Universitat Autònoma de Barcelona

Identifi cation of the personalized repertoire of conventional and non-canonical tumor antigens

Maria Lozano Rabella

Universitat Autònoma de Barcelona
Departament de Biologia Cel·lular, Fisiologia i Immunologia
Programa de doctorat en Immunologia avançada

Identification of the personalized repertoire of conventional and non-canonical tumor antigens

Maria Lozano Rabella
Doctoral thesis
Barcelona, 2022

Alena Gros — Director
Josep Tabernero — Director

Contents

Abbreviations

ACT	Adoptive T-Cell Transfer
ADAR	Adenosine Deaminase Acting on RNA
aORF	Alternative Open Reading Frames
APC	Antigen-Presenting Cells
Can	Canonical
CAR	Chimeric Antigen Receptor
CAR-T	Car-Transduced T Cell
CD	Cluster Of Diff Erentiation
aDNA	Complementary Dna
CEA	Carcinoembryonic Antigen
aTEC	Cortical Thymic Epithelial Cells
CTL	Cytotoxic T Lymphocytes
CTLA-4	Cytotoxic T-Lymphocyte Antigen 4
DC	Dendritic Cells
DRiP	Defective Ribosomal Products
EBV	Epstein-Barr Virus
ER	Endoplasmic Reticulum
ERAP	Er-Associated Aminopeptidases
ERE	Endogenous Retroelement
F	Phenylalanine
FDA	Food And Drug Administration
FDR	False Discovery Rate
GI	Gastrointestinal Cancers
Gyn	Gynecological Malignancies
HERV	Human Endogenous Retroviruses
HLA	Human Leukocyte Antigen
HPV	Human Papillomavirus
IFN	Interferon Gamma
IL-2	Interleukin-2
INDELS	Insertions or Deletions
IRES	Internal Ribosomal Entry Sites
IVS	In Vitro Sensitization
IVT	In Vitro Transcribed
LAK	Lymphokine-Activated Killer
LC-MS/MS	Liquid Chromatography Coupled Tandem Mass Spectrometry
LP	Long Peptides
LTR	Long Terminal Repeat
Mel	Melanoma
MEP	Minimal Epitope Pool
MHC	Major Histocompatibility Complex
mRNA	Messenger Rna
MS	Mass Spectrum
mTEC	Medullary Thymic Epithelial Cells

NB	Non Binders
ncRNA	Non-Coding Rna
NGS	Next-Generation Sequencing
nonC	Non-canonical
nonC-TL	Non-canonical HLA-I Tumor Ligand
NSM	Non-synonymous Mutations
ORF	Open Reading Frames
PBL	Peripheral Blood Lymphocytes
PBMC	Peripheral Blood Mononuclear Cells Lymphocytes
PCPS	Proteasome-Catalyzed Peptide Splicing
PD-1	Programmed Death 1
PD-L1	PD-1 Ligand
pHLA-I	HLA-I Complex
PLC	Peptide Loading Complex
PP	Peptide Pool
PRF	Programmed Ribosomal Frameshifting
PRM	Parallel Reaction Monitoring
PSA	Prostatic Acid Phosphatase
REP	Rapid Expansion Protocol
SB	Strong Binders
sgRNA	Single Guide RNA
SNV	Single Nucleotide Variants
TA	Tumor Antigen
TAP	Antigen Processing Proteins
TAPBPR	Tapasin-Related Protein
TCL	Tumor Cell Lines
TCR	T-Cell Receptors
TCRtd	TCR-transduced T cells
TE	Transposable Elements
TIL	Tumor-Infiltrating Lymphocytes
TMB	Tumor Mutational Burden
TMG	Tandem Minigenes
UGT1	Glycoprotein Glucosyltransferase 1
UTR	Untranslated Regions
W	Tryptophan
WB	Weak Binders
WES	Whole-Exome Sequencing
WT	Wild-Type
WT1	Wilms' Tumor Protein
B2M	B-2-Microglobulin

Summary

The identification of tumor antigens is critical to better understand the immune response to tumors and to develop more effective cancer immunotherapies. Immunological screenings have frequently identified natural T-cell responses to conventional tumor antigens such as neoantigens, cancer-germline, and differentiation antigens in cancer patients. Recent evidence suggests that peptides derived from nonC proteins are specifically presented on HLA-I from tumor cells, thus expanding the repertoire of targetable tumor antigens. One of the main concerns is whether those nonC HLA-I ligands can naturally elicit T-cell responses and whether they are specifically presented by tumor cells, with important implications for designing therapeutic interventions.

In this thesis, we have studied in detail the personalized repertoire of tumor antigens recognized by tumor-reactive T cells in 9 cancer patients with distinct tumor histologies including conventional tumor antigens, but also expanding the scope to non-canonical tumor antigen sources.

In the first part, we investigated whether tumor-reactive T cells recognize conventional tumor antigens through three different methods including personalized neoantigen screening with TMG, *in silico* prediction and pHLA-I immunopeptidomics. However, the specific antigen/s targeted by the majority of tumor-reactive lymphocytes could not be determined. In the second part, we used a proteogenomics approach to identify peptides derived from non-canonical proteins presented on HLA-I of patient-derived TCL (nonC-TL). We frequently identified nonC-TL which were mainly derived from 5'UTR regions and represented the most abundant source of candidate tumor antigens. However, pre-existing T-cell responses targeting nonC-TL were not detected in any patient studied. In contrast, tumor-reactive lymphocytes consistently displayed preferential recognition of neoantigens and also recognized cancer-germline and tissue differentiation antigens. Nonetheless, we showed that nonC-TL can elicit *de novo* T-cell responses via *in vitro* sensitization of donor PBL. We identified TCR specific to three nonC-TL peptides, two of which mapped to the 5' UTR regions of HOXC13 and ZKSCAN1 genes, and one mapping to a non-coding spliced variant of C5orf22C gene. Importantly, we found that these immunogenic nonC-TL display therapeutic potential, as they were expressed across diverse tumor types, but were barely detected in healthy cells. Altogether, our results show that nonC-TL antigens hold great promise as new therapeutic targets as they can be immunogenic, are frequently presented on HLA-I of tumor cells, shared across patients, and barely expressed on healthy cells.

However, despite the extensive immunological screening performed including nonC-TL, neoantigens, cancer-germline, and melanoma-associated antigens, the specificity of many tumor-reactive lymphocytes remains unknown and warrants further investigation.

Introduction

1. The immune system and cancer

Cancer arises from normal tissues through the accumulation of genetic and epigenetic alterations leading to uncontrolled cellular growth and invasion. In the human body with more than 10^{14} cells, billions of cells can experience mutations every day potentially disrupting cell cycle control. If a mutation gives a selective advantage allowing the cell to grow and divide slightly more vigorously and survive more readily than its neighbors, this cell can become a founder of a malignant clone (Albert, 2015). The immune system, which is constantly patrolling the body, can sense abnormal changes acquired by malignant cells and eliminate arising tumors.

In addition to the six **hallmarks of cancer** first proposed in 2000 (Hanahan & Weinberg, 2000), the evasion of the immune system was introduced as an emerging hallmark of cancer in 2011 together with the deregulation of the metabolism (Hanahan & Weinberg, 2011). Nowadays it is inconceivable to exclude the immune system from the picture, as reflected by the extensive and increasing number of cancer immunotherapies to treat cancer, and it has been sufficiently validated to be considered part of the core set of the eight hallmarks of cancer (Hanahan, 2022) (Figure 1).

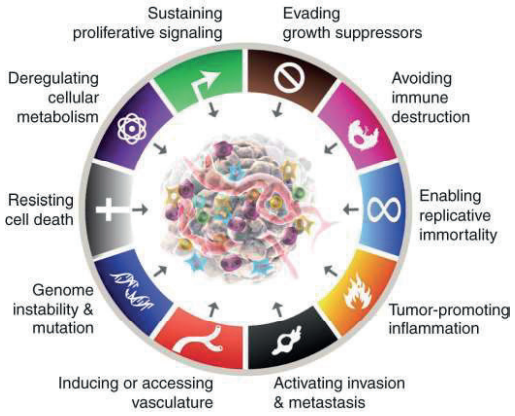


Figure 1. Current validated Hallmarks of Cancer. The eight core hallmarks of cancer and two enabling factors. Image from Hanahan, 2022.

The concept that the immune system plays a major role in the protection against cancer was proposed more than one century ago (Erhlich.P, 1909). It was further postulated that the immune system has the capacity of recognizing newly arising tumors through the expression of tumor-specific antigens on tumor cells and eliminating them(Burnet, 1970; Thomas, 1959).

Although the concept of **cancer immunosurveillance** was initially controversial because of the capability of nude mice to reject tumors (Rygaard & Povlsen, 1974), the development of genetically engineered mouse models with deeper and better characterized immunodeficiencies enabled experiments that demonstrated the role of the immune system in controlling cancer(Dunn et al., 2002). Deficiencies in key immunologic molecules such as IFN-γ (Kaplan et al., 1998) or Perforin were shown to enhance the susceptibility to both chemically induced and spontaneous tumors(Dunn et al., 2004a). The most convincing evidence demonstrating that cancer immune surveillance is mediated by lymphocytes was derived from experiments

using mice in which the recombinase activating gene-2 (RAG-2) had been selectively inactivated (Shankaran et al., 2001). These RAG2-deficient mice cannot rearrange lymphocyte antigen receptors and therefore lack T, B, and NKT cells. The results demonstrated, beyond any doubt, that lymphocytes protect mice against the development of tumors. Clinical evidence has accumulated supporting that immunosurveillance is also operative in humans (Dunn et al., 2002, 2004a). The observations that patients with primary (Gatti & Good, 1971) and acquired immunodeficiencies (Boshoff & Weiss, 2002) or therapeutically immunosuppressed after organ transplantation are at increased risk of viral-induced cancer (Penn, 1999) have been widely confirmed. Moreover, the identification of naturally occurring T cells capable of recognizing autologous tumors demonstrates that patients can spontaneously mount a response against cancer cells (Muul et al., 1987).

Despite this clear protective role of the immune system against tumors, other experiments suggested that lymphocytes might also contribute to the emergence of primary tumors with reduced immunogenicity potentially favoring immune evasion (Shankaran et al., 2001). These findings led to the concept of **cancer immunoediting** which is a dynamic process composed of three phases (Dunn et al., 2004b) (Figure 2). The first phase is elimination, whereby the immune system successfully eradicates the developing tumor. Secondly, the equilibrium phase, wherein the host immune system exerts potent and relentless selection pressure on the tumor cell variant/s surviving the first phase, which is enough to contain, but not fully eliminate the tumor. And last, the escape phase, wherein the tumor cell variants selected in the equilibrium phase now can grow, becoming clinically detectable tumors due to genetic/epigenetic changes conferring resistance to immune detection and/or elimination combined with induction of a favorable tumor microenvironment. Nowadays, the concept of immunoediting is widely accepted as a continual process that also occurs during immunotherapy in human patients with cancer (Gubin & Vesely, 2022).

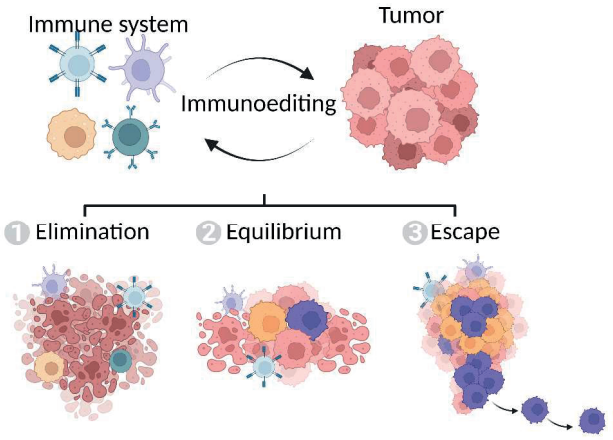


Figure 2. The cancer immunoediting process. Cancer immunoediting consists of three phases. The first phase is elimination, whereby the innate and adaptive immune systems act to destroy the emerging tumor. The second phase is equilibrium, where the tumor outgrowth is immunologically restrained but cancer is not eliminated. Further immunologic sculpting of the tumor (resistant tumor clones) and establishment of a suppressive environment leads to the escape phase, where cancer becomes a clinically apparent disease and spreads to other organs. Adapted from Gubin & Vesely, 2022. Image created with BioRender.com.

1.1. Antigen recognition by T cells

Conventional T lymphocytes recognize peptides bound to the major histocompatibility complex (MHC) on the surface of target cells through their T cell receptors (TCR) (Strominger, 2002). The peptide sequences that are recognized are called "epitopes," while their parent proteins are referred to as "antigens". In the context of cancer, those proteins are known as tumor antigens.

There are two main types of T cells identified by the cluster of differentiation (CD) markers CD8 and CD4. These two types are known to exert different physiological functions and to recognize cognate epitopes in two different ways. $CD8^+$ cells recognize short peptides (8-12 amino acids) typically derived from cytoplasmic proteins via proteasomal digestion that are presented on MHC-I (Kloetzel, 2001). In contrast, $CD4^+$ T cells recognize longer peptides presented on MHC-II that are predominantly derived from both endosomal and ingested proteins via lysosomal digestion. While MHC-I is expressed in almost all nucleated cells in the body, except for germ cells and placental trophoblast, the basal expression of MHC class II molecules is restricted to antigen-presenting cells (APC), although it can be expressed on other cells in the context of stress or inflammation and by some tumoral cells (Neefjes et al., 2011). Either way, upon antigen recognition, the molecules associated with the TCR transmit the signal through their intracellular domains inducing the activation of different programs in the lymphocyte (Kumar et al., 2018). In the case of cytotoxic T lymphocytes (CTL), the activation eventually leads to the destruction of target cells (Golstein & Griffiths, 2018).

1.2. HLA-I molecules and pathway

1.2.1. Structure of HLA-I molecules

The molecules of major histocompatibility complex (MHC) class I, also known as human leukocyte antigen (HLA) class I in humans, is a family of surface proteins composed of a heavy polymorphic α-chain and an invariant chain, the β-2-microglobulin (β2m). Their function is to present peptides to $CD8^+$ T cells (Bjorkman et al., 1987) (Figure 3, left).

The HLA-I peptides bind into a cavity located on the α chain known as the peptide-binding groove. The peptides bound to this cavity are restricted in length (8-12 aa) and amino acid composition, which is mainly determined by peptide anchor residues that bind to the HLA pockets (Figure 3, right) (Kubo et al., 1994). There are three types of HLA-I α chains encoded by the HLA-A, HLA-B, and HLA-C genes and they are the most polymorphic molecules in humans, with 7.000 to 8.000 HLA alleles reported to date for each of the three genes (https://www.ebi.ac.uk/ipd/imgt/hla/about/statistics/). The polymorphisms concentrate in the peptide-binding groove that defines each HLA allele and have been found to determine peptide binding (van Deutekom & Keşmir, 2015). Hence, the specific HLA allele limits the peptide sequences that can be presented to the immune system. As humans can be heterozygous for the three loci, up to six distinct forms of HLA class I molecules can be expressed in each cell of the body (except for the above-mentioned exceptions).

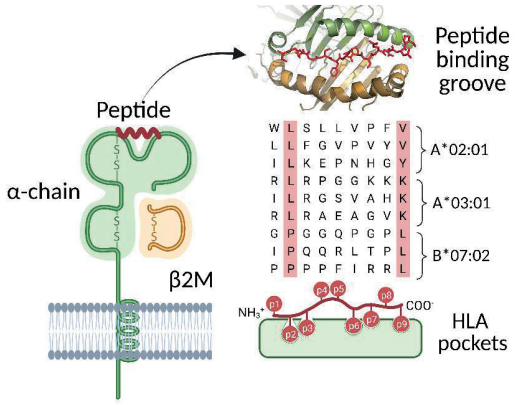


Figure 3. Structure of HLA-I molecules. The pHLA-I complex, composed of a heavy polymorphic α-chain, an invariant β-2-microglobulin chain, and the loaded peptide is depicted on the left. A schematic representation of the peptides on the peptide-binding groove is depicted on the right. 3D structure of the peptide-binding groove (top), potential peptide sequences bound to each HLA allele specified (middle), and the HLA pockets (bottom). The anchor residues at positions p2 and p9 are highlighted in red. Image created with BioRender.com.

Due to the extreme HLA diversity and co-expression of up to 6 alleles, each individual will present a different set of peptides to their T cells, known as peptide repertoire. Therefore, the wide application and use of effective T cell-based therapies is very complex and largely requires personalization.

1.2.3. Conventional HLA-I antigen presentation pathway

The HLA-I molecules display small fragments of the proteome on the plasma membrane of cells which is crucial for T-cell-mediated immune recognition of pathogens, but also other abnormalities such as cancer. The antigen processing and presenting process is complex and comprises several steps and specific conditions (Figure 4).(Blum et al., 2013)

On one hand, peptides presented by HLA-I generally originate from proteasomal degradation of cytosolic proteins (Kloetzel, 2001) that are later transported to the endoplasmic reticulum (ER) by the transporter associated with antigen processing proteins (TAP1 and TAP2). Once in the ER, peptides can be further trimmed by ER-associated aminopeptidases (ERAP1 and ERAP2) to generate shorter peptides that fit into the HLA-I peptide-binding groove. On the other hand, both HLA-I α and β2m chains are translated and inserted into the ER, where they associate to form a heterodimer in a process facilitated by chaperones. Next, the HLA-I/β2m heterodimers are recruited to a multiprotein complex called the peptide loading complex (PLC) which stabilizes the heterodimer and provides a peptide-rich environment that enables peptide loading, generating the HLA-I-peptide complex (pHLA-I). Once loaded with an optimal peptide, pHLA-I complexes traffic via Golgi to the cell surface where they will be presented to the $CD8^+$ T cells. Other components such as a tapasin-related protein (TAPBPR) or glycoprotein glucosyltransferase 1 (UGT1) have been described in the last years, which add even more complexity to the antigen processing and presentation process. (Aflalo & Boyle, 2021)

Although all the expressed intracellular proteins have the potential to be presented on HLA-I, only a small fraction of the peptides will actually reach the plasma membrane for immune recognition. The relative amounts of the different HLA-I peptides presented at the

cell surface are influenced by the efficiencies of interaction with all the components above mentioned, but it is also highly dependent on the rate of degradation of their source proteins. Indeed, because HLA peptides are presented following the degradation of their proteins of origin, short-lived proteins are more represented on the immunopeptidome than long-lived proteins. (Milner et al., 2006).

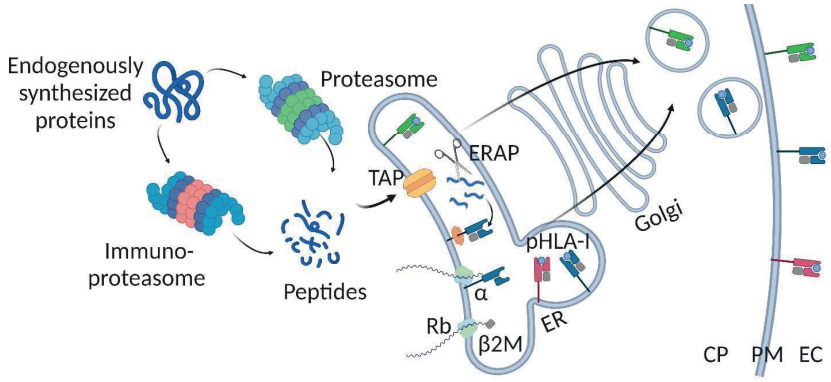


Figure 4. Conventional HLA-I antigen processing and presentation pathway. Endogenously synthesized proteins are degraded in the proteasome or immunoproteasome. The resulting peptides are transported to the ER through TAP and are further trimmed by ERAP to generate shorter peptides that fit into the HLA-I peptide-binding groove. HLA-I α and β2m chains are translated and inserted into the ER, where they are recruited to a multiprotein complex called PLC, enabling the peptide loading. The pHLA-I complex traffics via Golgi to the cell surface. Endoplasmic reticulum (ER), ER-associated aminopeptidases (ERAP), transporter associated with antigen processing proteins (TAP), ribosome (Rb), HLA-I α chain (α), β-2-microglobulin (β2m) HLA-I-peptide complex (pHLA-I), peptide loading complex (PLC), cytoplasm (CP), plasma membrane (PM), extracellular (EC). Image created with BioRender.com.

Due to the fact that the proteasome is the main responsible for intracellular proteolysis, it plays a key role in HLA-I presentation. Proteasomes are multicatalytic proteinase complexes composed by several subunits ($\alpha_{1\text{-}7}$ and $\beta_{1\text{-}7}$) that are constitutively expressed and abundantly located in nuclei and cytoplasm (Kloetzel, 2001). However, in some instances, the constitutive subunits β1, β2, and β5 can be replaced by inducible subunits (β1i, β2i, and β5i, respectively) to assemble a specialized form of proteasomes, known as immunoproteasomes (Basler et al., 2013). These are mainly expressed in lymphoid tissues, including thymus, lymph nodes, and spleen, but they can also be expressed in other cells upon cytokine stimuli such as IFN-γ. The incorporation of these subunits modifies the proteolytic specificity and is associated with increased efficiency of antigen processing for presentation on HLA-I. The immunoproteasome is crucially involved in mediating protective immunity against viral and bacterial antigens, and it is also involved in cancer immunity, although the role is controversial (Leister et al., 2022). While some studies suggest a pro-oncogenic role in the intestine due to inflammation-driven carcinogenesis (Leister et al., 2021), it has also been associated with better prognosis and response to checkpoint therapies in melanoma (Kalaora et al., 2020).

1.3. Central tolerance

Although peptide presentation on HLA-I is a requirement for detection by $CD8^+$ T cells, presentation does not necessarily ensure recognition. In fact, most of the peptides that are presented on HLA-I are derived from normal self-proteins that are typically not recognized by T cells, as opposed to peptides derived from foreign sequences. The ability of T cells to discriminate between self and non-self is one of the crucial properties of the immune system and is critical for both defense and tolerance.

During T-cell development, an extraordinary high diversity (in the order of 10-12 to 10-15) of TCR capable of recognizing virtually any pathogen but also self-antigens are generated in the thymus (Irla, 2022). Once TCR loci are rearranged, T-cell progenitors are first selected to recognize autologous HLA molecules presented by cortical thymic epithelial cells (cTEC), a process known as positive selection. Afterwards, immature T cells migrate to the thymic medulla where they undergo negative selection. The expression of AIRE on medullary thymic epithelial cells (mTEC) induces the transcription of a wide variety of genes otherwise restricted to specific tissues and contributes to an efficient negative selection, but other mechanisms are still under active investigation. As a result, a broad representation of the self-proteins expressed in the body are potentially presented to immature T cells which die by apoptosis upon strong TCR signaling. Hence, a large proportion of self-reactive cells are eliminated, thus reducing the possibilities of their potential harmful activation in the periphery. (Morris & Allen, 2012)

Nonetheless, despite this strict developmental control, some self-reactive T cells escape and migrate to the periphery. In fact, spontaneous anti-tumor T-cell responses recognizing certain self-proteins are relatively frequent in certain autoimmune pathologies and in cancer patients indicating that central tolerance, against a fraction of normal self-antigens, is incomplete.

1.4. Tumor antigens

Tumor antigens are defined as proteins expressed by tumor cells that are presented on HLA-I and that are able to elicit a T-cell response. Because T cells targeting tumor antigens can mediate tumor killing and are at the core of cancer immunotherapies, their identification has been of great importance in the field.

Since the first tumor antigen recognized by naturally occurring T-cell responses was discovered in humans (van der Bruggen et al., 1991), many studies have been focused on the identification of tumor antigens to exploit them therapeutically. The approaches used can be categorized into so-called “forward” and “reverse” antigen identification strategies. Forward strategies aim to identify tumor antigens that are recognized by antigen-specific T cells typically isolated from cancer patients. By contrast, reverse strategies aim to isolate antigen-specific T cells specific for a predicted or selected antigen, for instance through the immunization of mice (Parkhurst et al., 2009) or stimulation of naïve T-cell responses in healthy donors (Gerdemann et al., 2011).

Tumor antigens have been classically categorized according to tumor-specificity and have been divided into two main groups (Figure 5): 1) antigens of low tumor-specificity, also known as tumor-associated antigens, which include antigens that are also present in normal cells, and 2) antigens with high tumor-specificity, which include antigens whose expression is typically restricted to malignant cells (Coulie et al., 2014)

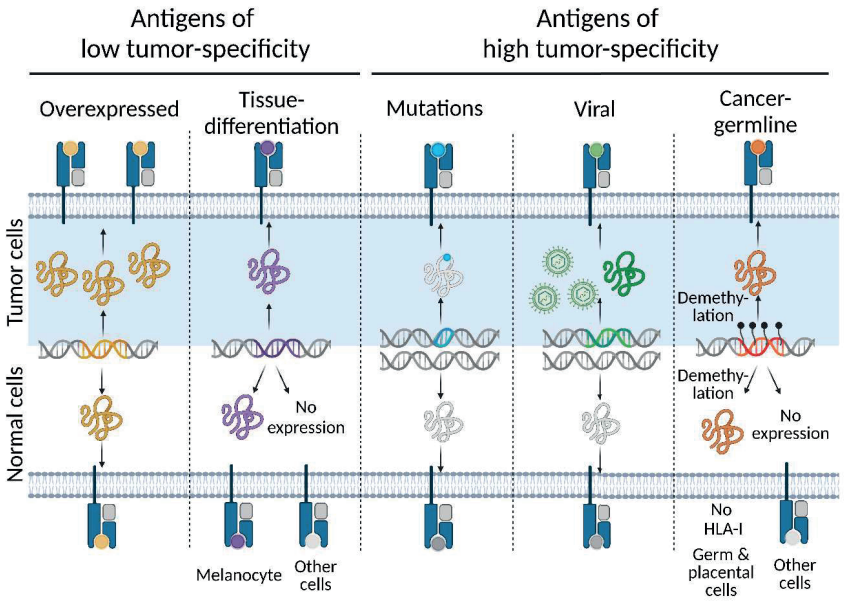


Figure 5. Classification of human tumor antigens. Expression and HLA-I presentation in tumor (top) and normal cells (bottom) of antigens with low or high tumor-specificity are depicted from the left to right, respectively. Antigens with low tumor-specificity include antigens that are overexpressed in tumor cells compared to normal cells and antigens derived from proteins involved in the differentiation of the specific tissue from which each tumor originates. Antigens with high tumor-specificity include antigens derived from tumor-specific mutations, viral proteins, and cancer germline antigens. The latter are exclusively expressed in germline cells and tumor cells as a result of gene demethylation, but are not presented on the cell surface of germline cells due to the lack of HLA-I molecules. Adapted from Coulie et al., 2014. Image created with BioRender.com.

1.4.1. Antigens of low tumor-specificity

This group comprises antigens encoded by non-mutated self-genes that may also be expressed in selected normal tissues. These include antigens derived from proteins that are overexpressed in tumors and tissue differentiation antigens.

Overexpressed antigens are products of normal genes that are expressed in several normal tissues but are overexpressed in certain types of cancer (Figure 5). For instance, the oncogene and growth factor receptor ERBB2 (also known as HER2) is overexpressed in many epithelial tumors including ovarian and breast carcinomas (Fisk et al., 1995). Another example is the transcription factor Wilms' tumor protein (WT1) which is more highly expressed in leukemic cells than in normal cells (Rezvani et al., 2005). T-cell responses targeting both HER2 and WT1 have been described in patients. In addition, antigenic peptides derived from carcinoembryonic antigen (CEA), which is expressed in a variety of epithelial cancers such as colorectal cancer, have been identified through reverse immunology and have also been explored as therapeutic targets (Parkhurst et al., 2009)

Differentiation antigens are expressed only in tumor cells and in the corresponding normal tissue of origin (Figure 5). Spontaneous T-cell responses to differentiation antigens have been well documented in patients with melanoma. The main differentiation antigens frequently recognized by naturally occurring T cells are derived from Tyrosinase encoded by TYR gene (Brichard et al., 1993), MART-1 encoded by MLANA gene (Coulie et al., 1994), and GP100 encoded by PMEL (H Bakker et al., 1995). In addition, antigenic peptides derived from the prostate-specific antigen (Correale et al., 1997) and prostatic acid phosphatase (PSA) (Olson et al., 2010), which are absent from other tissues have also been identified by reverse immunology and explored as therapeutic targets.

The fact that T-cell responses targeting these low tumor-specific antigens have been described in the literature indicates that central tolerance is not complete. Nevertheless, the affinity of TCR targeting these antigens is expected to be low since high-affinity TCR against normal cellular proteins are usually eliminated during negative selection in the thymus.

1.4.2. Antigens of high tumor-specificity

Three types of tumor antigens display high tumor specificity: antigens derived from mutations, viral proteins, and cancer-germline antigens.

Mutated genes accumulate during tumor development and have been shown to greatly contribute to the immunogenicity of human tumors (Coulie et al., 1995; Hanada et al., 2022; Kristensen et al., 2022; Parkhurst et al., 2019; Stevanović et al., 2017; Zacharakis et al., 2022). These antigens result from genomic perturbations that occur exclusively in tumor cells (Figure 5). They can be derived from single nucleotide variants (SNV), insertions or deletions (INDELS), and gene fusions that produce new antigenic peptides by changing one amino acid, altering the phase of the reading frame, by extending the coding sequence beyond the normal stop codon or by generating new DNA arrangements. Although they are truly tumor-specific, most of the immunogenic mutations (i.e. neoantigens) are consequence of "passenger mutations" and are typically private for each patient (Parkhurst et al., 2019). Only a small fraction of such neoantigens are shared across patients, which are frequently concentrated in positions of oncogenic driver genes known as "hotspot driver mutations" (Malekzadehet al., 2020).

A subset of cancers arise as a result of viral infection that leads to the expression of **viral proteins** (Figure 5) with oncogenic properties. For instance, human papillomavirus (HPV) is associated with a subset of head and neck, cervical, and anal cancers which express oncoproteins (E6 and E7) that can be recognized by patients' T cells (Eberhardt et al., 2021a). This has also been shown for other virus-driven solid cancers, including Merkel cell carcinomas resulting from polyomavirus infection (Iyer et al., 2011), and nasopharyngeal carcinomas that result from an Epstein-Barr virus (EBV) infection(Lee et al., 2000).

Due to their foreign nature, no central tolerance pressure against mutated and viral antigens is thought to exist. Therefore, T cells targeting them are not eliminated in the thymus, can be readily isolated from patients' tumors or peripheral blood, and typically bear high-affinity TCR (Aleksic et al., 2012; Bobisse et al., 2018).

In contrast, **cancer-germline antigens** are derived from non-mutated self-genes that are expressed during fetal development. These genes are epigenetically silenced in most normal adult tissues, except germ cells and placental trophoblast, which do not express HLA-I, thereby their presentation is generally lacking in normal adult cells (Figure 5)(Akers et al., 2010). As a result of DNA demethylation, they can be expressed in a variety of tumors and natural T cell responses in cancer patients are also frequently observed (Almeida et al., 2009; van der Bruggen et al., 1991)

1.5. Cancer Immunity Cycle

To develop an effective immune response against cancer, a series of events must be initiated and followed consecutively to kill cancer cells and ultimately lead to tumor rejection (Figure 6).(Chen & Mellman, 2013a).

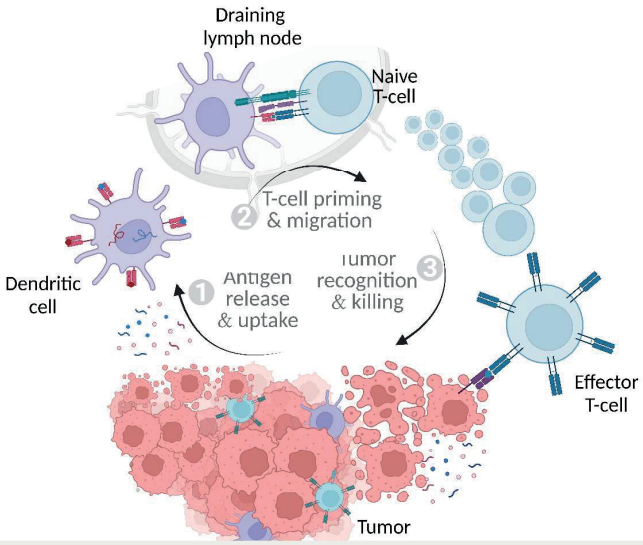


Figure 6. Cancer immunity cycle. (1) The first step is the release of tumor antigens that are captured by dendritic cells (DCs), which are then processed and cross-presented on HLA-I . (2) Next, these APC migrate to lymph nodes where they prime naïve $CD8^+$ T cells by presenting the antigen together with co-stimulatory signals. The stimulated T cells proliferate and differentiate into effector $CD8^+$ T cells, leave the lymph nodes and migrate to tumors. (3) Activated effector T cells specifically recognize and kill tumor cells that are presenting their cognate antigen on HLA-I. This, in turn, releases additional tumor antigens that increase the breadth and depth of the response in subsequent rounds of the cycle. Adapted from Chen & Mellman, 2013b. Image created with BioRender.com.

The first step is the release of tumor antigens that are captured by dendritic cells (DC), which are then processed and presented on HLA. Next, these APC migrate to lymph nodes where they present the antigens to naïve T cells. An important consideration at this point of the immune cycle is that the internalized antigens must be cross-presented to efficiently prime naïve $CD8^+$ T cells since these cells recognize the epitopes on HLA-I (Cruz et al., 2017). In the classical presentation pathway, the peptides presented on HLA-I (described in section 1.2.2) are derived from proteins synthesized by the cell itself, whilst the peptides derived from exogenous proteins are typically presented on HLA-II. In contrast, cross-presentation pathway, peptides derived from internalized proteins are presented on HLA-I (Norbury, 2016). Moreover, the presentation must be accompanied by co-stimulation which is induced by immunogenic signals. Otherwise, naïve CD8 T cells may be tolerized instead of activated (Lenschow

et al., 1996). DC play a key role in this step as they are professional APC that display a high ability to cross-present and express co-stimulatory molecules under inflammatory conditions. In fact, DC need to be pre-activated through a process that has been designated as "licensing" to be able to present antigens to CD8 T cells and induce an effector response. DC licensing signals are provided by CD4 T lymphocytes an/or innate immunity ligands.

