential arginine methylation of the G-protein pathway suppressor GPS-2 recognized by tumor-specific T cells in melanoma." Faseb j **24**(3): 937-946.
Kabsch, W. (2010). "XDS." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.
Malaker, S. A., et al. (2017). "Identification of Glycopeptides as Posttranslationally Modified Neoantigens in Leukemia." Cancer Immunol Res **5**(5): 376-384.
Mann, M. and O. N. Jensen (2003). "Proteomic analysis of post-translational modifications." Nature Biotechnology **21**(3): 255-261.
Manning, G., et al. (2002). "Evolution of protein kinase signaling from yeast to man." Trends Biochem Sci **27**(10): 514-520.
Marino, F., et al. (2017). "Arginine (Di)methylated Human Leukocyte Antigen Class I Peptides Are Favorably Presented by HLA-B*07." Journal of Proteome Research **16**(1): 34-44.
McCoy, A. J., et al. (2007). "Phaser crystallographic software." Journal of Applied Crystallography **40**(4): 658-674.
Mei, S., et al. (2020). "Immunopeptidomic analysis reveals that deamidated HLA-bound peptides arise predominantly from deglycosylated precursors." Molecular & Cellular Proteomics: mcp.RA119.001846.
Mikhail, F. M., et al. (2006). "Normal and transforming functions of RUNX1: A perspective." Journal of Cellular Physiology **207**(3): 582-593.
Mohammed, F., et al. (2008). "Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self." Nat Immunol **9**(11): 1236-1243.
Mohammed, F., et al. (2017). "The antigenic identity of human class I MHC phosphopeptides is critically dependent upon phosphorylation status." Oncotarget **8**(33): 54160-54172.
Mommen, G. P. M., et al. (2014). "Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD)." Proceedings of the National Academy of Sciences **111**(12): 4507-4512.
Niesen, F. H., et al. (2007). "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." Nature Protocols **2**(9): 2212-2221.
Ogawa, S. (2014). "Splicing factor mutations in AML." Blood **123**(21): 3216-3217.
Olsen, J. V. and M. Mann (2013). "Status of large-scale analysis of post-translational modifications by mass spectrometry." Mol Cell Proteomics **12**(12): 3444-3452.
Ouspenskaia, T., et al. (2020). "Thousands of novel unannotated proteins expand the MHC I immunopeptidome in cancer." bioRxiv: 2020.2002.2012.945840.
Pandey, K., et al. (2020). "In-depth mining of the immunopeptidome of an acute myeloid leukemia cell line using complementary ligand enrichment and data acquisition strategies." Mol Immunol **123**: 7-17.
Parham, P. and F. M. Brodsky (1981). "Partial purification and some properties of BB7.2. A cytotoxic monoclonal antibody with specificity for HLA-A2 and a variant of HLA-A28." Hum Immunol **3**(4): 277-299.
Petersen, J., et al. (2009). "Post-translationally modified T cell epitopes: immune recognition and immunotherapy." Journal of Molecular Medicine **87**(11): 1045.
Prange, K. H. M., et al. (2017). "MLL-AF9 and MLL-AF4 oncofusion proteins bind a distinct enhancer repertoire and target the RUNX1 program in 11q23 acute myeloid leukemia." Oncogene **36**(23): 3346-3356.
Robinson, J., et al. (2020). "IPD-IMGT/HLA Database." Nucleic Acids Res **48**(D1): D948-d955.
Shang, Y., et al. (2013). "Transcriptional corepressors HIPK1 and HIPK2 control angiogenesis via TGF-β-TAK1-dependent mechanism." PLoS Biol **11**(4): e1001527.

Uhlén, M., et al. (2015). "Tissue-based map of the human proteome." Science **347**(6220): 1260419.
van Alphen, C., et al. (2020). "Phosphotyrosine-based Phosphoproteomics for Target Identification and Drug Response Prediction in AML Cell Lines." Mol Cell Proteomics **19**(5): 884-899.
Virág, D., et al. (2020). "Current Trends in the Analysis of Post-translational Modifications." Chromatographia **83**(1): 1-10.
Visconte, V., et al. (2019). "Mutations in Splicing Factor Genes in Myeloid Malignancies: Significance and Impact on Clinical Features." Cancers (Basel) **11**(12).
Wang, F., et al. (2020). "Integrated transcriptomic and epigenetic data analysis identifiesaberrant expression of genes in acute myeloid leukemia with MLL-AF9 translocation." Mol Med Rep **21**(2): 883-893.
Wingelhofer, B. and T. C. P. Somervaille (2019). "Emerging Epigenetic Therapeutic Targets in Acute Myeloid Leukemia." Frontiers in Oncology **9**(850).
Yague, J., et al. (2000). "A post-translational modification of nuclear proteins, N(G),N(G)-dimethyl-Arg, found in a natural HLA class I peptide ligand." Protein Sci **9**(11): 2210-2217.
Zarling, A. L., et al. (2000). "Phosphorylated Peptides Are Naturally Processed and Presented by Major Histocompatibility Complex Class I Molecules in Vivo." J Exp Med **192**(12): 1755-1762.
Zarling, A. L., et al. (2014). "MHC-restricted phosphopeptides derived from Insulin receptor substrate-2 and CDC25b offer broad-based immunotherapeutic agents for cancer." Cancer Res **74**(23): 6784-6795.
Zarling, A. L., et al. (2006). "Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy." Proc Natl Acad Sci U S A **103**(40): 14889-14894.
Zhang, J., et al. (2012). "PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification." Mol Cell Proteomics **11**(4): M111.010587.
Zhao, Y. and O. N. Jensen (2009). "Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques." Proteomics **9**(20): 4632-4641.
Zhou, Y., et al. (2019). "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets." Nat Commun **10**(1): 1523.
Zhu, N., et al. (2015). "Mutations in tyrosine kinase and tyrosine phosphatase and their relevance to the target therapy in hematologic malignancies." Future Oncol **11**(4): 659-673.

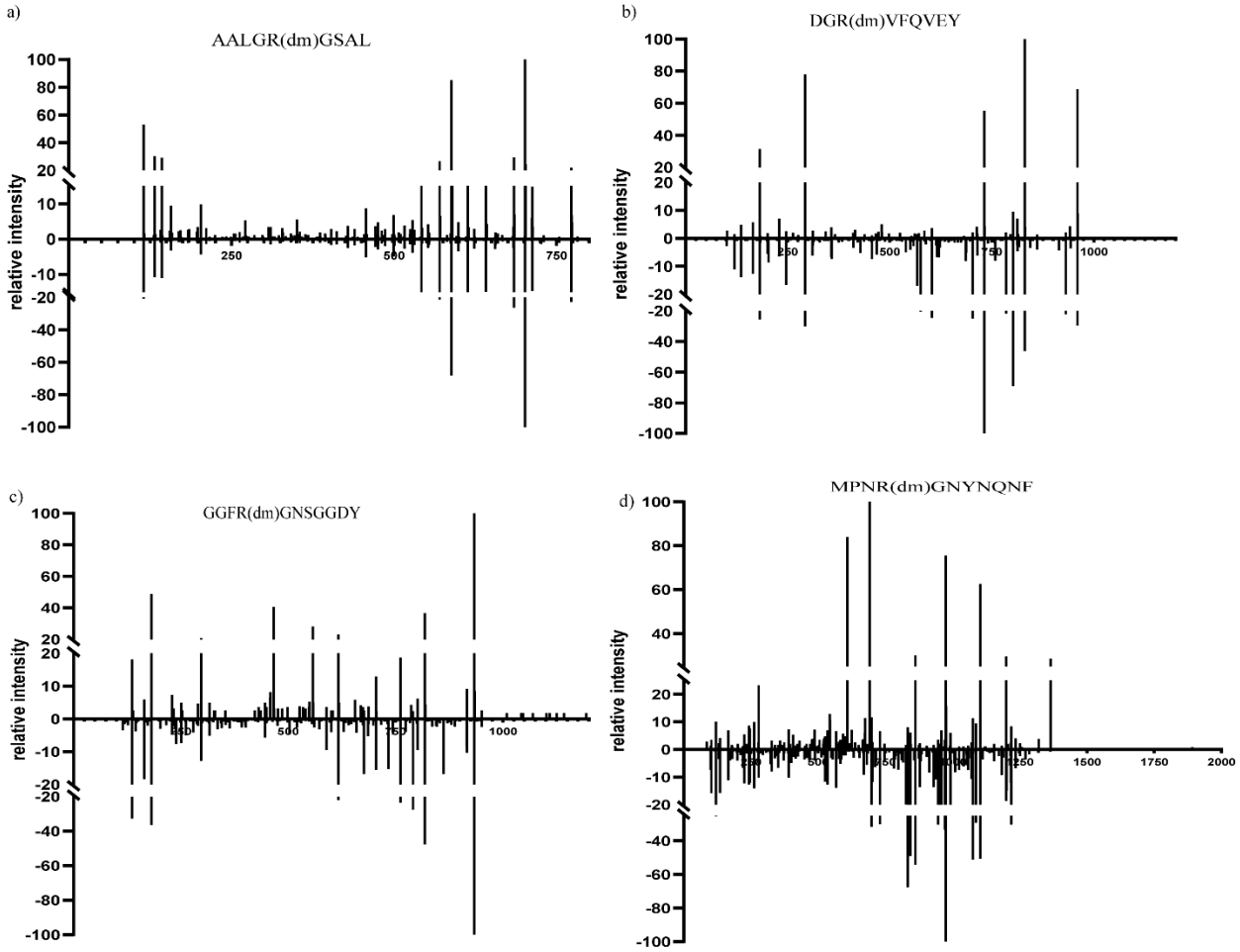
CHAPTER 4: SUPPLEMENTARY FIGURES



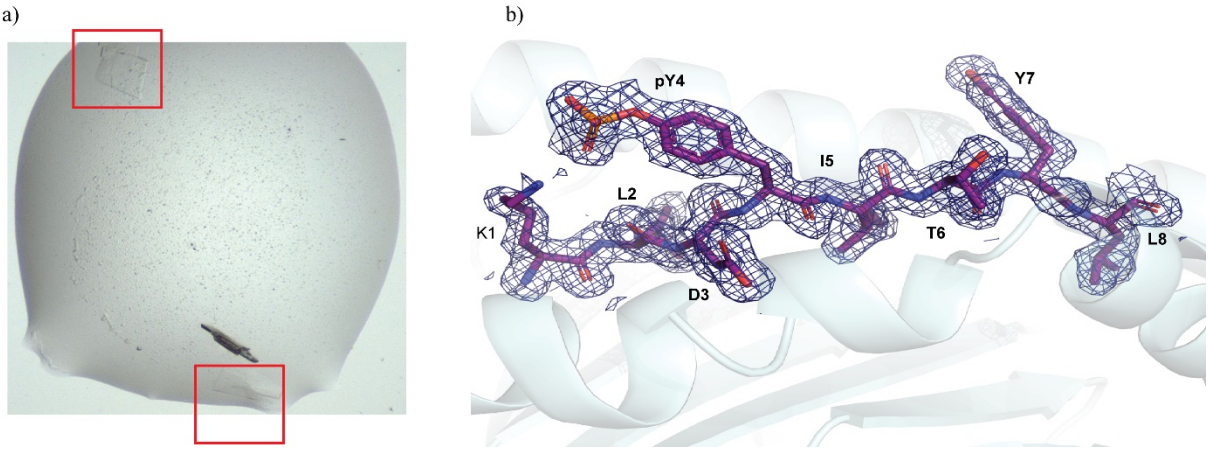
Supplementary Figure 1 – Unique and overlapping phosphopeptides identified in a) between cell lines and b) between cell lines and clinical samples including healthy and patient bone marrow aspirates.



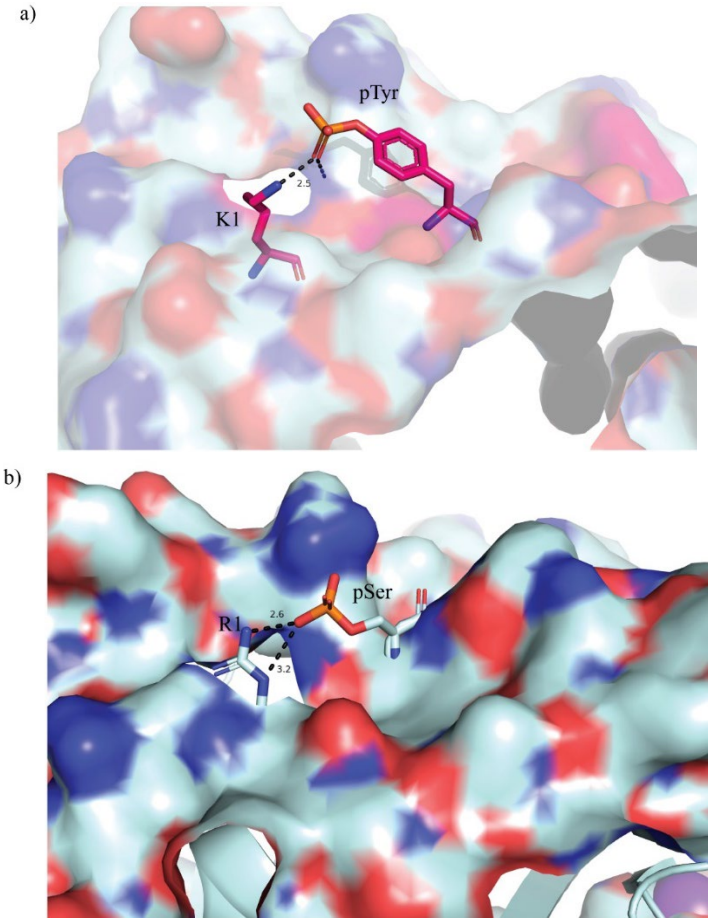
Supplementary Figure 2a – Similar MS/MS fragmentation pattern of endogenous (top) and synthetic (bottom) HLA A*02:01 restricted phosphopeptides. Three remaining phosphopeptides included in the study a) TLY(Ph)LTYEL b) KLVEQY(Ph)KL and c) KLY(Ph)SILKI generated similar series of b and y ions.



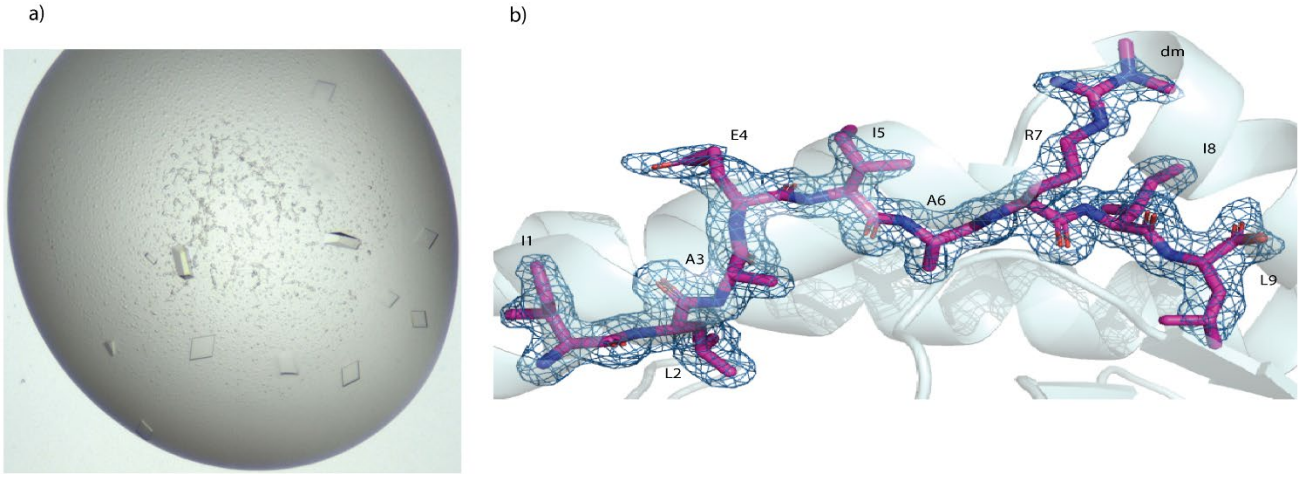
Supplementary Figure 2b – Similar MS/MS fragmentation pattern observed between endogenous (top) and synthetic (bottom) HLA class I restricted dimethyl arginine peptides. Four remaining dimethyl arginine peptides a) AALGR(dm)GSAL b) DGR(dm)VFQVEY c) GGFR(dm)GNSGGDY and d) MPNR(dm)GNYNQNF generated similar series of b and y ions.



Supplementary Figure 3 – a) A*02:01 - KLDY(Ph)ITYL crystal was resolved to 1.8 Å b) electron density map (2Fo-Fc) of pHLA complex with the phosphate moiety pointing upward.



Supplementary Figure 4 – Side view of pHLA complex a) KLDY(Ph)ITYL and RTYS(p)GPMNKV (PDB ID 3BH9) highlights the closeness of phosphate moiety to P1 Lys forming intra peptide hydrogen bonds.



Supplementary Figure 5 – a) HLA A*02:01-ILAEIAR(dm)IL crystal was resolved to 1.8 Å b) electron density map (2Fo-Fc) of pHLA complex with the asymmetric dimethyl moiety pointing upward.

CHAPTER - 5

Chapter 5: Investigating the immunopeptidome of diffuse intrinsic pontine glioma to identify immunotherapy targets

5.1 INTRODUCTION

Brain tumours are the leading cause of cancer-related mortality and morbidity in children (Fitzmaurice et al. 2017). According to the WHO, primary tumours of the central nervous system (CNS) are classified from grades I to IV, based on morphology, grade of malignancy, proliferative index, response to treatment and survival time (Urbanska et al. 2014, Louis et al. 2016). Diffuse midline gliomas encompass a heterogeneous disease spectrum that includes glioblastoma multiforme (GBM) and diffuse intrinsic pontine gliomas (DIPGs). DIPGs are categorised as a grade IV gliomas according to WHO criteria (Louis et al. 2016) and are spatially and temporally restricted, arising in the ventral pons region of infants as young as 3 years of age but peak incidence is observed between the ages of 6 to 9 years (Filbin et al. 2018). Previously, DIPG cases were classified along with their adult GBM counterparts, but in 2016 DIPG was reassigned to a separate category (Louis et al. 2016). Distinction of paediatric DIPGs were driven by mounting evidence that paediatric brain tumours differed from adult GBMs, not only in clinical presentation but also in their unique developmental origins and tumour microenvironment, which was a direct consequence of specific driver mutations (Jones and Baker 2014, Wang et al. 2019).

One of the key driver mutations correlated with DIPG pathogenesis is a missense mutation in histone 3.3 protein (H3.3), wherein a lysine at position 27 is mutated to methionine (hereafter referred to as K27M) (Schwartzentruber et al. 2012). Schwartzentruber *et al.* (Schwartzentruber et al. 2012) was the first large scale study to report the prevalence of the H3.3K27M mutation in DIPG (referred to as paediatric GBM in their study) using whole exome sequencing (WES). They reported the mutation in ~36% of cases (n=32/90) compared to only 3.4% (11/318) in adult GBM cases. Over the years several studies have not only recapitulated these findings but identified other proteins potentially dysregulated in DIPG including tumour protein 53 (TP53), alpha thalassemia/mental retardation syndrome X-linked (ATRX) protein, histone chaperone death-associated protein 6 (DAXX) and platelet derived growth factor receptor alpha (PDGFRA) (Khuong-Quang et al. 2012, Schwartzentruber et al. 2012, Wu et al. 2012, Buczkowicz et al. 2014). A clear outcome of these collective studies was that the survival rate of children with the H3.3K27M mutation was significantly reduced with median overall survival rate of 8 to 12 months post-diagnosis (Mathew et al. 2018), compared to children with a tumour expressing wild type histone protein who can live up to 4.5 years post diagnosis

(Khuong-Quang et al. 2012). At a cellular level, the H3.3K27M mutation results in loss of function by hindering methylation or acetylation at K27, which alters epigenetic regulation and transcriptional activation or repression of different genes downstream (Schwartzentruber et al. 2012, Wu et al. 2012).

Brainstem gliomas such as DIPG constitute ~75% of all CNS tumour burden in children (Warren 2012). Currently, the only treatments available include radiotherapy and chemotherapy, which alleviate symptoms but do not cure the cancer (Mathew and Rutka 2018). Therefore, alternate approaches to combat this disease are critically needed, such as targeted immunotherapy. This approach relies on T cell recognition of cancer-specific peptides directly presented on the surface of tumour cells by HLA molecules. These cancer-specific peptides may include differentially expressed tumour associated antigens (TAAs) derived from oncogenes or gonadotrophic proteins such as cancer testis antigens (CTAs), as well as unique mutated antigens, termed neoantigens. There is mounting evidence suggesting that HLA-restricted peptide-based cancer immunotherapy can induce specific immune responses and impact clinical outcomes (Neidert et al. 2018). Furthermore, efforts are underway to develop personalised cancer vaccines for different cancers based on immunopeptidome profiles (Dutoit et al. 2012). Thus, in-depth characterisation of the DIPG immunopeptidome may promote immunotherapeutics, which harness the specificity of T cells and redirect their effects towards targeted malignant cells to combat the disease (Rosenberg and Restifo 2015). Candidate immunotherapeutics could be in form of dendritic cell (DC) based vaccines (Mastelic-Gavillet et al. 2019), monospecific chimeric antigen receptor (CAR) T cells (Wang et al. 2019) or bispecific antibodies which engage with TAAs present on the surface of the tumour itself along with the TCR–CD3 complex, thereby redirecting T cell cytotoxicity to malignant cells (Dahlén et al. 2018).

This study utilised an in-depth immunopeptidomics approach coupled with high resolution MS to profile the HLA class I-restricted DIPG immunopeptidome. The resultant dataset was then interrogated for the presence of peptides derived from TAAs, CTAs, as well as neoepitopes derived from somatic mutations including the H3.3K27M driver mutation, TP53 and phosphoinositide 3 kinase (PI3K). Excitingly, several novel DIPG targets were unveiled that belonged to TAAs, CTAs, as well as a neoepitope associated with the H3.3K27M driver mutation restricted by different HLA alleles.