Lastly, the activated effector CD8 T cells leave lymph nodes and traffic to tumors where they can specifically recognize their cognate antigen that is presented on HLA-I from tumor cells and kill them. This, in turn, releases additional tumor antigens that increase the breadth and depth of the response in subsequent rounds of the cycle.

If any of these steps fails, the tumor can escape the immune system. Indeed, because in cancer patients this cycle is not optimal, the goal of cancer immunotherapy is to stimulate the anti-tumor response by initiating or reinitiating the cycle of cancer immunity, which can be pursued at any point of the cycle depending on the specific therapy. For instance, by inducing T-cell priming and proliferation, trafficking into the tumor, stimulating the effector function, or reducing the immunosuppressive microenvironment.

2. T-cell-based cancer immunotherapies

The development of the different modalities of cancer immunotherapy has reached an important inflection point in the history of cancer therapy. These therapies attempt to mobilize the immune system to recognize and kill cancer cells through different mechanisms and T cells play a central role in most of them (Table 1). They can be broadly divided depending on whether the antigen-specificity of the T cells is known before the treatment or not.

Table 1. Principal T-cell-based cancer therapies

	Therapy		Product	Mechanism of action
Not antigen-specific	IL-2		Human recombinant IL-2 cytokine	Stimulate proliferation of endogenous T cells potentially recognizing and killing the tumor
	Immune checkpoint inhibitors or co-stimulatory agonists		Monoclonal antibodies targeting inhibitory/stimulatory molecules	Block inhibitory signals or co-stimulate to activate T cells potentially recognizing and killing the tumor
Antigen-specific	Adoptive T-cell therapy	TILs	Autologous T cells expanded from tumor. Unselected or selected for tumor recognition	Transfer *ex vivo* expanded T cells of uknnown or known specificity potentially recognizing and killing the tumor
		TCR-transduced T cells	Autologous T cells genetically modified to express a TCR targeting a tumor antigen	Transfer modified T cells recognizing a specific antigen presented on HLA of tumor cells
		CAR-transduced T cells	Autologous T cells genetically modified to express a CAR targeting a surface tumor antigen	Transfer modified T cells recognizing a specific antigen expressed on the membrane of tumor cells independently from HLA
	Vaccines		Immunization with specific tumor antigens, administered as nucleic acid, peptides, inserted in viral vectors or loaded onto APCs	Induce endogenous T-cell responses to selected tumor antigens

Summary of distinct types of immunotherapies classified based on the knowledge of the specific antigens targeted and the mechanism of action. Immune-mobilizing monoclonal TCRs against cancer (ImmTACs), bispecific T-cell Engagers (BiTEs), and TCR mimics are not included.

2.1. IL-2

The systemic administration of interleukin-2 (IL-2) was the first type of non-specific immunotherapy applied in clinical trials and provided evidence that manipulation of the human immune system could reproducibly lead to durable tumor regressions (S. A. Rosenberg, 2014).

This cytokine was identified in 1976 as a T-cell growth factor produced by lymphocytes and capable of inducing the growth of T lymphocytes *in vitro*. A decade later, the administration of high doses of IL-2 in combination with lymphokine-activated killer (LAK) cells to a cohort of metastatic cancer patients demonstrated that it was capable of mediating tumor regression in humans(S. Rosenberg et al., 1985). Out of the total 25 patients treated, 11 objective responses were observed. Remarkably, 4 out of 7 patients with melanoma and 3 out of 3 patients with renal cancer experimented regression of metastatic lesions, and one patient had a complete response. Although generally reversible, early toxicities of high-dose IL-2 were notorious. They appeared to result from capillary leaking fluid specially to the lung interstitium sometimes compromising its function. Later on, it was shown that the efficacy of the therapy was mediated by IL-2 rather than LAK cells. These encouraging results led to an explosion of studies utilizing IL-2 in metastatic cancer patients. However, only rare responses were seen in patients with other tumor types. The US Food and Drug Administration (FDA) approved IL-2 for the treatment of metastatic renal cell carcinoma in 1992 and metastatic melanoma in 1998 (S. A. Rosenberg, 2014).

2.2. Immune checkpoint inhibitors

Studies of the immune response to chronic viral infection showed that after T-cell activation several inhibitory pathways are induced to prevent a potentially harmful prolonged T-cell response (Barber et al., 2006), which led to the concept of T-cell exhaustion. These molecules and pathways involved are known as immune checkpoints and their inhibition has been exploited in cancer immunotherapies, being PD-1 (programmed death 1) and CTLA-4 (Cytotoxic T-Lymphocyte Antigen 4) the most widely exploited so far (Sharpe, 2017).

Immune checkpoint blockade removes inhibitory signals of T-cell activation, which enables tumor-reactive T cells to overcome regulatory mechanisms and mount an effective antitumor response. However, the specific mechanisms of action of PD-1 and CTLA-4 blockade are highly complex and not fully understood(Wei et al., 2018). CTLA-4 blockade is thought to induce tumor rejection through distinct mechanisms, including the increased priming and activation of the effector compartment but also depletion of regulatory T cells(Quezada & Peggs, 2019). In contrast, PD-1 blockade is thought to reinvigorate CD8 T cells by restoring the activity of "exhausted" CD8 effectors despite continued PD-L1 (PD1 ligand) expression in the tumor microenvironment. However, the mechanism is also controversial since recent findings demonstrate that CD28 co-stimulation is necessary for responses to PD-1(Kamphorst et al., 2017). This suggests that PD-1 blockade acts not only in tumors but might also act in sites of priming. More recently, it has been proposed that PD-1 therapy bursts stem-like cells that express CD28 and have a self-renewal capacity (Eberhardt et al., 2021b; Im et al., 2016).

Although monotherapies using immune checkpoint inhibitors have achieved remarkable success, especially in melanoma, there is a need to improve efficacy across tumor types. A primary resistance mechanism is compensatory upregulation of additional immune checkpoint molecules. Therefore, several strategies combining targeting of inhibitory (such as LAG3, TIM3, TIGIT, VISTA, and) and activation molecules (such as ICOS, OX40, GITR, 4-1BB, CD40) are being currently tested in clinical trials. Moreover, despite regulatory approval of immune checkpoint inhibitors in a diverse range of human solid malignancies, most patients do not benefit from current cancer immunotherapies. Compiling data suggest that there is a correlation between tumor mutational burden (TMB) with better response to checkpoint blockade(-Yarchoan et al., 2017) and survival across multiple cancer types(Samstein et al., 2019). This supports a model in which neoantigens are a major driver of tumor immunogenicity.

2.3. Adoptive T-cell transfer

Adoptive T-cell transfer (ACT) is therapy that consists of the infusion of high numbers of tumor-reactive lymphocytes (up to 10^{11}) (Figure 7). The efficacy of this therapy relies on the ability of the infused T-cell product to recognize and kill cancer cells. ACT can use T cells naturally exhibiting antitumor activity, as tumor-infiltrating lymphocytes (TIL), or genetically engineered T cells modified to express receptors (TCR or CAR) targeting specific tumor antigens. A non-myeloablative lymphodepleting conditioning regimen prior to cell transfer is typically required, which eliminates endogenous T regulatory cells as well as endogenous lymphocytes that compete with the transferred cells for growth-promoting cytokines. In some instances, IL-2 is given following cell transfer to support the growth and survival of the cells infused.

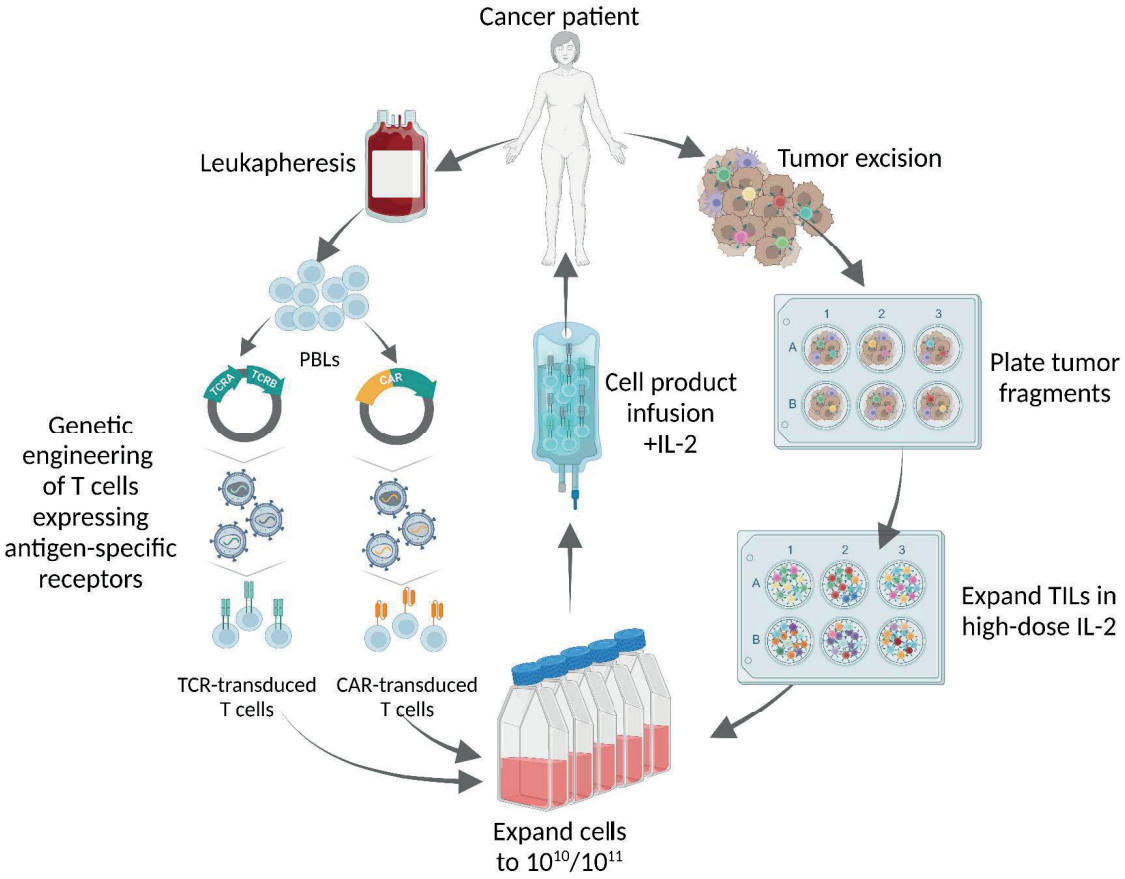


Figure 7. Adoptive T-cell transfer (ACT). Autologous lymphocytes are expanded up to high numbers (10^{10}- 10^{11}) and infused back into patients together with co-administration of IL-2 following a lymphodepleting chemotherapy. For ACT with TIL, a tumor biopsy or surgically excised tumor specimen is cut into fragments, and cultured in T-cell media containing high doses of IL-2 (right). For ACT with genetically engineered T cells, autologous PBL are obtained by leukapheresis and transduced with viral vectors encoding receptors (TCR or CARs) targeting specific tumor antigens (left). Peripheral blood lymphocytes (PBL), tumor-infiltrating lymphocytes (TIL), chimeric antigen receptor (CAR). Image created with BioRender.com

The advantages of ACT over immune checkpoint blockade include first, the ability to overcome the limited capacity of T-cell activation and expansion in an immunosuppressive tumor microenvironment by infusing high numbers of *ex-vivo* expanded T cells. Second, the lymphodepleting preparative regimen may remove certain T-cell suppressive mechanisms. And third, the use of peripheral blood lymphocytes (PBL) engineered to express a tumor-specific TCR obviates the requisite of finding pre-existing TCR targeting a tumor antigen in the patient's repertoire. One clear disadvantage is that ACT requires individual preparation of the patients' autologous T cells for infusion, and thus it can be performed only in a few specialized centers. This challenge is currently being addressed through the transfer of allogeneic T-cell products or the generation of a genetically modified PBL that can be used as a universal donor for all patients.

2.3.1. ACT of tumor-infiltrating lymphocytes (TIL)

The demonstration that TIL obtained from melanoma lesions and grown in IL-2 were able to kill autologous human tumors was an important step in the development of cell transfer therapy for the treatment of human cancers (Muul et al., 1987).

A clinical trial in 1988 with autologous TIL was the first report showing the ability of adoptive cell transfer to induce the regression of large established cancers in patients with metastatic melanoma (S. Rosenberg et al., 1988). Twenty patients with metastatic melanoma were treated with *ex vivo* expanded autologous TIL together with IL-2 and objective tumor regression was seen in 11 patients. Since then, many patients with melanoma, but also other tumor types have been treated with ACT of TIL (S. A. Rosenberg & Restifo, 2015).

Retrospective studies in melanoma patients have shown that the T cells infused frequently recognize differentiation or cancer-germline antigens, and in most cases, tumor-specific mutations. In a recent study, the frequency of neoantigen-specific $CD8^+$ T cells in TIL infusion products has been correlated with increased survival in patients treated with TIL (Kristensen et al., 2022).

Although TIL can be grown *in vitro* from virtually all tumors, long-term durable regressions have been mostly achieved in melanoma rather than other tumor types. This can be explained, at least in part, because TIL with high antitumor reactivity are consistently expanded most frequently from melanomas. Nonetheless, T cells recognizing neoantigens can be found, albeit probably at lower frequencies, in patients with other epithelial cancers (Hanada et al., 2022; Stevanović et al., 2017; Tran et al., 2015; Zacharakis et al., 2022), and encouraging case reports have shown that ACT can mediate effective responses when TIL products are selected

for tumor-recognition (Tran et al., 2016; Zacharakis et al., 2018). In one report, a patient with metastatic cholangiocarcinoma was initially treated with bulk autologous TIL without objective clinical response (Tran et al., 2014). Thereafter, the patient received a second infusion product with TIL selected to contain more than 95% of mutation-reactive T cells which led to a dramatic regression of metastases. In another report, a metastatic breast cancer patient was treated with anti-PD-1 combined with TIL selected to contain T cells recognizing peptides derived from 3 tumor-specific mutations (Zacharakis et al., 2018). At 6 weeks after cell transfer, the tumor burden was reduced by 51%, and after 22 months, all target and nontarget lesions were resolved.

Altogether, the clinical experience of ACT with TIL has shown that this personalized therapy is feasible and can be highly effective leading to durable tumor regressions in selected patients, and when the tumor targets are identified and reactive TIL are enriched.

2.3.2. ACT of TCR-engineered T cells

Another strategy that can be used to increase the anti-tumor potential of the infused T-cell product is to genetically modify autologous T cells with virus encoding TCR recognizing specific tumor antigens. This therapy typically uses off-the-shelf TCR that were previously isolated and characterized. Because TCR recognize their cognate antigen in a particular HLA context, patients must be selected based on the HLA typing and based on the level of expression of the specific antigen targeted.

The first example of the successful use of this therapy was reported in 2006 in patients with melanoma treated with autologous lymphocytes encoding a TCR recognizing the melanoma-associated antigen MART-1 (MLANA gene), which was isolated from TIL of a patient responsive to ACT (Morgan et al., 2006). Out of the 31 patients treated, 4 (13%) experienced objective regression without apparent toxicity. The authors attributed the limited efficacy to the moderate ability of the TCR used to recognize low amounts of antigen and hypothesized that a more highly reactive TCR might be clinically more effective. This led to a second study using high-affinity TCR targeting the melanoma-associated antigens MART-1 and gp100 (PMEL gene) generated by immunizing transgenic mice (Johnson et al., 2009). However, although better objective antitumor responses were observed in 30% and 19% of patients, respectively, severe toxicities were also observed in the eye and ear associated with the destruction of healthy melanocytes present in these specific organs..

Later on, the administration of T cells transduced with TCR recognizing NY-ESO-1 showed that targeting cancer-germline antigens could induce tumor regression in synovial cell sarcomas as well as melanomas (P. F. Robbins et al., 2011). Importantly, although the TCR was modified to display high affinity, no severe toxicities were observed, indicating that the presentation of the antigen targeted was restricted to the tumor. In contrast, a TCR targeting the cancer germline antigen MAGE-A3 showed objective cancer regressions, but patients unexpectedly developed severe neurological toxicity. The TCR used was obtained from immunized mice and affinity-enhanced, and further studies showed that the toxicity was probably due to

the cross-recognition of a highly homologous peptide derived from MAGE-A12, a member of the MAGE family that is expressed in a small subset of cells in the brain. Another example of the use of an affinity-enhanced TCR targeting this same antigen MAGE-A3(Linette et al., 2013), led to a fatal event in the two patients treated due to cardiovascular toxicity. Further studies showed that the TCR recognized a completely unrelated protein (titin) expressed in the striated muscle. TCR targeting additional cancer germline antigens have shown promising clinical activity in patients with solid cancers and are actively being investigated in the clinic..

TCR targeting overexpressed antigens such as CEA have also been explored in clinical trials. Despite the use of affinity-enhanced TCR targeting this antigen resulted in objective regression of tumors in one of the three metastatic colorectal cancer patients treated, all of them suffered from severe transient inflammatory colitis (Parkhurst et al., 2011).

Although neoantigens derived from tumor-specific mutations have great potential as therapeutic targets due to their high tumor-specificity, the fact that most mutations are private has hampered the development of TCR-engineered T cells targeting them. Nonetheless, a few clinical trials with TCR targeting neoantigens are currently are actively recruiting patients (NCT04520711, NCT04146298, NCT03745326, NCT03190941, NCT05194735, NCT03970382, NCT04102436, NCT03412877). Very recently, a patient with pancreatic ductal adenocarcinoma resistant to immunotherapy was treated with genetically modified T cells expressing two allogenic TCR targeting $KRAS_{G12D}$ previously reported (Tran et al., 2016), and an overall partial response of 72% was observed (Leidner et al., 2022). Moreover, a recent report of a chemorefractory breast cancer patient treated with an allogeneic TCR specific for $p53_{R175H}$ showed an objective tumor regression (55% reduction in size) that lasted 6 months (Kim et al., 2022).

The use of TCR-engineered T cells is a promising strategy that has significantly expanded the range of tumors that can be treated by adoptive cell therapy including less immunogenic tumor types such as pancreatic cancer. Nonetheless, these clinical data emphasize the critical importance of assessing potential TCR off-tumor toxicities prior to clinical development, both on-target or off-target due to cross-reactivity, especially when using affinity-enhanced TCR that have not undergone thymic selection.

2.3.3. ACT of CAR-engineered T cells

Chimeric antigen receptors (CAR) are synthetic receptors that exploit the antigen-specificity of the antigen-binding domains of monoclonal antibodies and combine it to the intracellular T-cell activation domains. Hence, CAR-transduced T (CAR-T) cells recognize the antigen as an antibody does. This means that they recognize the 3D structure of intact membrane-bound molecules on the cancer cell surface, independently of HLA molecules, in contrast to conventional T cells. (Maus & June, 2016)

The first report of ACT using a CAR targeting CD19 in a patient with lymphoma was published in 2010 and showed a remarkable long-lasting tumor regression (Kochenderfer et al., 2010). Although normal B cells were also eliminated because they also express CD19, periodic immunoglobulin transfusions partially overcame the B-cell aplasia. Subsequent studies

in B-cell lymphomas and leukemias have shown impressive regressions which have generated considerable enthusiasm and led to the FDA's first approval of a genetically engineered cell therapy in 2017. By contrast, results applying CAR- T to solid malignancies have been comparatively modest(Chandran & Klebanoff, 2019) . Several mechanisms can explain the lack of effectiveness such as inefficient trafficking into the tumor bed or a highly immunosuppressive tumor environment.

2.3.4. Vaccines

Therapeutic cancer vaccines apply the large experience of producing preventive vaccines to pathogens to generate therapeutic vaccines in cancer.

Anti-tumor vaccines consist in active immunization designed to treat growing tumors by the induction and amplification of anti-tumor T cell responses. They aim at activating and priming naïve antigen-specific T cells through the effective presentation of tumor antigens by APC, thus stimulating and expanding pre-existing TCR. Several strategies to deliver antigens exists such as viral vectors, peptides, DNA, mRNA directly administrated together with adjuvants, or loaded onto DC.

Since the first human tumor antigen was described(van der Bruggen et al., 1991) enthusiastic efforts were made to develop cancer vaccines. Hundreds of clinical trials had been carried out targeting melanoma-differentiation antigens such as MART-1, gp100, tyrosinase or TRP-2, cancer-germline antigens such as NY-ESO-1 or MAGE-A12, or overexpressed antigens such as Her2/neu or telomerase proteins. More recently, advances in next-generation sequencing and bioinformatics tools have enabled the systematic and rapid identification of tumor-specific mutations some of which give rise to neoantigens. However, despite generating high frequencies of reactive T cells in the blood in some cases, in the vast majority of treated patients these vaccines failed to induce tumor regressions, as the estimated objective response rate was as low as 3.6%(S. A. Rosenberg et al., 2004). This low clinical effectiveness has been attributed to the inability to overcome the immunosuppressive tumor environment. In the case of vaccines targeting self-antigens it is thought to result from a limited number of T cells targeting these antigens in the periphery together with the low affinity of their TCR due to negative thymic selection. Despite the limited efficacy, therapeutic vaccines targeting neoantigens, as well as other antigens, are being actively explored in clinical trials using various adjuvants and delivery approaches as well as in combination with immune checkpoint inhbitors.(Blass & Ott, 2021)

3. Identification of tumor-reactive T cells and their targets

Due to the central role of T cells in cancer immunotherapy, many efforts have focused on the identification of tumor-reactive T cells and their specific targets. This knowledge is essential not only to broaden our understanding of the immune response to tumors, but also to develop more effective therapies.

Current conventional strategies to identify the specificity of T cells can be divided into methods that evaluate the response to a specific antigen (functional assays) and methods that evaluate whether the TCR can bind to a specific antigen bound to HLA (pHLA multimers). In the case of CD8+ T cells, upon antigen recognition, the molecules associated with the TCR initiate a signaling cascade of reactions that ultimately lead to T-cell activation, acquisition of effector functions and, eventually, the destruction of target cells. Hence, after co-culture of T cells with the targets of interest, cell proliferation, cytolytic activity, cytokine expression or the upregulation of activation markers can be used as indicators of antigen recognition. In contrast, pHLA-I multimers consist of multimerized and labeled soluble pHLA-I complexes (containing α-chain, β2M and peptide) that can bind to TCR recognizing the specific peptides they are bound to. Since pHLA-I tetramers were described in 1996(Altman et al., 1996), these have become essential reagents for the visualization and isolation of antigen-specific T cells. Although this technology was initially technically demanding and low high-throughput, several advances have facilitated their generation and broad application (Bentzen & Hadrup, 2017). The main disadvantage is that this method requires previous knowledge of the ability to bind to a particular HLA allele of the peptide of interest and that these reagents are typically only available for the most frequent HLA molecules.

One critical step to determine the specific antigen recognized by a T-cell is the selection of the candidates to test in further immunological assays. The particular proteins expressed by tumor cells, the mutations accumulated, and the patient's HLA-I alleles will determine the landscape of antigens presented to T cells. Therefore, because each patient's tumor is unique, the effective identification of tumor-reactive T cells and their targets should be performed in a personalized fashion.

3.1. Tumor cDNA libraries

The use of cDNA libraries was the first strategy employed to identify tumor antigens in humans and involved the laborious screening of cytotoxic tumor-reactive lymphocytes clones for recognition of tumor cDNA libraries. This method led to the discovery of the first tumor antigen identified in humans, the cancer-germline antigen MAGE-A1 (van der Bruggen et al., 1991).

This method typically starts with the isolation of total RNA from tumor cells which is then converted in pools of cDNA plasmids that are further transfected into recipient cells together with plasmids encoding specific HLA molecules. Following isolation of tumor-reactive T-cell populations, those are co-cultured with the transfected cells and assessed for recognition

of cDNA pools and, subsequently, individual cDNA plasmids. Last, to validate and define the epitopes, peptide sequences encoded in the recognized plasmids are synthesized and pulsed into cells transfected with the HLA alleles.

Although this method potentially allows the identification of any antigen expressed on tumor cells, this method is extraordinarily laborious and time-consuming, and therefore inappropriate for high-throughput antigen screening. Moreover, it has some technical limitations related to the size of the transcript, expression levels, or GC-richness of transcripts and hence, it may be not sensitive enough to detect some tumor antigens (Garcia-Garijo et al., 2019).

3.2. Personalized neoantigen screening with tandem minigenes or long peptides

Due to the growing interest in neoantigens and the advances in next-generation sequencing (NGS) it has become possible to interrogate tumor-specific mutations in a high-throughput, unbiased and personalized manner (Lu et al., 2014; Tran et al., 2014) (Figure 8).

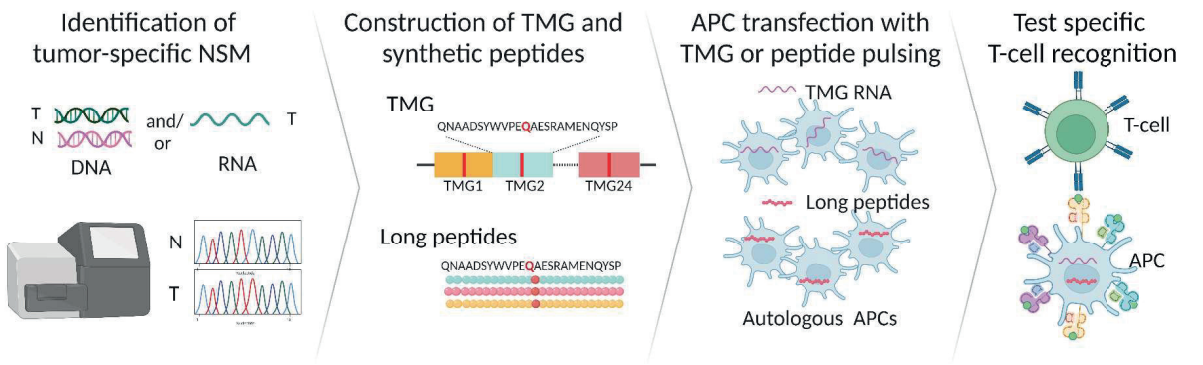


Figure 8. Personalized neoantigen screening using tandem minigenes (TMG) and long peptides. WES is performed on tumor and normal DNA to identify tumor-specific NSM. When available, RNAseq can also be used instead or in addition to WES to select mutations that are expressed. Once NSM are identified, minigenes encoding each of the NSM consisting of the mutation flanked by 12 amino acids (or more) of the wild-type protein on each side are designed, concatenated and cloned to construct TMG. Alternatively, long peptides encoding for the same sequences as minigenes can also be synthesized. Thereafter, TMG constructs are used to *in vitro* transcribe TMG RNA, which is electroporated into autologous APC, or long peptides are pulsed onto autologous APC. APC are used in co-culture experiments with T-cell populations to evaluate specific T-cell recognition. Whole exome sequencing (WES), non-synonymous somatic mutations (NSM), tumor (T), normal (N), tandem minigenes (TMG), antigen-presenting cell (APC). Image created with BioRender.com

This strategy starts with the identification of tumor-specific mutations by performing whole-exome sequencing (WES) of tumor and matched normal DNA. Although RNA sequencing can alternatively be used, the identification of non-synonymous somatic mutations (NSM) is less reliable, because of the poor quality of the RNA template but also due to misidentifications of germinal variants. Next, a minigene is generated for each NSM identified, consisting of the mutation flanked by 12 amino acids (or more) of the wild-type protein on each side. Tandem minigenes (TMG) are then constructed by concatenating the minigenes encoding the mutations and cloning these into plasmids that are used as templates to generate *in vitro* transcribed (IVT) RNA. Third, autologous APC are transfected with TMG RNA to allow expression, processing and presentation on the patient's HLA-I (and, potentially, HLA-II). Alternatively, the mutated sequences can be synthetized as long peptides (LP) and used to

generate peptide pools which are then pulsed into autologous APC, which is thought to better mimic the HLA-II pathway as opposed to TMG. Last, the T-cell populations of interest are co-cultured with the autologous APC expressing the mutations or pulsed with the peptides and the recognition is evaluated by functional assays. Peptide pools or TMG that elicit T-cell activation are further deconvoluted to identify the specific tumor antigens.

This method allows the detection of neoantigen-specific T cells targeting mutations in the exome with high efficiency, particularly for SNV but also frameshifts derived from short INDELs (Leko & Rosenberg, 2020). One advantage of this strategy is that it identifies neoantigens in a highly personalized and unbiased manner since it does not require previous knowledge of the specific epitopes presented, neither in length nor regarding the specific allele they can bind to. Because autologous APC are typically used, the mutations encoded in the TMG/LP can be processed and presented presumably on all the patient HLA-I alleles. However, the application of this approach is restricted to tumors with a limited number of mutations given the cost and effort of screening a large set of candidates. In tumors with a high mutation burden, it is necessary to further select candidates based on RNA expression level or HLA binding prediction algorithms.

3.3. HLA-I binding prediction

The advances in computational biology and immunology have led to the development of prediction algorithms that can be used to prioritize the candidate peptides that are more likely to be presented on a particular HLA-I allele based on biochemical and biophysical properties.

An extensive number of different algorithms have been developed over the last years, each of them has its own particularities as they take into account one or several of the steps along the antigen processing and presentation pathway (Mei et al., 2019). One of the most widely used is NetMHCPan (Reynisson et al., 2021), which is trained with large datasets of experimentally defined HLA ligands based on binding affinity data combined with mass spectrometry-eluted ligands.

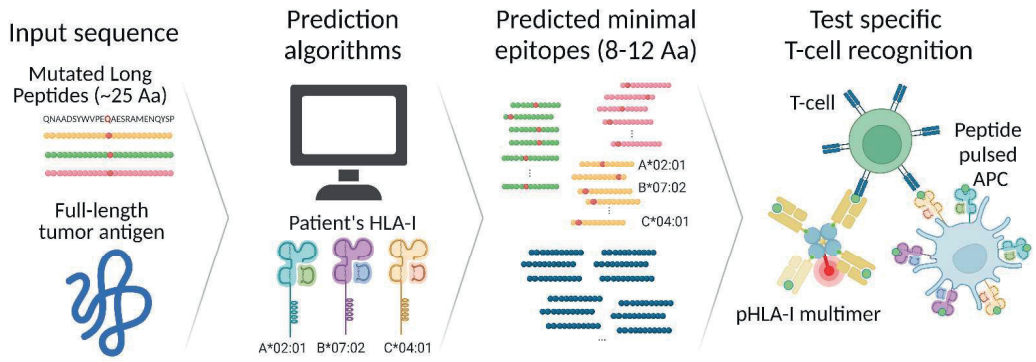


Figure 9. HLA-I binding prediction to prioritize candidate tumor antigens and screen for T-cell responses. Input Aa sequences of interest are fed to the HLA binding prediction algorithms. In the case of HLA class-I binding prediction algorithms, a list of 8-12mers that are predicted to bind to the HLA-I molecules specified is obtained. These short peptides are commonly prioritized based on binding affinity ranking, synthesized, and tested in functional assays using APC expressing the HLA of interest or through the generation of pHLA-I multimers. Antigen-presenting cell (APC). Image created with BioRender.com.

To predict which candidate antigens are more likely to bind to HLA, the prediction tools are commonly fed with a list of protein sequences as input (Figure 9). In the case of neoantigens, the input consists of long peptides containing the mutation flanked by the wild-type sequence (as explained in the previous section), for other types of antigens the full-length protein of interest is used as input. Because HLA-I binds short peptides typically from 8-12 aa, the algorithms generate small peptides from the input sequence and predict the likelihood of binding to the specified HLA alleles for each of them. Since several peptides derived from the same long peptide (8mer-12mer, containing the mutation at different positions) are generated, more than one sequence is likely to bind one or more alleles. Therefore, the potential number of candidate peptides can sometimes increase, rather than decrease. The resulting short peptides are commonly prioritized based on binding affinity ranking, synthesized and further tested in functional assays or with pHLA-I multimers (Figure 9).

One of the main advantages of these tools is that the predicted candidate peptides are the potential minimal epitopes being presented and recognized by T cells. However, is important to emphasize that they do not predict the immunogenicity but only the likelihood of binding to a given HLA allele. Nonetheless, new algorithms to predict the immunogenicity are now currently being developed (Schmidt et al., 2021; Smith et al., 2019).

3.4. Mass spectrometry-based immunopeptidomics: naturally presented HLA-I ligands

Tumor antigens can be identified by direct interrogation of the endogenous peptides that are presented by HLA molecules on the cell surface (HLA ligands), known as the tumor immunopeptidome or ligandome (Freudenmann et al., 2018).

Using a laborious task combining mass spectrometry with different chromatography separation methods together with immunogenicity assays, a peptide derived from the melanoma-associated antigen PMEL was identified in 1994 as the minimal epitope recognized by several human CTL clones (Cox et al., 1994). Since then, extraordinary advances in mass spectrometers and bioinformatics tools have made possible the study of the immunopeptidome in a more comprehensive manner and with increased sensitivity.

This method starts with the lysis of the tumor material followed by immunoprecipitation of the pHLA-I complexes (Figure 10). After elution and purification, the peptides bound to HLA molecules are separated and acquired by liquid chromatography coupled tandem mass spectrometry (LC-MS/MS), typically in discovery mode (non-targeted). Next, the amino acid sequence of each mass spectrum (MS) is usually inferred by database search. To identify tumor-specific peptides including mutations, customized databases are commonly generated by combining NGS data from patients' tumors with the reference proteome.

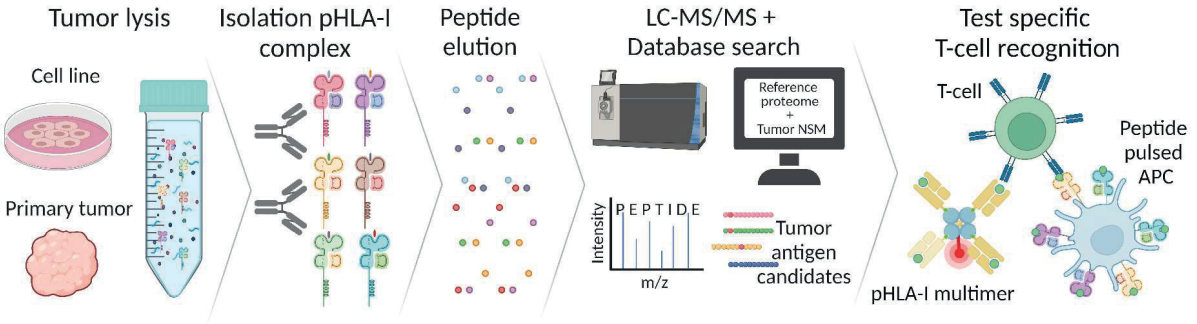


Figure 10. MS-based immunopeptidomics strategy to identify naturally presented HLA-I tumor antigens. Tumor cell lines or tumor samples are lysed and homogenized. Following isolation of pHLA-I complexes by affinity chromatography, the peptides bound to HLA-I molecules are eluted, purified, and analyzed by LC-MS/MS. Next, the amino acid sequence is usually inferred by searching each experimental MS2 spectra to a database containing the reference proteome and the mutated tumor-specific Aa sequences previously identified by WES or RNAseq. The candidate peptides derived from tumor antigens are synthesized and tested in functional assays using modified autologous APC or with pHLA-I multimers. Non-synonymous somatic mutations (NSM), antigen-presenting cell (APC). Image created with BioRender.com.

Cancer-associated HLA-I ligands have been identified by immunopeptidomics for several tumor types which have been used to develop cancer vaccines (Bassani-Sternberg & Coukos, 2016). But their identification remains challenging, particularly for neoantigens, which is reflected by the fact that very few are usually detected by this method. Nonetheless, private mutations presented on HLA-I have been identified from human tumor cell lines (Kalaora et al., 2016), directly from tumor samples (Bassani-Sternberg et al., 2016) and from patient-derived organoids (Newey et al., 2019).

Although the number of candidate tumor antigens identified is reduced as compared to the previously-mentioned strategies, the main advantage is that the candidates detected are certainly presented on HLA. However, as for HLA-I binding prediction (section 3.3), it the identification of peptides bound to HLA-I does not determine their immunogenicity. Hence, further evaluation of T-cell reactivity to such peptides is required.

4. Alternative sources of tumor antigens

The tumor antigens previously described in section 1.4 are considered conventional or canonical tumor antigens, as they are generated from coding regions of the genome via conventional transcription, translation, and proteasomal digestion. In contrast, unconventional or non-canonical (nonC) antigens, also known as cryptic, are those that arise from either non-coding regions of the genome or coding regions through aberrant transcription, translation, or post-translational modifications. Non-proteic antigens, though potentially important, are not discussed here.

Although the exome represents approximately 2% of the entire genome and it is generally accepted as protein-coding, in the last years it has been shown that almost 75% can still be transcribed and potentially translated (Djebali et al., 2012) (Figure 11). The development of new technologies and bioinformatics tools such as Ribo-seq analysis has enabled the discovery

of alternative open reading frames (aORF) substantially expanding the proteome. In fact, naturally occurring T-cell responses against some non-canonical antigens have been occasionally identified (mainly) using cDNA libraries and reported in the literature (Garcia-Garijo et al., 2019).

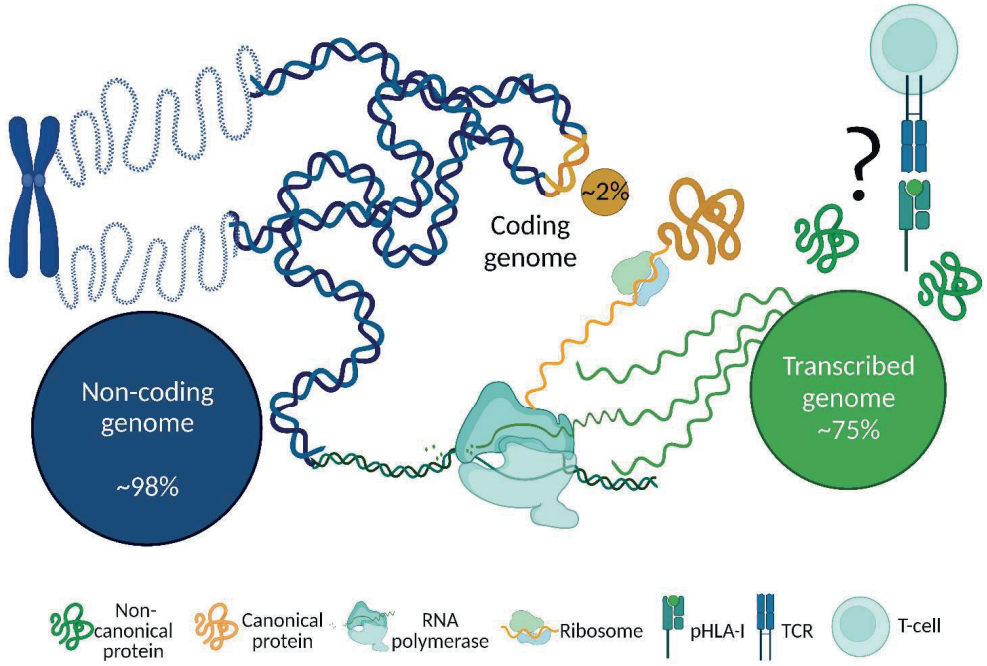


Figure 11. A big fraction of the presumably non-coding genome can be transcribed, and potentially translated. Whilst the exome (yellow) which represents approximately 2% of the entire genome and is generally accepted as truly protein-coding, the remaining 98% of the genome is considered non-coding (blue). In the last years, it has been shown that almost 70% can still be transcribed and thus possibly translated (green). Peptides derived from such regions can be presented on HLA-I and potentially recognized by T cells. Image created with BioRender.com.

The fact that non-canonical protein translation can be favored under conditions of cellular stress suggests that it may be enriched in tumors. Moreover, given that 99% of cancer-specific mutations are located in the non-coding regions of the genome (Khurana et al., 2016) it could greatly extend the number of neoantigen candidates.

4.1. Types of non-canonical antigens

4.1.1. Transposable elements

Transposable elements (TE) are highly repetitive DNA sequences in the human genome derived from previous retrotransposition events and comprise about 45% of the human genome, a sequence space that vastly eclipses that of the coding genome. All TEs can be broadly divided into two subgroups (Figure 12): Type I (retrotransposons) and type II (DNA transposons). Whilst retrotransposons act via RNA intermediates that are converted to DNA sequences before transposition (reverse transcription), DNA transposons do not require RNA. Type I TE are also known as endogenous retroelement (ERE) and can be divided into long terminal repeat (LTR) elements, which are mainly represented by human endogenous retroviruses (HERV) and non-LTR sequences that include short and long interspersed elements (SINEs, LINEs, respectively) and SINE-VNTR-Alu (SVA).

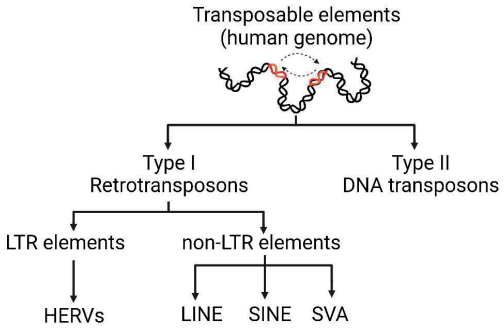


Figure 12. Classification of transposable elements in the human genome. Human transposable elements are categorized depending on whether they act via RNA to mediate the transposition in the genome (Type I; retrotransposons) or not (Type II; DNA transposons). Retrotransposons are further categorized based on the presence of long terminal repeat (LTR) elements.

The best studied TE in cancer are HERV which are remnants of ancient retroviral integration events and are located throughout the human genome constituting approximately 8.5% of the genomic DNA (Mayer et al., 2011). However, only a small fraction of these loci encode functional proteins since they are highly degenerated due to the accumulation of mutations or homologous recombination events between LTR, thus the proportion of transcribed HERV is unknown. Similarly to cancer-germline antigens, they are generally silenced in normal cells, but are more prone to be expressed in tumors due to DNA demethylation during malignant transformation. However, it has been shown that HERV can be expressed in normal tissues in a tissue-specific manner and that they are more highly expressed in mTEC and testis(Larouche et al., 2020). Nonetheless, they can be immunogenic and they can also induce a strong interferon response mediated by the formation of double-stranded RNA, a phenomenon called "viral mimicry". Several reports have shown that T cells targeting them can be found in healthy donors, suggesting an incomplete central tolerance (Attermann et al., 2018). More importantly, T-cell responses have been reported in some cancer patients with melanoma, breast, ovarian, and renal cell cancer, suggesting a role in cancer immune surveillance (Leko & Rosenberg, 2020). Although there is limited clinical experience of direct HERV targeting, a clinical trial with TCR-transduced T cells directed against HERV-E antigen is currently being tested in patients with metastatic renal cell cancer (NCT03354390).

Beyond HERV, the expression of transposable elements in tumors is controversial. On one hand they have been associated with immune infiltration and increased antigenicity(Kong et al., 2019), while recent reports have attributed the expression of TE to the T cells infiltrating tumors, rather than tumors themselves. Some resports have shown that they can be naturally presented on HLA-I in cell lines (Kong et al., 2019; Larouche et al., 2020) and primary melanoma tumors (Attig et al., 2019; Larouche et al., 2020). However, a major impediment to understanding their relevance in tumor immunity is the analytic challenge of accurately identifying these short-read sequences from such repetitive regions.

4.1.2. Aberrant mRNA Splicing and RNA editing

RNA transcripts can be altered by aberrant splicing or RNA editing, generating novel proteins that can potentially be detected by T cells (Figure 13).

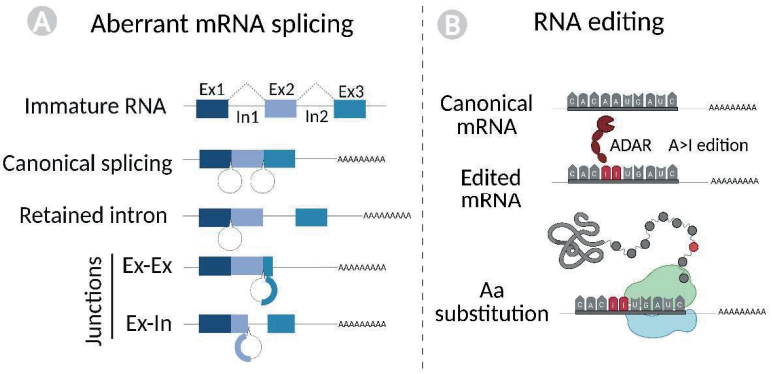


Figure 13. Novel peptides derived from aberrant mRNA splicing and RNA editing. (A) Scheme depicting mechanisms generating aberrant mRNA splicing (left). In a normal situation, the spliceosome machinery can detect and eliminate the introns to generate a mature mRNA (canonical splicing). In some instances, the introns may be aberrantly retained in the mature mRNA and translated (intron retention). Likewise, the splicing machinery can abnormally splice exons producing junctions derived from atypical exon-exon or exon-intron joins (junctions) potentially generating novel protein variants. (B) Scheme representing the most abundant type of RNA editing. RNA editing from adenosine-to-inosine (A-to-I) is catalyzed by ADAR. When edited RNAs are processed, the ribosomes interpret inosines as guanosines which can lead to non-synonymous substitutions potentially generating novel protein variants. Image created with BioRender.com.

Aberrant mRNA splicing in tumors may result in intron retention or novel exon/intron-exon/intron junctions which can be potentially recognized as non-self by T cells (Figure 13A) (Frankiw et al., 2019). Several reports have associated an increased alternative splicing usage in tumors with mutations in elements of the splicing machinery such as the splicing factors U2AF1 or SF3B1 (Sette & Paronetto, 2022) . More recently, in a study performed with RNAseq data from more than 8.000 patients, it was shown that tumors have up to 30% more alternative splicing events than normal samples (Kahles et al., 2018). In fact, in this report, the authors claimed that neoantigens derived from exon-exon junctions outnumbered the neoantigens derived from SNV, and thus they could be a better source of antigens for cancer immunotherapy. However, they did not investigate their HLA-I presentation or immunogenicity as the study was largely based on RNA expression and HLA-I binding prediction. Similar large-scale studies have been reported to date describing new pipelines for the identification of aberrant splicing events expressed(Jayasinghe et al., 2018) or even presented on HLA-I (Smart et al., 2018; Wang et al., 2021), yet the immunogenicity has not been analyzed experimentally.

Nonetheless, several intronic antigens recognized by T cells isolated from patients were identified early on using cDNA library screening approaches. Two examples are peptides derived from retained introns of gp100 gene (P. Robbins et al., 1997) and GnT-V gene (Guilloux et al., 1996) which although being highly expressed in melanoma cell lines were also expressed in melanocytes, among other studies (Godet et al., 2008; Harada et al., 2001; Lupetti et al., 1998; van den Eynde et al., 1999). More interestingly, the recognition of a peptide encoded by an exon-intron junction of MUM-1 gene containing a mutation was reported and conferred high tumor specificity (Coulie et al., 1995) in a patient with melanoma. Very recently, an elegant

study on uveal melanoma (Bigot et al., 2021) was reported. It is a rare tumor type typically displaying few somatic mutations but carrying mutations in the splicing factor SF3B1 in 20% of the patients. The authors showed that oligoclonal memory CD8⁺T cells targeting intron-exon junctions were frequently detected in SF3B1 mutated patients and that these T cells were able to recognize and kill uveal melanoma tumor cell lines.

Although this source of antigens is promising, especially in tumors containing mutations in splicing factors, the safety and efficacy of targeting antigens resulting from aberrant RNA splicing by ACT are still unknown.

Several types of RNA editing have been characterized so far (Eisenberg & Levanon, 2018) (Figure 13B). The most abundant is adenosine-to-inosine (A-to-I) editing, in which enzymes encoded by the adenosine deaminase acting on RNA (ADAR) gene family catalyze deamination of adenosine nucleotides to inosines. When edited RNAs are processed, the ribosomes and splicing machinery interpret inosines as guanosines instead of the adenosines encoded in the genome and it can lead to non-synonymous substitutions potentially generating novel protein variants. One report showed that HLA-I ligands derived from RNA editing were found in primary human melanoma tumors and a fraction of the healthy samples analyzed, albeit less frequently. Interestingly, TIL were able to specifically recognize the edited peptide but not the wild type counterpart, suggesting that edited peptides could be a source of tumor antigens (Zhang et al., 2018a).

4.1.3. Aberrant RNA Translation

Aberrant RNA translation comprises aORF or translation of allegedly non-coding sequences. These include the translation of regions entirely considered non-coding (such as non-coding RNA or intergenic regions) but also the translation of regions of coding transcripts such as 5′ or 3' untranslated regions (UTR) that presumably do not give rise to proteins (Figure 14A).

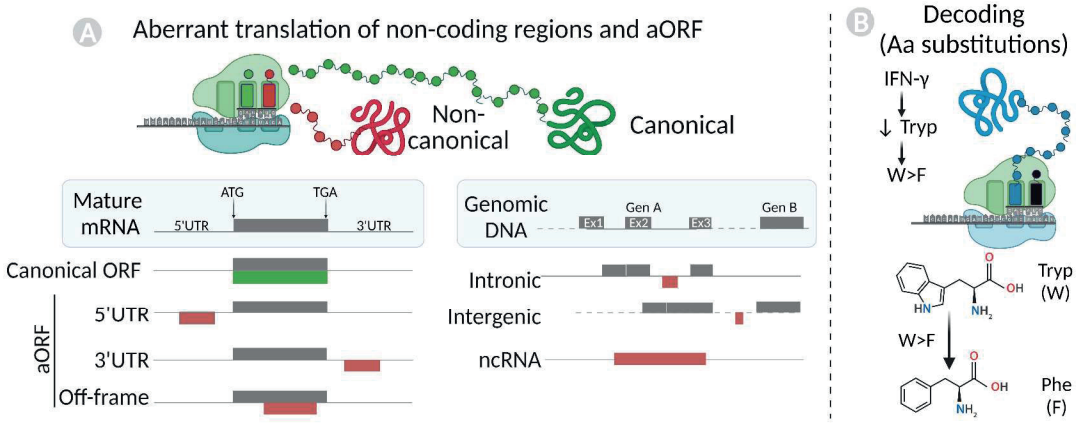


Figure 14. Novel peptides derived from aberrant translation. (A) Aberrant RNA translation of alternative ORF (aORF) encoded in normal mature mRNA transcripts (left) or translation of allegedly non-coding regions (right). Different mechanisms can induce the translation of 5′/3′ UTR (usually non-translated) or the initiation of the translation in a different frame generating completely novel proteins. Some allegedly non-coding regions of the genome can be transcribed and translated giving rise to novel proteins generated from introns, intergenic regions, or non-coding RNA. (C) Decoding. Prolonged IFN-γ stimuli can lead to tryptophan depletion and induce tryptophan to phenylalanine substitutions during ribosome translation.. Image created with BioRender.com.

A major cause of aberrant translation appears to be a phenomenon called ribosomal scan-through or "leaky scanning," where the ribosomes pass the first ATG and initiate translation from the next available ATG. Another relatively common mechanism is the initiation of the translation from non-ATG start codons (i.e. CTG, GTG, and TTG, which differ in one NT from the canonical start), or the translation from internal ribosomal entry sites (IRES sequence) that allow the internal initiation of translation in a cap-independent manner.

Other proposed mechanisms are stop codon read-through, where the ribosome proceeds to translate 3′- UTR down to the next in-frame stop codon, and programmed ribosomal frameshifting (PRF), where the ribosome encounters a "slippery site" and shifts the reading frame during translation. Although these mechanisms are thought to be a marginal cause of tumor antigen generation, a very interesting report was recently published where the authors used ribosome profiling in melanoma cells to investigate the effects of prolonged IFNγ treatment on mRNA translation (Bartok et al., 2021). They observed that the tryptophan depletion induced accumulations of ribosomes downstream of tryptophan codons which lead to ribosomal frameshifting events. Moreover, they identified aberrant trans-frame peptides (junction of wt – fs sequence) presented on HLA-I after treatment with IFNγ that were able to elicit T-cell responses induced by *in vitro* sensitization of PBL from healthy donors. Together, these results suggest that this mechanism may have a role in the immune recognition of melanoma cells *in vivo* in the context of inflammatory conditions by contributing to the diversification of the immunopeptidome landscape.

In fact, several tumor antigens encoded in aORF of previously described protein-coding genes recognized by T cells isolated from patients have been discovered using cDNA screening approaches. The first was a peptide encoded in an aORF of the TRP-1 gene (Wang et al., 1996), which encodes the tyrosinase-related protein 1 (gp75) and is important for melanocyte development and function. Two overlapping ORF could be translated from this gene but only one was recognized by TIL. This finding was followed by other similar examples, mainly in melanoma but also in other types of cancer (Aarnoudse et al., 1999; Poullion et al., 1999; Probst-Kepper et al., 2001; S. A. Rosenberg et al., 2002; Topalian et al., 1998).

Other mechanisms such as decoding can also alter the native sequence of the nascent proteins by substitutions of amino acids (Figure 14B). In a very recent report, it was shown that prolonged IFN-γ stimuli which leads to tryptophan (W) depletion induces phenylalanine (F) substitutions (Pataskar et al., 2022). Interestingly, they analyzed immunopeptidome data sets containing IFN-γ treated samples and found that peptides containing W>F substitution can be presented on HLA-I. Moreover, in an *in vitro* model, they showed that this can be sensed by T cells. Although interesting, the tumor specificity of such event was not demonstrated and natural T-cell responses in cancer patients have not been identified yet.

Although this source of antigens remarkably expands the repertoire of potential tumor antigens, the main concern, given the absence of relevant clinical experience, is whether these processes occur exclusively in cancer cells.

4.1.4. Post-translational changes

Post-translational changes can alter the native sequence of the protein which can potentially be detected by T cells (Figure 15).

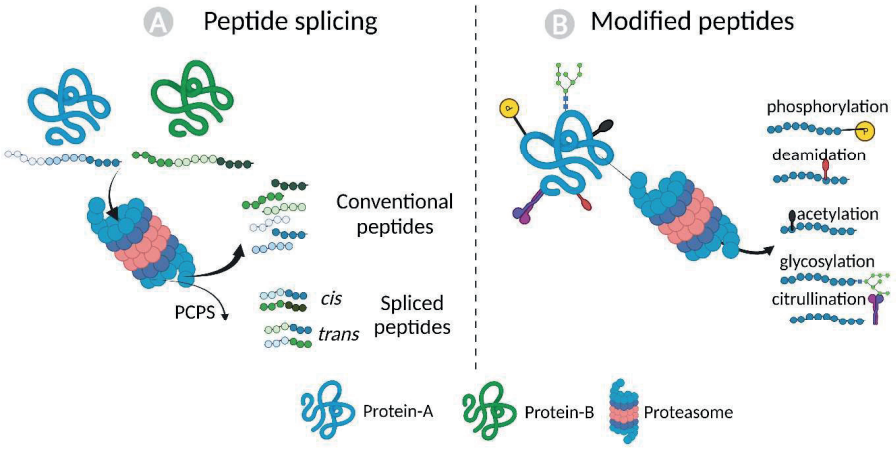


Figure 15. Novel peptides generated from post-translational changes. (A) Diagram depicting peptide splicing mechanisms. In addition to linear peptide fragments (conventional peptides), the proteasome can generate spliced peptides through PCPS. Such spliced peptides can originate from the same (*cis*-splicing) or distinct (*trans*-splicing) parental proteins. (B) Enzymatically modified proteins can be degraded by the proteasome and preserve these modifications as deamidation, phosphorylation, acetylation, citrullination, or glycosylation. Proteasome-catalyzed peptide splicing (PCPS). Image created with BioRender.com.

In addition to linear peptide fragments, it has been shown that the proteasome can generate spliced peptides through a process called transpeptidation or proteasome-catalyzed peptide splicing (PCPS) (Figure 15A). Such spliced peptides can originate from the same (*cis*-splicing) or distinct (*trans*-splicing) parental proteins or from reversed fragments. Whether these peptides are abundantly presented by HLA remains controversial. In 2016 Liepe and colleagues developed a computational algorithm for the detection of sliced peptides in the immunopeptidome and claimed that around 30% of HLA-I ligands were derived from PCPS. This observation was supported by other groups using alternative computational analysis (Faridi et al., 2018). In contrast, the reanalysis of the same data by other researchers indicated that the contribution of such spliced peptides is remarkably lower (less than 2 to 6%) and showed that a big proportion of the reported spliced peptides were due to false identifications resulting from methodological errors (Erhard et al., 2020; Mylonas et al., 2018). Even if PCPS is a rare event, some spontaneous T-cell responses to spliced peptides have been reported in cancer patients, typically identified through cDNA libraries (Dalet et al., 2011; Hanada et al., 2004; Michaux et al., 2014; Warren et al., 2006). However, the contribution of spliced peptides to the pool of immunogenic tumor antigens has been recently estimated to be marginal (Verkerk et al., 2022).

Another type of post-translationally modified tumor antigens consists of proteins with enzymatic modifications, such as deamidation, phosphorylation, acetylation, citrullination, or glycosylation (Figure 15B). The first case reported was a peptide derived from TYR and modified by deamidation of an asparagine into an aspartic acid (Skipper et al., 1996) that was recognized by TIL. Another example is a phosphorylated peptide derived from MART-1 that was found to elicit recognition by circulating $CD4^+$ T cells from a patient with melanoma (Depontieu et al., 2009). Although these modified tumor antigens can be immunogenic, their iden-

tification is challenging since the modification occurs after translation and hence it cannot be predicted at DNA or RNA level. Therefore, proteomics or immunopeptidomics-based studies are better suited to identify them.

4.2. MS-based immunopeptidomics studies to identify non-canonical antigens

All the above-reviewed studies suggest that alternative sources of antigens could expand the targets for cancer immunotherapy beyond neoantigens, cancer-germline or differentiation antigens. However, despite the enormous potential as a source of tumor antigens, their systematic identification in humans remains challenging.

One of the main hurdles to identify peptides derived for non-canonical proteins is that the reference proteome used to infer the amino acid sequences must contain the proteins of interest. The addition of all potential proteins originated from the 6-frame translation of any region of the genome into the database massively expands the search space, and the control of the false discovery rate (FDR) through standard methods becomes a hard task. Nonetheless, thanks to the advances in MS-based immunopeptidomics and bioinformatics tools, a growing number of proteogenomics-based pipelines have been developed to identify nonC HLA-I ligands (Table 2). To generate reduced databases, the inclusion of non-canonical proteins is typically restricted to candidates that are being transcribed or translated in a given sample by employing data from RNA-seq or Ribo-seq, respectively. Some strategies also compute the FDR separately for each class of non-canonical peptides enabling a more reliable identification. Indeed, several studies have demonstrated that these ligands represent a substantial proportion of the immunopeptidome (Erhard et al., 2020; Laumont et al., 2016; Ruiz Cuevas et al., 2021; Scull et al., 2021). A fraction of such aberrant translation events has been postulated to be specifically presented on tumor cells, representing an abundant source of tumor antigens greater than neoantigens or other conventional tumor antigens (Chong et al., n.d.; Cleyle et al., 2022; Laumont et al., 2018; Ouspenskaia et al., 2022; Zhao et al., 2020). However, the immunogenicity of nonC antigens identified through immunopeptidomics is largely unexplored.

The fact that peptides derived from nonC proteins are presented on HLA-I is not that surprising. The conventional pathway of antigen processing and presentation establishes that HLA-I peptides originate from degraded long-lived proteins (“retirees”). However, almost two decades ago, defective ribosomal products (DRiP) were shown to provide a great source of peptide ligands for HLA-I molecules (Yewdell, 2011). DRiP are a subset of unstable and rapidly degraded polypeptides including peptides produced as part of the pioneer round of translation, premature translation termination, and proteins failing to fold properly or assemble. This concept arises from observations in viral experiments, but it was later shown that cancer cells can increase the presentation of DRiP over retirees as compared to non-malignant cells (Bourdetsky et al., 2014). Hence, many of the non-canonical translation products are likely to be similar to DRiP: short, unstructured, and non-functional polypeptides that are efficiently presented on HLA-I presentation.

Table 2. Reported proteogenomics studies identifying nonC HLA-I ligands from several non-canonical sources

Publication	Approach basis	Samples	Immunogenicity
Laumont CM et al.,2016	Personalized DB RNAseq	EBV-transformed B cell line (non-tumoral)	Human helathy donor IVS
Laumont CM et al.,2018	Personalized DB RNAseq (aligment free) + mTEC exclusion (≈Tumor-specific)	B-lineage ALL samples lung tumor biopsies	Mice
Chong et al.,2020	Personalized DDBB: 1) RNAseq and WES 2) Ribo-seq	Melanoma samples (TCL and primary) Paired cancer and normal lung tissues	Cancer patient PBL IVS Cancer patient REPEd TIL
Erhard et al., 2020	Non-personalized RNAseq/Riboseq independent De novo sequencing+ mapping DDBB multiround	Melanoma Lung cancer Glioblastoma Triple-negative breast cancer Mantle cell lymphoma Fibroblasts	Not tested
Zhao et al., 2020	Same approach as in Laumont CM et al.,2018	Ovarian cancer samples	Not tested
Ruiz Cuevas et al., 2021	Personalized DB Ribo-seq combined with RNA-seq	Human B cell lymphomas	Not tested
Ouspenskaia et al., 2021	1) Non-personalized DB (nuORF DB) Ribo-seq across 29 tumor and healthy samples 2) Personalized DB: WGS tumor vs healthy + Riboseq	B721.221 cells CLL Glioblastoma Melanoma Ovarian cancer ccRCC	Not tested
Scull et al., 2021	Personalized DB RNAseq	THP-1 cell line	Not tested
Cleyle J et al., 2022	Same approach as in Laumont CM et al.,2018	Colorectal cancer-derived cell lines colon adenocarcinoma tumor and normal adjacent tissue(NAT)	Not tested
Xiang H, et al., 2022	Personalized DB WES + pNovo3	Colorectal cancer cell line HCT116	Not tested

The author and year of publication, brief description of the databases generated, samples analyzed and whether the immunogenicity was studied is described. Pipelines designed to identify spliced peptides are not included.

Their non-mutated nature and occasionally shared presentation across different patients have further attracted attention to these nonC antigens as new therapeutic targets. However, to what extent nonC HLA-I ligands are actually tumor-specific and contribute to cancer immune surveillance in patients remains poorly understood. These features of nonC antigens should be further elucidated before or in parallel to the development of immunotherapies targeting these antigens.

Hypothesis and objectives

The identification of tumor-reactive T cells and their specific targets is essential to broaden our understanding of the immune response to tumors and to develop more effective cancer immunotherapies. Recent evidence suggests that peptides derived from non-canonical proteins are specifically presented on HLA-I from tumor cells, thus expanding the repertoire of targetable tumor antigens.

We hypothesize that current methods frequently used to identify tumor antigens fail to identify a fraction of antigens targeted by tumor-reactive lymphocytes and that non-canonical tumor antigens detected through proteogenomics could contribute to tumor immunogenicity thus making them attractive targets for cancer immunotherapies.

The **general objective** of this thesis is:

To identify and characterize the personalized repertoire of tumor antigens recognized by tumor-reactive T cells in cancer patients including canonical or conventional tumor antigens, but also expanding the scope to non-canonical tumor antigen sources by proteogenomics.

The **specific objectives** are:

1. Evaluate the recognition of **conventional tumor antigens** (neoantigens, cancer-germline antigens or tissue differentiation antigens) by tumor-reactive $CD8^+$ T cells derived from cancer patients

1.1. Generate autologous patient reagents to identify tumor-reactive T cells.

1.2. Identify tumor antigens targeted by tumor-reactive lymphocytes using several methods including:

- Personalized neoantigen screen with tandem minigenes
- *In silico* minimal epitope prediction
- pHLA-I immunopeptidomics

1.3. Determine which fraction of the $CD8^+$ tumor-reactive lymphocytes recognize conventional tumor as compared to unknown tumor antigens.

2. Explore the immunogenicity of **non-canonical tumor antigens** identified by proteogenomics to address whether they can be exploited therapeutically

2.1. Identify tumor-specific HLA-I ligands derived from non-canonical proteins in patient-derived tumor cell lines in a personalized fashion using Peptide-PRISM pipeline.

2.2. Detect pre-existing T-cell responses to non-canonical HLA-I ligands in cancer patients.

2.3. Induce *de novo* T-cell responses to non-canonical HLA-I ligands through *in vitro* sensitization of donor peripheral blood lymphocytes.

2.4. Determine the tumor specificity of non-canonical HLA-I ligands and their potential use as new therapeutic targets.

Materials and methods

Patient samples

All patient samples were obtained as part of a project approved by the Institutional Review Board at Vall D'Hebrón Hospital and patients signed an informed consent.

Establishment of autologous tumor cell line

To establish patient-derived tumor cell lines a small fragment of tumor biopsies or surgically resected tumor was cultured in RPMI 1640 plus (Lonza) containing 10% FBS Hyclone (GE Healthcare), 100 U/mL penicillin (Lonza), 100 µg/mL streptomycin (Lonza) and 25 mM HEPES (Thermo Fisher Scientific) at 37°C in 5% CO2. The medium was replaced once every month until the cell line was established and then further expanded in T2 media containing RPMI 1640 plus (Lonza), 10%-20% FBS (Gibco), depending on the TCL, 100 U/mL penicillin (Lonza), 100 µg/mL streptomycin (Lonza) and 25 mM HEPES (Thermo Fisher Scientific) or cryopreserved until used. TCL were regularly tested for mycoplasma and were authenticated based on the identification of patient-specific somatic mutations and HLA molecules.

TIL expansion

For tumor-infiltrating lymphocyte (TIL) expansion, small tumor fragments were cultured in individual wells of a 24-well plate in T-cell media consisting of RPMI 1640 plus (Lonza) supplemented with 10% human AB serum (BST), 100 U/mL penicillin (Lonza), 100 µg/mL streptomycin (Lonza), 2mM L-Glutamine (Lonza), 25 mM HEPES (Thermo Fisher Scientific) and 6e6 IU IL-2 (Proleukin) at 37°C and 5% CO2. Fresh media containing IL-2 was added on day 5 and media was changed or TIL were split when confluent every other day thereafter. T cells were expanded independently for 15-30 days and cryopreserved until use. In some cases, T cells underwent a rapid expansion protocol (REP), as explained below.

Rapid expansion protocol (REP)

T cells underwent a rapid expansion protocol (REP) for 14 days using 30 ng/ml anti-CD3 (OKT3, Biolegend), 3e3 IU of interleukin IL-2 (Proleukin) and irradiated allogeneic PBMC (50 Gy) pooled from three donors as feeder cells in T-cell medium RPMI 1640 plus (Lonza):AIM-V (Gibco) containing 5% human AB serum (BST), 100 U/mL penicillin (Lonza), 100 µg/mL streptomycin (Lonza), 2mM L-Glutamine (Lonza), 12.5 mM HEPES (Thermo Fisher Scientific). After day 6, half of the medium was replaced with fresh T-cell medium containing IL-2 every other day. Cells were split when confluent, harvested on day 14, and cryopreserved until use.

PBMC isolation

Peripheral blood mononuclear cells lymphocytes (PBMC) were obtained using a Ficol density gradient (Lymphoprep, Stem cell) from pheresis or whole blood and cryopreserved for cell sorting, DNA extraction for WES and to expand B cells *ex vivo*.