5.2. MATERIALS AND METHODS

5.2.1 Cell lines and culturing

DIPG cell lines (DIPG58 and SF7761) were sourced from Dr Jason Cain (Hudson Institute of Medical Research) and other set of DIPG cell lines (DIPG27, DIPG35, DIPG38 and DIPG43) were obtained from Dr Misty Jenkins (Walter and Eliza Hall Institute, Victoria, Australia). Briefly, cell lines were maintained in working Tumour Stem Medium (TSM). First, Tumour Base Media (TBM) was prepared by mixing 1:1 ratio of Neurobasal-A medium (Gibco, ThermoFisher, USA) and D-MEM/F-12 (Gibco, ThermoFisher, USA), supplemented with 10mM HEPES (Gibco, ThermoFisher, USA), 1mM MEM sodium pyruvate (Gibco, ThermoFisher, USA), 0.1mM MEM nonessential amino acids (Gibco, ThermoFisher, USA), GlutaMAX-I supplement (Gibco, ThermoFisher, USA) and 1X Antimycotic (Gibco, ThermoFisher, USA). Second, TBM was supplemented with B27 supplement (minus Vitamin A, Invitrogen, USA), H-EGF (20ng/ml, Shenandoah Biotech, USA), H-FGF basic 154 (20ng/ml, Shenandoah Biotech, USA), H-PDGF-AA (10ng/ml, Shenandoah Biotech, USA), H-PDGF-BB (10ng/ml, Shenandoah Biotech, USA) and 0.2% Heparin solution (2mg/mL, StemCell Technologies, USA). HLA class I typing of DIPG58, DIPG38 and SF7761 was performed by the Victorian Transplantation and Immunogenetics Service and whilst for DIPG27, DIPG35 and DIPG43 was performed by PathWest (Perth, Australia). The H3.3 K27M mutation for SU-DIPG27, SU-DIPG35 cell lines was confirmed using RNA sequencing (data provided by Dr Jason Cain), whilst SU-DIPG38 and SU-DIPG43 cell lines have been previously reported as H3.3 K27M and H3.3 WT positive respectively [2, 3].

Human embryonic kidney 293 cells (HEK293) expressing wild type H3.3 histone protein (herein called HEK WT H3.3) and HEK293 cells transfected with H3.3K27M mutant histone protein (herein called HEK K27M H3.3) were obtained from Dr Jason Cain. Cell lines were cultured in Dulbecco's modified essential media (DMEM, Gibco, ThermoFisher, USA) supplemented with 2mM MEM nonessential amino acid solution (Gibco, ThermoFisher, USA), 100mM HEPES (Gibco, ThermoFisher, USA), 2mM L-glutamine (Gibco, ThermoFisher, USA), 100 U/mL/100 µg/mL penicillin/streptomycin (Gibco, ThermoFisher, USA), 50µM 2-mercaptoethanol (Sigma-Aldrich, USA) and 10% heat inactivated FCS (Sigma-Aldrich, USA). Puromycin (3µg/mL, Invivogen, USA) was added as selection antibiotic. HLA class I typing was confirmed as A*02:01, A*03:01, B*07:02 and C*07:02 (performed by VTIS). For both the HEK293 WT and mutant cell lines, HLA class I expression was measured by flow cytometry following surface staining using neat hybridoma supernatants for anti-HLA-

A2 (BB7.2 [HB-82; ATCC] (Parham et al. 1981); produced in-house) and pan-HLA class I (W6/32 [HB-95; ATCC] (Brodsky et al. 1982); produced in-house) as primary antibodies, with a goat anti-mouse IgG PE (Southern Biotech, Birmingham, AL, USA) as the secondary antibody.

H3.3 K27M transfected C1R A*03:01 cells were cultured in RF10 (RPMI 1640 (Gibco) supplemented with 2mM MEM nonessential amino acid solution (Gibco, ThermoFisher, USA), 100mM HEPES (Gibco, ThermoFisher, USA), 2mM L-glutamine (Gibco, ThermoFisher, USA), 100 U/mL/100 µg/mL penicillin/streptomycin (Gibco, ThermoFisher, USA), 50µM 2-mercaptoethanol (Sigma-Aldrich, USA) and 10% heat inactivated FCS (Sigma-Aldrich, USA). Puromycin (3µg/mL, Invivogen, USA) and G-418 (0.5mg/ml, Roche, USA) were added to obtain cells expressing high levels of both mutant histone protein and HLA A*03:01, respectively. HLA A*03:01 surface expression was monitored by staining using hybridoma supernatants for both anti-HLA-A3 (GAPA3 [HB-122; ATCC]] (Berger et al. 1982) produced in-house) and pan-class I (W6/32) as primary antibodies and goat anti-mouse IgG PE as the secondary antibody. The level of expression of the H3.3K27M mutant histone protein was confirmed by western blot by virtue of a Haemagglutinin (HA) tag at the C-terminus of the protein. $1x10^5$ cells of both C1R-A3 WT and H3.3K27M were lysed and separated by SDS-PAGE using gradient (4%-18%) Mini-PROTEAN® TGX precast gels (Biorad) under reducing conditions by addition of 100mM dithiothreitol (DTT) (Sigma-Aldrich, USA). The resolved proteins were transferred onto nitrocellulose membranes by western blotting at 110 mA for 1 hr using transfer buffer and the membranes were blocked using 5% (w/v) skim milk powder in TBS-T. The blocked membranes were probed with the anti-HA antibody, followed by an anti-rat horse raddish preoxidase (HRP) conjugated secondary antibody. Membranes were scanned with a ChemiDoc Touch Imaging System (Biorad), using Western Lightning Plus-ECL Enhanced Chemiluminescence reagent (Perkin Elmer). Image was visualised using the Biorad ImageLab Software (v5.2.1) in combination with Image J (v1.5i).

5.2.2 Purification of HLA-peptide complexes from small size pellets

Small scale immunoaffinity purifications were performed for the four DIPG cell lines (DIPG27, DIPG35, DIPG38 and DIPG43). Cells (~$2x10^7$) were lysed with 100µl of lysis buffer (0.5% IGEPAL, 50mM Tris (pH 8.0), 150 mM NaCl and 1X protease inhibitor tablet), mixed gently and left rolling at 4°C for 1 hr. 2.8mL of protein A resin (PAS, CaptivA®, Repligen, USA) was prepared by washing with 3 CV of 1X PBS and then divided into two; 400µL was

used for the pre-column and the 1.2mL PAS was incubated with 1.2 mg BB7.2 (anti-HLA-A2) and 1.2 mg W6/32 (pan-HLA class I) in tandem.

Following lysis, samples were centrifuged at 3724g, 15 mins, 4°C. Each lysate was first incubated with only the pre-column for 1 hr at 4°C to remove non-specific binding material, followed by serial affinity capture of HLA A*02:01 peptides (BB7.2 column) and the remaining class I peptides (W6/32 column). All three columns including the pre-column and affinity columns BB7.2 and W6/32 were washed with 5 CV of 1X PBS to remove unbound antibody, detergent, and other salts from the sample. Bound HLA-peptide complexes were eluted using 150µl of 10% acetic acid. The eluate was warmed to 70°C for 10 mins to disassociate any remaining pHLA complex and then left to cool at room temperature. The eluate was passed through a 5 kDa molecular weight cut off (MWCO) filter (Amicon, Sigma-Aldrich, USA) and centrifuged at 16,060g, 30 mins at room temperature. Another 50µl of 10% acetic acid was added to the filter and centrifugation was repeated. The filtered sample was centrifugally evaporated and desalted by reverse phase C18 stage tips (Omix, Agilent). The C18 stage tips were washed with 0.1% Trifluroacetic acid (TFA, ThermoFisher Scientific, USA) and peptides were eluted with 40% acetonitrile (ACN, ThermoFisher Scientific) in 0.1% TFA. Samples were interrogated by LC-MS/MS using a Q-Exactive Plus (QE Plus) mass spectrometer (Thermo Scientific) and subsequently interrogated for specific peptides using targeted LC-multiple reaction monitoring (MRM) on QTRAP 6500+ mass spectrometer (Sciex).

5.2.3 Purification of HLA-peptide complexes from large size pellets

DIPG cell lines SF7761 and DIPG58 were culture in TSM working meida and expanded to $2x10^8$. HEK H3.3 WT and HEK H3.3K27M mutant cells were cultured as mentioned 5.2.1 and expanded to $6x10^8$, whilst C1R.A*03:01/H3.3K27M cells were cultured as mentioned in 5.2.2 and expanded to $1x10^9$ in roller bottles. Cells were harvested by centrifugation (3724g, 15 mins, 4°C), pellets snap frozen in liquid nitrogen and stored at -80°C until required.

For immunoaffinity purification appropriate affinity columns were used for each cell line. Purified antibodies (10 mg) were cross-linked to 1 mL of PAS resin as previously described (Pandey et al. 2020) and used for the immunoaffinity capture of solubilised native pHLA. Briefly, frozen cell pellets were pulverised using a cryogenic mill (Retsch Mixer Mill MM 400), reconstituted in Lysis Buffer [0.5% IGEPAL (Sigma-Aldrich, USA), 50mM Tris pH 8, 150mM NaCl and protease inhibitors (Complete Protease Inhibitor Cocktail Tablet, 1 tablet per 50 mL solution; Roche Molecular Biochemicals, Switzerland)] and incubated for 1 hr at

4°C with rotation. The supernatant was passed through a PAS pre-column (500µL), followed by serial affinity capture. For HEK cell lines the lysate was first passed through a BB7.2 column (to pull down HLA A*02:01 complexes) followed by a W6/32 column (to pull down remaining A*03:01, B*07:02 and C*03:03 complexes).

For the C1R A*03:01/ H3.3K27M cell line, lysate was passed sequentially through a GAPA3 column (to pull down HLA A*03:01 complexes) followed by a W6/32 column (to pull down remaining HLA I complexes). Bound complexes were eluted from the individual columns using 5 CV of 10% acetic acid. Eluates of all transfected cell lines were fractionated by RP-HPLC on a 4.6 mm internal diameter 100 mm long RP monolithic C18 HPLC column (Chromolith Speed Rod, Merck-Millipore, Germany) using an ÄKTA micro HPLC system (GE Healthcare) running a mobile phase consisting of buffer A 0.1% TFA and buffer B (80% ACN). A flowrate of 2 mL/min was used to separate the peptides into 1 mL fractions across a changing gradient; 0-15% buffer B over 0.25 mins, 15-30% buffer B over 4 mins, 30-40% buffer B over 8 mins, 40-45% buffer B over 10 mins, and 45-99% buffer B over 2 mins. A total of 43 peptide-containing fractions were collected, vacuum concentrated and concatenated into three peptide-containing pools for each antibody and were analysed by LC-MS/MS on a high-resolution QE Plus Orbitrap MS.

5.2.4 Analysis of HLA class I bound peptides by LC-MS/MS

The Q-Exactive plus LC-MS was run in data dependent acquisition (DDA) mode coupled online to a RSLC nano HPLC (ultimate 3000 UHPLC, ThermoFisher Scientific). The samples were injected onto a 100 µm, 2 cm nanoviper Pepmap100 trap column prior to separation on a 75 µm x 50 cm, Pepmap100 C18 analytical column (ThermoFisher Scientific). The eluate was nebulised and ionised using a nano electrospray source (ThermoFisher Scientific) with a distal coated fused silica emitter (New Objective). The capillary voltage was set at 1.7 kV. The Q-Exactive Plus mass spectrometer was operated in the DDA mode to automatically switch between full MS scans and subsequent MS/MS acquisitions. Survey full scan MS spectra (*m/z* 375–1600) were acquired in the Orbitrap with 70,000 resolution after accumulation of ions to a 5e5 target value with a maximum injection time of 120 ms. For MS/MS dynamic exclusion was set to 15 s to avoid resequencing the same analyte. The 12 most intense charged ions ($z \geq +2$) were sequentially isolated and fragmented in the collision cell by higher-energy collisional dissociation (HCD) with a fixed injection time of 120 ms, 35,000 resolution and automatic gain control (AGC) target of 2e5 with isolation window of 1.8 *m/z*, intensity threshold 1.7e4 and scan range of 200 – 2000 *m/z*.

5.2.5 Synthetic peptides

Synthetic A2 and A3 mutant H3.3K27M peptides RMSAPSTGGV and RMSAPSTGGVK respectively, with heavy Proline at P5 (13C5, 15N1) were synthesised by Mimotopes (Australia) with >95% pure as indicated by analytic HPLC and MS analyses. Peptide stocks were dissolved in DMSO (Sigma-Aldrich, USA) at a concentration of 10 mg/ml and stored at −80°C until use.

5.2.6 MRM quantification of HLA bound peptides

The A2 positive DIPG cell line eluates were spiked with the A2 isotopic heavy peptide. Samples were assessed for the presence of peptide by MRM using SCIEX QTRAP 6500+ mass spectrometer, equipped with an on-line Ekspert nanoLC 415 (Eksigent) autosampler utilising Analyst 1.6 (SCIEX) software. Samples were injected and loaded onto a trap column (ChromXP C18, 3 µm 120 Å, 350 µm × 0.5 mm [SCIEX] at a flow rate of 5µL/min in 98% buffer A (0.1% FA), 2% buffer B (95% can, 0.1% FA) for 10 min. Samples were eluted from the trap column and separated using an analytical column (ChromXP C18, 3 µm 120 Å, 75 µm × 15 cm [SCIEX]) at 300 nL/min for 90 mins using the following gradient conditions: 0–1 min 2% B, 2–75 min 3–50% B, 75-79 min 35–80% B, 79–87 min hold at 80% B, 87–88 min 80–2% B, followed by equilibration at 2% B for 8 min. The dwell time on each transition was 20secs.

The QTRAP 6500+ was operated in MRM mode in unit resolution for Q1 and Q3, coupled to an IDA criterion with a triggered enhanced product ion scan (EPI) (10 000 Da/s; rolling CE; unit resolution) following any MRM transition exceeding 1000 counts. Triggering MRM transitions were ignored for the subsequent 30s.

MRM data visualised on Skyline software 64-bit v4.1.0 (MacCoss Laboratory). The heavy labelled peptides were detected by modifying the endogenous peptide by adding the 13C5 15N1 modification on Pro.

5.2.7 Mass spectrometry data analysis

For the Q-Exactive plus generated MS/MS raw files were exported and analysed using Peaks 8.5 software (Bioinformatics solutions) searching against the human proteome (Uniprot 15/06/2017; 20,182 entries) wherein the H3.3K27M protein sequence was appended. For MS data analysis, a 5% false discovery rate (FDR) cut off was applied. For data processing, firstly iRT and decoy peptides were removed along with any duplicate peptides. Secondly, only peptides with 8-12 amino acids were retained. Thirdly, peptide scans with alternative Ileu-Leu explanations were removed from the dataset and finally tryptic contaminants were removed

from the dataset by performing Gibbs clustering for 8-12mer peptides. For detecting the heavy peptide at MS/MS level a variable modification which caused mass shift of 6.0138 in +1 charge state was used. MS1 information (from raw files) visualised on Skyline software 64-bit v4.1.0 (MacCoss Laboratory). MS/MS data from Peaks searches (peptide.csv) for each cell line is provided as supplementary data files, as mentioned in the appendix.

5.3 RESULTS

5.3.1 LC–MS/MS profiling of HLA class I ligands from paediatric DIPG cell lines

An in-depth immunopeptidomics study was undertaken to identify the repertoire of endogenously derived HLA class I peptides in six DIPG cell lines. Four patient derived DIPG cell lines expressed the H3.3 K27M mutation and two cell lines were H3.3 WT (Table 1). A total of 21,281 unique 8-12 amino acid long peptides were identified across the six cell lines (Table 1), which were collectively restricted by 22 HLA class I allotypes (five HLA-A, nine HLA-B and eight HLA-C alleles, Table 1). There were several common HLA allotypes across different samples including HLA A*02:01 (n = 4), A*01:01 (n = 2), A*68:01 (n = 2), B*44:02 (n = 2) and C* 04:01 (n = 2) which represent frequent Caucasian allomorphs (Robinson et al. 2015). For the four HLA A*02:01 positive cell lines, namely DIPG27, DIPG35, DIPG38 and DIPG43, the HLA A*02:01 complexes were pulled down first using an anti-HLA-A2 column followed by a pan class I antibody to capture the remaining HLA complexes.

Table 1 –Total number of peptides (8-12mers) identified across four DIPG cell lines.

Cell line	K27M Mutation	HLA class I typing	HLA A*02 :01 Total peptides (8-12mer)	Pan HLA class I Total peptides (8-12mer)	Total HLA class I peptides identified
DIPG27	Present	A*02:01, A*68:01, B*35:12, C*04:01	1,557	1,103	2,660
DIPG35	Present	A*02:01, A*30:01, B*13:02, B*40:01, C*03:04, C*06:02	2,256	1,749	4,005
DIPG38	Present	A*02:01, A*01:01, B*35:03, B*44:02, C*04:01, C*05:01	1,260	642	1,902
SF7761*	Present	A*29:02, A*68:01, B*44:03, B*51:02, C*08:01, C*16:01	-	12,598	
DIPG43	Wild type	A*02:01, B*44:02 B*45:01, C*05:01 C*16:01	1,494	288	1,782
DIPG58*	Wild type	A*01:01, A*03:01, B*07:02, B*08:01, C*07:01, C*07:02	-	GAPA3 – 2,186 W632 – 1,533	3,719
Total			6,567	20,025	26, 592

*Pellet size of $2x10^8$ whilst the other were $2x10^7$

5.3.2 The HLA A*02:01 immunopeptidome of DIPG cell lines

A total of 6,567 peptides restricted to HLA A*02:01 were identified across the four DIPG cell lines (Table 1). The identified peptides followed a canonical length distribution for HLA class I ligands, with the majority of peptides (~63-70%) being nonamers (Figure 1a). Of these total peptides, 3,501 were unique with most peptides identified in SU-DIPG35 (n=973) and the least peptides identified in DIPG43 (n=354) (Figure 1b). Only 389 peptides (~10%) were common amongst all four DIPG cell lines (Figure 1b). These nonamers were used to confirm consensus HLA-binding motifs. Not surprisingly, the binding motifs of these peptides matched the existing HLA A*02:01 motif, with small hydrophobic amino acids Leu/Met at P2 and Leu/Val at P9 (Figures 1c – 1f) for all four DIPG cell lines.

A previous study by Chheda *et al.* (Chheda et al. 2018), reported a HLA A*02:01 restricted H3.3K27M neoepitope, RMSAPSTGGV containing the K27M mutation. Surprisingly, with our global LC-MS/MS approach no evidence of this neoepitope was detected in any of the four DIPG cell lines even though the transcript was present in abundance as evident from RNA

sequencing data (for cell lines DIPG27 and DIPG35 (provided by Dr Jason Cain as personal communication)). Interestingly, when investigated at source protein level, histone H3.3 contributed eleven peptides, with 8 of these peptides previously reported as HLA A*02:01 ligands and 3 of them common amongst all four cell lines (Supplementary Table 1). Additionally, several HLA A*02:01-restricted peptides were also identified from other histone proteins namely H1, H2 and H4 (Supplementary Table 2).

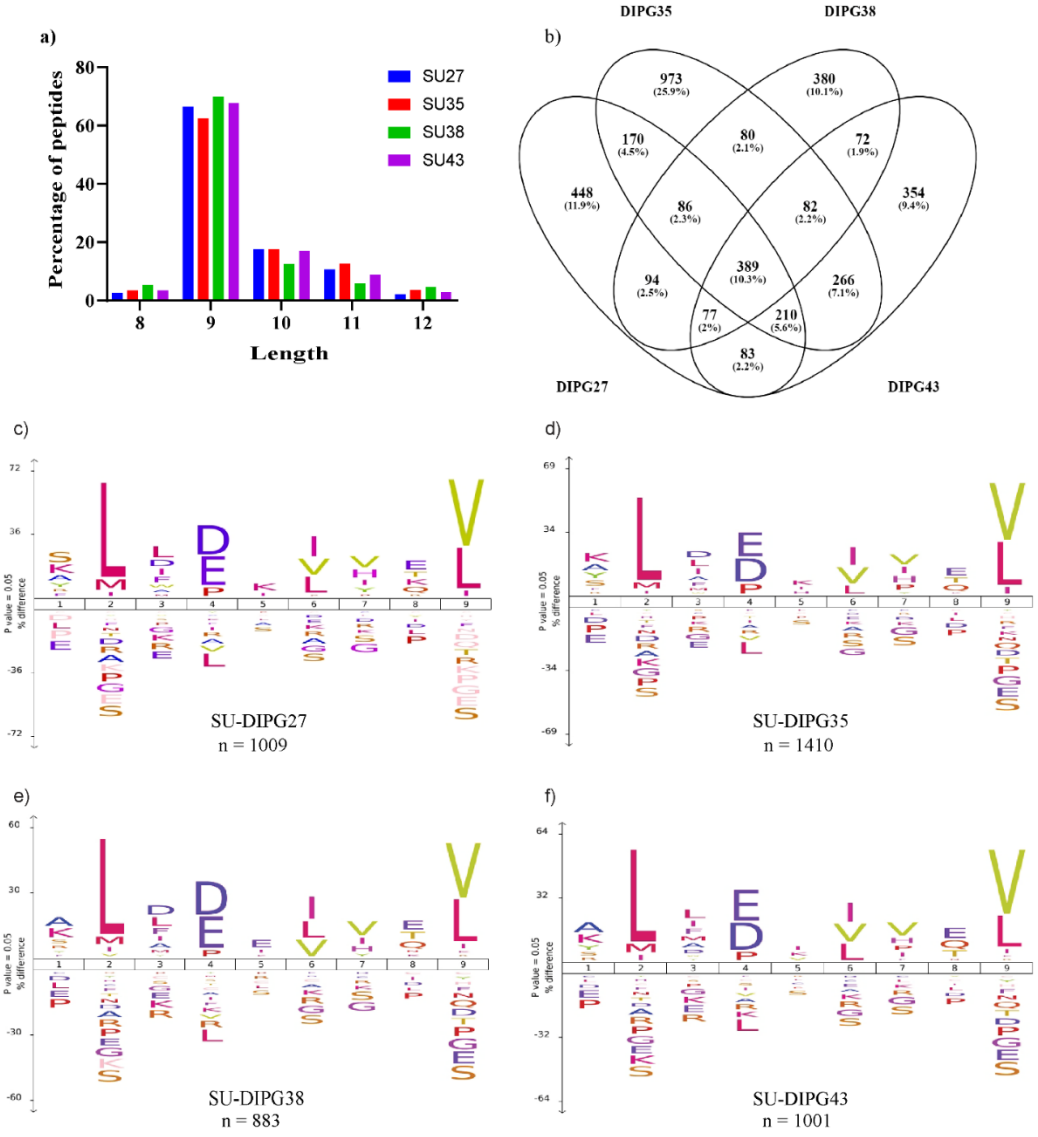


Figure 1 – HLA A*02:01 derived DIPG peptides follow canonical peptide length distribution and consensus binding motif. a) **The** majority of peptides identified in each cell line were nonamers. b) Venn diagram depicting unique and overlapping peptides in the dataset. Binding motifs of 9mer peptides identified in c) DIPG27, d) DIPG35, e) DIPG38 and f) DIPG43 plotted using IceLogo (Colaert et al. 2009) follow canonical HLA-A2 motif with **n** depicting number of peptides used to plot the motif.

5.3.3 The HLA pan class I immunopeptidome of DIPG cell lines

A total of 20,025 peptides were identified in the HLA pan class I fraction across all cell lines (Table 1). As expected, there were fewer overlapping peptides amongst the 6 cell lines in the dataset with overall 17,321 unique peptides, reflecting the genetic diversity and non-overlapping HLA alleles in these cell lines. When compared with the HLA A*02:01 dataset there were only 612 peptides overlapping between the two datasets. Similar to HLA-A2 bound peptides, the majority of the peptides were nonamers (Figure 2).

To delineate peptides restricted by the different HLA allotypes expressed by each cell line, NetMHC 4.0 (Jurtz et al. 2017) was used. Peptides were identified as either strong binders (SB, based on rank threshold of 0 to 0.5) or weak binders (WB, based on rank threshold of 0.5 to 2.0). All allotypes contained both SB and WB which was used to plot the motif for each allele (Figure 3a – 3f).

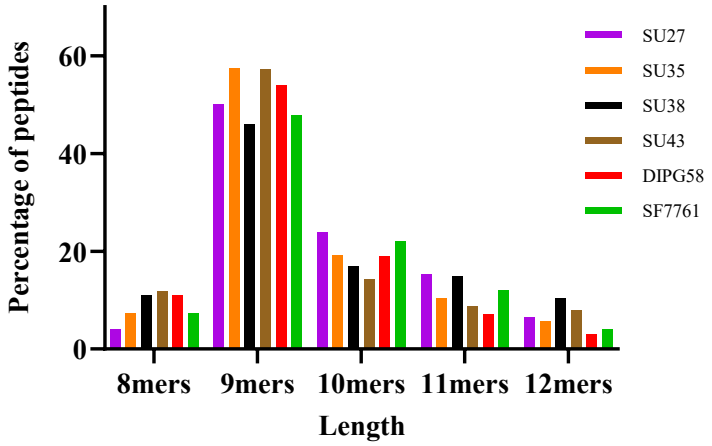


Figure 2 – HLA pan class I derived peptides follow canonical peptide length distribution. **The** majority of peptides identified in each cell line were nonamers followed by longer peptides of 10 and 11mers.

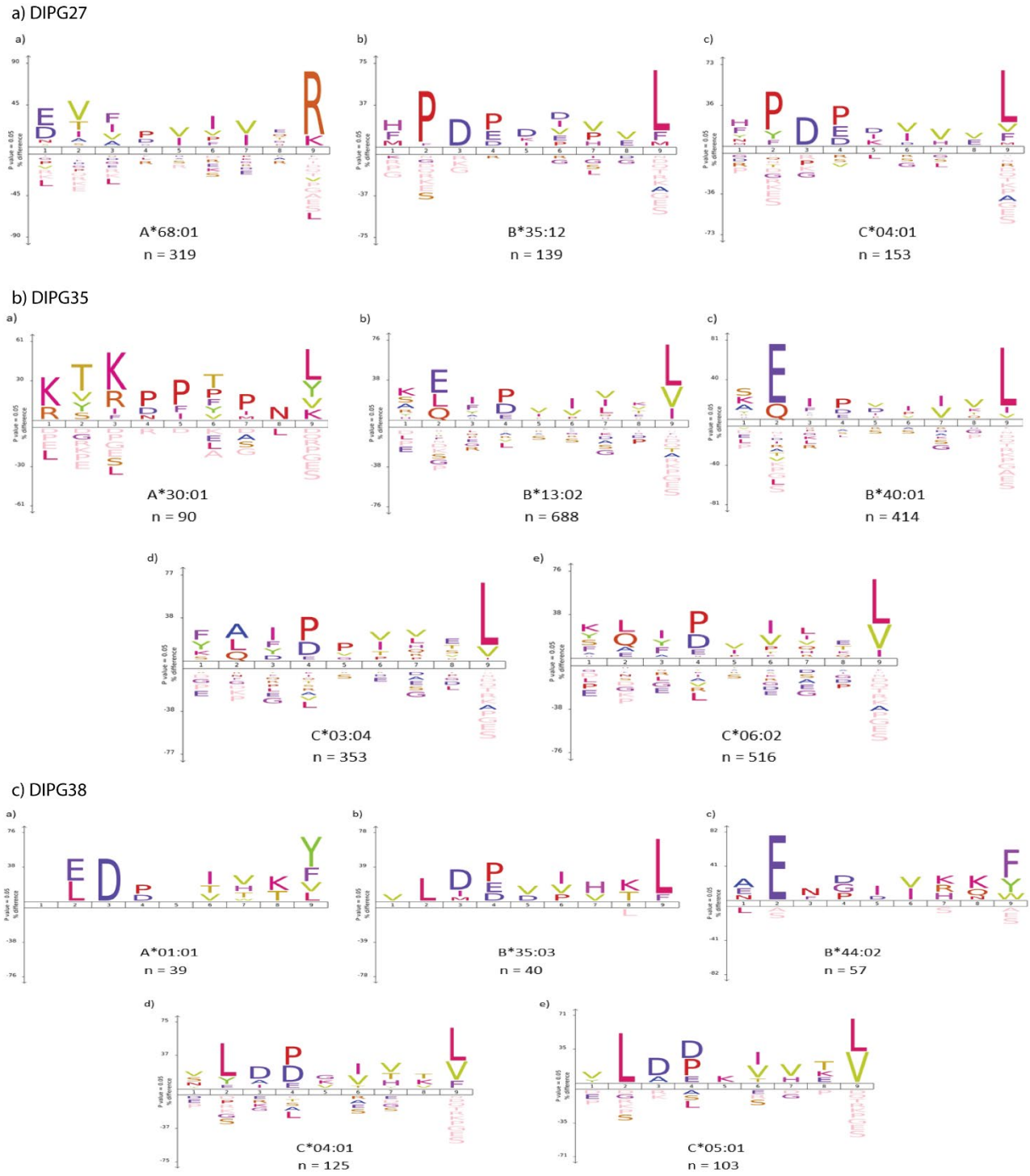


Figure 3a-3c – Nonamers identified from DIPG H3.3 K27M positive cell lines have canonical binding motifs. DIPG H3.3 K27M positive cell lines including a) DIPG27, b) DIPG35 and c) DIPG38. Nonamers were selected, with modifications and duplicates being removed, and their binding affinity was predicted using NetMHC Pan 4.0. Icelogo (Colaert et al. 2009) representation of motifs of peptides predicted to bind to HLA class I alleles expressed in the cell lines with **n** representing number of binders identified for each allele.

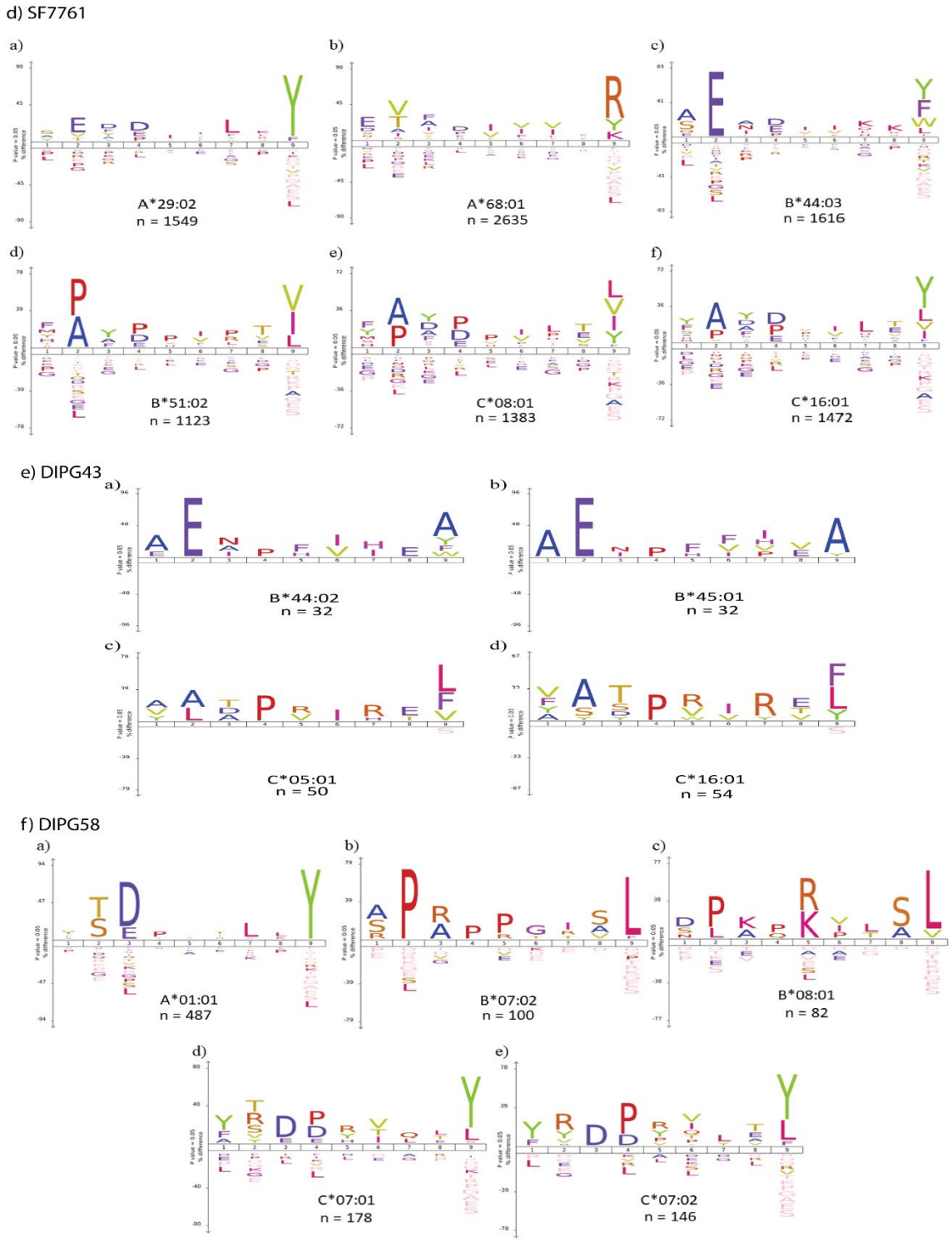


Figure 3d-3f – Nonamers identified from DIPG H3.3 K27M positive and DIPG WT cell lines have canonical binding motifs. DIPG H3.3 K27M positive cell lines including d) SF7761 and DIPG WT cell lines including e) DIPG43 and f) DIPG58. The binding affinity of nonamers was predicted using NetMHC Pan 4.0. Icelogo (Colaert et al. 2009) representation of motifs of peptides predicted to bind to HLA class I alleles expressed in the cell lines with **n** representing number of binders identified for each allele.

5.3.4 Mining the DIPG immunopeptidomes for cancer antigen-derived peptides

The immunopeptidomes from the six DIPG cell lines were interrogated for presence of TAAs and CTAs using two different human cancer databases; (i) Tumour T-cell antigen database (TANTIGEN, http://cvc.dfci.harvard.edu/tantigen) (Olsen et al. 2017) and (ii) CTdatabase (http://www.cta.lncc.br/) (Almeida et al. 2009). A total of 1,123 unique TAA and CTA-derived peptides were identified across the six DIPG cell lines. Of these 1,041 unique TAA (Supplementary file 1), 74 unique CTA (Supplementary file 2) and 8 peptides were present in both databases. The maimum number of unique TAA (543) and CTA-derived peptides (51) were identified in the SF7761 cell line. In the entire dataset, only 20 peptides have been previously reported as epitopes restricted by HLA-A2 or HLA-A11 in the TANTIGEN database found in ovarian (Ramakrishna et al. 2003) and renal cancer (Weinschenk et al. 2002, Stickel et al. 2009). Notably, 9 of the 20 peptides have been reported in previous GBM immunopeptidome studies (Table 2) (Dutoit et al., Neidert et al).

Table 2 – Previously reported tumour associated antigens identified in the study

Peptide identified	HLA restriction	Database/ Study	Source Protein	Type of cancer peptide identified in
LLDRFLATV	A2	TANTIGEN	Cyclin-1 (CCNI)	Ovarian ≠
ILDDIGHGV	A2	TANTIGEN	Abl interactor 2 (ABI2)	Ovarian ≠
KLDVGNAEV	A2	TANTIGEN	B-cell receptor-associated protein 31 (BAP31)	Ovarian ≠
KLQELNYNL	A2	TANTIGEN	Signal transducer and activator of transcription 1(STAT1)	Ovarian ≠
RLYPWGVVEV	A2	TANTIGEN	Septin-2 (SEPT2)	Ovarian ≠
FLYDDNQRV	A2	TANTIGEN	DNA topoisomerase 2-alpha (TOP2A)	Ovarian ≠
SLLQHLIGL	A2	TANTIGEN	Preferentially expressed antigen of melanoma (PRAME)	Ovarian ≠
KVLEYVIKV	A2	TANTIGEN	Melanoma-associated antigen 1 (MAGA1)	Ovarian ≠
FLDPRPLTV	A2	TANTIGEN	Cytochrome P450 1B1 (CP1B1)	Ovarian ≠
ETIPLTAEKL	A*68:01	TANTIGEN	Cyclin-D1 (CCND1)	Renal^
VGLIRNLAL	B8	TANTIGEN	Beta-catenin (CTNB1)	Renal^
ALFPERITV	A2	Neidert *et al*	(ATAT1)	GBM
SILDIVTKV	A2	Neidert *et al*	(RFTN2)	GBM
KIQEILTQV	A2	Dutoit *et al*	Insulin Growth factor 2 binding protein (IGF2BP3)	GBM
TMLARLASA	A2	Dutoit *et al*	Chondroitin sulphate proteoglycan 4 (CSP4)	GBM
LTFGDVVAV	A2	Dutoit *et al*	Fatty acid-binding protein 7, brain (FABP7)	GBM
NLDTLMTYV	A2	Dutoit *et al*	Neuroligin 4, X-linked (NLGN4X)	GBM

AIIDGVESV	A2	Dutoit *et al*	Protein tyrosine phosphatase, receptor-type Z polypeptide 1 (PTPRZ1)	GBM
KVFAGIPTV	A2	Dutoit *et al*	PTPRZ1	GBM
AMTQLLAGV	A2	Dutoit *et al*	Tenascin C (TNC)	GBM

^ (Weinschenk et al. 2002), ≠ (Ramakrishna et al. 2003)

The source proteins of the peptides included well known TAAs and CTAs such as MAGEA1, PRAME, Cyclin and Catenin β (CTNNB1). The peptides also previously identified in ovarian cancer cell lines (SKOV3 and OVCAR3) by MS and peptide specific $CD8^+$ T cells expanded *in vitro* from healthy donores were capable lysing tumour cells expressing the peptides. The peptide ETIPLTAEKL was restricted to HLA A*68:01 and has been identifed in renal carcinoma and melanoma(Heinzel et al. 2001, Weinschenk et al. 2002).

To verify if the source proteins of the TAA and CTA peptides identified in the immunopeptidome were expressed at transcriptomic level, RNAseq data for cell lines DIPG27 and DIPG35 was investigated. Transcriptomic evidence was found for a combined total of 52 source proteins in both the cell lines (Supplementary file 3 and file 4).

To investigate if the 52 proteins which had evidence at the mRNA level were being expressed at the protein level in DIPG patients, the immunopeptidome source protein data was compared data obtained from whole proteome analysis of CSF and FFPE of DIPG patients published by Saratsis *et al,* (Saratsis et al. 2012) (Saratsis et al. 2014). Some of these source proteins were present in FFPE samples (n = 25) and CSF samples (n = 12). Between the 2 cell lines, 15 common proteins were found (Table 3), which were present in FFPE and CSF samples as well. Peptides belonging to these proteins are most attractive candidates and ideal targets for immunotherapy as their presence has been confirmed at all 3 levels, transcriptomic, proteomic and immunopeptidomic.

The proteins contributing to TAA and CTA-derived peptides were also cross referenced to 2,587 proteins reportdly found elevated in the brain as per human brain protein atlas (HBPA) (http://www.proteinatlas.org). Thirteen source proteins from the DIPG immunopeptidome were common to the HBPA. A total of 65 unique peptides were identified from these 13 proteins which included proteins such as SOX family, MAGE and cyclins (Supplementary file 5).

Table 3 – Source proteins from cell lines DIPG27 and DIPG35 identified from transcriptomic and proteomic data.

Proteins present in immunopeptidome and RNAseq data	Present in DIPG patients FFPE Samples	Present in DIPG patients CSF samples
ACTN4	Yes	CSF tumour only
ANXA2	Yes	CSF upregulated
CDK1	Yes	NA
CSPG4	Yes	CSF tumour only
LRP1	Yes	CSF upregulated
MACF1	Yes	NA
PCBP2	Yes	NA
PGK1	Yes	CSF tumour only
PPIB	Yes	CSF downregulated
PRDX5	Yes	CSF tumour only
RTCB	Yes	NA
STAT1	Yes	NA
STRAP	Yes	NA
TOP2A	Yes	NA
XPO1	Yes	NA

5.3.5 Absence of a reported HLA-A2-restricted H3.3K27M neoepitope in HLA-A2 positive DIPG cell lines

We used a discovery-based approach to characterise the repertoire of peptides presented by DIPG cell lines. We anticipated detecting the reported H3.3K27M neoepitope (RMSAPSTGGV) (Chheda et al. 2018) within our A2/H3.3 K27M positive cell line dataset, but the neoepitope was not detected. In order to investigate the sample more sensitively, multiple monitoring reaction (MRM) assays were developed for peptide identification. Post-elution, samples were spiked with a heavy labelled version of RMSAPSTGGV. The peptide (RMSA**P**STGGV) carried a substitution of heavy Pro at P5 (13C5, 15N1), which introduced a 6 Da mass difference. The advantage of using a heavy labelled peptide is that it co-elutes with the endogenous peptide, but differs exclusively by the introduced mass change, enabling confident identification of the endogenous peptide. In the study by Chheda *et al* (Chheda et al. 2018), only the modified version of the peptide with an oxidised Met at P2 (R**M**SAPSTGGV) was identified. Thus, a total of 24 MRM transitions were selected which comprised of 6 transitions for each, native (RMSAPSTGGV) and modified peptides (R**Mox**SAPSTGGV) including both heavy and endogenous peptide version (Table 4). The limit of quantitation for

both the peptides showed that heavy labelled peptide could be detected even at low concentrations of 5femtomole/µL although for experimental purpose we used 50femtomole/µL.

Alongside, MRM transitions were also established for 6 constitutively presented HLA A*02:01 restricted peptides namely TLADLVHHV, ATYVFLHTV, VMDSKIVQV, YLGRLAHEV, RMLPHAPGV and VLYDQPRHV. The criteria for peptide selection were 1) previously identified in the C1R.A*02:01 immunopeptidome (Pandey et al. 2020), 2) belonged to proteins that were constitutively expressed or were highly abundant and 3) predicted strong binders by NetMHC pan 4.0 (based on rank threshold of less than 0.5).

The BB7.2 eluates of the HLA A*02:01 positive DIPG cell lines DIPG35, DIPG38 and DIPG43 were examined using these MRM-assays. Samples were interrogated for all the transitions including the H3.3 K27M neoepitope and constitutive HLA-A2 peptides. The heavy labelled A2 peptide MRMs for both native and oxidised forms were identified in all samples (Figure 4), but the 10mer endogenous neoepitope was not detected in any of the H3.3K27M mutant cell lines (DIPG-35 and DIPG-38) as a native (Figures 4) or oxidised species. All 6 constitutive HLA-A2 control peptides were identified in two H3.3K27M positive cell lines (DIPG35 and DIPG38) with similar retention time, whilst only 4 peptides were identified in H3.3K27M WT cell line DIPG43 (Supplementary Figure 1).