T-cell sorting from PBL

PBL were sorted based on the expression of surface markers previously described to be enriched for tumor-reactive T cells in peripheral blood such as PD-1 (Gros et al., 2016) and other markers reported to be expressed by tumor-reactive TIL (Duhen et al., 2018). Briefly, PBMC were thawed and rested overnight without cytokines. Following $CD8^+$ enrichment using CD8 microbeads (Miltenyi Biotec), the Fc receptor was blocked (Miltenyi Biotec) and cells were stained for 30 minutes at 4°C with the following antibodies, CD3-PECy7 (BD, clone SK7, 0.5:50), CD8-APCH7 (BD, clone SK1, 1:50), PD-1-PE (Biolegend, clone EH12.2H7, 0.75:50), CD38-APC (Biolegend, clone HIT2, 0.5:50) and HLA-DR BV605 (Biolegend, clone L243, 0.75:50). $CD3^+CD8^+$ cells expressing PD-1^{hi} alone or in combination with HLA-DR and CD38 were sorted in BD FACS AriaTM and expanded using a REP as previously specified.

Generation of autologous antigen-presenting cells

To generate autologous antigen-presenting cells (APC), B cells were isolated from cryopreserved PBMC by positive selection using $CD19^+$ microbeads (Miltenyi Biotec) and expanded through CD40-CD40L stimulation by culturing cells for 4-5 days with irradiated NIH3T3 feeder cells constitutively expressing CD40L at 37°C in 5% CO2 in B cell medium. B cell medium consisted of Iscove's IMDM media (Gibco) containing 10% human AB serum (Biowest), 100U/ml Penicillin and 100 µg/mL streptomycin (Lonza), 2mM L-Glutamine (Lonza), and supplemented with 200 U/ml IL-4 (Peprotech). Up to three rounds of stimulation and expansion were performed consecutively. B cells were cryopreserved on day 5-6 after stimulation until use. When used after cryopreservation, B cells were thawed in B cell medium containing DNAse (Pulmozyme, Roche) 20 h before use in co-culture assays.

Alternatively, $CD4^+$ T cells were isolated from PBMC by positive selection using $CD4^+$ microbeads (Miltenyi Biotec) or FACS sorting and subsequently expanded through a REP for 14 days as explained above.

Whole-exome sequencing and RNA sequencing

To identify the tumor-specific NSM, genomic DNA was purified from a cell pellet of patient-derived TCL and matched PBMC. WES libraries were generated by exome capture of approximately 20,000 coding genes using SureSelect human All exon V6 kit (Agilent Technologies) and paired-end sequencing was performed on a HiSeq sequencer (Illumina) at Macrogen. The average sequencing depth ranged from 100-150 for each of the individual libraries generated.

Alignments of WES to the reference human genome build hg19 were performed using novoalign MPI from novocraft. Duplicates were marked using Picard's MarkDuplicates tool. Insertion and deletion (indel) realignment and base recalibration were performed according to GATK best-practices. Samtools was used to create tumor and normal pileup files. Four independent mutation callers (Varscan, SomaticSniper, Mutect and Strelka) were used to call somatic NSM according to the following criteria: minimum coverage of 10 reads, minimum 4 variant reads, greater than 7% VAF and called by 2 or more callers (single nucleotide variants) or 1 for insertions and deletions, and annotated as coding mutation in 2 of the three annotation data sets (Refgene, UCSC, Ensembl).

An mRNA sequencing library was also prepared from patient-derived TCL and healthy human cell lines using Illumina TruSeq RNA library prep kit. RNA alignment was performed using STAR. Duplicates and marked using Picard's Mark Duplicate tools. Transcripts per million (TPM) normalization was applied to the raw gene counts obtained from feature counts and used to evaluate the gene expression

Cloning, *in vitro* transcription of RNA, and electroporation

TMG were cloned into pcDNA3.1 with HindIII-BamHI and constructed as previously described(-Lu et al., 2014). For each nonsynonymous variant identified by WES, we constructed a minigene, consisting of the mutant amino acid flanked by 12 amino acids of the wild type protein sequence. Up to 24 minigenes were concatenated to generate a TMG construct.

The HLA sequences of interest or predicted ORF of the immunogenic nonC-TL peptides were cloned into pcDNA3.1 using BamH1 and EcoR1. HLA-I sequences were obtained from IPD-IMGT/HLA and codon-optimized. The predicted ORF were constructed from the second nearest upstream in-frame start codon (ATG, CTG, or GTG) to the last in-frame stop codon downstream; the sequence was not codon optimized nor additional start codons were added. As all expression constructs, they contained a Kozak motif upstream of the canonical or predicted start codon. All the plasmids were synthesized by Genscript.

For in vitro transcription (IVT) of RNA the plasmids were linearized with Not-I followed by phenol-chloroform extraction and precipitation with sodium acetate and ethanol. Next, 1 µg of DNA was used as a template to generate RNA by IVT using HiScribe# T7 ARCA mRNA Kit with tailing (New England) following manufacturer's instructions. RNA was precipitated using $LiCl_2$, resuspended at 1µg/ul in molecular grade H_2O, and stored at -80º until use.

From 0.5-1e6 TCL, healthy human cells, and B cells were harvested and resuspended in 100 µl of Opti-MEM media (Gibco) and transferred into a sterile 0.2 cm cuvette (VWR electroporation cuvettes). From 4 to 8 ug of RNA encoding for the sequence of interest were added for electroporation. Cells were electroporated at 150 V, 20 ms, and 1 pulse using an ECM 830 BTX-Electroporator. After electroporation, cells were resuspended in pre-warmed specific media containing DNAse (Pulmozyme, Roche). B cells were cultured in 5 mL polystyrene tubes. TCL and human healthy cells were cultured in tissue-treated 6 well plates with the corresponding media in a temperature and CO_2 controlled humidified incubator. After 20 h cells at 37ºC and 5% CO_2, cells were harvested, washed with PBS, and used in co-culture assays. A GFP RNA electroporation control was included for each cell line and assessed by Flow cytometry as a transfection control.

Peptides

The amino acid sequences of the identified tumor antigen HLA-I ligands were obtained from JPT Peptide Technologies (Berlin, Germany) as crude and used for screening, IVS, and MS validation with synthetic peptides. HPLC peptides were supplied by JPT Peptide Technologies (Berlin, Germany) and used in co-culture experiments to confirm the reactivities. Selected endogenous HLA-I ligands were ordered from Thermo Fisher Scientific as crude (PePotec grade 3) with one stable isotope-labeled amino acid and used for PRM validation.

Co-culture assays: IFN-γ enzyme-linked immunospot (ELISPOT) assay and detection of activation marker 4-1BB by flow cytometry.

Effector T-cell populations of interest were co-incubated with target cells and T-cell reactivity was evaluated after 20-24 h by measuring IFN-γ release by ELISPOT and by detection of 4-1BB (CD137) expression on the $CD8^+$ T cells by flow cytometry. T cells were thawed into T-cell medium supplemented with 3,000 IU IL-2 (Proleukin) and DNAse (Pulmozyme, Roche) three to four days before coincubation with target cells. All co-cultures were performed in the absence of exogenously added cytokines. Cells were stained with CD3-APCH7 (BD, clone SK7, 0.3:40), CD8-PECy7 (BD, clone RPA-T8, 0.1:40), CD4-PE (BD, RPA-T4, 0.3:40) and CD137-APC (BD, clone 4B4-1, 0.5:40), and in some cases mTRB-FITC (eBiosciences, clone H57-597, 0.2:40) for 30 minutes at 4°C, washed with staining buffer containing PI (1:2000) and acquired in BD FACSLyric™, BD FACSCanto™ or BD FACSLyric™. In parallel, IFN-γ secretion was detected using IFN-γ capture and detection antibodies (MABtech technologies) assessed by ELISPOT assay following manufacturer instructions. ELISPOT plates were analyzed and counted in ELISPOT reader. For all the assays, plate-bound OKT3 (1 µg/mL; Biolegend) was used as a positive control. Media, and/or autologous APC pulsed with irrelevant peptides were used as negative controls. In some cases, MHC-I blocking antibody (W6/32) was added to the target cells at a concentration of 100 µg/mL for 2 h and then diluted to 50 ug/mL during the co-culture assay.

For the detection of re-call T-cell responses, from 2e4 to 5e4 T *ex vivo* expanded TIL, sorted PBL or enriched populations of tumor-reactive lymphocytes were co-cultured with 1e5 to 2e5 peptide-pulsed autologous APC (either B cells, or $CD4^+$). T-cell reactivities were considered positive if the number of IFN-γ spots were greater than double the amount of the irrelevant control condition and greater than 40 spots. Additionally, reactivities had to be observed in at least two independent experiments. Crude peptides were preparations were used for screening, and the reactivities were further confirmed with HPLC grade peptides. Experiments were performed at least twice.

Enrichment of tumor-reactive and antigen-specific T cells

To enrich for tumor-reactive or peptide-reactive T cells, either expanded TIL, sorted PBL, or IVS T cells were co-cultured with an autologous tumor cell line or peptide-pulsed autologous APC for 20 h, $CD3^+CD8^+$ cells expressing 4-1BB were sorted in BD FACS AriaTM or BD Influx™ and expanded using a REP as previously specified. The same antibodies and dilutions used for co-cultures described above were scaled up for sorting $4\text{-}1BB^+$ T cells.

HLA-I expression and restriction

The membrane expression of HLA-I was determined by flow cytometry with HLA-ABC-APC antibody (Biolegend, clone W6/32, 0.5:50).

HLA-I alleles were cloned into pcDNA3.1 as described above (Cloning, *in vitro* transcription of RNA, and electroporation). To determine which HLA alleles presenting the immunogenic peptides identified, COS7 cells were transfected with plasmids encoding the individual HLA molecules using Lipofectamine 2000 (Life Technologies). After resting overnight, cells were harvested and pulsed with the corresponding peptides for 2 h, washed, and used as targets in co-culture assays.

In vitro sensitization of PBL

HLA-A*11:01 donor PBMC were stimulated with 5 independent peptide pools (PP) each containing up to 35 nonC-TL peptides selected by the prediction score to bind to HLA-A*11:01 according to NetMHCPan4.0. Cells underwent three consecutive rounds of stimulation every 7 days with 0.25 ug/ml per peptide and a combination of IL-21, IL-7 and IL-2. More specifically, at day 0, 5e6 donor PBMC were cultured in 24-well plates with Opti-MEM media (Gibco) containing IL-21 (Peprotech 25 ng/mL) and the corresponding PP at 0.25 ug/ml per peptide. On day 6, IL-2 (Proleukin 18 IU/mL) and IL-7 (Peprotech 10 ng/mL) were added. For STIM2 (day7) and STIM3 (day14), T cells from the previous STIM were harvested, counted, and re-stimulated with autologous irradiated PBMC (50Gy) pulsed with the corresponding PP at 1:10 ratio. Thereafter, fresh Opti-MEM media (Gibco) containing IL-2 (18 IU/mL) and IL-7 (10 ng/mL) was replaced when medium looked acidified or cells required splitting because of confluence.

De novo T-cell responses were evaluated after three stims by co-culturing IVS T cells with autologous B cells pulsed with the corresponding PP and analyzing 4-1BB upregulation by flow cytometry as described above (Co-culture assays). T cells recognizing the corresponding PP were sorted based on 4-1BB expression and expanded for 14 days REP (Enrichment of antigen-specific T cells). To identify the specific peptide recognized within the PP, sorted populations were co-cultured with B cells pulsed with individual peptides. The recognition was confirmed using HPLC purified peptides.

TCR sequencing and PBL transduction

The TCR locus was sequenced by multiplex single-cell RNA sequencing of enriched antigen-specific T-cell populations. The samples were multiplexed using TotalSeq™ barcodes. Sequencing was done on an Illumina NS6000 with an S1 flowcell and v1 chemistry. Mapping, quantification, and clonotype definitions were done using cell ranger multi software (version 6.1.1 using the reference vdj_GRCh38_alts_ensembl-5.0.0). Demultiplexing and subsequent analysis was done in R using the packages Seurat (version 4.0.3) and scRepertoire (version 1.3.5); Seurat::HTODemux was run using default parameters to obtain singlets.

TRA V-J-encoding sequences and TRB V-D-J-encoding sequences were combined to sequences encoding the mouse constant TRA and TRB chains (Cohen et al., 2006), respectively. Mouse constant regions were modified, as previously described (Cohen et al., 2007; Haga-Friedman et al., 2012). The full-length TRB and TRA chains were cloned, in this orientation and separated by a furin SGSG P2A linker, into pMSGV1 retroviral vector (GenScript). The use of the modified mouse TCR constant regions promotes pairing of the introduced TCR and also facilitates the identification of positively transduced human T cells by flow cytometry using an antibody specific for the mouse TCR-β chain.

To generate transient retroviral rich supernatants, the retroviral vector MSGV1 encoding the antigen-specific TCR and the envelope encoding plasmid RD114 were co-transfected into the retroviral packaging cell line 293GP (plated the day prior to transfection) using Lipofectamine 2000 (Life Technologies). Retroviral supernatants were collected at 42-48 h after transfection, diluted 1:1 with DMEM media (Gibco), and then centrifuged onto Retronectin-coated (10 μg/ml, Takara), non-tissue culture-treated 24-well plates at 2,000 g for 2 h at 32°C. Donor PBMC activated 2 days before with 50 ng/ml soluble OKT3 in T-cell media, were harvested, added onto the retrovirus plates, and centrifuged for 10 min at 300 g. Activated T cells were transduced overnight, removed from the plates, and further cultured in IL-2 containing T-cell media. Cells were typically assayed 10-14 days post-retroviral transduction.

CRISPR/Cas9 Knock out

Single guides RNA (sgRNA) targeting the predicted ORF for 5'U-HOXC13 were designed with CRISPOR tool http://crispor.tefor.net/. sgRNA specifically binding the genomic peptide location or upstream with the highest predicted KO efficiency were selected. Next, sgRNA were cloned into lenti-Cas9-v2 (Addgene, #52961) with BsmBI.

To generate lentiviral supernatants, the lentiviral vector lenti-Cas9-v2 encoding the sgRNA of interest, as well as packaging psPAX2 and envelope pMD2G plasmids, were co-transfected into HEK293 (plated the day prior to transfection) with PEI (Sigma) and non-supplemented DMEM media (Gibco). Media was replaced after 12h with Opti-MEM (Gibco) containing 2% FBS (Gibco) supplemented with Pen/Strep (Lonza). Lentiviral supernatants were collected at 42-48 h after transfection, and filtered through a 0.45 um low-protein binding filter (Millipore). Patient-derived TCL plated one day before were infected with filtered lentiviral supernatants containing polybrene (Sigma) final concentration of 8 ug/mL) followed by spinoculation 900 g, 32ºC, 50 min. Five days post-infection, T2 media (RPMI containing 10% FBS supplemented with Pen/Strep and L-Glut) was replaced with T2 media containing puromycin 1 µg/mL to select the cells efficiently transduced. Thereafter, media was replaced with fresh media containing puromycin when acidified or cells split when confluent. To evaluate the KO efficiency, puromycin-resistant cells were used as targets in co-culture assays with 5'U-HOXC13 specific T cells.

Healthy human cell lines

Healthy human cell lines were ordered from Promocell. Cells were thawed and cultured following manufacturer's instructions in the recommended media without antibiotics. Cells were split when confluent with Dettaching kit (Promocell). Cells were cultured at the recommended concentration and expanded no more than 4 passages until use. HCM-c (Cat: C-128810) were cultured in myocyte growth medium (Cat nº: C-39275). HREpC-c (CatC-12665) cultured in Renal Epithelial Cell GM media (Cat nº: C-39606). HSAEpC-c (Cat: C-12642) cultured in Small Airway Epithelial cell GM (Cat nº: C-39175). NHEM.f-c (Cat: C-12400) cultured in Melanocyte growth medium (Cat nº: C-39415).

Purification of HLA-I peptides

Purified anti-HLA-I clone W6/32 (ATCC® HB95) antibodies were cross-linked to protein-A Sepharose 4B conjugate beads (Invitrogen) with dimethyl pimelimidate dihydrochloride (Sigma-Aldrich) in 0.2 M Sodium Borate buffer pH 9 (Applichem). From 5e7 to 3e8 tumor cells were snap-frozen, thawed, and lysed with PBS containing 0.6% CHAPS (Applichem) and Protease inhibitor Cocktail Complete (Roche). The cell lysates were sonicated (Misonix 3000) and cleared by centrifugation for 1h at max speed to obtain the soluble fraction containing the pHLA complexes. The HLA-I affinity chromatography was performed using a 96-well single-use micro-plate with 3 µm glass fiber and 10 µm polypropylene membranes (Agilent). Sep-Pak tC18 100 mg Sorbent 96-well plates (Waters) were used for peptide purification and concentration as previously described (Chong et al., 2018). Peptides were eluted with 500 µl of 32,5% ACN in 0.1% TFA, lyophilized, and further cleaned and desalted with TopTips (PolyLC Inc.)

LC-MS/MS acquisition

Purified HLA-I peptides were diluted in 3% ACN, 1% FA. Samples were loaded onto a 300 µm × 5 cm Acclaim Pep-Map nanoViper, C18 (Thermo Scientific) at a flow rate of 15 µl/min using a Thermo Scientific Dionex Ultimate 3000 chromatographic system (Thermo Scientific). Peptides were separated using a C18 analytical column of 75 µm × 250 mm, 1.8 µm, 100Å (Waters) or 25 µm × 250 mm, 1.8 µm, 100Å (Waters). Orbitrap Fusion Lumos™ Tribrid (Thermo Scientific) mass spectrometer was operated in data-dependent acquisition (DDA) mode. Survey MS scans were acquired in the orbitrap with the resolution (defined at 200 m/z) set to 120,000. The top speed (most intense) ions per scan were fragmented in the linear ion trap (CID) and detected in the Orbitrap with the resolution set to 30,000. Quadrupole isolation was employed to selectively isolate peptides of 400-600 m/z. Included charged states were 2 and 3. Target ions already selected for MS/MS were dynamically excluded for 10 s.

Mass spectrometry data analysis of HLA-I peptides

To identify HLA-I ligands derived from conventional tumor antigens, ProteinScape3 was used to identify the amino acid sequences through Mascot search engine. The FDR was set at 5% and the Mascot ion score at >20. We created personalized databases for each patient including the reference proteome (Swissprot, downloaded 23/01/2019, containing 20416 entries), common contaminants in immunopeptidomics (CRAP and MaxQuant, 339 entries), and the personalized list of NSM identified by WES. Nt-Acetylation, Methionine oxidation and Serine, Threonine, Tyrosine phosphorylation were set as variable modifications. All the analyses showed were performed with R studio.

To identify HLA-I ligands derived from nonC proteins through proteogenomics, Peptide-PRISM was used as previously described (Erhard et al., 2020) without including random substitutions nor proteasome-spliced peptides. Briefly, for each identified fragment ion mass spectrum the Top10 candidates were first identified by de novo sequencing with PEAKS X and later aligned to a database containing a 3-frame transcriptome (Ensembl90) and 6-frame genome (hg38). Additionally, vcf files from whole-exome sequencing were used to interrogate non-synonymous somatic mutations (NSM) in a personalized fashion. All identified string matches were categorized into CDS (in-frame with annotated protein), 5'-UTR (contained in annotated mRNA, overlapping with 5'-UTR), Off-frame (off-frame contained in the coding sequence), 3'-UTR (all others that are contained in an mRNA), ncRNA (contained in annotated ncRNA), Intronic (intersecting any annotated intron) or Intergenic. Then, for each fragment ion mass spectrum, the category with the highest priority (CDS>5'UTR>Off-frame>3'UTR>ncRNA>Intronic>Intergenic) was identified, and all other hits among the 10 de novo candidates were discarded. The FDR was calculated for each category in a stratified mixture model taking into account the peptide length and database size. All the analyses were performed with R studio, the FDR was set at 1% and ALC at 30. The same pipeline was applied to immunopeptidomics data obtained from HLA ligand atlas (Marcu et al., 2021) including several tissues and HLA allotypes. The ORF from the nonC HLA-I ligands identified in the healthy immunopeptidome were retrieved and used to filter out the nonC HLA-I ligands from our tumor samples derived from the same ORF. The Peptide-PRISM pipeline and the filtering of the raw data was performed by Andreas Schlosser, University of Wurzburg. All the downstream analyses were performed at VHIO using with R studio, the FDR was set at 1% and ALC at 30.

HLA-I typing and prediction of binding to patient-specific HLA-I alleles

HLA typing was determined from the WES data using the PHLAT algorithm. Eluted ligand likelihood (ELL) percentile rank scores for binding to the patient's HLA molecules were obtained for all unique peptides ≥8 Aa eluted from TCL using NetMHCpan 4.0. The threshold for binding was set to <2%-tile rank.

Validation of HLA-I peptides with synthetic peptides

Spectrum validation of the experimentally eluted HLA-I ligand tumor antigen candidates was performed by computing the similarity of the spectra acquired in the sample with the corresponding non-labeled synthetic peptide from the library. Briefly, synthetic crude peptides obtained from JPT were acquired in a pool using LC-MS/MS in conditions similar to those previously used to analyze LC-MS/MS in conditions similar to those previously used to analyze samples to generate a spectral library. Peptide sequences were identified by database search with PEAKS-X Pro using a database containing Swiss-Prot as well as all the tumor antigen candidates interrogated. The search was exported as "for third party" format and imported into Skyline software to generate the library. The experimentally acquired HLA-I tumor samples were uploaded into Skyline and the similarity of the fragments (b/y ions) from the library (synthetic) vs. endogenous (sample) were analyzed considering library dot product (dotp) values, which range from 0 to 1 and , being dotp=1 the closest match.

Validation of HLA-I peptides with isotope-labeled peptides

For each selected peptide, a synthetic isotope-labeled peptide at one chosen amino acid was spiked into the samples and used as an internal standard for Parallel Reaction Monitoring (PRM) detection. The amount of internal standard peptide to be spiked in each sample was evaluated using dilution

curves and the final concentration was chosen based on a good chromatographic signal and no trace detectable of potential unlabeled traces from the synthetic internal standard. For Mel-3 TCL 30% of the sample (total of 5e7 cells) and for Mel-1 50% of the sample (total 1e8) was analyzed by PRM using different MS machines.

Orbitrap Eclipse (Thermo Fisher Scientific) coupled to an EASY-nanoLC 1000 UPLC system (Thermo Fisher Scientific) with a 50 cm C18 chromatographic column. The chromatographic gradient started at 95% buffer A and 5% buffer B with a flow rate of 300 nL/min and going up to 25% buffer B and 75% A in 52 min and to 40% B and 60% A in 8 min (Buffer A: 0.1% formic acid in water and Buffer B: 0.1% formic acid in acetonitrile). The Orbitrap Eclipse was operated in positive ionization mode with an EASY-Spray nanosource at 2.4 kV and at a source temperature of 305 ºC. A PRM method was used for data acquisition with a quadrupole isolation window set to 1.4 *m/z* and MS2 scans over a mass range of *m/z* 250-1800, with detection in the Orbitrap mass analyzer at a 240 K resolution. MS2 fragmentation was performed using HCD fragmentation at a normalized collision energy of 30%, the AGC was set at 100,000, and the maximum injection time at 502 ms. All data were acquired with XCalibur software.

For data analysis fragment ion chromatographic traces corresponding to the targeted precursor peptides were evaluated with Skyline software v.21.2. Verification of the endogenous peptides was based on: i) the number of detected traces, ii) co-elution of endogenous traces, iii) co-elution of endogenous and internal standard peptides, iv) correlation of the fragment ions relative intensities between endogenous and internal standard peptides and, v) expected retention time.

Data analysis and visualization

All flow cytometry data were analyzed with FlowJo software v7.6.5 and v10 (Tree Star). Mass spectrometry data were analyzed as specifically described with ProteinScape3, Peptide-PRISM, and Skyline software. Data was represented with GraphPad PRISM 6 software and R studio version 4.1.2.

Results

1. Identification of conventional tumor antigens

To identify the personalized repertoire of tumor antigens that elicit CD8⁺ T-cell responses in cancer patients we used the strategy depicted in Figure 16.

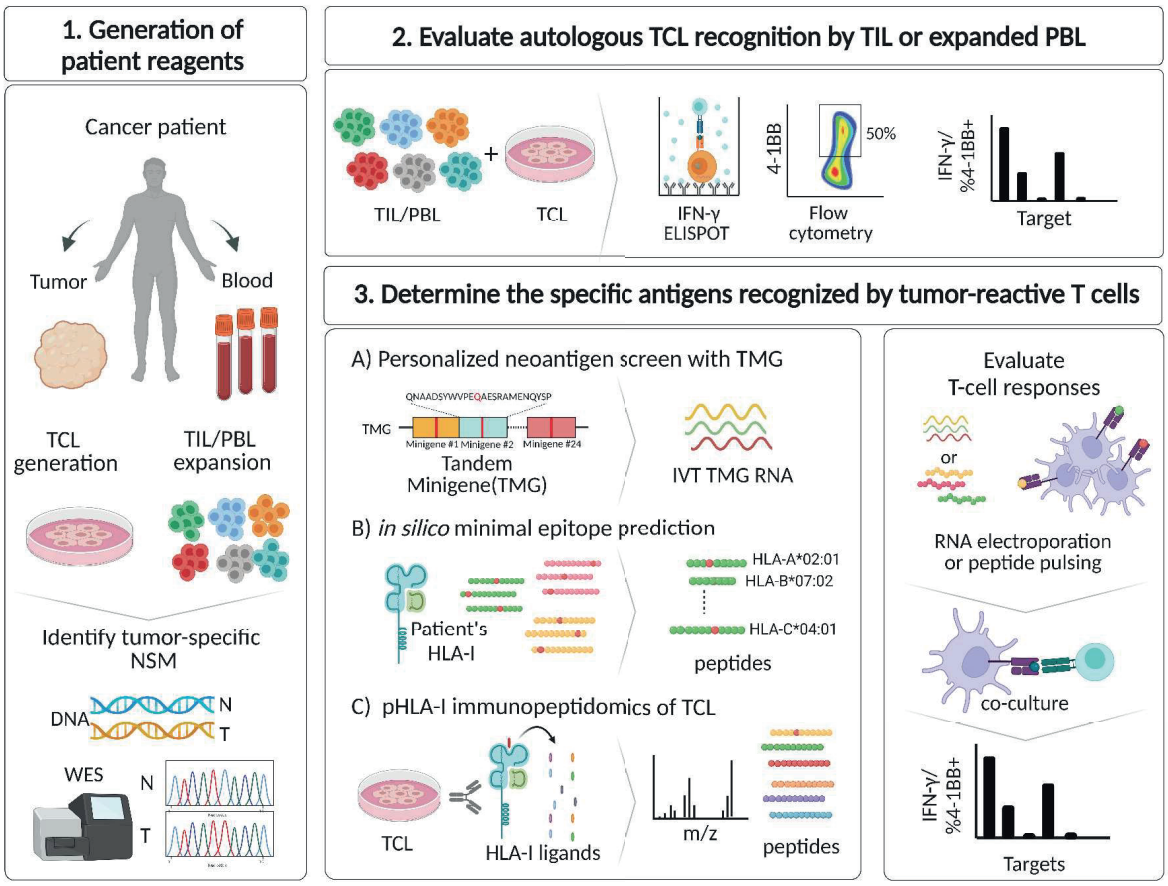


Figure 16. Strategy to identify the personalized tumor antigens recognized by T cells in cancer patients. (1) Tumor biopsies and blood samples were obtained from cancer patients. Lymphocyte populations were expanded from TIL or sorted from PBL and biopsies were cultured to generate autologous patient-derived TCL. WES was performed from matched tumor or TCL and healthy PBMC to identify tumor-specific NSM. (2) Lymphocyte populations were screened for tumor recognition after 20 h of co-culture with the autologous TCL measuring IFN-γ secretion by IFN-γ ELISPOT and upregulation of 4-1BB by FACS analysis. (3) Three distinct approaches were used to determine the specific antigens recognized by tumor-reactive T cells: (A) Personalized neoantigen screen using TMG constructs encoding the NSM identified. (B) Prediction of minimal epitopes from mutated peptides based on *in silico* prediction algorithms. (C) Immunopeptidomics of patient-derived TCL to identify the peptides naturally processed and presented on HLA-I molecules. Specific T-cell responses were evaluated by co-culturing tumor-reactive T cells with autologous APC either electroporated with RNA encoding TMG or pulsed with the synthetic peptides. T-cell activation was measured as mentioned above (right panel). Tumor-infiltrating lymphocytes (TIL), peripheral blood lymphocytes (PBL), tumor cell lines (TCL), Whole-exome sequencing (WES), non-synonymous somatic mutations (NSM), tandem minigenes (TMG), antigen-presenting cell (APC), in vitro transcription (IVT). Image created with Biorender.

First, we derived tumor cell lines (TCL) from tumor biopsies. In parallel, we obtained T cells from the same patient either through the ex vivo expansion of TIL or by sorting specific T cells subsets from peripheral blood (PBL) based on specific surface markers. WES from patient-derived TCL and matched PBMC was carried out to detect tumor-specific NSM (Figure 16.1).

Next, we evaluated autologous tumor recognition in 9 cancer patients (see Table 3) by co-culturing the patient-derived TCL with lymphocytes (TIL and/or PBL) and measuring T-cell responses through the detection of cytokine release (i.e. IFN-γ) and the upregulation of activation surface markers (i.e. 4-1BB) (Figure 16.2).

Last, we investigated the specific antigens recognized by tumor-reactive T cells in three patients using three different methods: A) personalized neoantigen screen with tandem minigenes (TMG), B) *in silico* minimal epitope prediction, and C) pHLA-I immunopeptidomics of TCL. Using the TMG approach, we sought to interrogate the immunogenicity of all the NSM identified by WES without any filter or prediction algorithm, since we constructed concatenated minigenes encoding for each NSM identified (Figure 16.3A). In parallel, we synthesized the top 20 mutated minimal epitopes (8-12 Aa) predicted to bind to the patient's HLA-I molecules according to NetMHCPan4.0 (Figure 16.3B). We also eluted the peptides naturally bound to HLA-I molecules from patient-derived TCL and analyzed the amino acid sequence by LC-MS/MS using personalized databases containing the NSM to identify the tumor antigen candidates (Figure 16.3C).

1.1. Generation of patient reagents

In order to study the personalized repertoire of tumor antigens, we first generated patient-derived TCL from tumor biopsies and expanded lymphocytes from either tumor or blood from cancer patients.

1.1.1. Generation of patient-derived TCL

To generate patient-derived TCL, we cultured a small fragment of tumor derived from a core biopsy or surgical resection in medium containing high concentrations of serum and regularly assessed tumor cell growth by monitoring colony formation.

Out of the 60 biopsies and 68 primary resections processed, we established three and four patient-derived TCL, respectively. These TCL derived from patients with four gynecological malignancies (Gyn), one melanoma (Mel), and two head and neck (H&N) cancer patients. In addition, to increase the number of patients studied, we included two additional patient-derived melanoma cell lines previously established by collaborators from the NCI (Table 3).

We performed WES from the TCL and matched PBMC to identify the tumor-specific NSM and to determine the HLA-I typing. As summarized in Table 5, the number of mutations varied widely among patients. While in Gyn-4 only 14 mutations were detected, Gyn-2 and Gyn-3 displayed an exceptionally high mutational load (7512 and 14765 NSM, respectively). These two TCL harbored mutations in the POLE gene which encodes the catalytic subunit of the DNA polymerase epsilon and is associated with hypermutated cases of endometrial cancer and other malignancies (Mur et al., 2020). Furthermore, the measurement of HLA-I expression by flow cytometry revealed that the expression varied greatly among TCL. Nevertheless, all patient-derived TCL expressed HLA-I on the cell surface which was requisite to further develop this project.

1.1.2. Lymphocyte expansion

Following standard protocols, TIL were expanded from several small tumor fragments and cultured independently in high doses of IL-2 to maintain the tumor heterogeneity. As summarized in Table 6, following this strategy we successfully expanded TIL from at least 3 tumor fragments in all patients studied, with a total of 65 independently expanded TIL populations (Table 3).

For Mel-1 and Mel-2, since only one or two TIL populations were available (a generous gift from our collaborators from the NCI), we took advantage of PBL as an alternative source of tumor-reactive T cells. We sorted the T cells by flow cytometry-based sorting based on the expression of PD-1, a co-inhibitory marker that has been used to identify tumor-reactive lymphocytes from TIL (Gros et al., 2014) and peripheral blood (Gros et al., 2016, 2019), either alone or in combination with other markers that are currently being investigated in our laboratory such as CD39, CD38 or HLA-DR. Figure 17 shows the gating strategy used to sort PBL for Mel-1 as an example. In this case, $CD8^+$ T cells were sorted based on different expression levels of PD-1 (negative, dim, high), either alone or in combination with CD39, a marker that has been described to be preferentially expressed by tumor-reactive T cells in tumors (Duhen et al., 2018). In total, we sorted 12 $CD8^+$ populations for Mel-1 and 16 $CD8^+$ populations for Mel-2 from peripheral blood (Table 5).

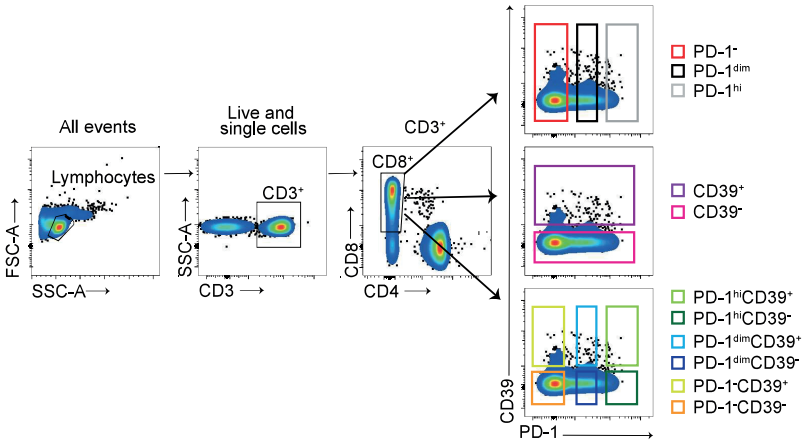


Figure 17. Flow cytometry-based sorting of Mel-1 PBL to identify circulating tumor-reactive T cells. Representative FACS plots displaying the gating strategy used for sorting $CD8^+$ T-cell populations from peripheral blood of Mel-1 based on the expression of PD-1 (negative, dim, high), CD39 (+/-) alone or the combined expression of PD-1 and CD39. Following cell separation, cells were expanded for 14 days through a rapid expansion protocol (REP). The gates used for sorting lymphocytes with either unique or combined expression of PD-1 and CD39 are displayed in different colors.