Table 4 – MRM transitions for HLA A*02:01-restricted K27M 10mer neoepitope

Epitope	Sequence	Q1 *m/z* (charge)	Q3 *m/z* (ion)
A2 – K27M	RMSAPSTGGV (Native Light)	481.7 (+2)	543.2
			630.3
			731.3
			675.3
			588.2
			517.2
A2 – K27M	RMSAPSTGGV (Native Heavy)	484.7 (+2)	549.2
			636.3
			737.3
			681.3
			594.3
			523.2
A2 – K27M	RMSAPSTGGV (Met-Ox Light)	489.7 (+2)	559.2
			646.2
			747.3
			675.3
			588.2
			517.2
A2 – K27M	RMSAPSTGGV (Met-Ox Heavy)	492.7 (+2)	565.2
			652.3
			753.3
			681.3
			594.3
			523.2

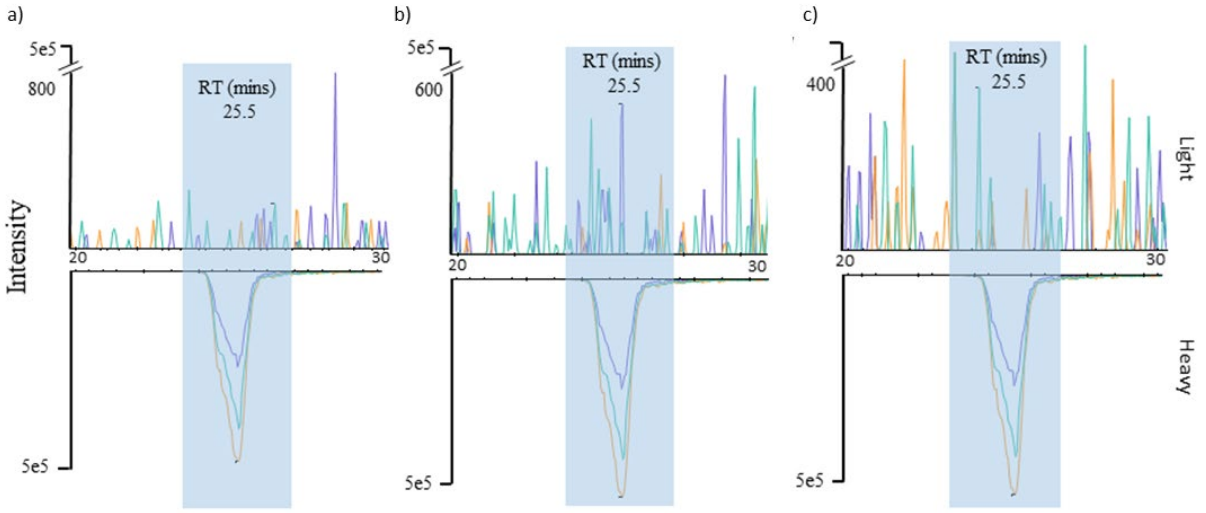


Figure 4 - MRM transition profiles reveal absence of HLA A*02:01 neopeptide RMSAPSTGGV in DIPG cell lines a) DIPG35, b) DIPG38 and DIPG43. To design the MRMs, Skyline software (64-bit v4.1.0) was used (Maclean et al. 2010). Peptide sequence was entered and using predictive algorithms, the software generated information regarding the charge states and fragment ions for each charge state. MRMs were designed in an instrument specific manner for QTRAP 6500+. LC-MS/MS acquisition was performed on both native and oxidised version of the peptide to determine the best collision energy and to obtain the full fragment ion spectrum; the six highest-intensity peaks were selected to be built into MRM for the native peptide comprising of b and y fragment ions. Once the acquisition method was optimised, the DIPG cell line samples were spiked with the heavy labelled peptide and acquired on the QTRAP 6500+. The heavy labelled peptide was detected across samples (a-c), but the endogenous peptide was not detected in any of the sample.

5.3.6 Identification of the H3.3K27M neoepitope

Since the HLA-A2 neoepitope was not detected in the DIPG cell lines, investigations for the 10mer were performed using an alternate HLA A*02:01 positive reporter cell line, HEK293 (HLA class I typing: A*02:01, A*03:01, B*07:02, C*03:03). The HEK293 parental (WT) cells were transfected with a plasmid encoding the H3.3K27M histone to facilitate high level expression of the mutant protein. Both the HEK293 parental and mutant lines were cultured, and HLA expression was monitored using anti-HLA A2 BB7.2 and anti-pan class I W632 antibody. Both cell lines HEK293 WT (Figure 5a) and HEK293 H3.3K27M (Figure 5b) had high HLA class I expression. Immunoaffinity purification was performed to capture HLA-A2-peptides and the remaining HLA class I-peptide complexes.

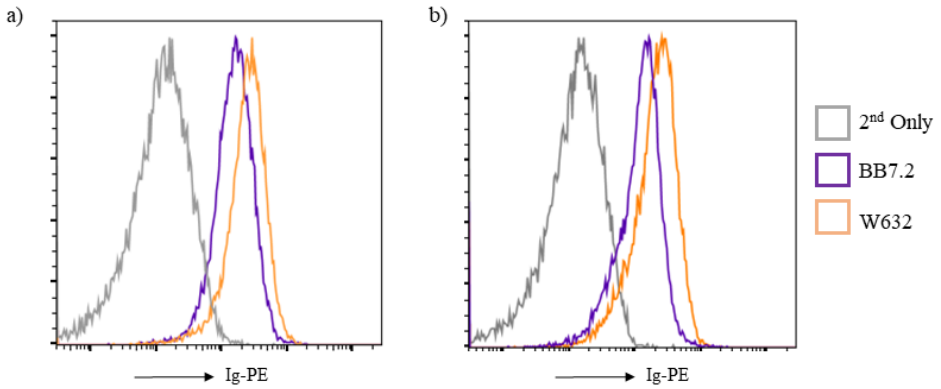


Figure 5 – Similar HLA-A2 and pan class HLA cell surface expression of a) HEK293WT and b) HEK293 transfected with H3.3K27M mutant cell lines. Cells were stained with either BB7.2 (anti-HLA-A2) or W6/32 (pan HLA class I) antibodies followed by secondary antibody (goat anti-mouse IgG PE). (A) Cells were acquired on LSR IIa (BD Biosciences) using the FACSDiva software. (B) Analysis of the cell population was performed using FlowJo software (Treestar Inc, USA.). The level of HLA cell surface expression was observed using histograms. HLA expression on the surface was seen as shift in antibody stain when compared to control cells

For the HEK H3.3 K27M mutant cell line, a total of 3,880 peptides (8-12 mers) were identified, with 1,005 peptides restricted by HLA-A2 and the remaining 2,875 peptides bound to the other HLA class I allotypes (A*03:01, B*07:02, C*03:03). The length distribution of peptides identified in both HLA A*02:01 and pan class I eluates revealed that 9mers were the dominant peptide length (Figures 6a and 6b). The nonamers identified from HLA-A2 eluates matched the canonical binding motif (Figure 6c). The pan class I motif was dominated by a mixture of HLA A*03:01 and B*07:02 peptides (Figure 6d). Once again, the previously reported HLA-A2 neoepitope, RMSAPSTGGV, was not detected depsite identifying a total of 12 HLA bound peptides from the mutant H3.3K27M protein resulting in covering ~75% of the entire protein (Figure 7). Six of these peptides have previously been reported as HLA-A2 binders in the Immune epitope database (IEDB) (Abelin et al. 2017, Faridi et al. 2018, Vita et al. 2019).

Importantly, a novel 11mer peptide (RMSAPSTGGVK) encompassing the K27M mutation was identified in the pan HLA class I (*i.e.* W6/32) eluate. Further investigations were performed to define the HLA-restriction of this potential neoepitope. The peptides from the W6/32 eluate of HEK293 H3.3K27M cells was deconvolved using unsupervised Gibbs clustering into A*03:01 and B*07:02 binders (Figure 8). The neopeptide sequence

(RMSAPSTGGVK) matched the binding motif of HLA A*03:01 with Arg at P1 and Lys at P11 (Figure 8).

The presence of the RMSAPSTGGVK peptide was verified at the MS level, with the correct *m/z*, distinct isotopic peaks of high intensity eluting at a retention time of 20 mins. At the MS/MS level a full sequence ladder of b and y ions was generated (Figure 9). Moreover, this candidate neoepitope was detected only in the HEK H3.3K27M mutant cells and was absent from the WT cells. A modified version of the peptide with oxidation of Met at P2 (R**M**SAPSTGGVK) was also detected. This modification has been reported for the HLA-A2 peptide spanning this region of H3.3 (R**M**SAPSTGGV) (Chheda et al. 2018), with oxidation likely to have occurred during sample preparation. Together, these observations demonstrate that the novel HLA A*03:01-restricted H3.3K27M$_{26–36}$ peptide is naturally presented by HEK293 cells transfected with the H3.3K27M protein.

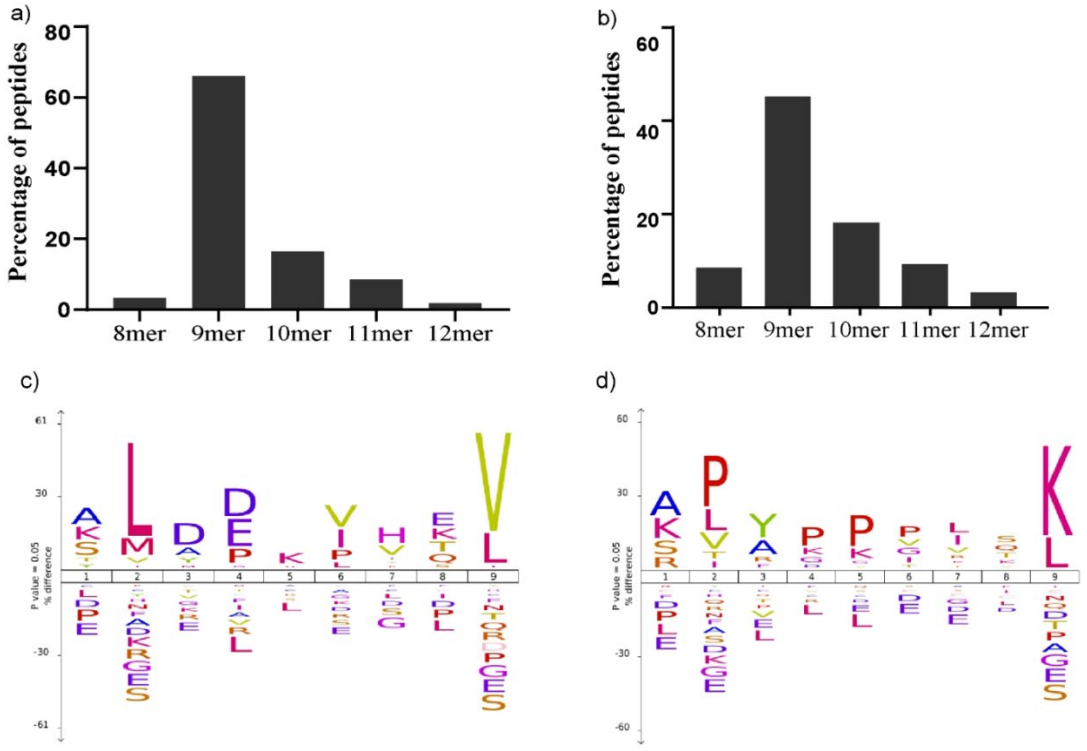


Figure 6 –Peptide length distribution and binding motifs of HEK293 H3.3K27M mutant cells. For peptide length distribution, the percentage of peptides was calculated for both a) HLA A*02:01 and b) pan HLA class I and plotted. The majority of peptides identified for all alleles were 9mers and their binding motif was interpreted using IceLogo (Colaert et al. 2009) for both c) HLA A*02:01 (n=1604) with a hydrophobic amino acid at P2 and P9 and d) pan HLA class I (n=1554) with a mixed motif observed comprised of A*03:01, B*07:02 and C*07:02, as evident from amino acids presented at P2 and P9.

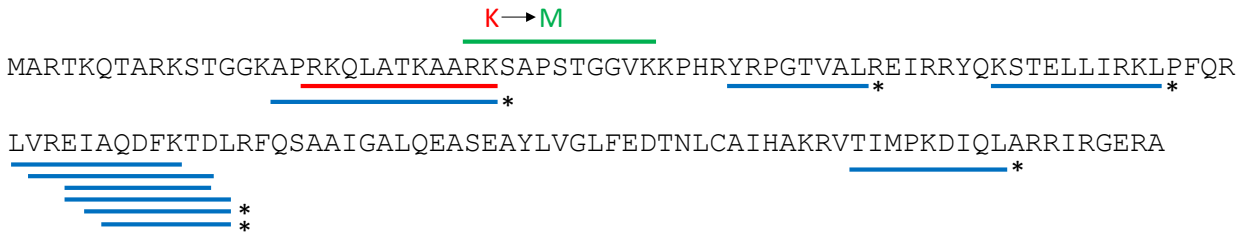


Figure 7 - H3.3 histone protein sequence and location of HLA-A2-restricted peptides derived from HEK293 cells (both WT and K27M mutant). The peptides underlined in blue are common to both WT and mutant histone, whilst the peptides marked with * are previously reported (verified in IEDB). The peptide underlined in red RKQLATKAARK was found exclusively in HEK293 WT cells. However, in the K27M mutant cells, peptide RMSAPSTGGVK is presented instead of the WT peptide.

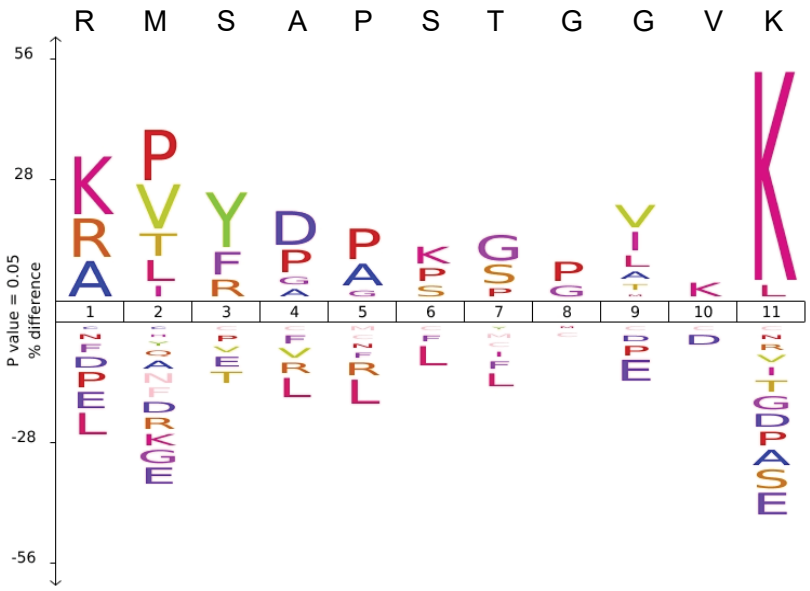
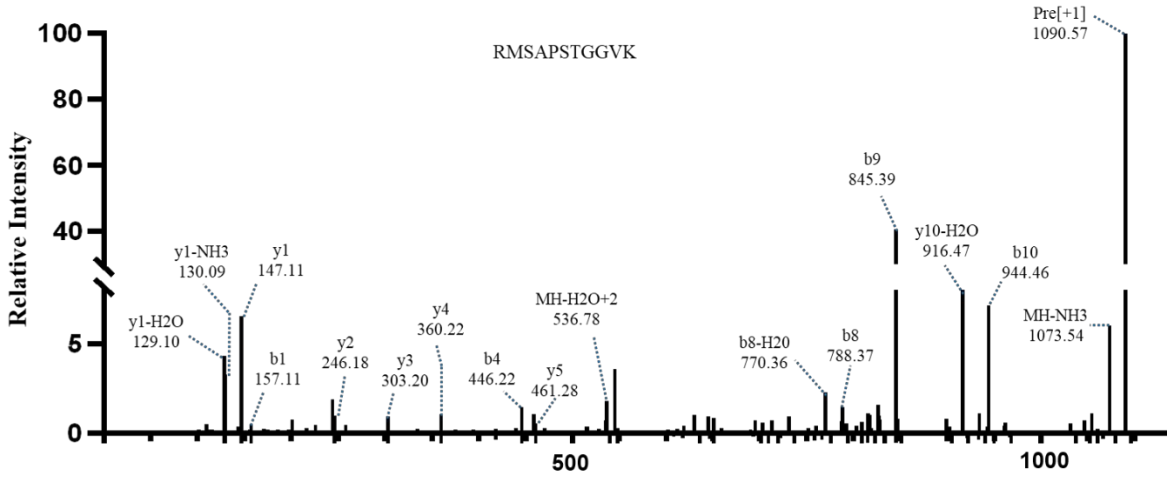


Figure 8 – A novel HLA A*03:01-restricted peptide from mutant H3.3K27M.
Gibbs clustering was performed for all 11mer peptides present in the pan HLA class I (W6/32) fraction derived from the HEK293 H3.3 K27M mutant cells. The data showed that the HLA mixed motif comprised of peptides restricted to either A*03:01 or B*07:02, as evident from the IceLogo representation of binding motif. The RMSAPSTGGVK neoepitope was detected in the A*03:01 cluster with amino acid matches at P1, P4, P5, P6 and P11.



Amino acid	b	b-H2O	b-NH3	b (2+)	y	y-H20	y-NH3	y(2+)
R	157.11	139.10	140.08	79.05				
M	288.15	270.14	271.12	144.57	934.47	916.47	917.44	467.73
S	375.18	375.18	358.15	188.09	803.43	785.42	786.40	402.24
A	446.22	428.21	429.19	223.61	716.40	698.38	699.37	358.70
P	543.27	525.33	526.24	272.14	645.36	627.35	628.33	323.18
S	630.31	612.30	613.28	315.65	548.30	530.29	531.28	274.65
T	731.36	713.34	714.32	366.18	461.28	443.26	444.24	231.15
G	788.38	770.36	771.35	394.69	360.23	342.21	343.20	180.61
G	845.39	827.38	828.37	423.20	303.21	285.19	286.18	151.79
V	944.46	926.45	927.44	472.28	246.18	228.17	229.15	123.59
K					147.11	129.10	130.09	74.06

Figure 9 – The MS/MS fragmentation pattern of the HLA A*03:01-restricted peptide RMSAPSTGGVK. X-axis represents mass over charge ratio (*m/z*) and y-axis represents the intensity of ions generated from HCD fragmentation of the peptide and sequence ladder of b and y ions identified for the neoepitope RMSAPSTGGVK.

5.3.7 Validation of HLA A*03:01-restricted neoepitope in the HEK293 system

To validate the HLA A*03:01 (HLA-A3) neoepitope (RMSAPSTGGVK) encompassing the H3.3K27M mutation we used a heavy labelled peptide approach similar to that used for the HLA-A2-rstricted determinant. The heavy peptide also carried a substitution at Pro in position 5 (13C5, 15N1), thereby introducing a 6 Da mass difference. A fixed amount of both HLA-A2 and -A3 heavy peptides were spiked into the BB7.2 and W6/32 fractions eluted from HEK H3.3K27M mutant, respectively and the samples were acquired on a Q-Exactive Plus LC-MS system. In the BB7.2 fractions, only the heavy HLA-A2 neoepitope was detected, with no trace of the endogenous peptide even at the MS1 level. This confirmed the absence of the A2 neoepitope in the HEK model. Importantly, we confidently identified the oxidised version of HLA-A3 neoepitope (R**Mox**SAPSTGGVK) in the W6/32 fraction (Figure 10a) as both the

endogenous and heavy peptides co-eluted together at 20 mins with *m/z* of 553.7848 (endogenous, Figure 10b) and 556.7917 (heavy labelled, Figure 10c) with ~ 6 copies/cell for the endogenous peptide.

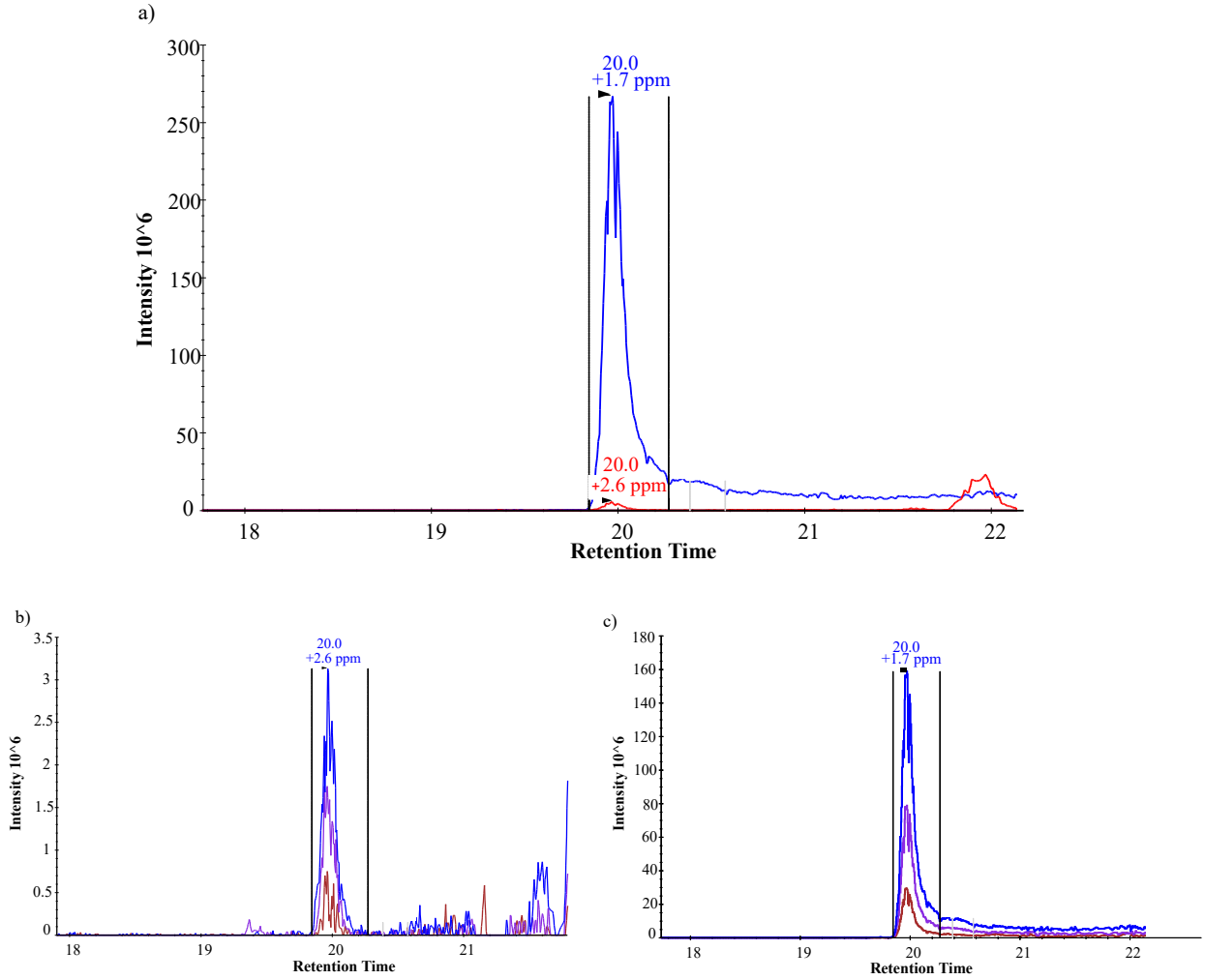


Figure 10 – HLA-A3-restricted RMoxSAPSTGGVK neoepitope is identified in in HEK293 H3.3K27M W6/32 fraction with co-elution of both endogenous and heavy peptides at 20.0 mins. Heavy RMoxSAPSTGGVK (heavy Pro in P5) was spiked into HEK293 H3.3K27M W6/32 fractions and the samples were acquired on the QE Plus. Extracted ion chromatograms (XICs) of precursors indicate in a) co-elution of both the heavy (blue) and the endogenous (red) peptides in the sample with same chromatographic retention times but *m/z* variable intensities. b) Shows three extracted molecular ion isotope peaks M, M+1, and M+2 for the endogenous peptide at *m/z* of 553.7848. c) Three extracted molecular ion isotopic peaks for the heavy peptide at *m/z* of 556.7917 with parts per million (ppm) for both the peptides. Figure was generated using Skyline software (64-bit v4.1.0) (Schilling et al. 2012)

5.3.8 Verification of the HLA-A3-restricted RMSAPSTGGVK neoepitope in a mono allelic system

Since the putative HLA-A3-restricted neoepitope (RMSAPSTGGVK) was identified in a polyallelic system (HEK293 cells) we wanted to confirm the presentation of the neoepitope by HLA-A3. Thus, we decided to investigate its presentation in a simpler, monoallelic CIR A*03:01 cell line transfected with H3.3K27M histone protein. The cells were grown under selection pressure of both G-418 and Puromycin to obtain cells expressing high levels of HLA A*03:01 (Figure 11a) and H3.3K27M mutant histone which was verified by western blot (Figure 11b). HLA class I expression of cell lines was measured routinely by flow cytometry by using hybridoma supernatants for both anti-HLA-A3 (GAPA3) and pan-class I (W6/32) and compared with C1R A*03:01 WT. Immunoaffinity purification for peptide HLA complexes was performed using the GAPA3 antibody.

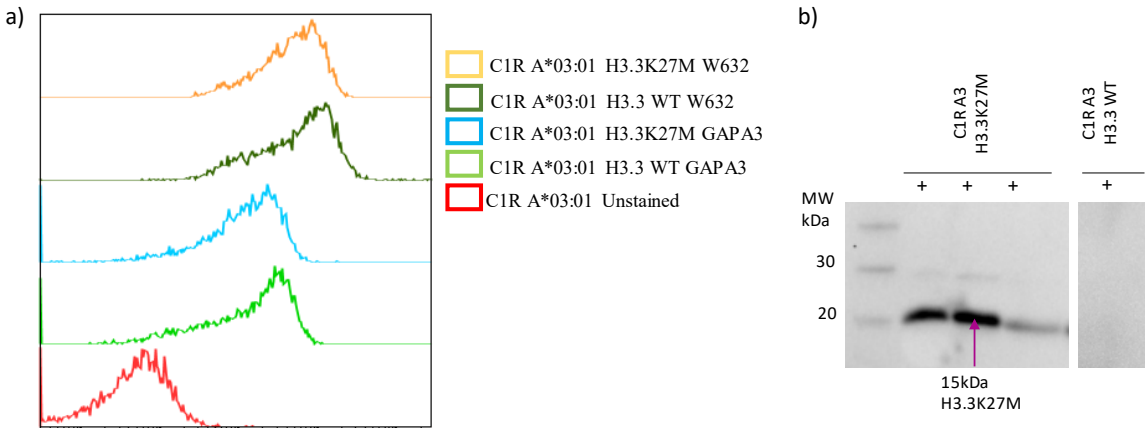


Figure 11 – Transfected C1R A*03:01 expresses high levels of HLA A*03:01 and H3.3K27M mutant histone. a) The HLA expression on C1R*03:01 H3.3K27M cell lines was checked routinely and compared with C1R A*03:01 WT using both anti-HLA-A3 antibody GAPA3 and anti-pan class I antibody W632 followed by secondary antibody (goat anti-mouse IgG PE). HLA expression on the surface was seen as shift in antibody stain when compared to control unstained cells. **b)** Expression of H3.3K27M was regularly checked in both C1R A*03:01 parental and C1R A*03:01 H3.3K27M cell lines by western blot.

A total of 9,935 HLA A*03:01 peptides were identified from the system which followed the canonical length distribution, with 48% being nonamers and remainder of peptides being longer (Figure 12a). Since the neoepitope identified in HEK293 H3.3K27M cells was an 11mer, all 11mer peptides were selected (n = 1,529) to determine their binding motif using Icelogo. These

peptides were observed to fit the HLA-A3 consensus binding motif with Lys/Arg at P1 and Lys at PΩ, with the RMSAPSTGVVK neoepitope being identified in this group (Figure 12b).

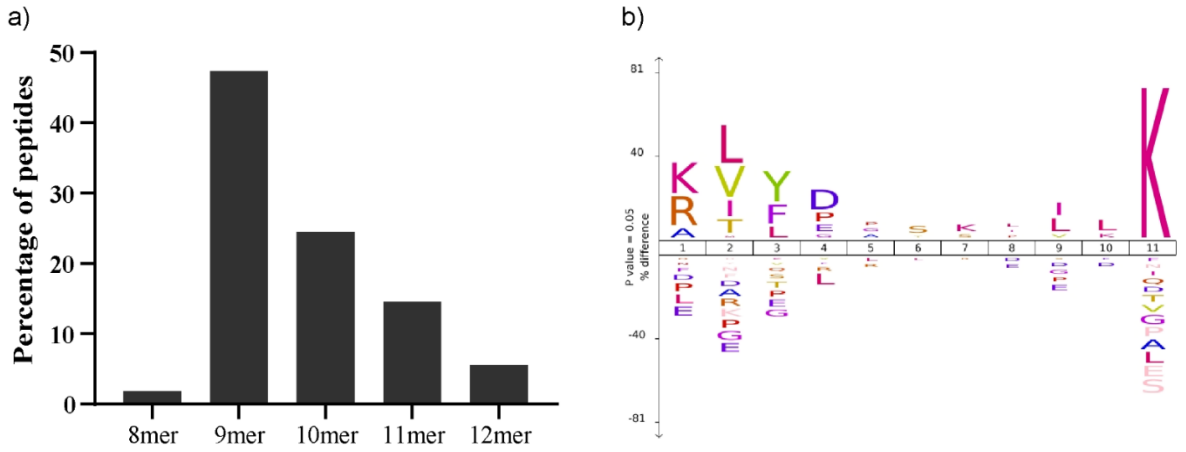


Figure 12 - Peptide length distribution and binding motif of peptides eluted from mono allelic C1R.A*03:01/H3.3K27M cell line. a) Percentage of peptides present in GAPA3 fraction followed the canonical length distribution seen for HLA A*03:01, wherein 48% of peptides were nonamers and remainder were longer peptides (10-12mers). b) The binding motif of endogenous 11mer peptides as determined from IceLogo also matched the known HLA-A3 motif.

The presence of the both the native and modified version of the HLA-A3-restricted putative neoepitope was verified at the MS1 and MS/MS level using heavy peptide standards. Both peptides were detected at MS1 level, with the P2 Met oxidised peptides eluting earlier at 19.5 mins (Figure 13a) with the *m/z* of 553.7873 (endogenous peptide) and 556.7917 (heavy labelled peptide). The native peptide co-eluted at 21.4 mins (Figure 13b) with *m/z* of 545.7873 (endogenous peptide) and 548.7942 (heavy peptide). The area under the curve (AUC) of heavy spiked in peptide being higher than the endogenous peptide in both instances with similar copy number for both native and endogenous peptides of ~3 copies of the pHLA/cell. All of the peptides were also detected using MS/MS (Figure 14a and 14b), with good fragmentation as evident from generation of b and y ion series, with the fragmentation spectra of endogenous peptides matching the synthetic heavy peptides.

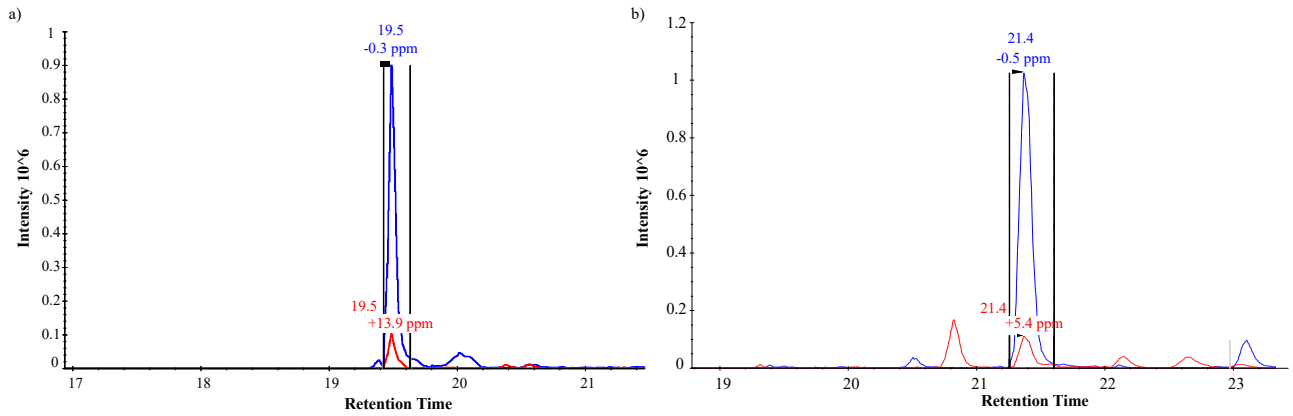


Figure 13 – Both peptide versions a) modified and b) native were identified in C1R.A*03:01/H3.3K27M cells. Extracted ion chromatograms (XICs) of precursors indicate a) co-elution of modified peptide RMoxSAPSTGGVK both the **heavy (blue)** and the **endogenous (red)** at 19.5 mins and b) co-elution of native peptide RMSAPSTGGVK both the heavy (blue) and the endogenous (red) at 21.4 mins.

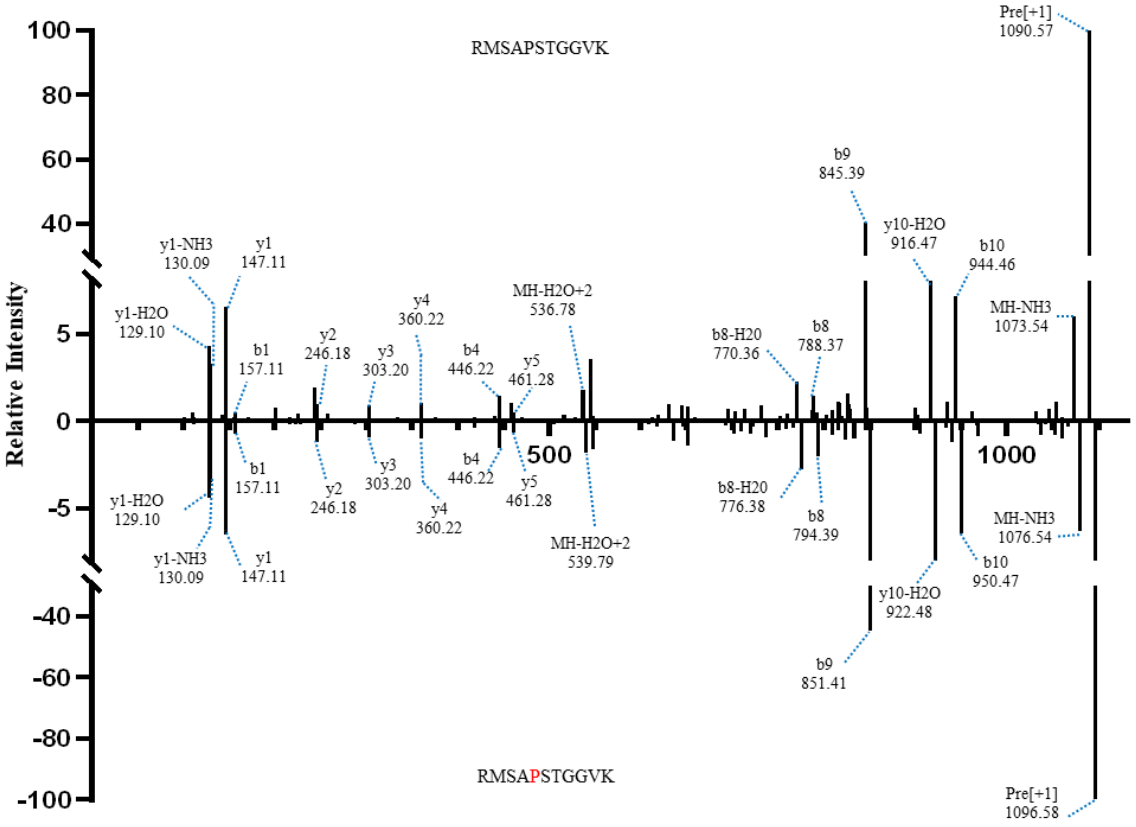


Figure 14a – Representative MS/MS fragmentation spectra of HLA A*03:01-restricted RMSAPSTGGVK (native version) peptide with a) spectra of endogenous peptide and b) spectra of synthetic Pro heavy labelled peptide with all b and y ions. X-axis represents mass over charge ratio (*m/z*) and y-axis representing the intensity of peptides

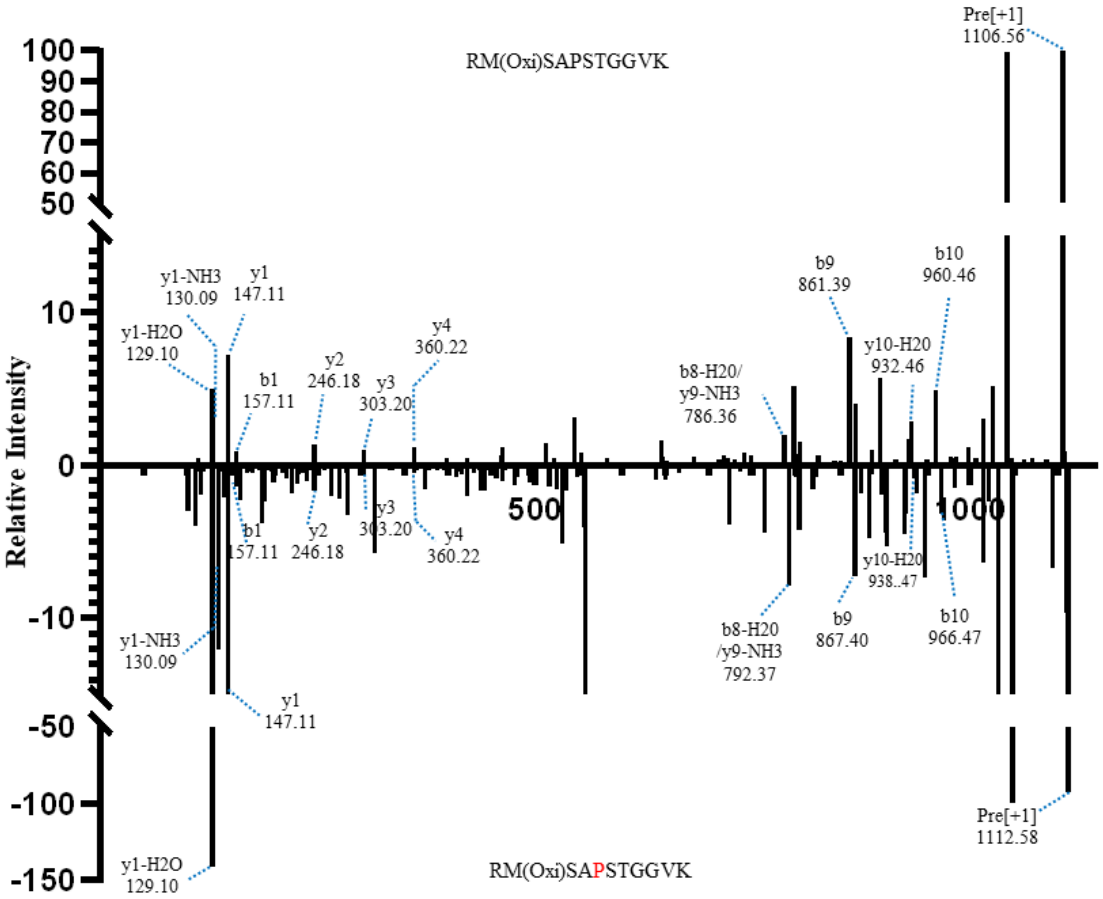


Figure 14b - Representative MS/MS fragmentation spectra of HLA A*03:01-restricted RMoxSAPSTGGVK (modified version) peptide with a) spectra of endogenous peptide and b) spectra of synthetic Pro heavy labelled peptide with all b and y ions. X-axis represents mass over charge ratio (*m/z*) and y-axis representing the intensity of peptides.

5.4 DISCUSSION

Recent immunopeptidomics studies undertaken in adult GBM and other cancers have demonstrated that direct analysis of the endogenous peptide repertoire of the primary tumour or patient derived cell lines aids in identification of tumour-associated peptides or neoepitopes associated with improved clinical outcome (Walter et al. 2012, Berlin et al. 2015, Neidert et al. 2018). The vast majority of DIPG studies have focussed on genomics, with only a few exploring their proteomic landscape (Saratsis et al. 2012, Saratsis et al. 2014) and to date there has been no study looking into the endogenous peptide repertoire presented by HLA class I molecules. Hence immunotherapy, particularly T cell mediated therapy, remains untranslated at the clinical level due to a dearth of DIPG specific peptides. The current study bridges this missing knowledge by identifying different types TAA and CTA-derived peptides as well as

neoantigen derived peptides, via high-resolution MS, that can be used in rational design of DIPG targeted immunotherapeutics.

There are several challenging aspects to DIPG research. First, the cell lines grow as neurospheres, which have slow doubling time and hence require prolonged *in vitro* culturing periods. Second, unlike most cancer lines which have high or medium HLA expression, these cell lines express HLA at very low levels. Despite these challenges, we successfully identified 21,101 unique endogenous HLA class I peptides (at 5% FDR) displayed on the surface of the DIPG cell lines. Importantly, we identified peptides restricted to not only HLA A*02:01, but other frequent class I alleles thereby covering most prominent HLA class I alleles found in the population. This can be particularly helpful as the peptides identified are applicable for HLA mediated immunotherapy for DIPG.

Our DIPG immunopeptidomics dataset was compared with a previous GBM immunopeptidomics study by Neidert *et al* (Neidert et al. 2018) which comprised of HLA restricted peptides identified in glioblastoma stem cell lines (GSC), GBM patients autopsy samples and healthy brain samples. There was an overlap of 1950 peptides between the DIPG data and combined GSC and GBM patient data (Supplementary Figure 2a). When compared to the healthy adult brain immunopeptidome, there was an overlap of only 554 peptides (Supplementary Figure 2b). This reiterates that paediatric brain is quite distinct to adult brain and tumours. To identify DIPG specific peptides we compared our data to a proteomic study by Saratsis *et al*. (Saratsis et al. 2014) wherein they identified 2,061 proteins from 14 formalin fixed paraffin embedded (FFPE) DIPG samples. When compared with our immunopeptidome source protein datasets, an overlap of 1,257 proteins was found (Supplementary Figure 3a), with a maximum overlap of 891 proteins in SF7761 (Supplementary Figure 3b) and the least overlap with ~450 to 460 proteins in H3.3 WT cell lines DIPG43 and DIPG58 (Supplementary Figure 3c and 3d), respectively. Since cerebrospinal fluid (CSF) is most intimately associated with the brain, our data was also compared to another study (Saratsis et al. 2012), wherein whole proteome analysis of CSF from DIPG patients resulted in identification of 507 unique proteins. The proteome was divided into three subsets, a) 177 proteins found exclusively in the tumour, b) 95 upregulated CSF proteins and c) 235 downregulated CSF proteins. In our DIPG immunopeptidome dataset, a total of 110 proteins were found to be overlapping with tumour only proteins and 51 proteins overlapping with upregulated CSF proteins and 63 proteins were found overlapping with the downregulated CSF protein category. The presence of HLA class I restricted peptides from these source proteins provides evidence that our in depth

immunopeptidomics study has yielded valid peptide targets which can explored further for immunotherapy.

The aim of our study was to identify TAA- and CTA-derived peptides which might be used for T cell mediated immunotherapy. Excitingly we identified 1,123 peptides, out of which 304 and 11 were HLA A*02:01-restricted TAA and CTA-derived peptide antigens, respectively. The frequency of CTAs in the dataset was quite low consistent with previous studies (Dutoit et al. 2012). Some proteins of interest, whose peptides have previously been shown to be immunogenic in glioblastoma patients, included Interleukin-13 receptor subunit alpha-2 (IL13Rα) in DIPG58 restricted by HLA A*01:01. The role of IL13Rα2-targeted chimeric antigen receptor (CAR) T cells is well documented in glioblastoma tumour regression (Brown et al. 2016), supporting the utility of this protein as a potential for DIPG targeted immunotherapy. In addition, other TAA-derived peptides were derived from other well-known antigen sources including Insulin Growth factor 2 binding protein (IGF2BP), Nucleophosmin (NPM) and STAGA complex 65 subunit gamma (ST65G), which are found specifically in gliomas were also identified. Interestingly, the H3.3 WT cell lines (DIPG43 and DIP58) had very low numbers of both TAA-derived and CTA-derived peptides, which could be due to either a lower peptide repertoire (when compared to H3.3 K27M cell lines) or more likely is that the H3.3K27M histone mutation tends to upregulate expression of TAAs and CTAs (Deng et al. 2018)

The key to cancer immunotherapy is identification of immunogenic neoepitopes found exclusively in the cancer. The H3.3K27M is a tumour-specific and highly penetrant mutation reportedly found in up to 70% of DIPG cases (Khuong-Quang et al. 2012), representing an attractive target for immunotherapy. Since the previously reported HLA-A2-restricted R**M**SAPSTGGV neoepitope was not detected in our global LC-MS/MS analysis of DIPG primary cell lines, we used a sensitive MRM approach to detect this neoepitope with improved selectivity and sensitivity over LC-MS/MS. Despite these more stringent investigations, the HLA-A2 neoepitope was still not detected. This might be attributed to either the low copy number of the A2 peptide in the HEK293 H3.3K27M system (as the A3 neoepitope itself had ~ 6 copies/cell) or the small sample size which might have not allowed for in-depth exploration of the immunopeptide of DIPG cell lines. Interestingly, the HLA-A2 restricted neoepitope was also not detected in the stably transfected HLA-A2 expressing HEK293 H3.3K27M mutant

cell line despite high HLA A*02:01 expression as evident from FACS analysis and the detection of several other H3.3 derived peptides.