Table 3. Autologous TCL and matched lymphocytes generated for each patient

Patient ID	Tumor type	[a]HLA typing	Autologous TCL			Autologous lymphocytes	
			[f]Tumor-specific NSM	[g]HLA expression	[h]Relevant mutated genes	TILs expanded/ cultured	PBLs sorted
Gyn-1	Endometrial cancer, CNL	A*03:01 B*47:01 B*51:01 C*02:02 C*06:02	110	HLA-I ++	PPP2R1A	9/12	--
Gyn-2	Endometrial cancer, POLE	A*02:01 B*51:01 C*02:02 C*15:02	7512	HLA-I +	POLE, PTEN, PIK3CA, CTNNB1, ARID1A, FGFR2, B2M	11/12	--
Gyn-3	Endometrial cancer, POLE	A*02:01[b] B*07:02[c] B*44:02[c] C*05:01 C*07:02	14765	HLA-I ++	POLE, PTEN, PIK3CA, CTNNB1, ARID1A, FGFR2	9/12	--
Gyn-4	Ovary small cell carcinoma	A*01:01 A*02:01 B*27:05 B*51:01 C*01:02 C*15:02	14	HLA-I ++	SMARCA4	3/6	--
H&N-1	Hypopharyngeal carcinoma	A*02:01[d] A*03:01[b] B*07:02[b] B*18:01[e] C*05:01 C*07:02	2484	HLA-I +	TP53, PIK3CA	9/12	--
H&N-2	Oropharyngeal squamous cell carcinoma	A*24:02 A*29:02 B*14:02 B*44:03 C*02:02 C*16:01	266	HLA-I +++	TP53, TP63	9/18	--
***Mel-1**	Melanoma	A*02:01 A*11:01[b] B*35:01 B*51:01 C*04:01 C*14:02	2951	HLA-I ++	CDKN2A	2*	12
***Mel-2**	Melanoma	A*01:01 A*30:02 B*15:01 B*18:01 C*03:03 C*05:01	255	HLA-I +++	BRAF	1*	16
Mel-3	Melanoma	A*11:01 A*29:02 B*35:01 B*35:02 C*04:01	283	HLA-I +++	CDKN2A	12/12	--

[a]HLA-I typing was evaluated for TCL and matched PBMC from WES data. Specific mutations found in TCL are noted. [b]Missense mutations. [c]Missense in one of the HLA-B alleles*. [d]Loss of heterozigosy. [e]Nonsense mutations. f Non-synonymous somatic mutations (NSM) identified by WES. [g]HLA-I expression was evaluated by flow cytometry using W6/32 antibody (PanHLA-I). [h]Relevant mutated genes depending on the tumor type.*Gift from our collaborators from the NCI.

1.2. Identification of tumor-reactive lymphocytes

Next, we sought to identify tumor-reactive lymphocytes in the 9 cancer patients for which both patient-derived tumor cell lines (TCL) and matched lymphocytes were generated, as explained in the previous section. We first assessed whether the independently expanded TIL or sorted PBL could recognize their corresponding autologous TCL. In some instances, where tumor-reactive cells were detected at a low frequency, the lymphocyte populations displaying tumor recognition were enriched by flow cytometry-based sorting.

1.2.1. Detection of tumor-reactive TIL or PBL by co-culture

To identify tumor-reactive lymphocytes, patient-derived TCL were co-cultured with autologous T-cell populations and tumor recognition was evaluated after 20 h by measuring IFN-γ secretion by ELISPOT and the upregulation of the activation marker 4-1BB on $CD8^+$ T cell by FACS. IFN-γ ELISPOT is a highly sensitive technique capable of detecting very infrequent T-cell responses, however it does not determine whether the T cells producing the cytokine are $CD8^+$ or $CD4^+$. Although FACS analysis is less sensitive, it allows to determine the T-cell subset activated regardless of the specific cytokine produced. Hence, by combining these two techniques we can study tumor-reactive $CD8^+$ T cells with higher accuracy.

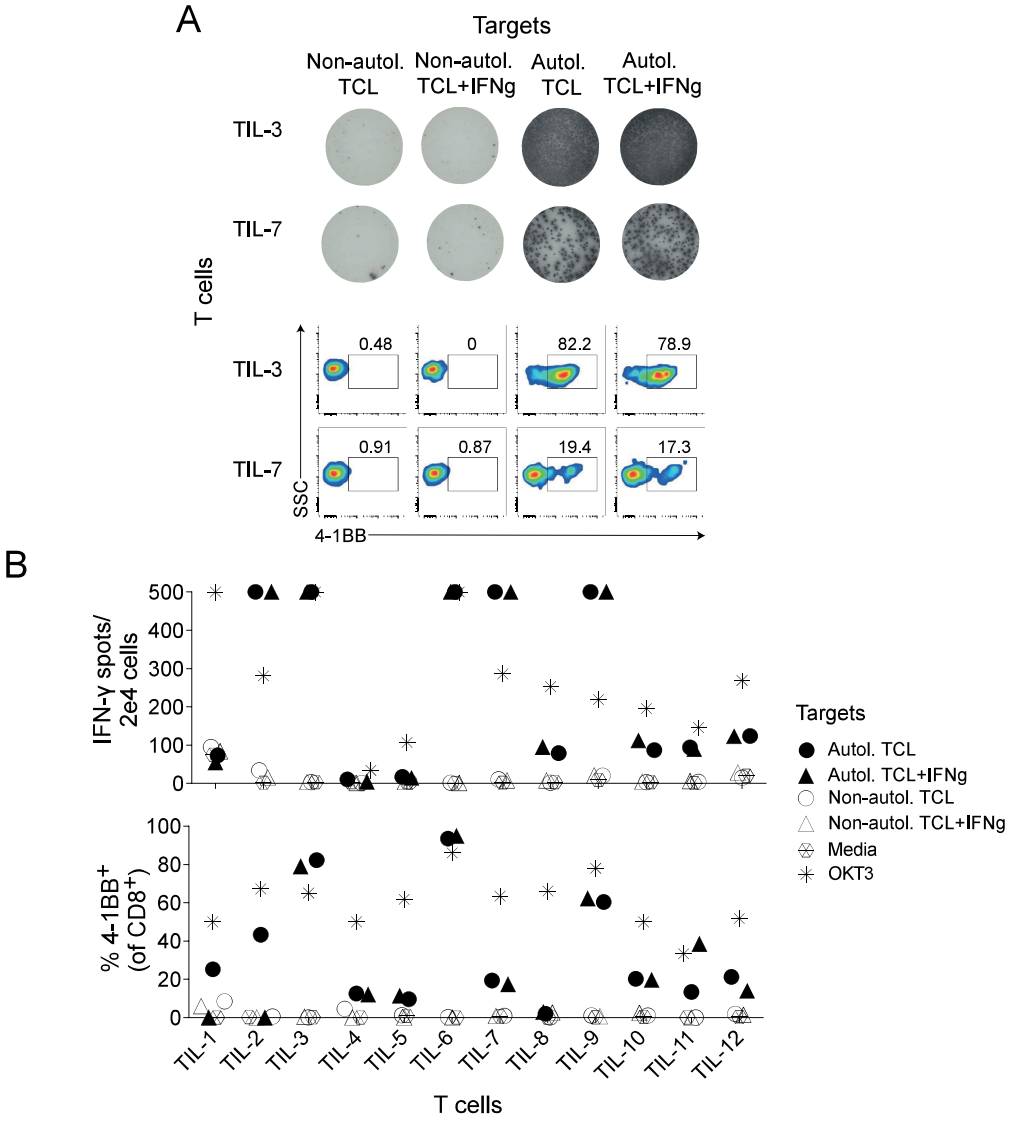


Figure 18. Autologous TCL recognition by Mel-3 TIL. The autologous Mel-3 (Autol.) and an irrelevant non-autologous (Non-autol., HLA-I unmatched) patient-derived TCL were co-cultured independently with 12 different populations of *ex vivo* expanded Mel-3 TIL. After 20 h tumor recognition was evaluated by measuring the number of IFN-γ spots using an ELISPOT assay and the upregulation of 4-1BB by flow-cytometry. Prior to the co-culture, TCL were either left untreated or pretreated with 10 ng/mL of IFN-γ. After ~24h TCL were harvested, washed, and 4e4 cells were seeded with 2e4 T cells. (A) Representative IFN-γ ELISPOT (top) and flow cytometry plots (bottom) for TIL-3 and TIL-7 derived from Mel-3 following co-culture with the targets specified. (B) Tumor reactivity of Mel-3 TIL. The number of IFN-γ spots per well (top panel) and the percentage of $CD8^+$ T cells expressing 4-1BB (bottom panel) are shown. OKT3 and media were used as positive and negative controls, respectively. Plotted cells were gated on live $CD3^+CD8^+$ lymphocytes. '>' denotes greater than 500 spots/2e4 cells. Experiments were performed twice.

Figure 18 shows the results of the TIL expanded for Mel-3 as a representative example of the 9 patients included in the study. As measured by IFN-γ ELISPOT, 11 out of the 12 *ex vivo* expanded TIL recognized the autologous TCL, but the frequency of tumor-reactive T cells varied from very low (<1% in the case of TIL-5, confirmed in section 1.2.2) to very high (80% for TIL-6), as shown by 4-1BB upregulation on $CD8^+$ cells. IFN-γ pre-treatment did not markedly influence TCL recognition by TIL for this particular patient, but it greatly enhanced the recognition for others (not shown). This could be attributed to an increase in HLA-I expression following incubation with IFN-γ, which was more evident for those TCL expressing low basal

levels of HLA-I. Alternatively, it could also be explained by changes in the pHLA-I repertoire due to reshaping of the antigen processing machinery upon IFN-γ stimuli, such as the induction of the immunoproteasome, which can occur *in vivo* in an inflammatory context.

For patients Mel-1 and Mel-2, we also sorted PBL based on the expression of specific surface markers as explained in section 1.1.2 (Figure 17). Indeed, co-culture experiments with the autologous TCL showed that specific circulating T-cell subsets also recognized the autologous tumor (Figure 19). More specifically, the population of circulating $CD8^+$ T cells expressing high or intermediate levels of PD-1 from Mel-1 patient were enriched in tumor recognition, as previously described (Gros et al., 2016). The expression of CD39 also identified circulating tumor-reactive T cells as shown by IFN-γ ELISPOT, but PD^{hi} alone or the combination of both markers ($PD\text{-}1^{hi/\,dim}$ and $CD39^+$) substantially increased the frequency of tumor-reactive T cells in Mel-1, as shown by 4-1BB upregulation. In total, 6 of the 12 sorted PBL populations were capable of recognizing the autologous TCL (Figure 19).

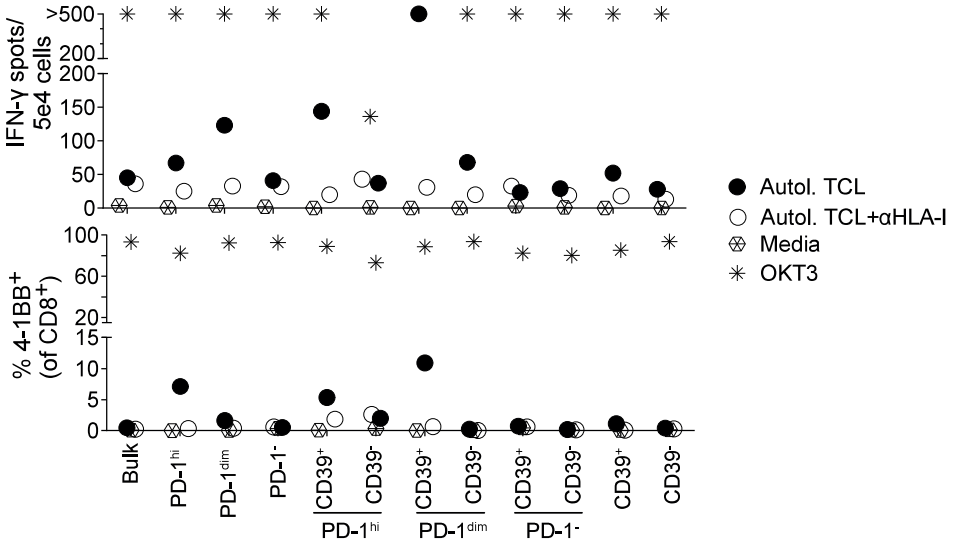


Figure 19. Tumor reactivity of Mel-1 ex vivo expanded PBL populations sorted based on PD-1 and/or CD39 expression. TCL recognition was evaluated by co-culturing the different sorted and expanded populations (described in Figure 17) with the autologous TCL. After 20 h, T-cell activation was measured by IFN-γ ELISPOT (top) and by measuring 4-1BB upregulation (bottom). OKT3 and media were used as positive and negative controls, respectively. In addition, the autologous TCL was treated with an HLA-I blocking antibody to evaluate whether TCL recognition was HLA-I -restricted. '>' denotes greater than 500 spots/2e4 cells

We studied T-cell responses against autologous TCL using the same strategy for all nine patients and detected tumor-reactive $CD8^+$ TIL in all of them regardless of the tumor type. For Mel-2, circulating tumor-reactive T cells could also be detected (Table 4). Altogether, these results indicate that tumor-reactive T cells are frequently detected in cancer patients, not only at the tumor site, but also circulating in peripheral blood.

1.2.2. Enrichment of tumor-reactive T cells

To enrich for infrequent tumor-reactive T-cell populations, we sorted CD8⁺ T cells expressing 4-1BB following TCL recognition by flow cytometry-based sorting and expanded them using a rapid expansion protocol (REP). Thereafter, the cells obtained generally exhibited increased recognition of the autologous TCL, as shown for TIL-5 and TIL-7 from Mel-3, as representative examples (Figure 20). In total, we successfully enriched tumor-reactive lymphocytes for 20 populations, as summarized in Table 4.

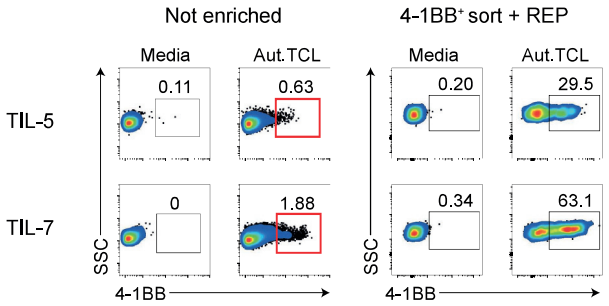


Figure 20. Enrichment of tumor-reactive T cells through selection of 4-1BB+ T cells by flow cytometry. Tumor-reactive CD8⁺ T cells detected at a low frequency within TIL-5 (top) and TIL-7 (bottom) from Mel-3 were enriched by flow cytometry sorting based on 4-1BB expression after a 20 h co-culture with the autologous TCL (Aut. TCL) and expanded for 14 days through a REP. Plots show 4-1BB expression by CD3⁺CD8⁺ T cells immediately prior to cell sorting and 4-1BB expression after selection and expansion of the sorted T cells through a REP (right). Gates used for sorting are depicted in red. Media was used as a negative control of T-cell activation.

We used this strategy to prevent loss of tumor recognition that can sometimes occurs after long-term culture of T cells. This is particularly useful when cells are expanded through a rapid expansion protocol with non-specific stimulation since non-reactive T cells frequently expand more efficiently. Nonetheless, it is also important to mention that at the same time, the TCR diversity is reduced and although it is presumed that all the T cells recognizing the tumor are selected, we cannot exclude that some tumor-reactive clones are lost during this selection and expansion.

Table 4. Tumor-reactive lymphocytes identified in cancer patients

	Tumor-reactive CD8⁺ lymphocytes		
Patient ID	**TIL reactive/ expanded**	**PBL reactive/ sorted**	**Enriched by FACS sorting**
Gyn-1	1/9	---	---
Gyn-2	10/11	---	---
Gyn-3	5/9	---	---
Gyn-4	2/3	---	---
H&N-1	6/9	---	4
H&N-2	5/9	---	4
Mel-1	2/2	6/12	2
Mel-2	1/1	6/16	2
Mel-3	11/12	---	8

Recognition of the autologous TCL was evaluated as explained in Figure 3 (TIL) and Figure 4 (PBL). Tumor enriched populations were sorted as explained in Figure 5.

1.3. Personalized neoantigen screening using TMG and *in silico* prediction

Because we were able to find T cells recognizing the autologous TCL, we next wanted to identify the specific antigens recognized by these T cells. We selected the first three patients for which we established tumor cell lines and detected autologous tumor-reactive TIL (H&N-1, H&N-2 and Mel-3) to perform a detailed characterization of the personalized tumor antigens targeted by T cells. Since neoantigens have been described as frequent drivers of T-cell responses in cancer patients (Hanada et al., 2022; Kristensen et al., 2022; Parkhurst et al., 2019; Stevanović et al., 2017; Zacharakis et al., 2022), we first investigated whether neoantigens derived from NSM could be recognized.

1.3.1. Generation of TMG and selection of candidate neoantigens by *in silico* prediction

We first used WES information from TCL and matched normal DNA to identify tumor-specific NSM in H&N-1, H&N-2, and Mel-3 patients. While patient H&N-1 exhibited a high number of mutations (n=2484 NSM), patients H&N-2 and Mel-3 carried a moderate mutational load (n=266 and 283, respectively) (Figure 21A). We also evaluated *in silico* the number of potential neoantigens using the NetMHCPan4.0, one of the publicly available algorithms which can be used to predict the minimal mutated epitopes presented on the patient's HLA-I molecules (Figure 21B). As a result, the number of candidate neoantigens increased from approximately 300 to about 1.200 candidates for the patients with a moderate mutational load, whereas it exceeded 16.000 candidates for H&N-1 the prediction exceeded 16.000 candidates.

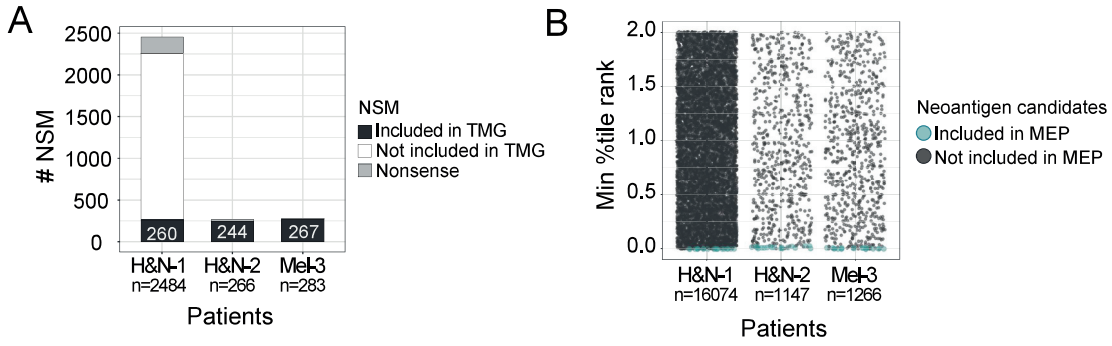


Figure 21. Tumor-specific NSM identified by WES and candidate neoantigens according to *in silico* prediction in patients H&N-1, H&N-2 and Mel-3.(A) Number of tumor-specific NSM identified by WES from the TCL for each patient. The NSM potentially giving rise to a mutated epitope and selected for TMG design and construction are depicted in dark grey and the ones not included are shown in white. NSM that do not generate mutated epitopes (nonsense mutations) are depicted in light grey. The number of NSM included in the TMG, and the total number NSM are indicated for each patient. (B) Minimal mutated epitopes from 8 to 12 Aa predicted to bind to the patient's HLA-I molecules according to NetMHCPan4.0. Y-axis shows the minimal percentile rank (%tile rank) for each peptide. Peptides were considered binders if %tile rank≤2. Each dot represents one peptide and the total number of predicted binders, per patient, is noted. The top 20 best binders for each patient, highlighted in turquoise, were selected and synthesized to test their immunogenicity and were included in a minimal epitope pool (MEP).

In order to screen for neoantigen recognition, we designed and generated 12 TMG containing all or a fraction of the NSM identified by WES for each patient to interrogate the tumor-specific mutations (Figure 16.3A). Each TMG was constructed by concatenating up to 24 minigenes, each encoding the mutated amino acid flanked by 12 amino acids of the wild-type

protein on each side. For patients H&N-2 and Mel-3 all missense mutations were included. In contrast, due to the exceptionally high number of NSM in patient H&N-1, we were forced to limit the candidates to 260 by selecting the mutations with a better binding prediction to the patient's HLA-I (Figure 21A).

The TMG constructed were used as templates to generate *in vitro* transcribed RNA, which were then electroporated into autologous APC. By expressing the minigenes encoding for all or a large fraction of NSM, we enable the APC to express, process and present the neoantigens in HLA-I and, potentially, HLA-II. The main advantage of the TMG strategy is that it bypasses *in silico* prediction, which can frequently miss immunogenic candidates. To further test this, we also selected and synthesized the top 20 predicted mutated minimal epitopes for each patient (Figures 16.3B, and 21B).

1.3.2. T-cell responses to neoantigens were detected for H&N-1, H&N-2 and Mel-3

Next, we tested whether the selected candidate neoantigens could elicit T-cell responses by co-culturing TIL with autologous APC either electroporated with RNA encoding the TMG or pulsed with the predicted mutated epitopes included in one pool (minimal epitope pool or MEP). After 20 h, T-cell activation was measured by IFN-γ release and 4-1BB upregulation on $CD8^+$ T cells (the latter is not shown).

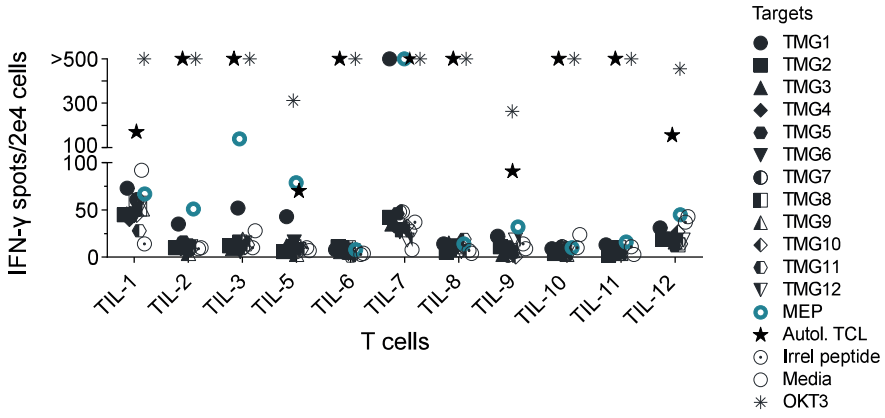


Figure 22. TIL derived from Mel-3 react to neoantigens encoded by TMG1 and contained in the minimal epitope pool. Recognition of neoantigens was evaluated by co-incubating 2e4 lymphocytes with 4e5 autologous APC electroporated with TMG1-12 RNA encoding for all the NSM identified by WES, or 2e5 APC pulsed with 1 μg/mL of the minimal epitope pool (MEP) containing the top 20 best predicted binders. Y-axis shows the number of IFN-γ spots per well measured by ELISPOT for each TIL co-cultured independently with the targets specified. Autologous *ex vivo* cultured B cells were used as APC. T cells cultured with media or B cells pulsed with an irrelevant peptide were used as negative controls. OKT3 and autologous TCL stimulated T cells were used as positive controls.

For patient Mel-3, we detected reactivity against TMG1 and the MEP in TIL-2, TIL-3, TIL-5, and TIL-7, as shown by IFN-γ ELISPOT (Figure 22). To determine the specific mutation recognized, long peptides encoding the individual mutations contained in TMG1 (Figure 23A) or the minimal epitopes contained in the MEP (Figure 23B) were independently tested. Interestingly, TIL-3 and TIL5 both recognized the same NSM derived from $ETV1_{p.E455K}$ which was contained in both TMG1 (p9) and the MEP (p13). Moreover, TIL7 recognized $GEMIN5_{p.S1360L}$,

which was also present in both TMG1 (p5) and the MEP (p14) (Figure 23). Hence, for patient Mel-3, both strategies led to the detection of the same neoantigens, as the minimal mutated epitopes recognized by tumor-reactive TIL were included in the top 20 best HLA-I binders according to *in silico* prediction.

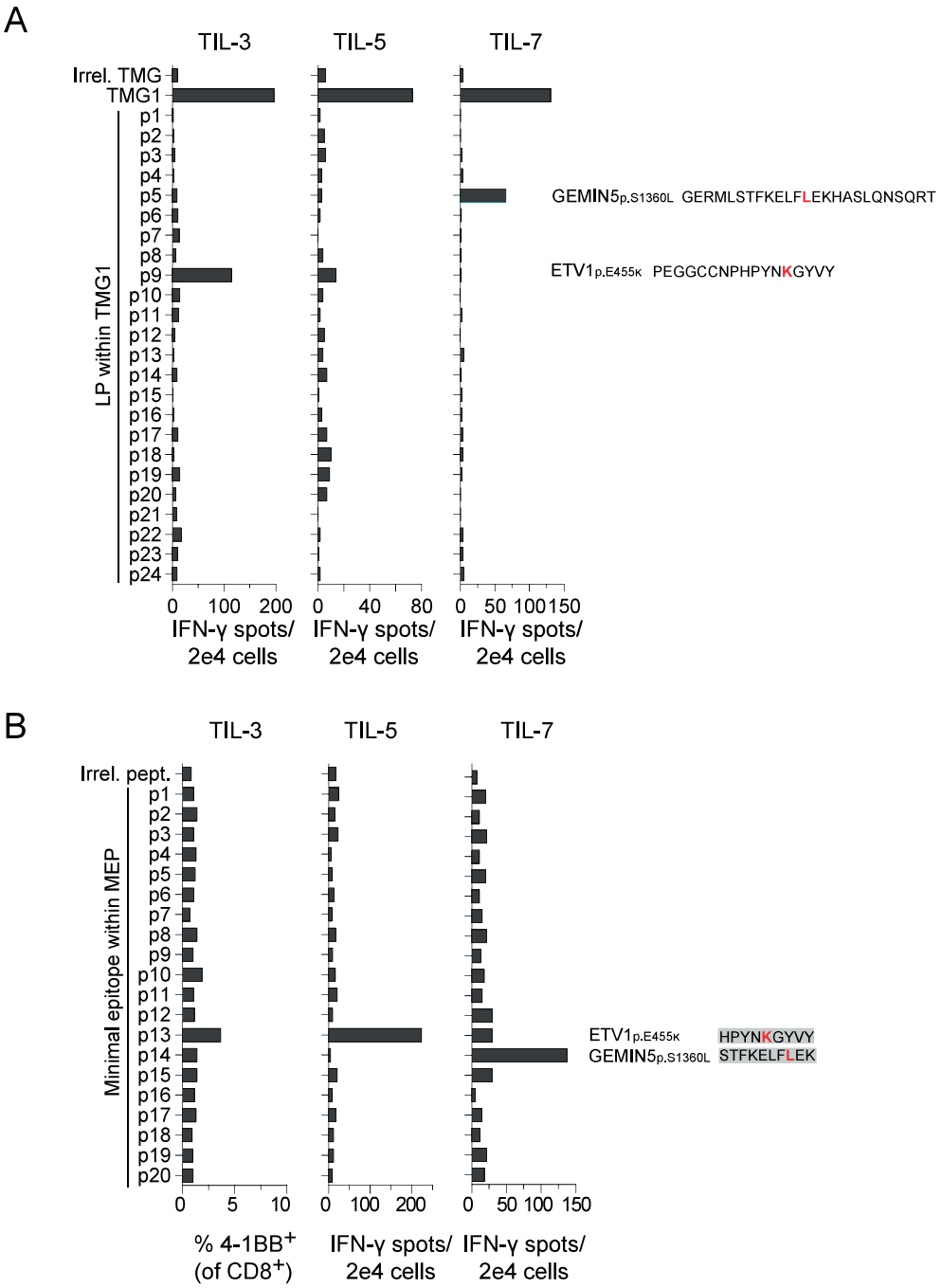


Figure 23. Mel-3 TIL recognize ETV1$_{p.E455K}$ and GEMIN5$_{p.S1360L}$ neoantigens identified by tumor WES. To determine the specific mutation recognized, (A) long peptides (LP) encoding the individual NSM contained in TMG1 or (B) the minimal epitopes contained in the MEP were independently pulsed onto autologous APC at 5 µg/mL O/N or 1 µg/mL for 2 h, respectively. APC were then washed and co-cultured with REPed TIL-3, TIL-5, and TIL-7. After 20 h the reactivity was measured by IFN-γ ELISPOT assay or upregulation of 4-1BB by flow cytometry as indicated in the x-axis. The specific mutated protein, Aa change, and sequence of the reactive peptides are noted on the right. The mutated Aa is highlighted in red and bold and the recognized minimal epitope is depicted in grey. Autologous B cells either electroporated or pulsed with an irrelevant TMG or peptide were used as negative controls.

To further confirm and better characterize T-cell responses detected at a low frequency, neoantigen-specific T cells were enriched by flow cytometry based-sorting (Figure 24A, left) and expanded using a REP (Figure 24A, right). As shown by peptide titration experiments, the different neoantigen-specific T-cell populations exhibited variable functional avidity to the specific neoantigen recognized, but were capable of recognizing down to 0.1 ng/mL of peptide (approximately equivalent to 0.1 nM). In addition, they also displayed different levels of recognition of the wild-type (wt) counterparts (Figure 24B). $ETV1_{p.E455K}$-specific T cells were markedly mutation-specific, while $GEMIN5_{p.S1360L}$-specific T cells also recognized the wt peptide to some extent. Nonetheless, considerably lower concentrations of the mutated minimal epitope were required to activate T cells compared to the wt counterpart (Figure 24B). Moreover, the recognition of $ETV1_{p.E455K}$ and $GEMIN5_{p.S1360L}$ by T cells was restricted to HLA-B*35:01 and HLA-A*11:01, respectively (Figure 24C). Importantly, co-culture experiments showed that the enriched populations of neoantigen-specific T cells were able to recognize the autologous TCL, ultimately demonstrating that these peptides are genuine tumor antigens (Figure 24A).

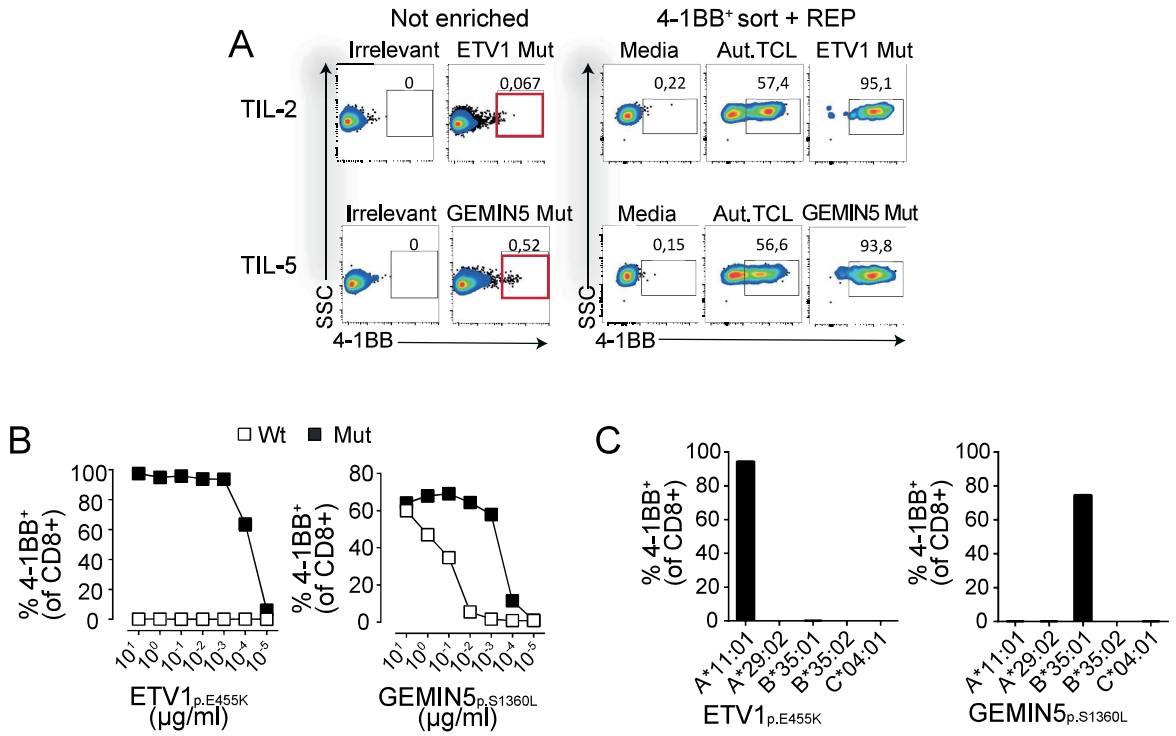


Figure 24. Enrichment and characterization of Mel-3 $ETV1_{p.E455K}$ and $GEMIN5_{p.S1360L}$ specific T cells. (A) T cell populations containing a limited frequency of reactivity to the mutated HLA-I peptides indicated were enriched by flow cytometry-based sorting of $CD8^+4\text{-}1BB^+$ T cells after co-culture with B cells pulsed with the peptides followed by *ex vivo* expansion for 14 days. Plots show the reactivity of the T cell populations to the peptides specified or to the autologous (Aut.) TCL by measuring 4-1BB expression on $CD3^+CD8^+$ T cells immediately prior to cell sorting (left) and after selection and expansion of the sorted T cells (right). Gates used for cell sorting are depicted in red. Media or an irrelevant peptide were used as negative controls of T-cell activation. (B) B cells were pulsed with serial dilutions of the wild-type (Wt) or mutant (Mut) $ETV1_{p.E455K}$ and $GEMIN5_{p.S1360L}$ peptides and co-cultured with the corresponding neoantigen-enriched T-cell population. T-cell reactivity was evaluated by measuring upregulation of 4-1BB expression after 20 h. (C) COS7 cells were co-transfected with the indicated individual HLA-I alleles and pulsed with the corresponding peptides to determine the restriction element. After 20 h of co-culture with neoantigen-enriched T cells, T cell activation was measured by IFN-γ ELISPOT (left axis) and upregulation of 4-1BB by flow cytometry (right axis).