Excitingly our investigations have led to the identification of a novel HLA A*03:01 restricted candidate neoepitope, RMSAPSTGGVK. This 11mer peptide was identified in the pan class I fraction of HEK293 H3.3K27M. It was further validated using an H3.3K27M transfected monoallelic cell line, C1R.A3/H3.3K27M and spiked heavy synthetic peptide standards. Whilst immunogenicity studies are forthcoming, we believe the identification of this DIPG-derived neoepitope provides a framework for T cell-mediated immunotherapeutics not only to HLA A*03:01, but potentially also for other allotypes belonging to HLA-A3 supertype (includes A*11:01, A*30:01 and A*74:01), as this peptide was predicted to be a strong binder to these related allotypes using NetMHC Pan 4.0. We also interrogated our immunopeptidomics dataset for neoepitopes arising from other mutations including TP53 and PI3K, but no additional neoepitopes were identified.

To summarise, by adopting the HLA-centric antigen discovery approach implemented in our study, we have for the first time identified the repertoire of targetable antigens for DIPG. This includes not only a novel HLA A*03:01-restricted H3.3K27M neoepitope but also an assortment of HLA A*02:01-restricted TAA- and CTA-derived peptide antigens. Further validation of the potential of these peptides to be recognised as immunogenic T cell targets is required to formulate a T cell-based immunotherapeutic for DIPG.

REFRENCES

Abbey, J. L. and H. C. O'Neill (2007). "Expression of T-cell receptor genes during early T-cell development." Immunol Cell Biol **86**(2): 166-174.

Abelin, J. G., et al. (2017). "Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction." Immunity **46**(2): 315-326.

Abelin, J. G., et al. (2015). "Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry." Nat Protoc **10**(9): 1308-1318.

Adams, P. D., et al. (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallographica Section D **66**(2): 213-221.

Aki, M., et al. (1994). "Interferon-gamma induces different subunit organizations and functional diversity of proteasomes." J Biochem **115**(2): 257-269.

Alatrash, G., et al. (2018). "Targeting the Leukemia Antigen PR1 with Immunotherapy for the Treatment of Multiple Myeloma." Clinical Cancer Research **24**(14): 3386-3396.

Albitar, M., et al. (2007). "Levels of soluble HLA-I and β2M in patients with acute myeloid leukemia and advanced myelodysplastic syndrome: association with clinical behavior and outcome of induction therapy." Leukemia **21**(3): 480-488.

Almeida, L. G., et al. (2009). "CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens." Nucleic Acids Res **37**(Database issue): D816-819.

Amir, A. L., et al. (2011). "PRAME-specific Allo-HLA-restricted T cells with potent antitumor reactivity useful for therapeutic T-cell receptor gene transfer." Clin Cancer Res **17**(17): 5615-5625.

Arango Duque, G. and A. Descoteaux (2014). "Macrophage Cytokines: Involvement in Immunity and Infectious Diseases." Front Immunol **5**.

Arvelo, F., et al. (2016). "Tumour progression and metastasis." Ecancermedicalscience **10**: 617.

Backert, L., et al. (2017). "A meta-analysis of HLA peptidome composition in different hematological entities: entity-specific dividing lines and "pan-leukemia" antigens." Oncotarget **8**(27): 43915-43924.

Barrett, J. C. (1993). "Mechanisms of Multistep Carcinogenesis and Carcinogen Risk Assessment." Environmental Health Perspectives **100**: 9-20.

Bassani-Sternberg, M., et al. (2010). "Soluble plasma HLA peptidome as a potential source for cancer biomarkers." Proc Natl Acad Sci U S A **107**(44): 18769-18776.

Belov, A. A. and M. Mohammadi (2012). "Grb2, a double-edged sword of receptor tyrosine kinase signaling." Sci Signal **5**(249): pe49.

Bennett, J. M., et al. (1976). "Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group." Br J Haematol **33**(4): 451-458.

Berlin, C., et al. (2015). "Mapping the HLA ligandome landscape of acute myeloid leukemia: a targeted approach toward peptide-based immunotherapy." Leukemia **29**(3): 647-659.

Bilich, T., et al. (2019). "The HLA ligandome landscape of chronic myeloid leukemia delineates novel T-cell epitopes for immunotherapy." Blood **133**(6): 550-565.

Birch, N. W. and A. Shilatifard (2020). "The role of histone modifications in leukemogenesis." Journal of Biosciences **45**(1): 6.

Bjorkman, P. J., et al. (1987). "Structure of the human class I histocompatibility antigen, HLA-A2." Nature **329**(6139): 506-512.

Blum, J. S. (2013). "Pathways of Antigen Processing." **31**: 443-473.

Bodmer, W. F. (1987). "The HLA system: structure and function." J Clin Pathol **40**(9): 948-958.

Bosman, F. T. (2019). Integrative Molecular Tumor Classification: A Pathologist's View. Encyclopedia of Cancer (Third Edition). P. Boffetta and P. Hainaut. Oxford, Academic Press**:** 279-285.
Bray, F., et al. (2018). "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA: A Cancer Journal for Clinicians **68**(6): 394-424.
Brodsky, F. M. and P. Parham (1982). "Monomorphic anti-HLA-A,B,C monoclonal antibodies detecting molecular subunits and combinatorial determinants." J Immunol **128**(1): 129-135.
Brown, C. E., et al. (2016). "Regression of Glioblastoma after Chimeric Antigen Receptor T-Cell Therapy." New England Journal of Medicine **375**(26): 2561-2569.
Buczkowicz, P., et al. (2014). "Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations." Nat Genet **46**(5): 451-456.
Campoli, M. and S. Ferrone (2008). "HLA antigen changes in malignant cells: epigenetic mechanisms and biologic significance." Oncogene **27**(45): 5869-5885.
Carreno, B. M., et al. (2015). "Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells." Science **348**(6236): 803-808.
Chase, A. and N. C. Cross (2006). "Signal transduction therapy in haematological malignancies: identification and targeting of tyrosine kinases." Clin Sci (Lond) **111**(4): 233-249.
Chen, Daniel S. and I. Mellman (2013). "Oncology Meets Immunology: The Cancer-Immunity Cycle." Immunity **39**(1): 1-10.
Chheda, Z. S., et al. (2018). "Novel and shared neoantigen derived from histone 3 variant H3.3K27M mutation for glioma T cell therapy." J Exp Med **215**(1): 141-157.
Chothia, C., et al. (1988). "The outline structure of the T-cell alpha beta receptor." The EMBO Journal **7**(12): 3745-3755.
Clements, C. S., et al. (2002). "The production, purification and crystallization of a soluble heterodimeric form of a highly selected T-cell receptor in its unliganded and liganded state." Acta Crystallographica Section D **58**(12): 2131-2134.
Cobbold, M., et al. (2013). "MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia." Sci Transl Med **5**(203): 203ra125.
Colaert, N., et al. (2009). "Improved visualization of protein consensus sequences by iceLogo." Nat Methods **6**(11): 786-787.
Collaborative (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallographica Section D **50**(5): 760-763.
Collins, E. J. and D. S. Riddle (2008). "TCR-MHC docking orientation: natural selection, or thymic selection?" Immunologic Research **41**(3): 267.
Connors, J. M., et al. (2018). "Brentuximab Vedotin with Chemotherapy for Stage III or IV Hodgkin's Lymphoma." N Engl J Med **378**(4): 331-344.
Couch, D. B. (1996). "Carcinogenesis: basic principles." Drug Chem Toxicol **19**(3): 133-148.
Cresswell, P. (2013). "Pathways of Antigen Processing." Annu Rev Immunol.
Cresswell, P., et al. (1999). "The nature of the MHC class I peptide loading complex." Immunol Rev **172**: 21-28.
Cresswell, P., et al. (1973). "Papain-Solubilized HL-A Antigens from Cultured Human Lymphocytes Contain Two Peptide Fragments." Proceedings of the National Academy of Sciences **70**(5): 1603-1607.
Dahlén, E., et al. (2018). "Bispecific antibodies in cancer immunotherapy." Ther Adv Vaccines Immunother **6**(1): 3-17.
Davis, M. M. and P. J. Bjorkman (1988). "T-cell antigen receptor genes and T-cell recognition." Nature **334**: 395.

De Kouchkovsky, I. and M. Abdul-Hay (2016). "'Acute myeloid leukemia: a comprehensive review and 2016 update'." Blood Cancer J **6**(7): e441.
Demine, R. and P. Walden (2005). "Testing the Role of gp96 as Peptide Chaperone in Antigen Processing." Journal of Biological Chemistry **280**(18): 17573-17578.
Deng, H., et al. (2018). "Histone H3.3K27M Mobilizes Multiple Cancer/Testis (CT) Antigens in Pediatric Glioma." Mol Cancer Res **16**(4): 623-633.
Depontieu, F. R., et al. (2009). "Identification of tumor-associated, MHC class II-restricted phosphopeptides as targets for immunotherapy." Proc Natl Acad Sci U S A **106**(29): 12073-12078.
DiNardo, C. D. and J. E. Cortes (2016). "Mutations in AML: prognostic and therapeutic implications." Hematology Am Soc Hematol Educ Program **2016**(1): 348-355.
Döhner, H., et al. (2015). "Acute Myeloid Leukemia." New England Journal of Medicine **373**(12): 1136-1152.
Dong, G., et al. (2009). "Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer." Immunity **30**(1): 21-32.
Drake, J. M., et al. (2012). "Oncogene-specific activation of tyrosine kinase networks during prostate cancer progression." Proceedings of the National Academy of Sciences **109**(5): 1643-1648.
Dreno, B., et al. (2018). "MAGE-A3 immunotherapeutic as adjuvant therapy for patients with resected, MAGE-A3-positive, stage III melanoma (DERMA): a double-blind, randomised, placebo-controlled, phase 3 trial." The Lancet Oncology **19**(7): 916-929.
Driscoll, J., et al. (1993). "MHC-linked LMP gene products specifically alter peptidase activities of the proteasome." Nature **365**(6443): 262-264.
Dudley, D. D., et al. (2005). "Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences." Adv Immunol **86**: 43-112.
Dufva, O., et al. (2020). "Immunogenomic Landscape of Hematological Malignancies." Cancer Cell.
Dutoit, V., et al. (2012). "Exploiting the glioblastoma peptidome to discover novel tumour-associated antigens for immunotherapy." Brain **135**(Pt 4): 1042-1054.
Elliott, T., et al. (1990). "Naturally processed peptides." Nature **348**(6298): 195-196.
Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallographica Section D **60**(12 Part 1): 2126-2132.
Escher, C., et al. (2012). "Using iRT, a normalized retention time for more targeted measurement of peptides." Proteomics **12**(8): 1111-1121.
Estey, E. (2016). "Acute myeloid leukemia: 2016 Update on risk-stratification and management." Am J Hematol **91**(8): 824-846.
Estey, E. H. (2009). "Treatment of acute myeloid leukemia." Haematologica **94**(1): 10-16.
Estey, E. H. (2018). "Acute myeloid leukemia: 2019 update on risk-stratification and management." Am J Hematol **93**(10): 1267-1291.
Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallographica Section D **62**(1): 72-82.
Evavold, B. D. and P. M. Allen (1991). "Separation of IL-4 production from Th cell proliferation by an altered T cell receptor ligand." Science **252**(5010): 1308-1310.
Evavold, B. D., et al. (1993). "Tickling the TCR: selective T-cell functions stimulated by altered peptide ligands." Immunology Today **14**(12): 602-609.
Falini, B., et al. (2005). "Cytoplasmic Nucleophosmin in Acute Myelogenous Leukemia with a Normal Karyotype." New England Journal of Medicine **352**(3): 254-266.
Falk, K., et al. (1991). "Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules." Nature **351**(6324): 290-296.

Faridi, P., et al. (2018). "In Immunopeptidomics We Need a Sniper Instead of a Shotgun." Proteomics **18**(12).
Faure, O., et al. (2004). "Inducible Hsp70 as target of anticancer immunotherapy: Identification of HLA-A*0201-restricted epitopes." Int J Cancer **108**(6): 863-870.
Fazio, V. J., et al. (2014). "A drunken search in crystallization space." Acta Crystallogr F Struct Biol Commun **70**(Pt 10): 1303-1311.
Ferguson, T. A., et al. (2011). "Armed response: how dying cells influence T-cell functions." Immunol Rev **241**(1): 77-88.
Filbin, M. G., et al. (2018). "Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq." Science **360**(6386): 331-335.
Fitzmaurice, C., et al. (2017). "Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study." JAMA Oncol **3**(4): 524-548.
Gaczynska, M., et al. (1993). "Gamma-interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes." Nature **365**(6443): 264-267.
Gajewski, T. F., et al. (2006). "Immune resistance orchestrated by the tumor microenvironment." Immunol Rev **213**: 131-145.
Gambacorti-Passerini, C., et al. (1993). "Human CD4 lymphocytes specifically recognize a peptide representing the fusion region of the hybrid protein pml/RAR alpha present in acute promyelocytic leukemia cells." Blood **81**(5): 1369-1375.
Garboczi, D. N., et al. (1996). "Structure of the complex between human T-cell receptor, viral peptide and HLA-A2." Nature **384**(6605): 134-141.
Garcia, K. C., et al. (1996). "An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex." Science **274**(5285): 209-219.
Gordon, S. (2002). "Pattern recognition receptors: doubling up for the innate immune response." Cell **111**(7): 927-930.
Graf, C., et al. (2007). "A neoepitope generated by an FLT3 internal tandem duplication (FLT3-ITD) is recognized by leukemia-reactive autologous CD8+ T cells." Blood **109**(7): 2985-2988.
Grafone, T., et al. (2012). "An overview on the role of FLT3-tyrosine kinase receptor in acute myeloid leukemia: biology and treatment." Oncol Rev **6**(1): e8.
Grimwade, D., et al. (1998). "The Importance of Diagnostic Cytogenetics on Outcome in AML: Analysis of 1,612 Patients Entered Into the MRC AML 10 Trial." Blood **92**(7): 2322-2333.
Guillaume, P., et al. (2018). "The C-terminal extension landscape of naturally presented HLA-I ligands." Proceedings of the National Academy of Sciences **115**(20): 5083-5088.
Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.
Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.
Harris, N. L., et al. (1999). "World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997." J Clin Oncol **17**(12): 3835-3849.
He, X., et al. (2019). "PRMT1-mediated FLT3 arginine methylation promotes maintenance of FLT3-ITD+ acute myeloid leukemia." Blood **134**(6): 548-560.
Heinzel, S., et al. (2001). "The self peptide annexin II (208-223) presented by dendritic cells sensitizes autologous CD4+ T lymphocytes to recognize melanoma cells." Cancer Immunol Immunother **49**(12): 671-678.
Hollingsworth, R. E. and K. Jansen (2019). "Turning the corner on therapeutic cancer vaccines." NPJ Vaccines **4**: 7.

Hornbeck, P. V., et al. (2012). "PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse." Nucleic Acids Res **40**(Database issue): D261-270.
Hou, H. A., et al. (2016). "Splicing factor mutations predict poor prognosis in patients with de novo acute myeloid leukemia." Oncotarget **7**(8): 9084-9101.
Housset, D. and B. Malissen (2003). "What do TCR–pMHC crystal structures teach us about MHC restriction and alloreactivity?" Trends in Immunology **24**(8): 429-437.
Hung, C. F., et al. (2008). "Therapeutic human papillomavirus vaccines: current clinical trials and future directions." Expert Opin Biol Ther **8**(4): 421-439.
Hunt, J. S., et al. (2005). "HLA-G and immune tolerance in pregnancy." The FASEB Journal **19**(7): 681-693.
Jarmalavicius, S., et al. (2010). "Differential arginine methylation of the G-protein pathway suppressor GPS-2 recognized by tumor-specific T cells in melanoma." Faseb j **24**(3): 937-946.
Jones, C. and S. J. Baker (2014). "Unique genetic and epigenetic mechanisms driving signatures of paediatric diffuse high-grade glioma." Nat Rev Cancer **14**(10).
Jorgensen, J. L., et al. (1992). "Molecular components of T-cell recognition." Annu Rev Immunol **10**: 835-873.
Juratli, T. A., et al. (2018). "Molecular pathogenesis and therapeutic implications in pediatric high-grade gliomas." Pharmacology & Therapeutics **182**: 70-79.
Jurtz, V., et al. (2017). "NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data." J Immunol **199**(9): 3360-3368.
Kabsch, W. (2010). "XDS." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.
Kantarjian, H. M., et al. (1992). "Prognostic significance of elevated serum beta 2-microglobulin levels in adult acute lymphocytic leukemia." Am J Med **93**(6): 599-604.
Keskinen, P., et al. (1997). "Regulation of HLA class I and II expression by interferons and influenza A virus in human peripheral blood mononuclear cells." Immunology **91**(3): 421-429.
Khuong-Quang, D. A., et al. (2012). "K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas." Acta Neuropathol **124**(3): 439-447.
Kjer-Nielsen, L., et al. (2002). "The 1.5 Å Crystal Structure of a Highly Selected Antiviral T Cell Receptor Provides Evidence for a Structural Basis of Immunodominance." Structure **10**(11): 1521-1532.
Klar, R., et al. (2014). "Therapeutic targeting of naturally presented myeloperoxidase-derived HLA peptide ligands on myeloid leukemia cells by TCR-transgenic T cells." Leukemia **28**(12): 2355-2366.
Knutson, K. L. and E. A. Mittendorf (2015). "Cancer vaccines in the new era of cancer immunotherapy." Vaccine **33**(51): 7376.
Korshunov, A., et al. (2015). "Integrated analysis of pediatric glioblastoma reveals a subset of biologically favorable tumors with associated molecular prognostic markers." Acta Neuropathol **129**(5): 669-678.
Kowalewski, D. J., et al. (2015). "HLA ligandome analysis identifies the underlying specificities of spontaneous antileukemia immune responses in chronic lymphocytic leukemia (CLL)." Proceedings of the National Academy of Sciences **112**(2): E166-E175.
Krishnadas, D. K., et al. (2015). "A phase I trial combining decitabine/dendritic cell vaccine targeting MAGE-A1, MAGE-A3 and NY-ESO-1 for children with relapsed or therapy-refractory neuroblastoma and sarcoma." Cancer Immunol Immunother **64**(10): 1251-1260.
Krueger, K. E. and S. Srivastava (2006). "Posttranslational protein modifications: current implications for cancer detection, prevention, and therapeutics." Mol Cell Proteomics **5**(10): 1799-1810.

Kumar, C. C. (2011). "Genetic Abnormalities and Challenges in the Treatment of Acute Myeloid Leukemia." Genes Cancer **2**(2): 95-107.
Lee, Y. J., et al. (1997). "TGF-beta suppresses IFN-gamma induction of class II MHC gene expression by inhibiting class II transactivator messenger RNA expression." The Journal of Immunology **158**(5): 2065-2075.
Lin, M. H., et al. (2019). "Immunological evaluation of a novel HLA-A2 restricted phosphopeptide of tumor associated Antigen, TRAP1, on cancer therapy." Vaccine X **1**: 100017.
Louis, D. N., et al. (2016). "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary." Acta Neuropathol **131**(6): 803-820.
Maclean, B., et al. (2010). "Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry." Anal Chem **82**(24): 10116-10124.
Maecker, B., et al. (2005). "Identification of a new HLA-A*0201-restricted cryptic epitope from CYP1B1." International Journal of Cancer **115**(2): 333-336.
Maeda, Y., et al. (2002). "Detection of peptide-specific CTL-precursors in peripheral blood lymphocytes of cancer patients." Br J Cancer **87**(7): 796-804.
Malaker, S. A., et al. (2017). "Identification of Glycopeptides as Posttranslationally Modified Neoantigens in Leukemia." Cancer Immunol Res **5**(5): 376-384.
Maleno, I., et al. (2011). "Frequent loss of heterozygosity in the β2-microglobulin region of chromosome 15 in primary human tumors." Immunogenetics **63**(2): 65-71.
Mann, M. and O. N. Jensen (2003). "Proteomic analysis of post-translational modifications." Nature Biotechnology **21**(3): 255-261.
Manning, G., et al. (2002). "Evolution of protein kinase signaling from yeast to man." Trends Biochem Sci **27**(10): 514-520.
Marino, F., et al. (2017). "Arginine (Di)methylated Human Leukocyte Antigen Class I Peptides Are Favorably Presented by HLA-B*07." Journal of Proteome Research **16**(1): 34-44.
Mastelic-Gavillet, B., et al. (2019). "Personalized Dendritic Cell Vaccines—Recent Breakthroughs and Encouraging Clinical Results." Frontiers in Immunology **10**(766).
Mathew, R. K. and J. T. Rutka (2018). "Diffuse Intrinsic Pontine Glioma : Clinical Features, Molecular Genetics, and Novel Targeted Therapeutics." J Korean Neurosurg Soc **61**(3): 343-351.
Maupin-Furlow, J. (2011). "Proteasomes and protein conjugation across domains of life." Nat Rev Microbiol **10**(2): 100-111.
McCoy, A. J., et al. (2007). "Phaser crystallographic software." Journal of Applied Crystallography **40**(4): 658-674.
McGranahan, N., et al. (2017). "Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution." Cell **171**(6): 1259-1271.e1211.
Medzhitov, R. and C. A. Janeway (2002). "Decoding the Patterns of Self and Nonself by the Innate Immune System." Science **296**(5566): 298-300.
Mei, S., et al. (2020). "Immunopeptidomic analysis reveals that deamidated HLA-bound peptides arise predominantly from deglycosylated precursors." Molecular & Cellular Proteomics: mcp.RA119.001846.
Meyer, V. S., et al. (2009). "Identification of natural MHC class II presented phosphopeptides and tumor-derived MHC class I phospholigands." J Proteome Res **8**(7): 3666-3674.
Mikhail, F. M., et al. (2006). "Normal and transforming functions of RUNX1: A perspective." Journal of Cellular Physiology **207**(3): 582-593.
Miyamoto, T., et al. (2002). "Myeloid or Lymphoid Promiscuity as a Critical Step in Hematopoietic Lineage Commitment." Developmental Cell **3**(1): 137-147.

Mohammed, F., et al. (2008). "Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self." Nat Immunol **9**(11): 1236-1243.
Mohammed, F., et al. (2017). "The antigenic identity of human class I MHC phosphopeptides is critically dependent upon phosphorylation status." Oncotarget **8**(33): 54160-54172.
Mommen, G. P. M., et al. (2014). "Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD)." Proceedings of the National Academy of Sciences **111**(12): 4507-4512.
Morra, E., et al. (2009). "Clinical management of primary non-acute promyelocytic leukemia acute myeloid leukemia: Practice Guidelines by the Italian Society of Hematology, the Italian Society of Experimental Hematology, and the Italian Group for Bone Marrow Transplantation." Haematologica **94**(1): 102-112.
Muenst, S., et al. (2016). "The immune system and cancer evasion strategies: therapeutic concepts." J Intern Med **279**(6): 541-562.
Narayan, R., et al. (2019). "Acute myeloid leukemia immunopeptidome reveals HLA presentation of mutated nucleophosmin." PLoS One **14**(7): e0219547.
Neidert, M. C., et al. (2018). "The natural HLA ligandome of glioblastoma stem-like cells: antigen discovery for T cell-based immunotherapy." Acta Neuropathologica **135**(6): 923-938.
Nie, Y., et al. (2001). "DNA hypermethylation is a mechanism for loss of expression of the HLA class I genes in human esophageal squamous cell carcinomas." Carcinogenesis **22**(10): 1615-1623.
Niesen, F. H., et al. (2007). "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." Nature Protocols **2**(9): 2212-2221.
Obara, W., et al. (2018). "Present status and future perspective of peptide-based vaccine therapy for urological cancer." Cancer Science **109**(3): 550-559.
Ochs, K., et al. (2017). "K27M-mutant histone-3 as a novel target for glioma immunotherapy." Oncoimmunology **6**(7): e1328340.
Ogawa, S. (2014). "Splicing factor mutations in AML." Blood **123**(21): 3216-3217.
Oliveros, J. "VENNY. An interactive tool for comparing lists with Venn diagrams." BioinfoGP, CNB-CSIC.
Olsen, J. V. and M. Mann (2013). "Status of large-scale analysis of post-translational modifications by mass spectrometry." Mol Cell Proteomics **12**(12): 3444-3452.
Olsen, L. R., et al. (2017). "TANTIGEN: a comprehensive database of tumor T cell antigens." Cancer Immunol Immunother **66**(6): 731-735.
Pandey, K., et al. (2020). "In-depth mining of the immunopeptidome of an acute myeloid leukemia cell line using complementary ligand enrichment and data acquisition strategies." Mol Immunol **123**: 7-17.
Panter, M. S., et al. (2012). "Dynamics of major histocompatibility complex class I association with the human peptide-loading complex." J Biol Chem **287**(37): 31172-31184.
Papaemmanuil, E., et al. (2016). "Genomic Classification and Prognosis in Acute Myeloid Leukemia." N Engl J Med **374**(23): 2209-2221.
Parham, P. and F. M. Brodsky (1981). "Partial purification and some properties of BB7.2. A cytotoxic monoclonal antibody with specificity for HLA-A2 and a variant of HLA-A28." Hum Immunol **3**(4): 277-299.
Parnes, J. R. and J. G. Seidman (1982). "Structure of wild-type and mutant mouse β2-microglobulin genes." Cell **29**(2): 661-669.
Patel, J. P., et al. (2012). "Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia." New England Journal of Medicine **366**(12): 1079-1089.
Petersen, J., et al. (2009). "Post-translationally modified T cell epitopes: immune recognition and immunotherapy." Journal of Molecular Medicine **87**(11): 1045.

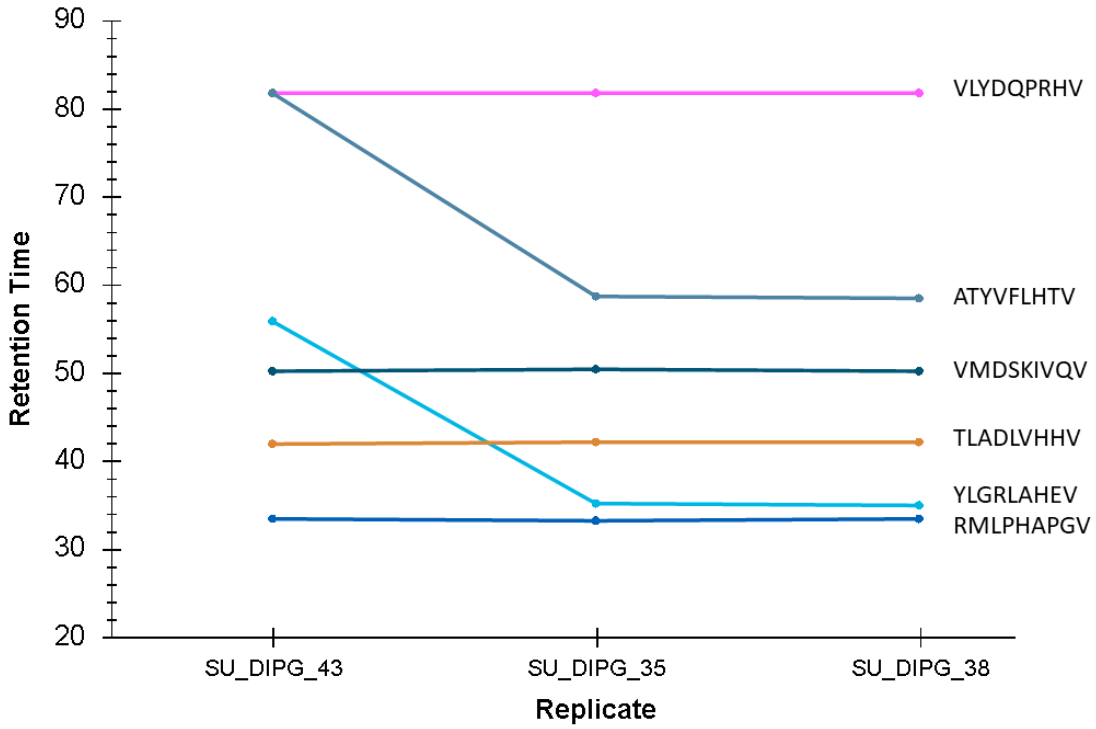
Petersson, M., et al. (1998). "Constitutive IL-10 production accounts for the high NK sensitivity, low MHC class I expression, and poor transporter associated with antigen processing (TAP)-1/2 function in the prototype NK target YAC-1." J Immunol **161**(5): 2099-2105.
Pitot, H. C. (1993). "The molecular biology of carcinogenesis." Cancer **72**(3 Suppl): 962-970.
Prange, K. H. M., et al. (2017). "MLL-AF9 and MLL-AF4 oncofusion proteins bind a distinct enhancer repertoire and target the RUNX1 program in 11q23 acute myeloid leukemia." Oncogene **36**(23): 3346-3356.
Przepiorka, D., et al. (2015). "FDA Approval: Blinatumomab." Clinical Cancer Research **21**(18): 4035-4039.
Purcell, A. W., et al. (2019). "Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics." Nature Protocols.
Pymm, P., et al. (2017). "MHC-I peptides get out of the groove and enable a novel mechanism of HIV-1 escape." Nat Struct Mol Biol **24**(4): 387-394.
QuintásCardama, A., et al. (2008). "Randomized Phase II Study of Proteinase 3-Derived PR1 Peptide Vaccine and GM-CSF with or without PEG-Interferon ALFA-2B to Eradicate Minimal Residual Disease in Chronic Myeloid Leukemia." Blood **112**(11): 3219-3219.
Rabinovich, G. A., et al. (2007). "Immunosuppressive strategies that are mediated by tumor cells." Annu Rev Immunol **25**: 267-296.
Rajasagi, M., et al. (2013). "Tumor Neoantigens Are Abundant Across Cancers." Blood **122**(21): 3265-3265.
Rajasagi, M., et al. (2014). "Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia." Blood **124**(3): 453-462.
Ramakrishna, V., et al. (2003). "Naturally occurring peptides associated with HLA-A2 in ovarian cancer cell lines identified by mass spectrometry are targets of HLA-A2-restricted cytotoxic T cells." Int Immunol **15**(6): 751-763.
Ramarathinam, S. H., et al. (2018). "Employing proteomics in the study of antigen presentation: an update." Expert Review of Proteomics **15**(8): 637-645.
Raval, A., et al. (1998). "Cytokine regulation of expression of class I MHC antigens." Exp Mol Med **30**(1): 1-13.
Reiser, J. B., et al. (2002). "A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex." Immunity **16**(3): 345-354.
Robinson, J., et al. (2015). "The IPD and IMGT/HLA database: allele variant databases." Nucleic Acids Res **43**(Database issue): D423-431.
Rock, K. L. and A. L. Goldberg (1999). "Degradation of cell proteins and the generation of MHC class I-presented peptides." Annu Rev Immunol **17**: 739-779.
Rock, K. L., et al. (2004). "Post-proteasomal antigen processing for major histocompatibility complex class I presentation." Nat Immunol **5**(7): 670-677.
Rodríguez, J. A. (2017). "HLA-mediated tumor escape mechanisms that may impair immunotherapy clinical outcomes via T-cell activation." Oncol Lett **14**(4): 4415-4427.
Roerden, M., et al. (2019). "Neoantigens in Hematological Malignancies-Ultimate Targets for Immunotherapy?" Front Immunol **10**: 3004.
Rosenberg, S. A. and N. P. Restifo (2015). "Adoptive cell transfer as personalized immunotherapy for human cancer." Science **348**(6230): 62-68.
Rovatti, P. E., et al. (2020). "Mechanisms of Leukemia Immune Evasion and Their Role in Relapse After Haploidentical Hematopoietic Cell Transplantation." Frontiers in Immunology **11**(147).
Ruddock, L. W. and M. Molinari (2006). "N-glycan processing in ER quality control." Journal of Cell Science **119**(21): 4373-4380.

Ruella, M. and M. V. Maus (2016). "Catch me if you can: Leukemia Escape after CD19-Directed T Cell Immunotherapies." Computational and Structural Biotechnology Journal **14**: 357-362.
Saratsis, A. M., et al. (2014). "Comparative multidimensional molecular analyses of pediatric diffuse intrinsic pontine glioma reveals distinct molecular subtypes." Acta Neuropathol **127**(6): 881-895.
Saratsis, A. M., et al. (2012). "Insights into pediatric diffuse intrinsic pontine glioma through proteomic analysis of cerebrospinal fluid." Neuro Oncol **14**(5): 547-560.
Sarkizova, S., et al. (2020). "A large peptidome dataset improves HLA class I epitope prediction across most of the human population." Nat Biotechnol **38**(2): 199-209.
Schilling, B., et al. (2012). "Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation." Mol Cell Proteomics **11**(5): 202-214.
Schlenk, R. F., et al. (2008). "Mutations and Treatment Outcome in Cytogenetically Normal Acute Myeloid Leukemia." New England Journal of Medicine **358**(18): 1909-1918.
Schuster, S. J., et al. (2011). "Vaccination with patient-specific tumor-derived antigen in first remission improves disease-free survival in follicular lymphoma." J Clin Oncol **29**(20): 2787-2794.
Schwartzentruber, D. J., et al. (2011). "gp100 Peptide Vaccine and Interleukin-2 in Patients with Advanced Melanoma." New England Journal of Medicine **364**(22): 2119-2127.
Schwartzentruber, J., et al. (2012). "Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma." Nature **482**(7384): 226-231.
Scull, K., et al. (2012). "Secreted HLA recapitulates the immunopeptidome and allows in-depth coverage of HLA A*02:01 ligands." Molecular immunology **51**: 136-142.
Shang, Y., et al. (2013). "Transcriptional corepressors HIPK1 and HIPK2 control angiogenesis via TGF-β-TAK1-dependent mechanism." PLoS Biol **11**(4): e1001527.
Shukla, S. A., et al. (2015). "Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes." Nature Biotechnology **33**(11): 1152-1158.
Stickel, J. S., et al. (2009). "HLA ligand profiles of primary renal cell carcinoma maintained in metastases." Cancer Immunology, Immunotherapy **58**(9): 1407-1417.
Tarafdar, A., et al. (2017). "CML cells actively evade host immune surveillance through cytokine-mediated downregulation of MHC-II expression." Blood **129**(2): 199-208.
Taylor, K. R., et al. (2014). "Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma." Nat Genet **46**(5): 457-461.
Terwilliger, T. and M. Abdul-Hay (2017). "Acute lymphoblastic leukemia: a comprehensive review and 2017 update." Blood Cancer J **7**(6): e577.
Toyonaga, B., et al. (1985). "Organization and sequences of the diversity, joining, and constant region genes of the human T-cell receptor beta chain." Proc Natl Acad Sci U S A **82**(24): 8624-8628.
Trombetta, E. S. and I. Mellman (2005). "Cell biology of antigen processing in vitro and in vivo." Annu Rev Immunol **23**: 975-1028.
Turvey, S. E. and D. H. Broide (2010). "Chapter 2: Innate Immunity." J Allergy Clin Immunol **125**(2 Suppl 2): S24-32.
Tyagi, P. and B. Mirakhur (2009). "MAGRIT: the largest-ever phase III lung cancer trial aims to establish a novel tumor-specific approach to therapy." Clin Lung Cancer **10**(5): 371-374.
Uhlén, M., et al. (2015). "Tissue-based map of the human proteome." Science **347**(6220): 1260419.
Urbanska, K., et al. (2014). "Glioblastoma multiforme - an overview." Contemp Oncol (Pozn) **18**(5): 307-312.

van Alphen, C., et al. (2020). "Phosphotyrosine-based Phosphoproteomics for Target Identification and Drug Response Prediction in AML Cell Lines." Mol Cell Proteomics **19**(5): 884-899.
van der Lee, D. I., et al. (2019). "Mutated nucleophosmin 1 as immunotherapy target in acute myeloid leukemia." J Clin Invest **129**(2): 774-785.
Vassilakopoulos, T. P., et al. (2019). "Immunotherapy in Hodgkin Lymphoma: Present Status and Future Strategies." Cancers (Basel) **11**(8).
Végh, Z., et al. (1993). "Increased expression of MHC class I molecules on human cells after short time IFN-γ treatment." Molecular Immunology **30**(9): 849-854.
Virág, D., et al. (2020). "Current Trends in the Analysis of Post-translational Modifications." Chromatographia **83**(1): 1-10.
Visconte, V., et al. (2019). "Mutations in Splicing Factor Genes in Myeloid Malignancies: Significance and Impact on Clinical Features." Cancers (Basel) **11**(12).
Vita, R., et al. (2019). "The Immune Epitope Database (IEDB): 2018 update." Nucleic Acids Res **47**(D1): D339-d343.
Walter, S., et al. (2012). "Multipeptide immune response to cancer vaccine IMA901 after single-dose cyclophosphamide associates with longer patient survival." Nat Med **18**(8): 1254-1261.
Wang, F., et al. (2020). "Integrated transcriptomic and epigenetic data analysis identifiesaberrant expression of genes in acute myeloid leukemia with MLL-AF9 translocation." Mol Med Rep **21**(2): 883-893.
Wang, S. S., et al. (2019). "Towards Immunotherapy for Pediatric Brain Tumors." Trends in Immunology **40**(8): 748-761.
Warren, K. E. (2012). "Diffuse intrinsic pontine glioma: poised for progress." Front Oncol **2**: 205.
Weinschenk, T., et al. (2002). "Integrated functional genomics approach for the design of patient-individual antitumor vaccines." Cancer Res **62**(20): 5818-5827.
Weinzierl, A. O., et al. (2007). "Distorted Relation between mRNA Copy Number and Corresponding Major Histocompatibility Complex Ligand Density on the Cell Surface." Molecular & Cellular Proteomics **6**(1): 102-113.
Wieczorek, M., et al. (2017). "Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation." Frontiers in Immunology **8**(292).
Wingelhofer, B. and T. C. P. Somervaille (2019). "Emerging Epigenetic Therapeutic Targets in Acute Myeloid Leukemia." Frontiers in Oncology **9**(850).
Wu, G., et al. (2012). "Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas." Nat Genet **44**(3): 251-253.
Yague, J., et al. (2000). "A post-translational modification of nuclear proteins, N(G),N(G)-dimethyl-Arg, found in a natural HLA class I peptide ligand." Protein Sci **9**(11): 2210-2217.
Yotnda, P., et al. (1998). "Cytotoxic T cell response against the chimeric ETV6-AML1 protein in childhood acute lymphoblastic leukemia." J Clin Invest **102**(2): 455-462.
Youn, B. S., et al. (2000). "Chemokines, chemokine receptors and hematopoiesis." Immunol Rev **177**: 150-174.
Yu, Y., et al. (2018). "Recent advances in CD8(+) regulatory T cell research." Oncol Lett **15**(6): 8187-8194.
Zarling, A. L., et al. (2000). "Phosphorylated Peptides Are Naturally Processed and Presented by Major Histocompatibility Complex Class I Molecules in Vivo." J Exp Med **192**(12): 1755-1762.

Zarling, A. L., et al. (2014). "MHC-restricted phosphopeptides derived from Insulin receptor substrate-2 and CDC25b offer broad-based immunotherapeutic agents for cancer." Cancer Res **74**(23): 6784-6795.
Zarling, A. L., et al. (2006). "Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy." Proc Natl Acad Sci U S A **103**(40): 14889-14894.
Zhang, H. and J. Chen (2018). "Current status and future directions of cancer immunotherapy." J Cancer **9**(10): 1773-1781.
Zhang, J., et al. (2012). "PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification." Mol Cell Proteomics **11**(4): M111.010587.
Zhang, X., et al. (2017). "Personalized cancer vaccines: Targeting the cancer mutanome." Vaccine **35**(7): 1094-1100.
Zhao, Y. and O. N. Jensen (2009). "Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques." Proteomics **9**(20): 4632-4641.
Zheng, P.-P., et al. (2018). "Approved CAR T cell therapies: ice bucket challenges on glaring safety risks and long-term impacts." Drug Discovery Today **23**(6): 1175-1182.
Zhou, Y., et al. (2019). "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets." Nat Commun **10**(1): 1523.
Zhu, N., et al. (2015). "Mutations in tyrosine kinase and tyrosine phosphatase and their relevance to the target therapy in hematologic malignancies." Future Oncol **11**(4): 659-673.

CHAPTER 5: SUPPLEMENTARY FIGURES AND TABLES



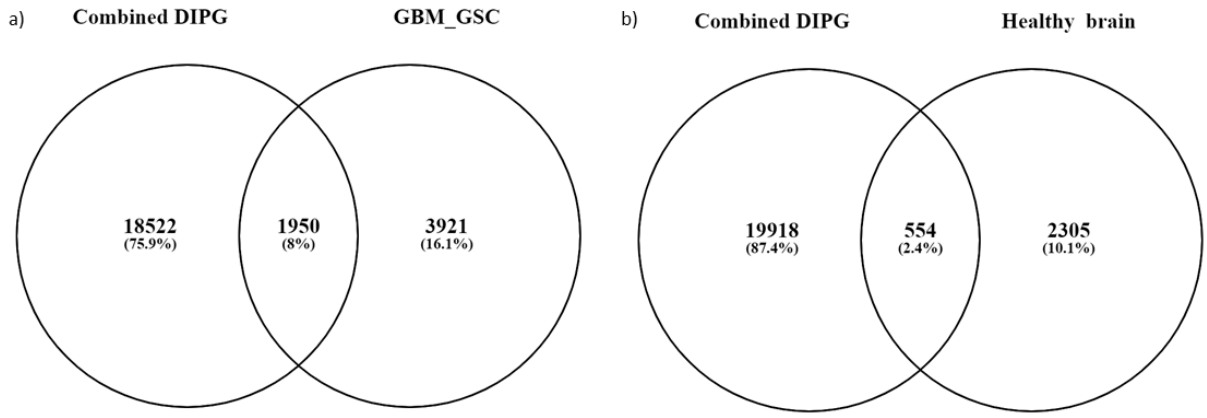
Supplementary figure 1 – Similar retention time of constitutive HLA*02:01 peptides across DIPG samples. The samples were acquired on QTRAP 6500+. MRMs for each peptide were designed using Skyline.

Supplementary Table 1 – Histone H3.3 peptides identified across different HLA A*02:01 positive DIPG cell lines. Peptides with * mark have previously been reported in IEDB and peptides in red were found in all samples.