The same strategy was used for patients H&N-1 and H&N-2, leading to the detection of 2 neoantigens recognized by TIL in H&N-1 and 2 neoantigens in H&N-2, as summarized in Table 5 . However, as opposed to Mel-3, in these two patients, the top 20 neoantigens selected through *in silico* prediction did not include the neoantigens identified using the TMG screening strategy nor any other immunogenic NSM.

In summary, we identified neoantigens recognized by pre-existing tumor-reactive T cells in all three patients studied. However, while the personalized high-throughput neoantigen screen with TMG, including a larger list of candidate neoantigens, enabled the detection of immunogenic mutations in all of them, *in silico* prediction only detected neoantigens in patient Mel-3, but failed for H&N-1 and H&N-2.

Table 5. Neoantigens identified through TMG and *in silico* minimal epitope prediction

Patient ID	[a]NSM	Strategy	[b]Immunogenic NSM/tested	TMG (TIL reactive)	Immunogenic NSM	[c]Immunogenic sequence	Restriction element	Wt recognition
H&N-1	2485	TMG	2/266	TMG4 (TIL3,TIL4) TMG3 (TIL5)	MAGEB2p.E167Q RPL14p.H20Y	KASEGLSVVFGL**Q**LNKVNPNGHTYT EVGRVAYVSFGP**Y**AGKLVAIVDVID	HLA-A*03:01 HLA-B*07:02	No No
		Minimal epitope prediction	0/20	--	--	--	--	--
H&N-2	261	TMG	2/261	TMG2 (TIL9) TMG8 (TIL10)	TP53RKp.S108C GBP5p.E359K	ISAPVVFFVDYA**C**NCLYMEEIEGSV TLQELLDLHRTS**K**REAIEVFMKNSF	HLA-A*29:02 HLA-B*14:02	n.d. No
		Minimal epitope prediction	0/20	--	--	--	--	--
Mel-3	283	TMG	2/283	TMG1 (TIL3,TIL5,TIL7)	GEMIN5p.S1360L ETV1p.E455K	GERMLSTFKELF**L**EKHASLQNSQRT PEGGCCNPHPYN**K**GYVY	HLA-A*11:01 HLA-B*35:01	Yes No
		Minimal epitope prediction	0/20	MEP (TIL3,TIL5,TIL7)	GEMIN5p.S1360L ETV1p.E455K	STFKELF**L**EK HPYN**K**GYVY	""	""

[a]Tumor specific non synonymous mutations (NSM) identified by WES. [b]Recognition of the NSM was evaluated as explained in Figures 6 and 7. [c]The mutated amino acids are shown in red bold letters

1.4. Identification of tumor antigens through HLA-I immunopeptidomics

To further investigate the specific antigens recognized by tumor-reactive T cells in the same three patients (H&N-1, H&N-2, and Mel-3) we sought to explore in greater detail the peptides presented by the HLA-I molecules on tumor cells through immunopeptidomics. In addition to neoantigens, we investigated whether other types of antigens such as cancer-germline antigens and melanoma-associated antigens were presented and recognized by tumor-reactive T cells.

1.4.1. Elution of peptides from HLA-I molecules

To identify candidate tumor antigens naturally presented on tumor cells, the peptides bound to HLA-I molecules (HLA-I ligands) were eluted and the amino-acid sequences were identified by LC-MS/MS.

Briefly, the TCL were first expanded up to high numbers depending on the expression of HLA-I (Figure 25A), harvested, and lysed. Next, the pHLA-I complexes were isolated by affinity chromatography, and the peptides were further purified and analyzed by LC/MS-MS. Thereafter, the amino acid sequences corresponding to each acquired mass-spectrum (MS) were inferred by database search. Importantly, to enable the identification of patient-specific neoantigens, we generated personalized databases including both the reference proteome and the NSM identified by WES (Figure 16.3C).

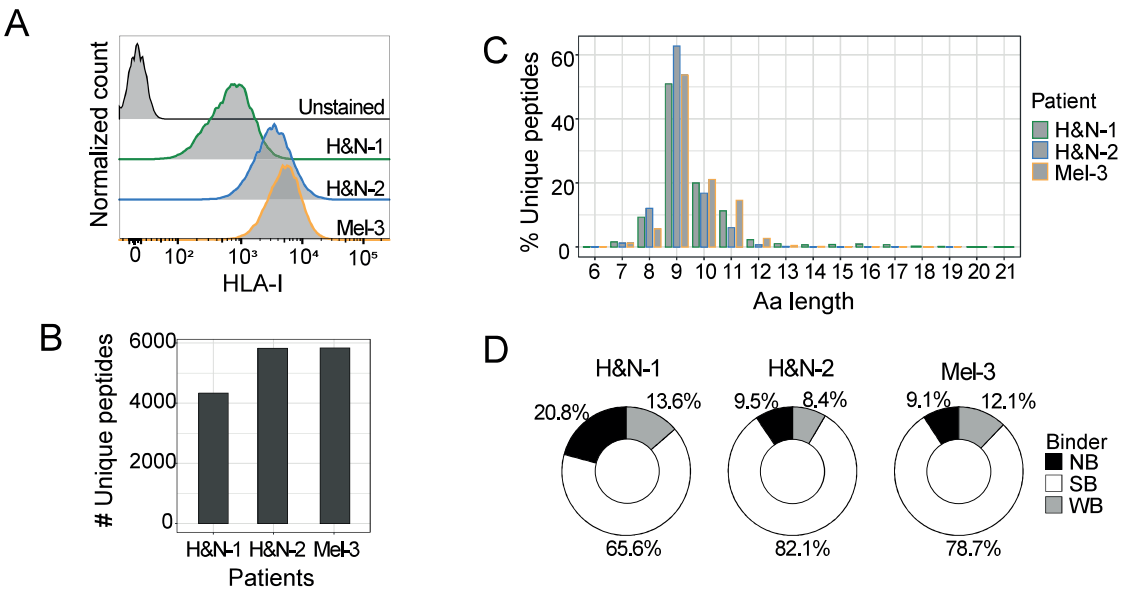


Figure 25. Peptides eluted from patient-derived TCL display typical HLA-I peptide features. (A) HLA-I expression by the TCL derived from the patients measured by flow cytometry using a PanHLA-I reactive antibody. Unstained cells were used as negative controls. Only H&N-1 TCL unstained control is shown, but all TCL exhibited similar autofluorescence. (B) Number of unique peptides eluted from HLA-I and identified by database search for each patient at 5% FDR using ProteinScape3. Only unique peptide sequences were considered. (C) Peptide length distribution of the total unique HLA-I peptide sequences identified for each patient. (D) Percentage of peptides predicted to bind to corresponding patient HLA-I molecules according to NetMHCPan4.0. Peptides are classified as strong binders (SB, %tile rank <0.5), weak binders (WB, %tile rank 0.5-2) and non-binders (NB, %tile rank >2). For each peptide, the minimal predicted %tile rank across all the patient's HLA molecules was considered.

Using this strategy, we identified from 4.000 to 6.000 unique peptide sequences for each TCL analyzed (Figure 25B). These peptides, displayed typical features of HLA-I ligands, as shown by the peptide length distribution (mainly 8-12 Aa, Figure 25C) and the great percentage of peptides predicted to bind to patient-specific HLA-I molecules (Figure 25D).

In addition, we carried out an unbiased clustering analysis of the peptides identified for each patient using GibsCluster2.0. This program aligns the peptides of interest and creates groups based on amino acid usage without any additional information other than the Aa sequence. We found that these peptides clustered in groups containing sequence motifs consistent with the conserved amino acids at anchor positions (p2 and p9). Of note, those groups displayed a remarkable resemblance to the consensus motifs of the patient's HLA alleles defined by NetMHCPan4.0. (Figure 26).

Altogether these analyses suggested that most identified sequences corresponded to peptides eluted from HLA-I rather than derived from contaminants or due to misidentifications.

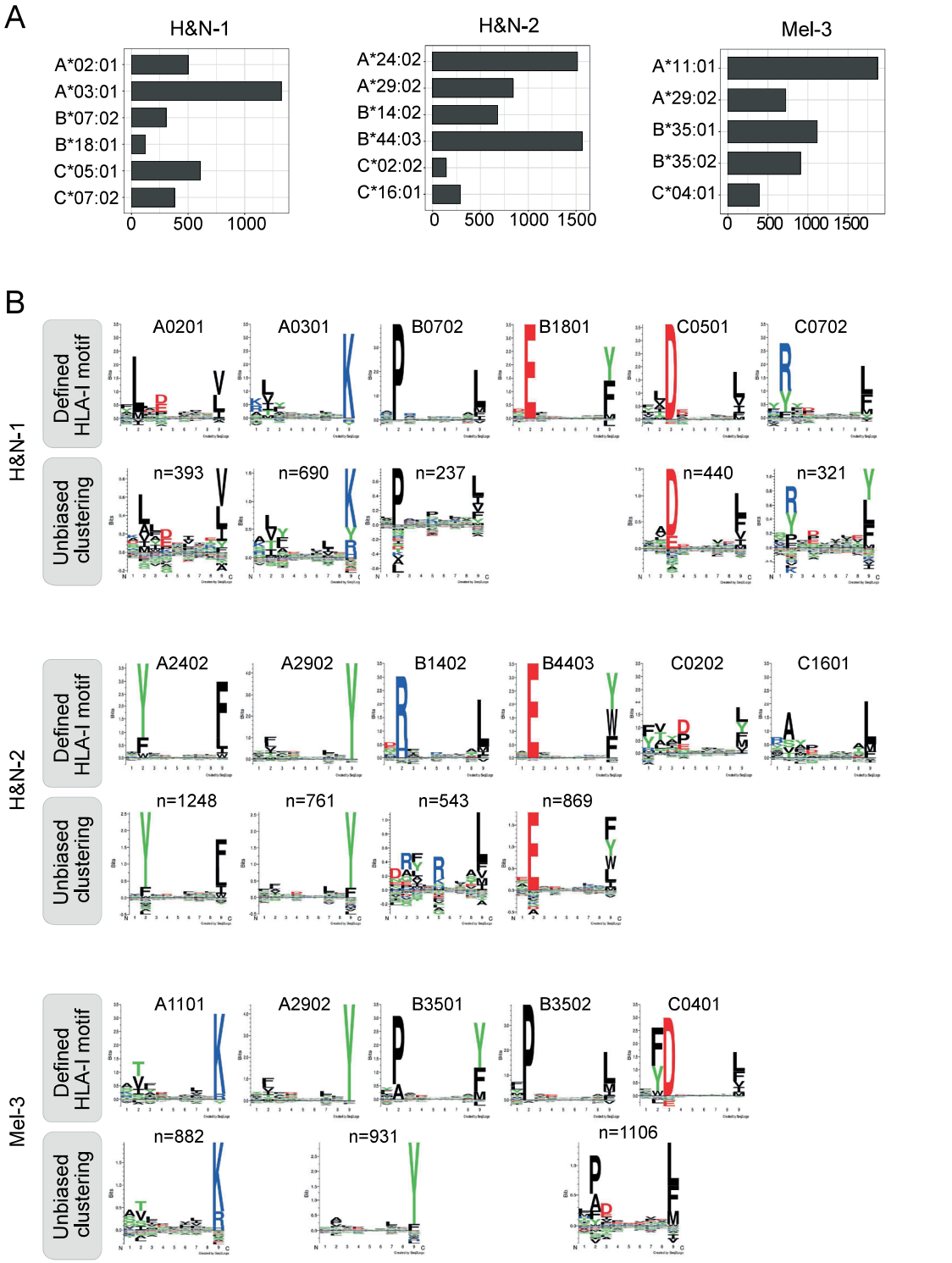


Figure 26. HLA-I binding preference and sequence motifs of the peptides eluted from HLA-I of patient-derived TCL. (A) HLA-I allele binding preference distribution of peptides for each patient-derived TCL studied. Only the minimal rank according to NetMHCpan4.0 for each peptide was considered. (B) Comparison of HLA-I binding motif for the patients' HLA-I alleles defined by NetMHCPan4.0 (top) with the motifs generated by unbiased grouping of peptides generated by GibsCluster2.0 (bottom). Only unique HLA-I peptides 9-mers identified for each TCL were selected for the analysis. The optimal solution according to GibsCluster2.0 analyzed by default settings is depicted. The number of sequences (n=) per group is indicated.

1.4.2. Candidate tumor antigens presented on HLA-I

Since all the proteins expressed in a cell can potentially be presented on HLA-I, we investigated whether tumor-specific peptides could be detected among the repertoire of eluted

HLA-I ligands. As expected, we found that the vast majority of the HLA-I ligands were derived from normal proteins (Figure 27A). In contrast, only 0.4% could be attributed to conventional antigens, including peptides derived from tumor-specific mutations and cancer-germline antigens. For patient Mel-3, we also considered some well-defined melanoma-associated antigens derived from tissue differentiation proteins known to frequently contribute to tumor immunogenicity in melanoma (Coulie et al., 2014), and found peptides derived from some of them (Figure 27A).

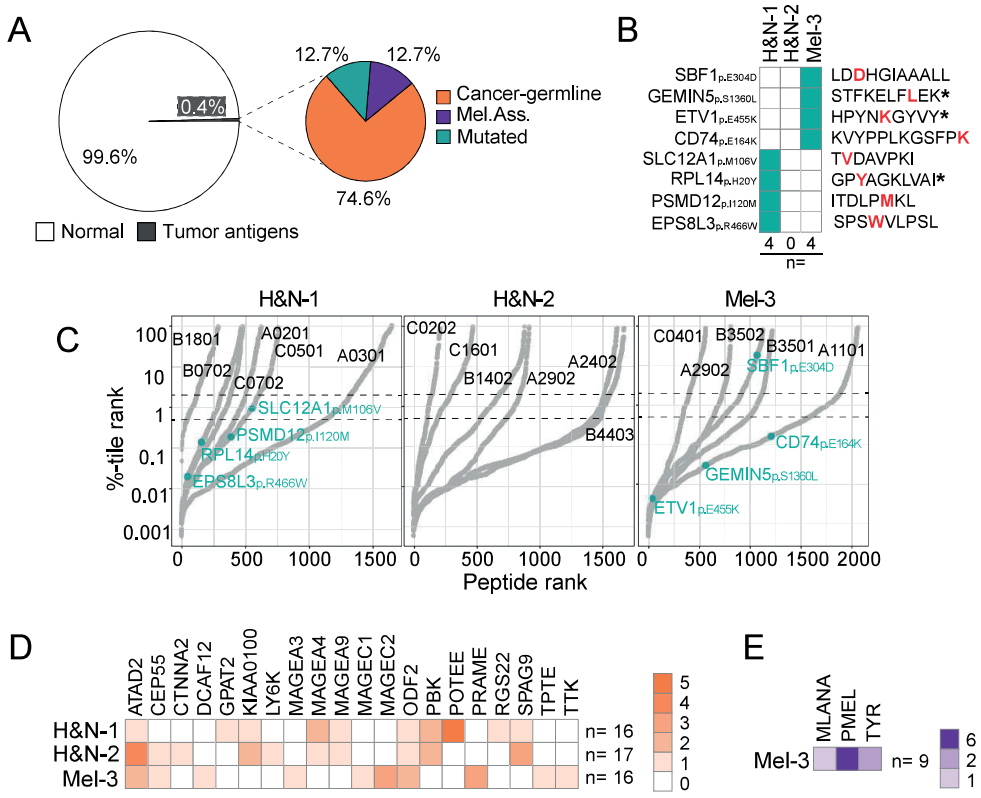


Figure 27. Candidate tumor antigens identified through pHLA-I immunopeptidomics. (A) Pie charts depict the percentage of peptides derived from normal self-proteins or derived from conventional tumor antigens (left), and the percentage of each of the tumor antigen categories (right). (B) HLA-I peptides derived from NSM. The specific mutated protein and Aa change and the peptide sequence are noted on the left and right, respectively. The mutated Aa is highlighted in red bold letters. Immunogenic mutated peptides identified in the previous section are marked with an *. (C) HLA-I binding prediction rank of HLA-I peptides eluted from TCL. HLA binding of all unique peptide sequences to the patient's HLA molecules was predicted using NetMHCPan4.0. Each dot represents a unique peptide. For each peptide, only the minimal percentile rank across all the patient-specific HLA alleles was considered and plotted accordingly. The HLA-I alleles are specified in black. Peptides derived from non-mutated and mutated peptides are shown in grey and turquoise, respectively. For mutated peptides, the mutated protein and Aa change is specified. D) Heat map displaying the number of HLA-I peptides derived from specific cancer-germline antigens per patient or E) derived from melanoma-associated antigens in Mel-3.

More specifically, we identified a total of 8 mutated HLA-I ligands in patients H&N-1 and Mel-3 (Figure 27B). Of note, in addition to the three *bona fide* neoantigens previously shown to be immunogenic as described in section 1.3.2 ($GEMIN5_{p.S1360L}$, $ETV1_{p.E455K}$ and $RPL14_{p.H20Y}$), five new candidate neoantigens were detected. This was particularly relevant for patient H&N-1, because not all the mutations were interrogated through the TMG approach due to the high number of NSM. In fact, the mutations $EPS8L3_{p.R466W}$ and $SLC12A1_{p.M106V}$ were not included in any of the TMG constructed for H&N-1, since they did not rank within the top 260 binders. However, the identified HLA-I ligands containing these two mutations were still predicted to bind to the patient's HLA as strong or weak binders, respectively (Figure 27C). It is also important to mention that not all the neoantigens identified for patients H&N-1 and H&N-2 through TMG strategy (Table 5) were detected by HLA-I immunopeptidomics.

An additional advantage of immunopeptidomics is that identifies the specific sequence and length of the HLA-I ligands potentially recognized by T cells, referred to as minimal epitopes. As opposed to TMG and LP, minimal epitopes are short peptides that do not require processing. At high concentrations, they directly bind to HLA-I by displacing the ligands presented. Consequently, the activation of antigen-specific T cells is typically stronger and their detection is more efficient. Therefore, although $CD74_{p.E164K}$ and $SLC12A1_{p.M106V}$ were previously interrogated through TMG screen in Mel-3 (section 1.3.2), testing these neoantigen candidates in minimal epitope format was still of interest. (Figure 27C).

Furthermore, we detected HLA-I ligands derived from cancer-germline antigens described in the CTdatabase (Almeida et al., 2009) in all three patients, as shown in Figure 27D. Finally, we also identified peptides derived from the melanoma-associated antigens MLANA, PMEL, and TYR in Mel-3 (Figure 27E).

In summary, HLA-I immunopeptidomics of patient-derived TCL led to the detection of a total of 66 tumor antigen candidates naturally presented on HLA-I. These included both new and previously identified candidate neoantigens, as well as candidates derived from cancer-germline or melanoma-associated antigens.

1.4.3. Immunogenicity of tumor antigen candidates identified by HLA-I immunopeptidomics

To determine whether the tumor antigen candidates identified through HLA-I immunopeptidomics were actually able to elicit a T-cell response, the peptides (n=66) were synthesized and pulsed onto autologous APC and co-incubated with tumor-reactive TIL populations. When several peptides derived from the same antigen were identified, they were pooled together (peptide pool, PP). T-cell activation was evaluated 20 h later by IFN-γ release and 4-1BB upregulation, as above.

For patient Mel-3, in addition to the previously identified neoantigens $GEMIN5_{p.S1360L}$ and $ETV1_{p.E455K}$ recognized by TIL-2, TIL-3, TIL-5, and TIL-7, we detected T-cell responses to the PMEL peptide pool (PP; TIL-3, TIL-7, and TIL10) as well as a peptide derived from MLANA (TIL-7) (Figure 28A). In fact, T-cell reactivity to PMEL was clearly dominant in TIL-3, since

80% of the T cells expressed 4-1BB upon antigen recognition. In contrast, the frequency of T cells reactive to $ETV1_{p.E455K}$ was far lower (<1%) (Figure 28A, bottom). However, no reactivity was detected to any of the cancer germline antigens tested.

We also further investigated the specific peptide contained in the PMEL PP recognized by TIL-3 and TIL-10 and observed that both TIL targeted the same epitope (GTATLRLVK) (Figure 28B).

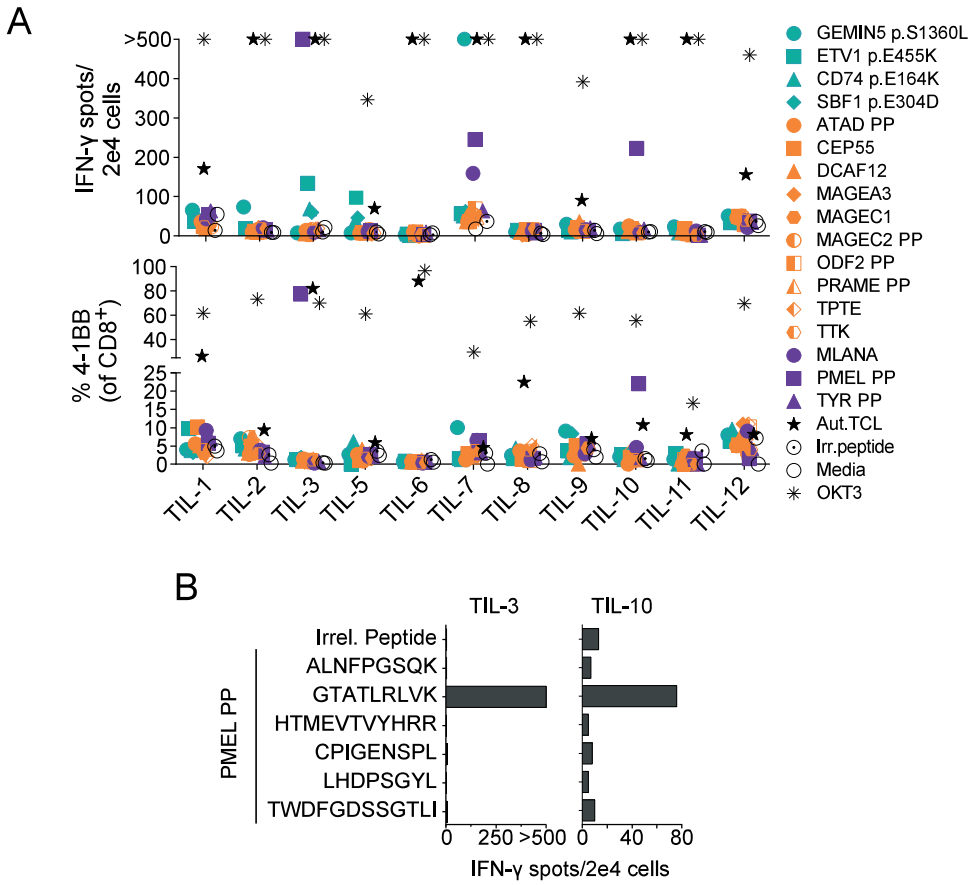


Figure 28. Mel-3 TIL recognize 2 neoantigens and 2 melanoma-associated antigens identified through HLA-I immunopeptidomics. (A) Recognition of tumor antigen candidates was evaluated by co-incubating 2e4 lymphocytes with 2e5 autologous APC pulsed with 1 µg/mL of synthetic peptides individually or pooled according to the protein of origin (PP). The number of IFN-γ spots per well measured by ELISPOT assay for each TIL co-cultured with the targets specified (top) and the percentage of $CD8^+$ cells expressing 4-1BB measured by flow cytometry (bottom) are shown. B cells were used as APC. Media and B cells pulsed with an irrelevant peptide were used as negative controls. OKT3 and autologous TCL were used as positive controls. (B) Individual peptides from PMEL PP were individually pulsed onto autologous APC at 1 µg/mL for 2 h, washed, and co-cultured with REPed TIL-3 and TIL-10. After 20 h the reactivity was measured by IFN-γ ELISPOT assay. > denotes greater than 500 spots/ well.

We tested the candidate tumor antigens identified through HLA-I immunopeptidomics using the same strategy in H&N-1 and H&N-2. Unfortunately, whereas in patient Mel-3 this technique enabled the identification of the neoantigens previously detected, but also two newly identified immunogenic melanoma-associated antigens, no additional tumor antigens were detected for patients H&N-1 and H&N-2 (Table 6), as compared to those detected using the TMG and minimal epitopes (Table 5).

Table 6. Immunogenic peptides identified through pHLA-I immunopeptidomics

Patient ID	TA type	Immunogenic / eluted	Tumor antigen	Sequence
H&N-1	NSM	1/4	$RPL14_{p.H20Y}$	GPYAGKLVAI
	CG	0/16	--	--
	Mel.Ass	--	--	--
H&N-2	NSM	0/0	--	--
	CG	0/17	--	--
	Mel.Ass	--	--	--
Mel-3	NSM	2/4	$GEMIN5_{p.S1360L}$ $ETV1_{p.E455K}$	STFKELFLEK HPYNKGYVY
	CG	0/16	--	--
	Mel.Ass	2/9	MLANA PMEL	MPREDAHF GTATLRLVK

Recognition of the tumor antigen candidates was evaluated as explained in Figure 12. TA, tumor antigen. The mutated amino acids are shown in red bold letters in the case of neoantigens.

1.5. The specific antigen recognized by tumor-reactive T cells could not be determined despite extensive screening

Collectively these results indicate that tumor-reactive $CD8^+$ lymphocytes can frequently be detected in cancer patients. Although the level of recognition varied widely among TIL derived from different fragments, we detected from 5 to 11 TIL capable of recognizing the autologous TCL in each of the three patients studied (Figure 29 bottom). Moreover, we detected $CD8^+$ T cells targeting neoantigens in all of them (two neoantigens in each patient). In addition, T-cell responses to two peptides derived from melanoma-associated antigens were also detected in patient Mel-3 (Figure 29 top).

However, it should be noted that the frequency of activated cells upon recognition of specific antigens tested was remarkably low as compared to the level of activation induced by the autologous TCL, with few exceptions. This observation suggested that these tumor-reactive T cells might be in fact recognizing additional tumor antigens. Likewise, we could not detect any specific target for many tumor-reactive TIL.

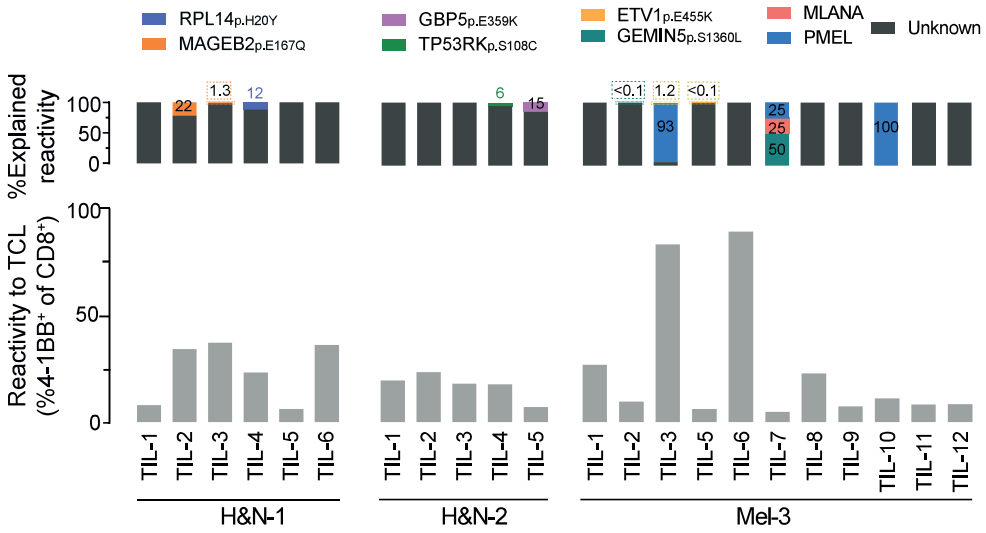


Figure 29. The specific antigens targeted by the vast majority of tumor-reactive T cells are still unknown. The percentage of tumor reactivity displayed by each tumor-reactive T-cell population and patient based on 4-1BB expression is depicted (bottom). The percentages of tumor reactivity explained by recognition of the tumor antigens identified are represented in different colors or grey when the target was not identified (top). The explained reactivity was calculated as a percentage taking into account the frequency of $CD8^+$ lymphocytes expressing 4-1BB after co-culture with the TCL and the percentage of $CD8^+$ T cells expressing 4-1BB upon specific antigen recognition.

Altogether, these results indicated that despite extensive screening using multiple experimental approaches, the specific antigen recognized by tumor-reactive T cells could not be determined.

2. Identification of non-canonical tumor antigens through proteogenomics

The antigens investigated in part 1 all derive from the annotated reference proteome encoded by the exome, including mutated as well as non-mutated proteins, referred to as conventional or canonical tumor antigens.

The exome constitutes only the ~2% of the entire genome and it has been commonly accepted as the protein-coding part of the genome, but almost 75% of the genome is actually transcribed and, potentially, translated (Djebali et al., 2012). Indeed, a growing number of studies have demonstrated that peptides derived from nonC ORF or from allegedly non-coding regions are presented on HLA-I (Erhard et al., 2020; Laumont et al., 2016; Ruiz Cuevas et al., 2021; Scull et al., 2021). A fraction of such considered aberrant translation events have been postulated to specifically occur in tumor cells, thus substantially expanding the repertoire of targetable tumor antigens (Laumont et al., 2018; Ouspenskaia et al., 2022; Zhao et al., 2020). In fact, spontaneous T-cell responses against peptides derived from nonC proteins have been occasionally identified by screening T cells using tumor cDNA antigen libraries (summarized in Garcia-Garijo et al., 2019). However, the systematic identification of truly tumor-specific nonC antigens in humans remains challenging. Most of the studies reported so far rely on the direct identification of HLA-I ligands through immunopeptidomics, but only a few have examined their contribution to tumor immunogenicity (Chong et al., 2020).

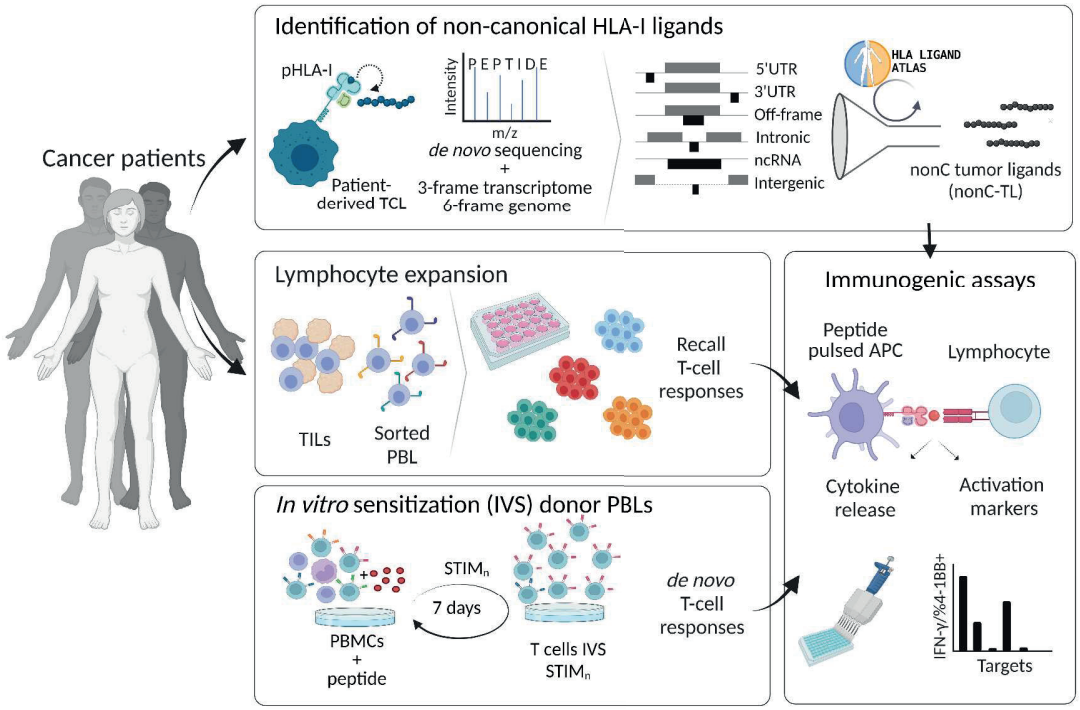


Figure 30. Experimental strategy to identify and determine the immunogenicity of non-canonical tumor antigens. HLA-I ligands derived from nonC proteins were identified through HLA-I immunopeptidomics of TCL through conventional procedures. The Aa sequence of each mass spectrum was first identified by *de novo* sequencing and later aligned to a database containing the 3-frame transcriptome and 6-frame genome. The nonC peptides preferentially presented on tumor cells (top) were further selected for immunological testing based on tumor specificity using data from the HLA ligand Atlas. To detect pre-existing T-cell responses in cancer patients, either TIL or sorted PBL were screened for peptide recognition (middle). To stimulate T-cell responses *de novo*, do-

nor PBMC were serially stimulated with nonC-TL every 7 days (bottom). T-cell recognition of the specific targets was assessed in both cases after a 20 h co-culture by IFN-γ ELISPOT and by measuring the upregulation of 4-1BB by flow cytometry (right panel). Image created with BioRender.com.

Given the potential of nonC HLA-I ligands as a new source of tumor antigens, in this section, we explored their presentation and immunogenicity to address whether they could explain, at least a fraction of the T-cell reactivity observed against the autologous TCL, and also whether they could be exploited therapeutically. To this end, we used a recently described proteogenomics pipeline (Erhard et al., 2020) to identify nonC HLA-I ligands derived from the off-frame translation of coding sequences and non-coding regions (UTR, ncRNA, intronic, and intergenic) presented by our patient-derived TCL. This pipeline was further modified to select the tumor-specific candidates by exploiting immunopeptidome data from healthy tissues. We aimed to better understand their natural or induced immunogenicity by testing for pre-existing or *in vitro* sensitized T-cell responses to non-canonical tumor HLA-I ligands (Figure 30).