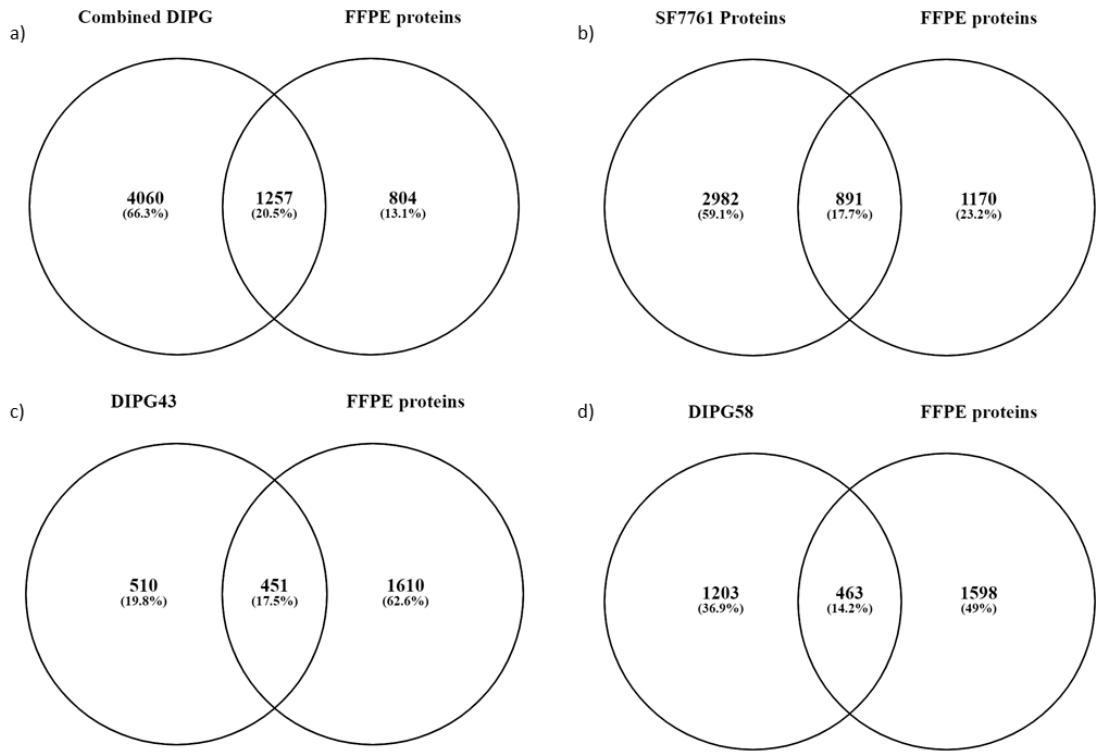
H3.3 Peptides	DIPG27	DIPG35	DIPG38	DIPG43
RMSAPSTGGV (Neoepitope)	No	No	No	No
YRPGTVALR*	Yes	Yes	Yes	Yes
YQKSTELLI*	No	Yes	No	No
STELLIR*	Yes	Yes	Yes	Yes
REIAQDFKTDL*	No	Yes	No	No
EIAQDFK*	Yes	Yes	Yes	Yes
AQDFKTDLRFQ	No	No	No	Yes
ALQEASEAYL*	No	Yes	No	No
ALQEASEAYLV*	Yes	Yes	No	Yes
RVTIMPK	Yes	Yes	Yes	Yes
TIM(+15.99)PKDI	Yes	No	Yes	No
TIM(+15.99)PKDIQL*	Yes	No	Yes	Yes

Supplementary Table 2 – HLA class I restricted peptides identified form other histone proteins in DIPG cells lines

Peptide	Protein
TVTAMDVVYALK	Histone H1.4
TVTAMDVVYALK	Histone H1.4
VLRDNIQGI	Histone H1.4
DVVYALKRQGR	Histone H1.4
ASGPPVSELITK	Histone H1.4
SGVSLAALK	Histone H1.4
GTGASGSFK LVREIAQDFK	Histone H1.4
LVREIAQDFK	Histone H3.1
AQDFKTDL	Histone H3.1
APRKQLATKAA	Histone H3.2
YRPGTVALR	Histone H3.2
YQKSTELLI	Histone H3.2
ISGLIYEETR	Histone H4
DAVTYTEHAK	Histone H4
VLKVFLENV	Histone H4
RISGLIYEETR	Histone H4
TAMDVVYAL	Histone H4



Supplementary Figure 2 – Very little overlap between the immunopeptidome of DIPG and a) GBM patients and glioblastoma stem cell like lines (GSC) and b) Healthy brain samples.



Supplementary Figure 3 – Overlap between source proteins of the DIPG immunopeptidome and a) proteins identified in FFPE samples of DIPG patients. The highest overlap was found between b) SF7761 cell line and least overlap was found in H3.3 WT cell lines c) DIPG43 and d) DIPG58

CHAPTER 6: General Discussion

6.1 INTRODUCTION

Cancer is classified as a chronic disease that presents a significant health and economic burden, particularly in western countries. To add to this, it also is a leading cause of mortality and morbidity all over the world. Over the years, as understanding of the immune system and the role it plays in cancer development and progression improved. It led to immunotherapy being catapulted into the forefront, as the next generation of cancer treatments. There are several success stories to support this, such as immune checkpoint inhibitors targeting PD1 and PD-L1 receptors present on T cells and APCs, respectively (Martín-Ruiz et al. 2020). Drugs such as Nivolumab and Pembrolizumab are highly effective in treating melanoma, NSCLC and Hodgkin's lymphoma (Martín-Ruiz et al. 2020). However, passive immunotherapy which involves administration of antibodies or genetically modified immune cells to patients, to activate an antitumor action does not always work. This can be attributed to lack of generation of immunological memory to fight the cancer, if and when it remerges.

Another arm of immunotherapy, which relies on stimulating the patient's own immune system to produce an antigen-specific antitumor effect is active immunotherapy. The effectiveness of it is on full display with success of preventative vaccines such as Gardasil and Heplisav-B which provide immunity against cervical cancer (caused by human papilloma virus) and liver cancer (caused by hepatitis B virus) (Knutson et al. 2015). Despite these advances, there are still cancers for which chemotherapy and radiotherapy are the mainstay of treatment. This can be partly attributed to their complexity or them being immunologically cold due to absence of effector CD8 T cells, pro-inflammatory cytokines and low HLA expression (Thorsson et al. 2018). Since most immunotherapy treatments target surface receptors, cancers such as Hodgkin's lymphoma are much easier to treat as they express several cell surface antigens including CD19/CD30 which can be targeted (Zheng et al. 2018). Most forms of cancers do not express surface markers which can be readily targeted by drugs. Similarly, preventive vaccines might not always work as their target would be self-proteins or peptides rather than oncogenic viruses which are foreign and more likely to generate a strong immune response.

Hence for a successful immunotherapy alternate targets must be explored which can be incorporated into either vaccines or modified T cell mediated immunotherapies such as CAR T cells. This would involve exploring the tumour landscape to identify tumour antigens capable

of activating the immune system. This thesis successfully dealt with this aspect. By utilising high resolution mass spectrometry, the thesis explores the immunopeptidome of primary tumour samples including AML and DIPG. This has resulted in generation of an extensive list of therapeutically actionable neopeptides, TAAs and CTAs, neopeptide and PTM-peptides.

6.2 Factors influencing immunopeptidome isolation and identification

The immunopeptidome is highly complex due to HLA polymorphism (Purcell et al. 2019). This results in different HLA class I allotypes, each of which binds to a distinct set of peptides for presentation on the surface of nucleated cells. MS because of its high sensitivity and rigor is the only technique capable of disentangling the different layers of complexity associated with the immunopeptidome (Purcell et al. 2019). In order to improve the depth and coverage of the immunopeptidome, a considerable part of this thesis was spent on optimising the workflow and MS based acquisition methods for immunopeptidomics. For separating the pHLA complexes, I explored two peptide enrichment methods namely the use of RP-HPLC and ultrafiltration using a 5kDa MWCO. This culminated in chapter 2 of the thesis wherein I demonstrated the power complementary peptide separation techniques can have on expanding the detectable immunopeptidome.

The study highlighted several key aspects influencing immunopeptidome isolation and identification. First, increased unique peptide identification was essentially dependent on the technique's ability to reduce the complexity of the samples. There are 2 factors contributing towards making a sample complex, these are 1) level of HLA expression on the cells and 2) sample size. RP-HPLC as a technique can negate the influence of both the factors. As evident from data in chapter 2 wherein the THP-1 eluate was split in two, with one half being processed by RP-HPLC whilst the other by ultrafiltration. The RP-HPLC fraction always generated ~1,500 to 7,000 more unique peptides than ultrafiltration across biological replicates, in both HLA A2 and pan class I fractions. Overall, in all three biological replicates, RP-HPLC was much preferred over MWCO as it not only yielded more unique peptide IDs but also resulted in identification of more TAAs and CTAs and even phosphopeptides.

Based on chapter 2 results, we decided to use RP-HPLC to separate pHLA across our remaining samples including cell lines and clinical samples. The 5kDa MWCO filter was used for less complex samples such as small cell pellets of size $2e^7$ (in chapter 5) or isolation of sHLA bound peptides from plasma samples (chapter 3).

Another important aspect of the workflow involved optimisation of the MS method used to acquire my data. This was critical to my work as I was also interested in identifying PTM-

peptides including Ser/Thr/Tyr phosphorylated and Arg/Lys dimethylated peptides. We decided to use a more flexible and sensitive Thermo Scientific Tribrid Fusion orbitrap mass analyser. This allowed us 1) to use high energy collisional dissociation (HCD) along with 2) a decision tree strategy to include +1 charge state peptides in the range of 800-1800 amu which influenced collection of MS^1 and MS^2 spectra. HCD gave us more spectral information particularly in lower mass ranges leading to confident identification of peptides. The inclusion of singly charged species lead to identification of an additional 12,093 class I peptides in our THP1 dataset (as mentioned in chapter 2). The +1 charge state peptides matched the motifs and length distribution of +2 charge state peptides. Notably, most +1 peptides identified were restricted to HLA A*02:01, B*15:11 and C*03:03 whilst allele B*27:05 did not contribute much to +1 charged peptides. This is because former alleles prefer peptides which have small hydrophobic residues in them such as Val, Leu, Ser, Ile whilst B*27:05 peptides are rich in basic amino acids such as Arg/Lys.

To conclude, the thesis very clearly highlights the role of an optimised workflow for samples based on HLA alleles expressed and how that can influence peptides identified by the MS.

6.3 Moving towards vaccine design for AML and DIPG

Considerable progress has been made in the field of immunotherapy especially in B cell leukaemia and lymphoma, involving not only checkpoint inhibitors but also antibody-drug conjugate treatments which target cell surface receptors such as CD30 and deliver a cytotoxic drug to B lymphoma cells. This has resulted in ~5% increased survival rate in patients by lowering combined risk of disease-free progression and death (Connors et al. 2018). Other emerging immunotherapy include genetically modified T cells including bispecific T cell engagers (BiTE) which is an anti-CD19/CD3 antibody (Przepiorka et al. 2015) and CAR T cell therapies which included the Kymriah™ and Yescarta™ (Zheng et al. 2018)targeting B cells in lymphomas and leukaemia. However, the effect of these treatment is not long lasting, with ~30%-55% of the patients treated demonstrating clinical relapse within a year.

Hence working towards either preventive or therapeutic vaccines for cancer which trains the patient's own immune system to fight cancer over a course of time is a much more useful approach. Another part of this thesis explored the endogenous peptide HLA class I repertoire naturally presented in AML and DIPG. In total we identified ~109,702 class I peptides presented by 33 unique alleles. This included alleles such as A*02:01, A*01:01, A*03:01, B*07:02 which are quite commonplace in caucasian population but also less frequent alleles

including A*74:01, A*33:03 and B*15:11 present in African and Asian population respectively.

In case of AML, initially samples only comprised of peptides identified were from both well-established cell lines such as leukemic cell lines THP-1, MV411.1 and HL60. Later the thesis incorporated both healthy and patient bone marrow aspirates. This was particularly exciting, as till date, there have been no study which has investigated the immunopeptidome from primary tumour material. Isolating pHLA complexes from clinical samples was perhaps most difficult as they were not only low cell numbers (~1-2e^{8}) but also precious. Although dealing with clinical samples provides a glimpse as to what the future holds, with immunopeptidomics studies slowly becoming mainstream and no doubt data generated from such studies will be incorporated in vaccine design efforts.

Post sample acquisition we performed rigorous analysis of our data and identified a total of 81,027 peptides across different samples of HM. With the aim of identifying leukaemia specific TAA and CTA we mined the 81,027 peptides for tumour antigens originating from ~700 oncogenes and TSG (using TANTIGEN) and CTAs originating from ~360 CTA proteins (using CTdatabase). We identified a total of 3,874 TAA and CTA with peptides originating from proteins such as NPM1, MPO, RUNX1, TP53, WT, MYC and ETV peptides along with MAGE and PRAME peptides.

There were a two limitation of the study, first, that no neopeptides were identified. This is because the data was not searched for mutations due to lack of NGS data. Second, that despite identifying several interesting candidates, due to limited time, T cells assays to test the immunogenic capability of the peptides could not be performed. This leaves immense scope and some of the key peptides especially the ones reported to be immunogenic in previous studies can be readily taken forward for T cell studies including ELISPOTS, tetramer based magnetic enrichment to identify specific T cell populations and intracellular staining (ICS) assays to measure cytokine release such as IFNγ and TNFα.

One of the most unexpected yet interesting data that came from this thesis was the potential role mutations and TME in AML might have on HLA class I expression. The results from chapter 3 proved that chromosomal aberrations such as MLL-AF9 and MLL-AF4 reportedly found in THP1 and MV11.1 cell line respectively did not alter HLA class I expression. However, the presence of mutation in NPM1 protein resulted in drastic decrease in the pHLA repertoire in an AML patient2 with only ~1,200 peptides identified from it compared to

~10,000 peptides identified from healthy and AML patient1 BM. The influence of NPM1 mutations on downregulating HLA class II expression is well known (Dufva et al. 2020), however its role in suppressing class I expression is previously unreported. A textbook example of how cancers reduce HLA class I to evade detection was seen in ALL patient3 which had only 95 class I peptides identified from its bone marrow. Together, this data highlights the obstacles vaccines might face not only in AML but also other cancers wherein the immune system is heavily compromised. However, it goes on to show how immunopeptidomics studies form a crucial part of vaccine design and can guide immunotherapy decisions such as incorporating drugs or cytokines treatment which can aide in generation of an appropriate immune response post vaccination.

Another major focus of this thesis was DIPG, a rare but aggressive form of paediatric brain cancer with only chemotherapy and radiotherapy as mainstay of treatment (Khuong-Quang et al. 2012). Till date, there has been no immunopeptidomics study for DIPG other than the work presented as a part of this thesis. The study of HLA class I peptide repertoire was undertaken using primary DIPG cell lines. A total of 26,592 peptides were identified from 6 cell lines. Similar to bone marrow, the data was mined for both TAA and CTA peptides and a total of 1,123 peptides were identified. Several therapeutically actionable peptides from proteins such as IL-13αR, IGF2BP3, CSP4 and MAGA and PRAME were identified. Peptides from protein such as IL-13αR have been incorporated into CAR T cell immunotherapy for GBM.

Excitingly, we also identified a neopeptide encompassing the highly penetrant mutation H3.3 K27M found in almost 70% of the cases was successfully identified. The data was also searched for a few other mutations including PIK3R1 M261I, PIK3CA E707K and TP53 R273C in DIPG27 cell line. However, no neopeptide was identified from these mutations. This might be due to small sample size ($2x10^7$) used to investigate the immunopeptidome or the low HLA expression on cell lines (since they are brain cell line).

6.4 Exploring the potential of PTM-peptides

With this thesis we decided to explore an emerging category of peptide antigens i.e. PTM-peptides. These peptides are of interest as unlike their native counterparts, they might be excluded from thymic selection and hence might have a naïve TCR capable of recognising them. We decided to investigate for both phosphorylated and methylated peptide. As mentioned in chapter 4, we identified a total of ~2,000 PTM peptides, making it one of the

largest PTM peptide data set yet reported. Several of these peptides came from known oncogenes and TSG such as RUNX1, MYC and NPM1.

There were several exciting key discoveries with regards to tyrosine phosphorylated (pTyr) peptides a subset of phosphorylated peptides not reported previously in any immunopeptidomics study. The thesis not only provides key features (regarding length and motif) pTyr peptides restricted to HLA A*02:01 might possess but also provides structural basis for their presentation. Crystallography studies undertaken elucidated for the first time how an 8mer, pTyr peptide stabilises HLA A*02:01 by interacting with key residues in α1 and α2 chains. The PTM is solvent accessible and therefore might act as a beacon for TCRs thereby altering the immunogenicity of pTyr peptides when compared with its native counterpart. This is particularly of interest as the same peptide was found in AML patient data.

In addition to this we also report first structure of HLA A*02:01 restricted dimethyl arginine (dmArg) peptide. Unlike the pTyr peptide, dmArg did not have any role to play in the stability of pHLA complex (as evident from thermal stability assay). The dm moiety was solvent exposed and would probably be only interacting with TCR and maybe influencing the immunogenicity of the peptide.

There has been only one study each highlighting the immunogenic potential of pSer/pThr peptides and dmArg peptides in healthy individuals. However due to time restriction this aspect of the study could not be completed.

6. 5 Future directions

Peptide-based vaccine or redirected T cell-based immunotherapy have showed their efficacy in other cancer such as melanoma, renal carcinoma and B cell lymphoma. As mentioned on different occasions in this thesis, both AML and DIPG are unique in their own way and hence present different challenges that have to be overcome in order to find an effective cure for either of them. The thesis is an attempt to address some of these questions. There is often a debate in the field of immunotherapy as to which peptide (TAAs, CTAs, neopeptides and PTM-peptides) is the most potent, there is no definitive answer to this question. This is because an immune response towards a peptide is decided by several different factors working together to produce an effective anti-tumour response. The way forward in cancer immunotherapy is having a multi-omics approach wherein immunopeptidomics data is incorporated along with WES/NGS data to make an informed decision on which antigens to be selected for a vaccine.

REFERENCES

Connors, J. M., et al. (2018). "Brentuximab Vedotin with Chemotherapy for Stage III or IV Hodgkin's Lymphoma." N Engl J Med **378**(4): 331-344.

Dufva, O., et al. (2020). "Immunogenomic Landscape of Hematological Malignancies." Cancer Cell.

Khuong-Quang, D. A., et al. (2012). "K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas." Acta Neuropathol **124**(3): 439-447.

Knutson, K. L. and E. A. Mittendorf (2015). "Cancer vaccines in the new era of cancer immunotherapy." Vaccine **33**(51): 7376.

Martín-Ruiz, A., et al. (2020). "Effects of anti-PD-1 immunotherapy on tumor regression: insights from a patient-derived xenograft model." Sci Rep **10**(1): 7078.

Przepiorka, D., et al. (2015). "FDA Approval: Blinatumomab." Clinical Cancer Research **21**(18): 4035-4039.

Purcell, A. W., et al. (2019). "Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics." Nature Protocols.

Thorsson, V., et al. (2018). "The Immune Landscape of Cancer." Immunity **48**(4): 812-830.e814.

Zheng, P.-P., et al. (2018). "Approved CAR T cell therapies: ice bucket challenges on glaring safety risks and long-term impacts." Drug Discovery Today **23**(6): 1175-1182.

APPENDIX

CHAPTER 2 -

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.molimm.2020.04.008.

CHAPTER 3

All supplementary files and immunopeptidome peptide.csv data can be found at (https://www.dropbox.com/sh/smfclqphm1tewdi/AABOysblsNGyTZ925md9HHO-a?dl=0)

A) Supplementary data files

1) Supplementary data file 1 – Overlapping peptides between samples
2) Supplementary data file 2 – AML specific cell line peptides
3) Supplementary data file 3 – AML specific patient sample peptides
4) Supplementary data file 4 – Other TAA clinical samples
5) Supplementary data file 5 – Other TAA cell lines
6) Supplementary data file 6 - CTA healthy BM
7) Supplementary data file 7 – CTA AML paitent1
8) Supplementary data file 8 – sHLA TAA AML patient2
9) Supplementary data file 9 – sHLA TAA ALL patient3
10) Supplementary data 10 – Overlapping peptides between Narayan *et al* and HM TAAs

B) **Haematological malignancy immunopeptidome peptide.csv data**

1) Supplementary data 1 - THP1 cell line (BR1 and BR2) –
2) Supplementary data 2 - MV411.1 cell line (BR1 and BR2)
3) Supplementary data 3 - HL60 cell line (BR1 and BR2)
4) Supplementary data 4 - Healthy bone marrow
5) Supplementary data 5 - AML Patient1 -
6) Supplementary data 6 - AML Patient2 -
7) Supplementary data 7 - ALL Patient3 -
8) Supplementary data 8 - sHLA collated patients

CHAPTER 4 – Data accessed through

(https://www.dropbox.com/sh/smfclqphm1tewdi/AABOysblsNGyTZ925md9HHO-a?dl=0)

A) Supplementary data files

1) Supplementary data file 1 – Phosphopeptides present in Phosphosite plus
2) Supplementary data file 2 – Phosphopeptides and oncogenes
3) Supplementary data file 3 – TAA peptides in Phosphosite plus
4) Supplementary data file 4 – Collated pTyr peptides
5) Supplementary data file 5 – Collated mono and dimethyl peptides

B) **Methylated peptides Peaks PTM.csv data**

1) Supplementary data 1 – Peaks PTM THP1
2) Supplementary data 2 - Peaks PTM MV411.1
3) Supplementary data 3 - Peaks PTM HL60
4) Supplementary data 4 - Peaks PTM Healthy BM
5) Supplementary data 5 - Peaks PTM AML Pt1

CHAPTER 5 – Data accessed through

(https://www.dropbox.com/sh/smfclqphm1tewdi/AABOysblsNGyTZ925md9HHO-a?dl=0)

A) Supplementary data files

1) Supplementary data file 1 – Collated DIPG Tantigens
2) Supplementary data file 2 – Collated DIPG CTA
3) Supplementary data file 3 – DIPG27 RNASeq TAA and CTA
4) Supplementary data file 4 – DIPG35 RNASeq TAA and CTA
5) Supplementary data file 5 – TAA and CTA in HBPA

B) DIPG, HEK293 and C1R A*03:01 immunopeptidome peptide.csv data

1) H3.3K27M DIPG cell line data –
   a) Supplementary data 1 - DIPG27
   b) Supplementary data 2 - DIPG35
   c) Supplementary data 3 - DIPG38

d) Supplementary data 4 - SF7761

2) H3.3 WT DIPG cell line data –

a) Supplementary data 5 - DIPG43

b) Supplementary data 6 - DIPG58

3) Supplementary data 7 - HEK293 H3.3K27M cell line data

4) Supplementary data 8 - C1R A*03:01 H3.3K27M cell line data