2.1. Identification of non-canonical HLA-I ligands in patient-derived TCL

We first determined whether we could detect HLA-I ligands originated from nonC proteins in the 9 patient-derived TCL that were recognized by autologous T cells, as described in section 1.2. This included 4 Gyn, 3 Mel, and 2 H&N cancer patients.

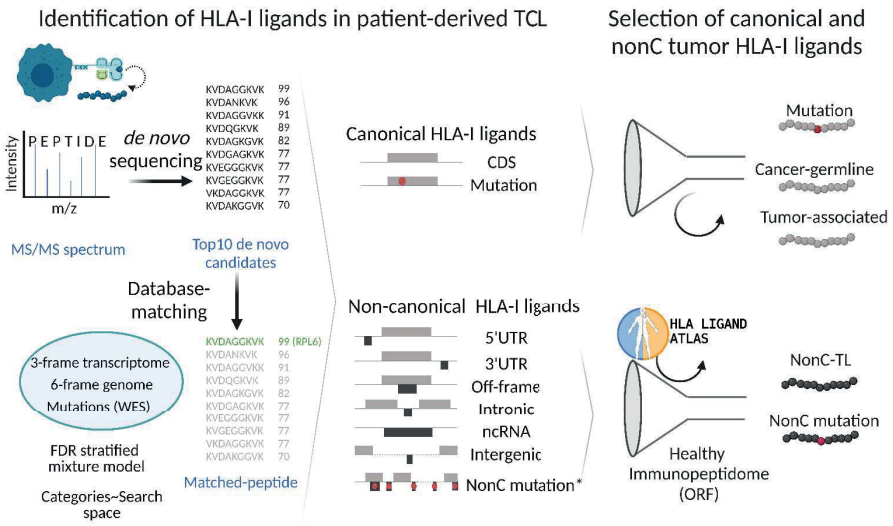


Figure 31. Identification of canonical and nonC HLA-I ligands presented by TCL using Peptide-PRISM. The Top 10 candidates for each mass spectrum were first identified by *de novo* sequencing and later aligned to a database containing the 3-frame transcriptome, and 6-frame genome including the mutations identified by WES. The FDR was calculated by each category in a stratified mixture model (left panel). All canonical peptides containing mutations and peptides derived from cancer-germline or tumor-associated antigens were selected for immunological screening. For nonC HLA-I ligands, an immunopeptidomics healthy dataset containing several tissues and HLA-I allotypes obtained from the HLA ligand atlas was used to eliminate those derived from normal ORF and to select for nonC tumor ligands preferentially found in tumor (nonC-TL) (right panel). Image created with BioRender.com.

2.1.1. NonC-TL are frequently identified in patient-derived TCL and display features similar to peptides derived from canonical proteins

First, peptides bound to HLA-I were isolated and analyzed by LC-MS/MS using state-of-the-art procedures, as described in section 1.4.1. Next, to identify peptides derived from nonC proteins, the amino acid sequences were inferred through Peptide-PRISM (Erhard et al., 2020) developed by our collaborator Andreas Schlosser from the University of Würzburg (Figure 31). Briefly, for each MS spectrum, the top 10 candidates were first identified by *de novo* sequencing and later mapped to a database including the human 3-frame transcriptome, as well as the 6-frame genome. Additionally, WES information from each TCL was included to interrogate personalized mutations. The FDR was calculated independently from each category considering the search space and peptide length in a stratified mixture model as previously described (Erhard et al., 2020).

Following this strategy and selecting a stringent 1% FDR, we identified both C, but also nonC HLA-I ligands in all studied patients, which ranged from 0.5% to 5.4% of the total eluted peptides (Figures 31 left, and 32A).

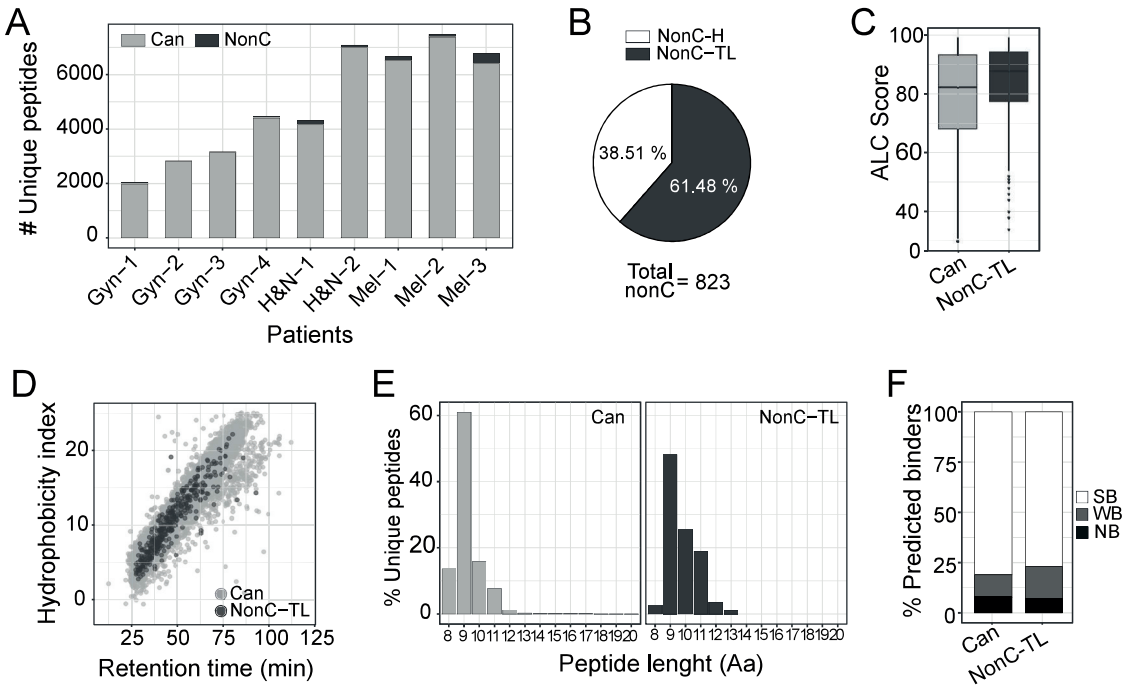


Figure 32. NonC-TL are frequently presented on patient-derived TCL and display features similar to HLA-I ligands derived from canonical proteins. (A) Number of canonical and non-canonical HLA-I peptides identified in each TCL. (B) Number of nonC ligands derived from ORF predicted to be present or absent from healthy tissue using a healthy tissue immunopeptidome data set (nonC-H and nonC-TL, respectively). (C) ALC identification score boxplot. Data from all patients was pooled together. (D) Predicted hydrophobicity index (y axis) and retention time (x axis). Each dot represents one unique HLA-I ligand sequence. Data from patients acquired in the same conditions (130min gradient run) is depicted. (E) Length distribution of unique HLA-I peptide sequences. Only peptides <20 Aa are depicted. (F) Percentage of peptides predicted to bind to the corresponding patient HLA alleles according to NetMHCPan4.0. Peptides were categorized into strong binders (SB; %-tile rank<=0.5), weak binders (WB; %-tile rank 0.5-2) or non-binders (NB; %-tile rank >2). Data from all patients was pooled together for B, C, D, E, F. Only unique peptide sequences were considered. In all the analyses shown, the FDR threshold was set at 0.01 and ALC score at 30.

In order to select nonC HLA-I ligands preferentially presented in tumor cells, we used immunopeptidomics data from several healthy tissues and samples available from the HLA ligand atlas (Marcu et al., 2021) to filter out peptides described to be presented in non-malignant tissues. To overcome a potential bias toward frequent alleles, the peptides were excluded at the ORF level rather than the amino acid sequence (Figure 2 right). For example, peptides were excluded if they derived from an ORF that was predicted to be translated based on the healthy ligandome dataset. Hence, we believe that although the alleles less frequently found in the population would probably be underrepresented in the immunopeptidomics dataset employed for filtering, the use of nonC ORF instead of the nonC HLA-I ligand sequences might reduce said bias. As a result, we found that from the total 823 unique nonC peptides detected in our tumor samples, 38.5% were derived from ORF also present in healthy tissue (nonC-H). Hence, 61.5% of the nonC HLA-I ligands were considered preferentially presented in tumors, from now on referred to as nonC HLA-I tumor ligands (nonC-TL) (Figure 32B).

Moreover, nonC-TL displayed characteristics similar to peptides derived from canonical proteins such as the MS identification score (ALC) or the correlation between the retention time and hydrophobicity index (Figures 32C-D). Additionally, nonC-TL displayed the typical HLA-I ligand features, as shown by the length distribution and the high percentage of peptides predicted to bind to the patient HLA-I molecules according to NetMHCPan4.0 (Figure 32E-F). Together, these analyses indicate that our approach accurately identifies the HLA-I ligand repertoire and showed that nonC-TL are frequently detected in patient-derived TCL.

2.1.2. NonC-TL are mainly derived from 5'UTR regions and preferentially bind to HLA-A*11:01 and A*03:01

Next, we evaluated the genomic origin of the identified nonC-TL. Consistent with previous studies (Erhard et al., 2020; Ouspenskaia et al., 2022; Ruiz Cuevas et al., 2021) , we found that translation of 5'UTR was the main source of nonC-TL, followed by off-frame and non-coding RNA (ncRNA) (Figure 33A). Peptides derived from 3'UTR, intronic and intergenic regions were less frequently detected. In addition, one nonC peptide derived from a 5'UTR containing a mutation was detected in patient Gyn-3 (Figure 33A). Importantly, the specific tumor type did not influence the frequency of nonC HLA-I ligands identified, but rather the expression of HLA-A*11:01 and HLA-A*03:01 was a major determinant (Appendix Figure 1). We noticed a bias in the HLA-I binding preference distribution of nonC peptides toward these alleles according to NetMHCPan4.0 (Figure 33B). Interestingly, both HLA alleles display a similar motif containing basic residues at p9 (Figure 33C), a unique feature among all the alleles studied (Appendix Figure 2). Indeed, the binding preference of nonC HLA-I ligands to HLA-A*11:01 and/or HLA-A*03:01 alleles has been previously reported in other immunopeptidomics studies(Erhard et al., 2020; Laumont et al., 2016). However, the exact mechanism underpinning this bias is unknown.

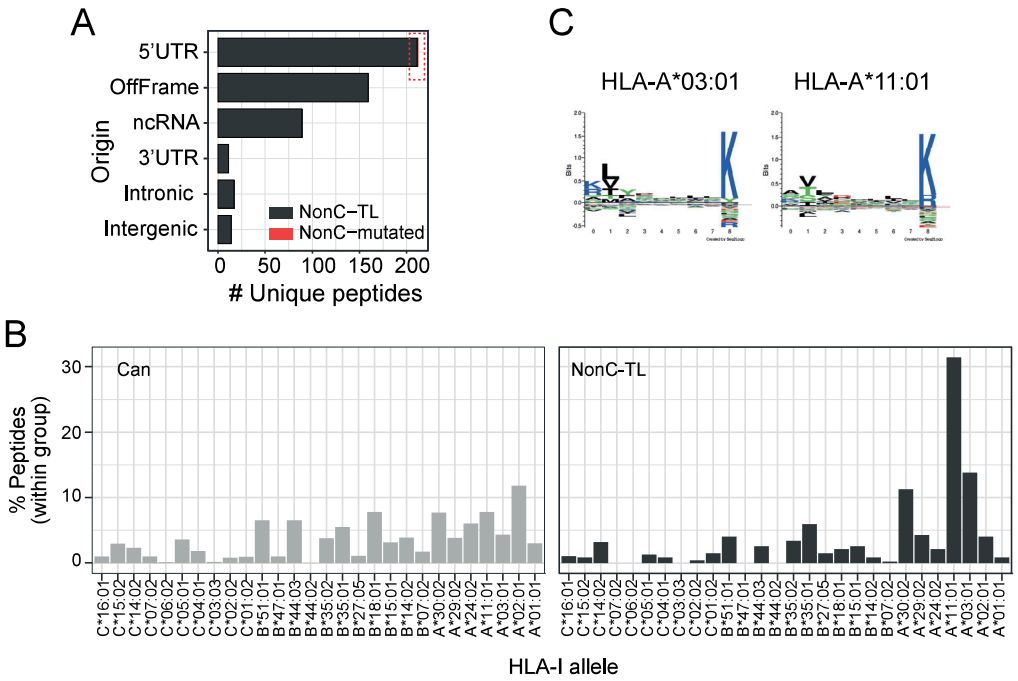


Figure 33. NonC-TL mainly derive from 5'UTR regions and are predicted to preferentially bind to HLA-A*11:01 and HLA-A*03:01. A) Number of nonC-TL originated from each ORF category. B) Allele binding preference of Can and nonC-TL peptides. For each peptide, only the min rank according to NetMHCpan4.0 was considered. C) Consensus binding motif of the two HLAs predicted to bind to the majority of the nonC peptides identified. Image downloaded from NetMHCpan motif viewer. Data from all patients was pooled together for A, B. Only unique peptide sequences were considered. The FDR threshold was set at 0.01 and ALC score at 30.

2.2. Distribution of candidate tumor antigens presented on HLA-I

To examine whether nonC-TL could be therapeutically exploited as a new source of tumor antigens we first compared them with clinically relevant tumor antigens, including peptides encoded in canonical coding regions either derived from mutations or derived from cancer-germline and tumor-associated antigens.

2.2.1. NonC-TL constitute an abundant source of candidate tumor antigens

First, we evaluated the number of peptides presented on HLA-I from patient-derived TCL for each antigen category. While the number of HLA-I ligands detected from canonical tumor antigens ranged from 24 to 36 peptides, nonC-TL outnumbered the other categories, exceeding a total of 500 unique tumor antigen candidates (Figure 34A). This observation was consistent across each of the patients studied, with few exceptions (Figure 34B). More specifically, a total of 33 peptides derived from mutations were detected in 6 out of 9 patients. Despite the number of eluted HLA-I ligands containing mutations was low compared to the total NSM identified by WES, these results are in line with previous immunopeptidomics studies where few mutated peptides are typically detected (Bassani-Sternberg et al., 2016; Kalaora et al., 2016; Newey et al., 2019) (Figure 34B and 34C). Moreover, 36 peptides derived from 12 cancer-germline antigens and 24 peptides derived from 5 melanoma-associated antigens were identified in 8 out of 9 patients (Figure 34D). As mentioned above, nonC-TL originated mainly from 5'UTR and off-frame translation, while nonC-TL derived from intergenic and intronic regions were not or barely detected, and this was consistent in most TCL (Figure 34E).

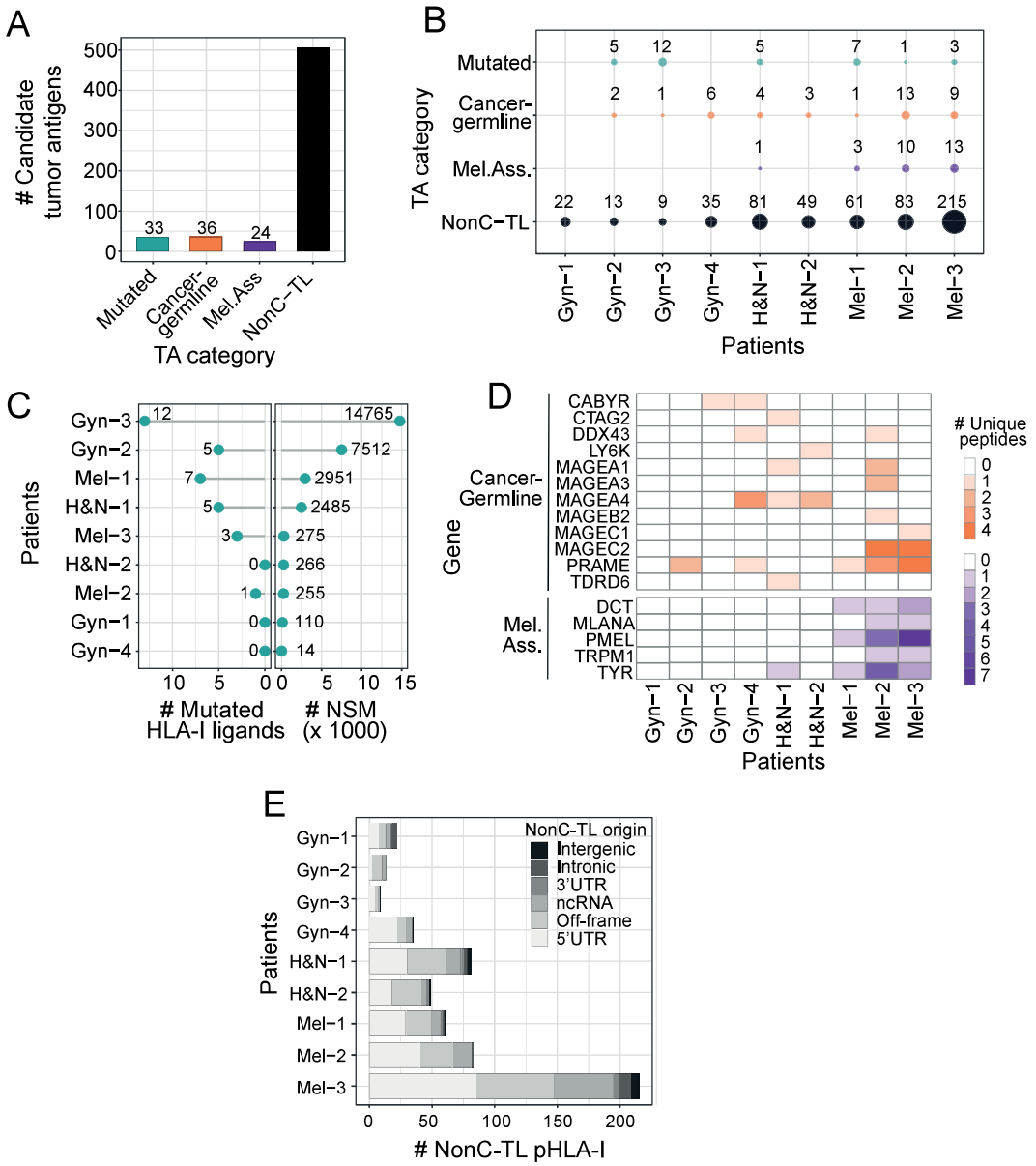


Figure 34. Candidate tumor antigens presented on HLA-I of patient-derived TCL. (A) Total number of unique peptide sequences derived from tumor antigens (TA) by category. Data from all patients was pooled together. Only unique peptide sequences were considered. (B) The number and category of tumor antigens are displayed for each TCL. (C) The number of mutated peptides eluted from HLA-I (left) and the number of NSM identified by WES (right) are displayed for each patient. (D) Heat map displaying the number of epitopes derived from specific cancer-germline or melanoma-associated antigens per patient. (E) Number of nonC-TL originated from each ORF category per patient. In all the analyses the FDR threshold was set at 0.01 and ALC score at 30.

2.2.2. NonC-TL are shared across cancer patients

One potential advantage of exploiting non-mutated antigens over private mutations as targetable tumor antigens is that they may be shared in different patients, facilitating the development of off-the-shelf vaccines or T-cell therapies. Similar to cancer-germline or melanoma-associated antigens, we observed that ~10% of nonC-TL were shared by at least 2 patients (Figure 35). Interestingly, nonC-TL included one and three peptide sequences identified in 5 and 4 patients, respectively, indicating that some can be widely shared. In contrast, all the mutated peptides detected derived from private mutations, and thus, were exclusively identified in a single patient. (Figure 35).

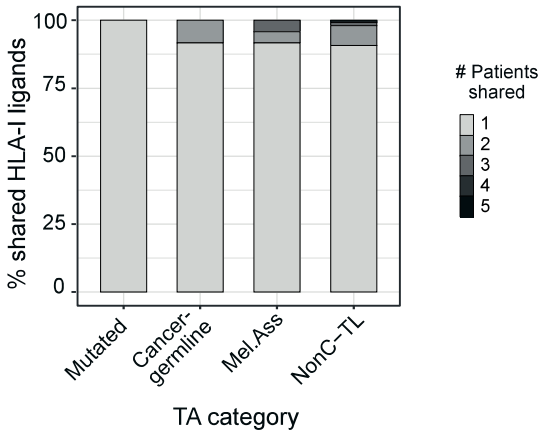


Figure 35. NonC-TL can be shared across patients. Percentage of nonC-TL uniquely identified in one patient or shared by tumor antigen category. Data from all patients was pooled together. Only unique peptide sequences were considered. The FDR threshold was set at 0.01 and ALC score at 30.

Altogether, this data highlights nonC-TL as promising targets for the development of therapeutic interventions since they constitute the most abundant source of candidate tumor antigens compared to peptides derived from mutations, cancer-germline, or melanoma-associated antigens, and can be shared across patients.

2.3. Validation of tumor antigen candidates

Although the analysis performed in section 2.1.1 indicate that the HLA-I peptides detected through Peptide-PRISM were accurately identified, we next wanted to verify the sequences of the candidates experimentally.

In view of the extensive number of candidates to test (600 peptides) and the complexity of the method typically used to accurately validate peptide sequences by mass spectrometry (targeted MS using internal standards), we used this approach for a selected list of peptides. We additionally decided to interrogate the maximum number of candidates through the comparison of the MS/MS fragmentation mass spectra with those of synthetic peptides.

2.3.1. Parallel reaction monitoring with internal standards

To determine whether the peptide sequences detected through this approach were accurately identified, we synthesized heavy versions of 15 selected peptides isotopically labeled at one chosen amino acid. Next, these isotope-labeled peptides were spiked into the corresponding pHLA-I sample (where they were previously identified), and both heavy (isotope-labeled, spiked in) and light (endogenous) peptides were monitored through parallel reaction monitoring (PRM). Since both versions have the same chemical characteristics, they should display similar features, chromatographic retention time and MS/MS fragmentation pattern. But they can be distinguished by LC-MS/MS thanks to their difference in mass.

Figure 36 shows the LC-MS/MS extracted ion chromatogram signals for one peptide of three different antigen categories (mutation, melanoma-associated, and nonC-TL) identified in Mel-3, as an example of validation with PRM and internal standards. The upper panels correspond to the endogenous peptides and the lower panels to the respective isotopically labeled

standards spiked into the samples. Taking into account the remarkable similarity in retention time, co-elution, peak shape, and MS transition rank, all three peptides were considered validated (Figure 36). Using this strategy we validated 14 out of the 15 HLA-I ligands tested. In addition to the examples shown in Figure 36, 6 out of the 7 mutations identified in Mel-1 as well as 2 nonC-TL, 2 mutated peptides, and 1 melanoma-associated peptide in patient Mel-3 were validated (not shown).

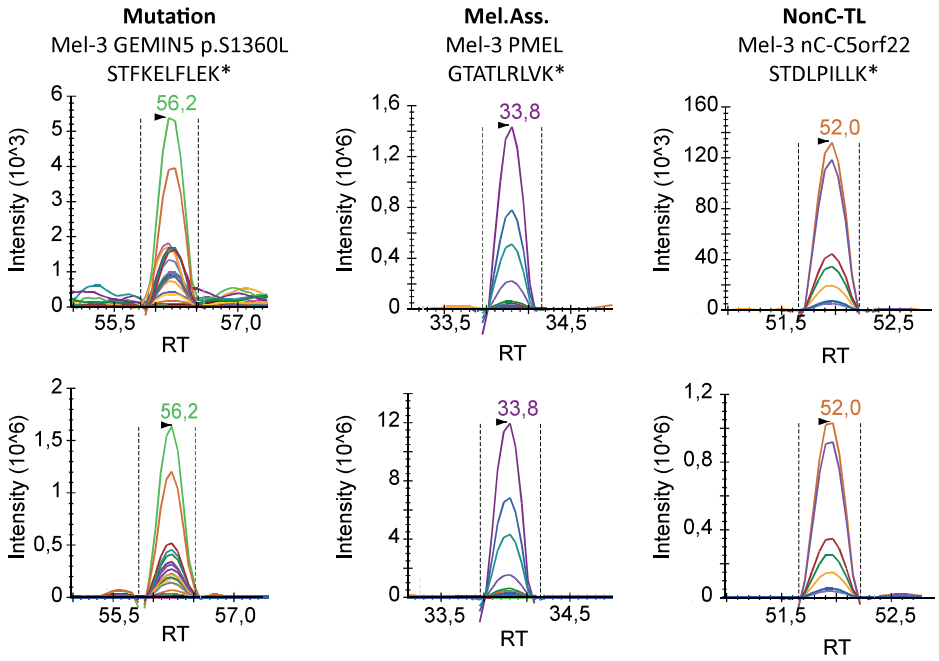


Figure 36. Validation of selected HLA-I ligand tumor antigen candidates identified in Mel-3 with isotope-labeled peptides. Isotope-labeled peptide versions for each selected HLA-I ligand sequence were spiked into the corresponding HLA-I peptide sample and both the heavy and the light peptides were monitored through parallel reaction monitoring. Representative co-elution profiles of transitions of endogenous (top) and isotope-labeled (bottom) peptides using Skyline software are displayed for three HLA-I ligand tumor antigen candidates. Tumor antigen category, patient ID, and the amino acid sequence validated are specified for each peptide. Isotope-labeled amino acids are marked with *.

2.3.2. Spectrum comparison with synthetic peptides

In order to increase the number of candidates evaluated, we explored an alternative method to validate the amino acid sequences in a more high-throughput manner. First, the candidate peptides were synthesized without labeling and shotgun LC-MS/MS data was acquired for a pool of the peptides using conditions comparable to the previously analyzed samples (non-targeted). Next, we determined the similarity of the fragmentation pattern (MSMS mass spectrum) obtained from the synthetic peptides to those previously assigned to the endogenous peptides identified. To this end, we generated a library containing the MSMS spectra of the synthetic peptides and computed the similarity to the MSMS spectra identified in the original sample by considering the library dot product (dotp) values provided by Skyline Software. This strategy enabled the comprehensive analysis of an extended list of candidates in an unbiased and automatized manner, without the need to manually review all mass spectra pairs one by one (Figure 37).

Figure 37A depicts the endogenous (top) and synthetic (bottom) mass spectrum of a

nonC-TL candidate and the corresponding relative intensities of each of the fragment ion signals integrated using Skyline (Figure 37B). As visually observed and also measured by dotp (0.94), both peptides displayed a very high similarity in the fragmentation pattern. A dotp close to 1, indicates that both MS are highly similar, suggesting that they probably derived from the same peptide sequence.

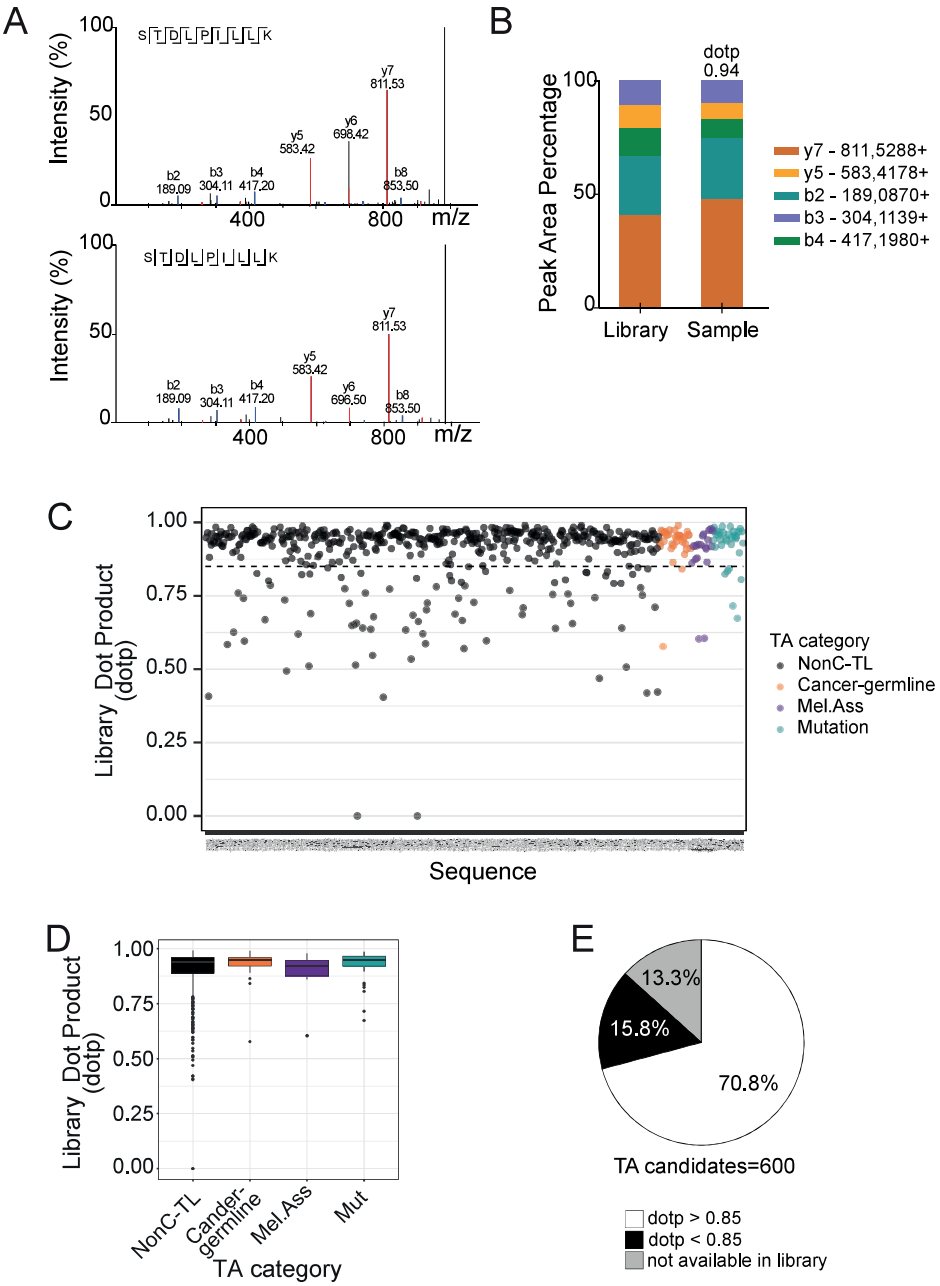


Figure 37. Validation of HLA-I ligand tumor antigen candidates with synthetic peptides. Spectrum validation of the experimentally eluted HLA-I ligand tumor antigen (TA) candidates was performed by computing the similarity of the spectra acquired in the sample with that of the corresponding non-labeled synthetic peptide from the library. A spectral library was generated with synthetic peptides acquired in a pool using comparable LC-MS/MS shotgun conditions. The similarity of the MS/MS fragmentation pattern from the library compared to the sample was evaluated using Skyline Software considering library dot product (dotp) values of the integrated peak areas for the observed fragment ions. (A) Representative example of an MS/MS fragmentation pattern of a nonC-TL eluted from Mel-3 TCL HLA-I (experimental identification, top) and its corresponding synthetic peptide (library, bottom), PEAKS XPro visualization. (B) b/y ions generated from the library (synthetic) compared to fragments acquired in the sample (experimental identification) with Skyline software. The similarity is measured according to dotp, which ranges from 0 to 1, being dotp=1 the closest match. Example of the same peptide represented on A.C) Overall acquisition data of tumor antigen (TA) candidates matched to synthetic peptides. The analysis of the library dot product (dotp) shows the level of agreement of each identified sequence (x-axis) to library →

→ values (y-axis). The threshold to consider a peptide validated was set at 0.85 dotp (dashed line). (D) Box and whisker plot for the dotp values for all the TA candidates identified in the library for each category. (E) Percentage of candidates validated according to library dot product (dotp >=0.85), not validated (dotp <0.85) or not interrogated (synthetic not detected, and thus not included in library).

Using this approach we analyzed all the peptides matching the library and observed that most peptide sequences displayed a dotp close to 1 (Figures 37C and 37D), independently of the tumor antigen type, suggesting that the peptides were properly identified.

Unfortunately, although we attempted to acquire all the TA candidates in the pool to evaluate the similarity of all of them, it was not possible for a small fraction (13.3%) since the synthetic peptides were not identified by MS. Hence, these candidates could not be included in the library, and thus, were not interrogated through this method. Nonetheless, considering a stringent dotp of 0.85, we were able to validate the amino acid sequence of at least 70.8% of the tumor antigen candidates, indicating that most of them were accurately identified, including tumor antigen candidates derived from canonical antigens (mutations, cancer-germline, and tumor-associated) but also nonC-TL.

2.4. Immunogenicity of nonC-TL in cancer patients

To evaluate the role of nonC-TL in cancer immune surveillance, we investigated the presence of pre-existing T-cell responses targeting candidate tumor antigens including all nonC-TL identified. In parallel, we also tested the immunogenicity of canonical HLA-I ligands derived from mutated, cancer-germline and tumor-associated antigens detected through the same pipeline. To this end, the *ex vivo* expanded lymphocytes recognizing the autologous TCL identified in section 1.2., were co-cultured with APC pulsed with the synthetic personalized candidate peptides. Again, T-cell reactivity was assessed by measuring IFN-γ release by ELISPOT and the upregulation of 4-1BB on $CD8^+$ T cells by flow cytometry. Given the larger number of nonC peptides tested, up to 35 peptides were pooled together (PP).

2.4.1. Mel-3 TIL recognize neoantigens and tumor-associated antigens, but not nonC-TL

For patient Mel-3, the recognition of at least one candidate peptide was detected in 5 out of the 11 tumor-reactive TIL populations tested. More specifically, TIL recognized the same two *bona fide* neoantigens ($ETV1_{p.E455K}$ and $GEMIN5_{p.S1360L}$) and the two immunogenic peptides derived from melanoma-associated antigens (PMEL and MLANA) which had been previously identified (section 1.3.3). However, no reactivity was detected against any of the 9 peptides derived from cancer-germline antigens nor any of the 215 nonC-TL candidates (Figure 38).

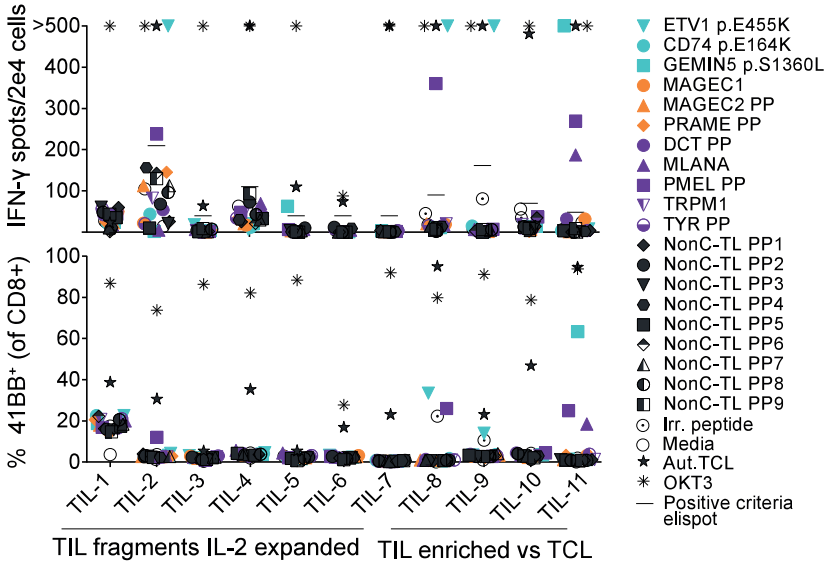


Figure 38. Pre-existing T-cell responses to candidate tumor antigens in Mel-3. Reactivity was evaluated by co-incubating 2e4 TIL with 2e5 autologous APC with 1 µg/mL of selected peptides either alone or in pools (PP). IFN-γ ELISPOT and 4-1BB upregulation by flow cytometry were used to measure T-cell responses after 20 h. The number of IFN-γ spots per well (top panel) and the percentage of cells expressing 4-1BB (bottom panel) are shown. Mutated peptides are plotted in turquoise, cancer-germline in orange, melanoma-associated in purple, and nonC-TL in black. PP stands for peptide pool. Plotted cells were gated on live $CD3^{+}CD8^{+}$ lymphocytes. '>' denotes greater than 500 spots/2e4 cells. B cells were used as APC. Experiments were performed twice.

2.4.2. Tumor-reactive T cells in cancer patients target neoantigens and tumor-associated antigens rather than nonC-TL

We applied the same strategy to identify pre-existing T-cell responses targeting candidate tumor antigens to all the patients included in the study (Figure 39). In total, the extensive screening of tumor-reactive T cells for recognition of 600 tumor antigen candidates led to the detection of 19 immunogenic peptides in the 9 cancer patients studied (Table 7, Figure 39).

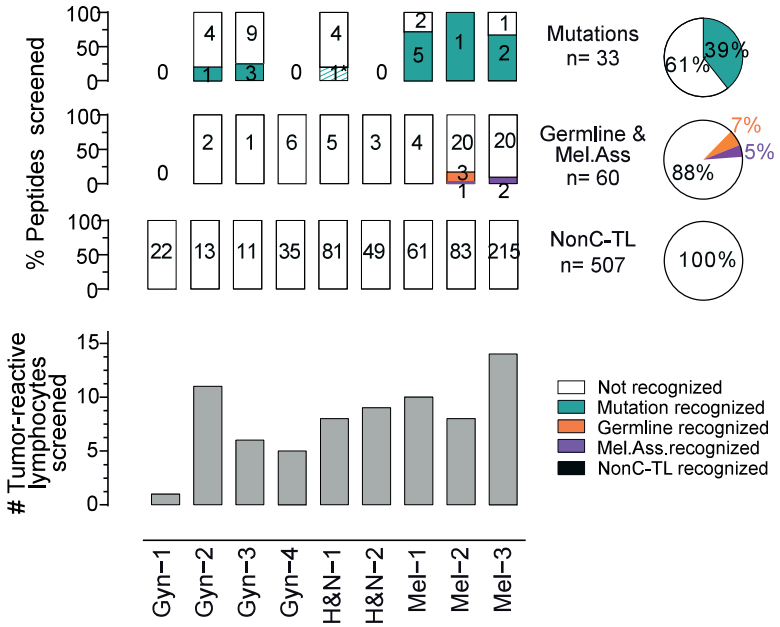


Figure 39. Pre-existing T-cell responses to candidate tumor antigens in cancer patients. For each patient, reactivity was evaluated by co-incubating 2e4 T cells (tumor-reactive TIL or PBL subsets), with 2e5 autologous APC pulsed with 1 µg/mL of selected peptides either alone or in pools (PP). IFN-γ ELISPOT and 4-1BB upregulation by FACS were used to measure T-cell responses after 20 h. The percentage and the absolute number of recognized

and non-recognized peptides within each category are shown per patient (bar plot) and for all the patients studied (pie chart). The number of tumor-reactive lymphocytes tested for each patient is shown on the bottom. Either B cells or $CD4^+$ lymphocytes were used as APC. Plotted cells were gated on live $CD3^+CD8^+$ lymphocytes. '>' denotes greater than 500 spots/2e4 cells. Experiments were performed twice.

In addition, we enriched the reactive lymphocyte populations by FACS sorting as explained in section 1.3.2 to confirm the reactivity and characterize the antigen-specific T cells in detail (Appendix Figure 3). All antigen-specific T cell populations isolated were capable of recognizing the autologous TCL where the ligand was originally identified (Appendix Figure 4. Moreover, neoantigen-specific T cells displayed preferential recognition of the mutated peptide over the wild type counterpart (Appendix Figure 5). In addition, we experimentally identified the restriction element for the majority of the antigen-specific T cells (Appendix Figure 4 and Table 7).

Table 7. Immunogenic tumor antigens identified in cancer patients

Patients	TA[a]	Symbol[b]	Sequence[c]	Predicted HLA allele[d]	Predicted binder (min rank)[e]	Restriction HLA allele[f]
Gyn-1		KIF2C$_{p.L175F}$	IPSKC**F**LLV	B*51:01	SB (0.3137)	n.a
Gyn-2		FUBP1$_{p.R365Q}$	IITDLL**Q**SV	A*02:01	SB (0.1403)	A*02:01
		LAMB3$_{p.D710A}$	**A**PSGAFRML	B*07:02	SB (0.0203)	B*07:02
		RPL19$_{p.I137V}$	**V**LMEHIHKL	A*02:01	SB (0.0034)	A*02:01
H&N-1		RPL14$_{p.H20Y}$	GP**Y**AGKLVAI	B*07:02	SB (0.1369)	B*07:02
Mel-1		CDKN2a$_{p.G74fs}$	**AVCPWTWLR**	A*11:01	SB (0.3405)	A*11:01
		DUSP3$_{p.S81F}$	FYKD**F**GITY	C*14:02	SB (0.0137)	C*14:02
		NAT10$_{p.A320V}$	A**V**IPLPLVK	A*11:01	SB (0.0057)	A*11:01
		SRRT$_{p.A434V}$	IAPNISR**V**	B*51:01	WB (0.5504)	B*51:01
		TRRAP$_{p.S2502F}$	AMLP**F**ITNV	A*02:01	SB (0.0092)	A*02:01
Mel-2		MAGEA6$_{p.E168K}$	**K**VDPIGHVY	A*30:02	SB (0.0045)	A*30:02
		MAGEA3	EVDPIGHLY	A*01:01	SB (0.0039)	n.a
		MAGEA3	MEVDPIGHLY	B*18:01	SB (0.0055)	n.a
		MAGEB2	KVNPNGHTY GVYAGREHFV	A*30:02	SB (0.0032)	n.a
		MAGEC2	Y	A*30:02	WB (0.5942)	n.a
		TYR	HEAPAFLPW	B*18:01	SB (0.0488)	B*18:01
Mel-3		ETV1$_{p.E455K}$	HPYN**K**GYVY	B*35:01	SB (0.0043)	B*35:01
		GEMIN5$_{p.S1360L}$	STFKELF**L**EK	A*11:01	SB (0.0295)	A*11:01
		MLANA	MPREDAHF	B*35:01	SB (0.4758)	n.a
		PMEL	GTATLRLVK	A*11:01	SB (0.0531)	n.a

[a]The tumor antigen (TA) category is depicted in colors; mutated in turquoise, cancer-germline in orange, and melanoma-associated in purple. [b]Gene symbol, the amino acid change and position in the protein are shown. [c]Mutated amino acids are highlighted in red letters. [d]Predicted HLA allele and [e]binding affinity using NetMHCpan4.0, only the allele with the minimal rank is shown. SB: Strong binders (%-tile rank ≤2); NB: non-binders (%-tile rank >2). [f]HLA restriction element was evaluated experimentally by co-incubating reactive lymphocytes with COS7 cells transfected with plasmids encoding for the individual HLA followed by peptide pulsing. T-cell responses were measured by IFNγ elispot and 4-1BB upregulation. n.a=non-assessed.

Of note, except for patient H&N-1, we detected at least one T-cell reactivity to neoantigen in every patient whose mutations were detected thorugh immunopeptidomics and interrogated. However, we previously detected reactivity in one out of the 5 mutations tested in H&N-1 (RPL14$_{p.H20Y}$), indicating that, although this neoantigen was no longer recognized by the expanded TIL, this mutation was immunogenic. Overall, 12 out of the 33 mutated HLA-I ligands tested were immunogenic. Taking into account the additional immunogenic mutation RPL14$_{p.H20Y}$ from H&N-I, this represents 39% of the total neoantigen candidates tested. Furthermore, three immunogenic peptides derived from cancer-germline antigens and three immunogenic peptides derived from melanoma-associated antigens were also recognized by naturally occurring T cells. In contrast, none of the 507 unique nonC-TL candidates investigat-

ed were able to evoke a recall immune response in the patients studied (Figure 39).

Altogether, these results reveal that even though nonC-TL were frequently detected in TCL, only ligands derived from canonical tumor antigens were recognized by pre-existing tumor-reactive lymphocytes isolated from cancer patients.

2.5. Induction and characterization of T-cell responses to nonC-TL by *in vitro* sensitization

Although we did not detect recall T-cell responses targeting nonC-TL, we reasoned that they could still be immunogenic. If so, these would still represent attractive targets for vaccines or T-cell therapies. To address this question, we sought to first enrich nonC-TL-specific T cells by *in vitro* sensitization (IVS) in a non-autologous setting and then isolate and clone the TCR into PBL to further characterize the targets (Figure 40).

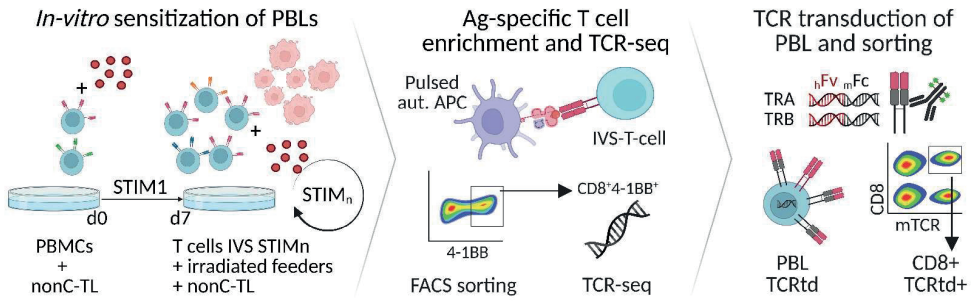


Figure 40. Experimental approach to identify and isolate T cells and TCR targeting nonC-TL through in vitro sensitization. Donor peripheral blood lymphocytes (PBL) were in vitro sensitized (IVS) via three consecutive rounds of stimulation with selected nonC-TL crude peptides predicted to bind to HLA-A*11:01. Reactive T cells were enriched through FACS sorting based on $CD8^+$ $4\text{-}1BB^+$ expression after 20 h co-culture with autologous B cells pulsed with the specific peptides and expanded for 14 days (Ag-specific T cells). Following TCR sequencing, the top 1 alpha-beta pairs were cloned into a retroviral vector used to transduce PBL (PBL TCRtd). The constructed TCR contained mouse constant regions, enabling the detection of the TCRtd lymhpocytes with antibodies specific to the mouse constant region ($TCRtd^+$). PBL TCRtd cells expressing CD8 and mTCRB were FACS sorted to obtain a pure $CD8^+$ transduced populations ($CD8^+TCRtd^+$).

2.5.1. Identification of three immunogenic nonC-TL through IVS

Out of the total 507 nonC-TL identified in the TCL, we selected 168 peptides predicted to bind to HLA-A*11:01 according to NetMHCpan4.0, an allele expressed in 14% of the Caucasian population. Next, PBMC obtained from an HLA-A*11:01 donor were pulsed with 5 pools, each containing up to 35 nonC-TL crude peptides, and cultured with a combination of cytokines, as previously described (Y. Li et al., 2005). After 7 days, cells were re-stimulated with irradiated autologous PBMC pulsed with the same PP and cultured again for 1 week.

After 3 rounds of stimulation, we detected T cells recognizing PP1, PP2, and PP3, as shown by 4-1BB upregulation (Figure 41A). Whereas PP1 and PP2 *in vitro* sensitized T cells (IVS-T cells) exclusively recognized the PP they were stimulated with, T cells sensitized with PP3 were reactive against both PP2 and PP3 (Figure 41A). Deconvolution experiments to determine the specific peptide recognized within each peptide pool (Figure 41B) showed that this cross-reactivity was due to a shared sequence contained in PP2 (PP2 p35, 10mer) and PP3 (PP3p18, 11mer). By contrast, T cells sensitized with PP2 recognized a distinct peptide

sequence from PP2 (PP2p22). Moreover, T cells sensitized with PP1 specifically recognized 3 different peptides contained in PP1 (PP1p19, PP1p33, and PP1p35) (Figure 41B).

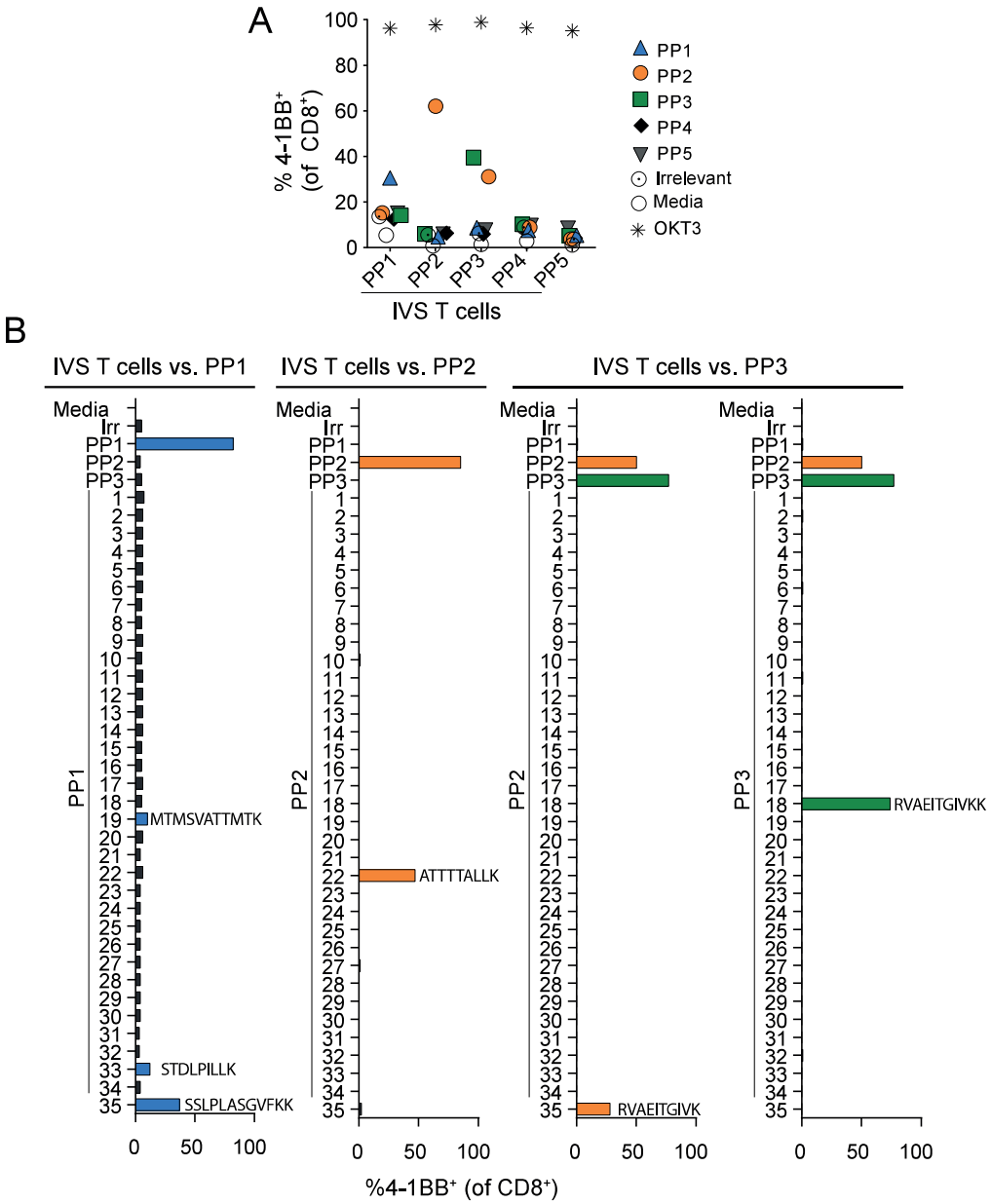


Figure 41. Identification of T cells targeting nonC-TL through in vitro sensitization. (A) Reactivity of IVS-T cells to the PP used for stimulation. PP1-5 stimulated T cells were co-cultured with autologous B cells pulsed with the PP and T-cell activation was evaluated by measuring the upregulation of 4-1BB on $CD8^+$ T cells by FACS. (B) Specific nonC peptide recognized by IVS-T cells. IVS-T cells were co-cultured with B cells pulsed with the individual peptides contained in the corresponding reactive PP. Reactive peptides are highlighted in blue for PP1, orange for PP2, and green for PP3. The amino acid sequences of each reactive peptide are indicated.

In order to characterize the T cells recognizing the 5 nonC-TL, we first enriched the antigen-specific T cells by FACS sorting (Figure 42A). Next, we determined the restriction element by co-culturing enriched T-cell populations with COS7 cells individually transduced with the donor alleles and pulsed with the corresponding peptides (Figure 42B). Of note, only two out of the 5 immunogenic nonC-TL were recognized in the expected allele HLA-A*11:01 (PP1p33 and PP1p35), while two were recognized in HLA-A*68:01 (PP1p19 and PP1p18) and one in HLA-B*35:03 (PP2p22).

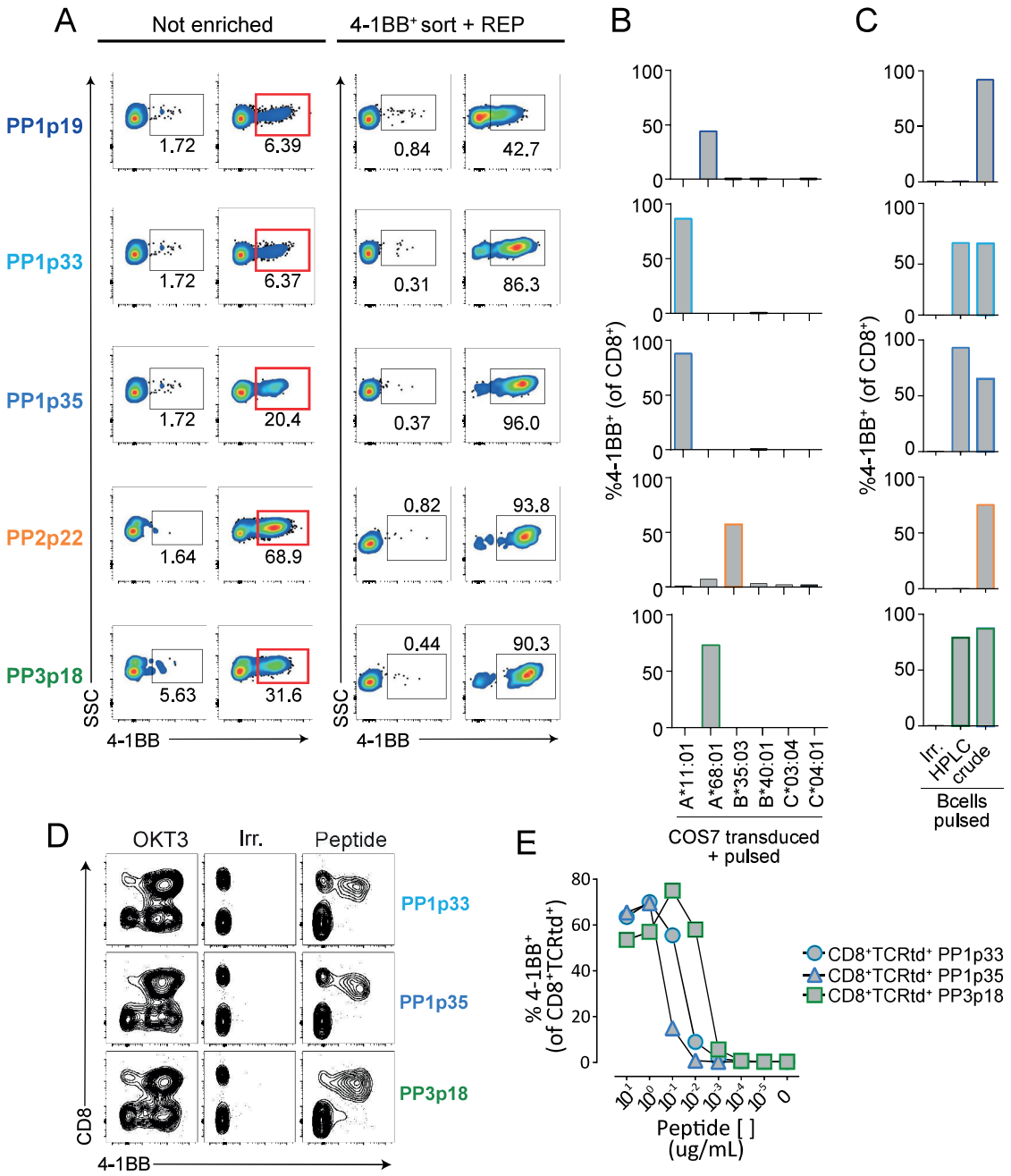


Figure 42. Isolation and characterization of T cells targeting nonC-TL. (A) Enrichment of nonC antigen-specific IVS-T cells. Reactive T cells (Figure 41) were enriched by flow cytometry-based sorting of 4-1BB⁺ cells and expanded for 14 days. FACS plots show the % of CD8⁺ cells expressing 4-1BB after co-culture with B cells pulsed with an irrelevant peptide (negative control) or the corresponding peptide before sorting (left) or after 4-1BB⁺ selection and expansion (right). (B) HLA restriction of immunogenic nonC antigens identified. Restriction element was evaluated by co-culturing enriched T-cell populations with COS7 cells transduced with the donor HLA alleles and pulsed with the corresponding peptides. (C) Reactivity of enriched populations to purified HPLC peptides. Purified HPLC peptides were synthesized and tested in parallel to crude versions to confirm the specificity of the enriched T-cell populations. (D) CD8 co-receptor activation-dependence of PBL transduced with antigen-specific TCR. FACS plots show the expression of 4-1BB by CD8 (gated on CD3⁺mTCR⁺) after 20 h co-culture with B cells pulsed with the corresponding peptides. OKT3 and B cells pulsed with an irrelevant peptide were used as positive and negative controls, respectively. (E) Functional avidity of sorted CD8⁺TCRtd⁺ cells. Sort purified TCR-transduced T cells (CD8⁺ TCRtd⁺) were co-cultured with B cells pulsed with serial dilutions of the corresponding peptide. The experiment was performed with technical duplicates, SD mean is plotted. PP1p19, PP1p33, PP1p35, PP2p22 and PP3p18-specific T cell populations are depicted in dark blue, light blue, blue, orange, and green, respectively.

To further confirm the specific recognition, the peptides were synthesized in HPLC grade to ensure a high quality and purity of the peptides tested (Figure 42C). Surprisingly, only 3 out of the initially 5 immunogenic nonC-TL could be confirmed, indicating that T cells were *in vitro* sensitized against an unknown immunogenic sequence contaminant present in the crude preparation, but not against our peptides of interest. Two of the immunogenic nonC-TL derived from 5'UTR regions of the canonical genes HOXC13 (5'U-HOXC13, PP1p35) and ZKSCAN1 (5'U-ZKSCAN1, PP3p18), and one derived from a non-coding spliced variant of C5orf22 gene (nc-C5orf22, PP1p33).

Next, the TCR αβ locus of the three confirmed antigen-specific IVS-T cells were sequenced by single cell sequencing and α-β pairs of the most frequent T-cell clone clone of each of the reactive populations were cloned into a retroviral vector that was used to transduce PBL. The TCR were constructed using the murine TRA and TRB constant region, to decrease mispairing with the endogenous TCR and enabling us to track the TCR-transduced (TCRtd$^+$) cells using an anti-mTRB antibody. The upregulation of 4-1BB on PBL showed that the isolated TCR required CD8 co-receptor to recognize their cognate antigen (Figure 42D). Given that the frequency of transduced cells was relatively low, we sorted CD8$^+$ cells expressing the transduced TCR to obtain an enriched population of CD8$^+$ T cells expressing the TCR of interest (CD8$^+$ TCRtd$^+$). Moreover, as shown by peptide titration experiments, CD8$^+$TCRtd$^+$ cells displayed a variable functional avidity, PP3p18 TCR exhibiting the highest whilst PP1p35 exhibited the lowest avidity for their corresponding cognate antigens.

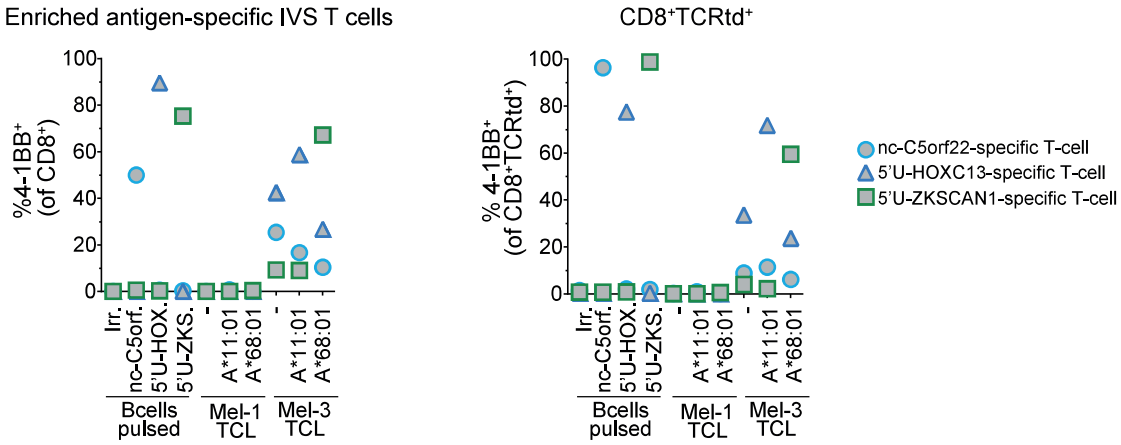


Figure 43. Enriched nonC-TL-specific T cells obtained through IVS and TCR-transduced cells recognize Mel-3 TCL. TCL were left untreated or electroporated with RNA encoding the specified HLA-I alleles and co-cultured with nonC-TL-specific T-cell populations (left) or TCR-transduced lymphocytes (right). After 20 h, the recognition was assessed by measuring 4-1BB upregulation by FACS. Mel-1 TCL was used as a negative control. Donor B cells (donor used for IVS) were pulsed with the specified peptides. Graphs show independent experiments using similar numbers of cells, from 1 to 2e5 TCL were co-cultured with 2e4 T cells per well.

Lastly, we investigated whether the enriched T-cell populations, as well as the isolated TCR, could recognize Mel-3 TCL from which all three nonC-TL were identified by HLA-I immunopeptidomics (Figure 43). Because Mel-3 did not express the HLA-I allele restricting PP3p18 presentation and recognition (HLA-A*68:01, with consensus motif highly similar to HLA-A*11:01 expressed by Mel-3), we decided to artificially express this HLA-I allele. This allowed us to test whether the antigen giving rise to PP3p18 could be expressed, processed

and presented on HLA-A*68:01. We found that both, enriched IVS-T cells and the PBL transduced with their corresponding TCR, were able to recognize Mel-3 TCL when the specific restriction allele was expressed (Figure 43). TCR targeting both PP1p33 and PP1p35 reacted to Mel-3 in absence of exogenous HLA expression, demonstrating that both of these peptides were processed and presented on HLA-A*11:01, naturally expressed by the cell line. Reactivity of PP1p33-specific T cells against Mel-3 TCL was substantially lower compared to PP135 and it could be enhanced, in some instances, by further forcing the expression of HLA-A*11:01. Finally, T cells specific for PP3p18 only recognized Mel-13 when the cell line was transfected with HLA-A*68:01

Altogether these results showed that T cells specifically recognizing three nonC-TL peptides could be detected, isolated, and expanded through IVS. Moreover, the three *in vitro* sensitized T cells as well as the isolated TCR were able to recognize Mel-3, indicating that these nonC-TL are tumor antigens.

2.5.2. The identified immunogenic nonC-TL are shared across patient-derived TCL

Although the three immunogenic nonC-TL were exclusively detected in Mel-3 TCL through immunopeptidomics, we investigated whether those non-mutated antigens could be shared by other patients.

To evaluate the expression and translation of these nonC-TL in a panel of 24 patient-derived TCL, we exploited the high sensitivity of antigen-specific T cells to detect presented HLA peptides (Figure 44). Hence, nonC-TL-specific T cells were co-cultured with TCL artificially expressing the restriction element of interest (i.e., HLA-A*11:01 for 5'U-HOXC13 and nc-C5orf22, and HLA-A*68:01, Figure 43B), irrespective of the endogenous patient HLA alleles.

Strikingly, we found that the three nonC-TL evaluated were frequently presented in patient-derived TCL as evidenced by 4-1BB upregulation on antigen-specific $CD8^+$ T cells (Figure 44) when the corresponding HLA restriction element was expressed. Of note, 5'U-HOXC13 and 5'U-ZKSCAN1 were largely detected by T cells in a number of melanoma TCLs, but also by less immunogenic tumor types such as gastrointestinal cancers (GI) or gynecological malignancies (Gyn). Furthermore, the recognition of several HLA-A*11:01$^+$ TCL without transfecting the allele (Mel-3, -4, -6 and -7), showed that nc-C5orf22 and 5'U-HOXC13 can be naturally processed and presented on HLA-I in multiple patient-derived TCL.

Altogether, these results show that the three immunogenic nonC-TL are shared across tumor types and can be naturally presented, making them attractive targets for off-the-shelf vaccines or T-cell-based therapies.

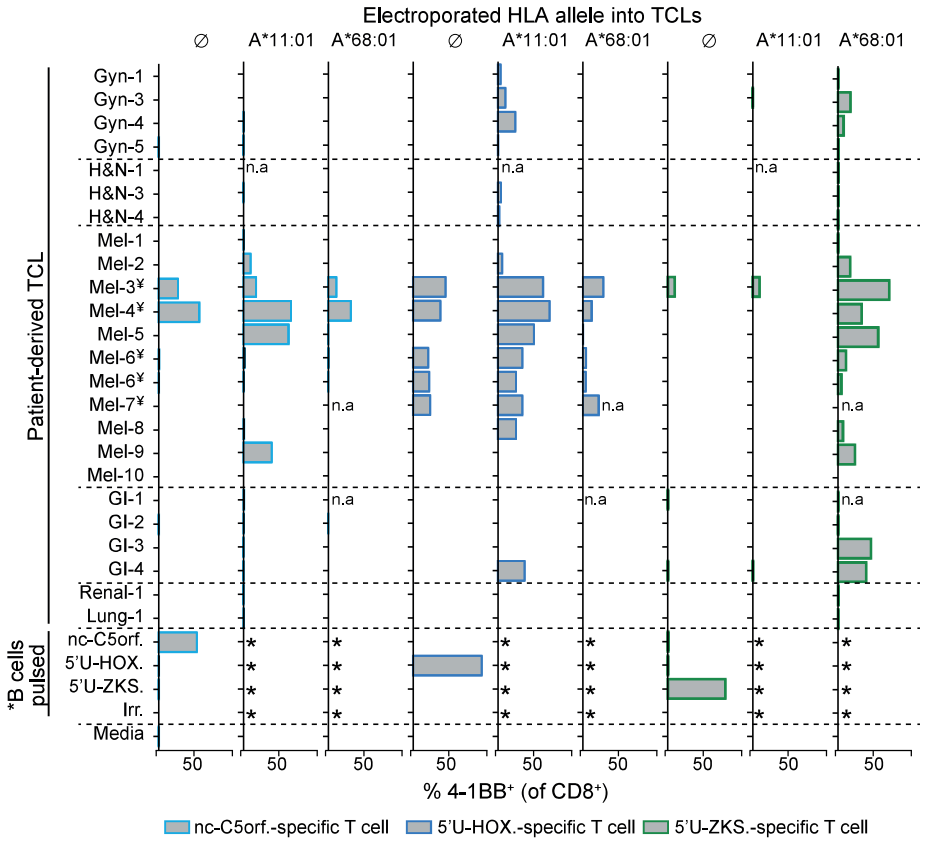


Figure 44. Immunogenic nonC-TL are shared across patient-derived TCL. Indirect analysis of expression and translation of the three immunogenic nonC-TL in a panel of patient-derived TCL through detection of T cell activation of nonC antigen-reactive T cells. TCL were left untreated or electroporated with RNA encoding the specified HLA alleles and co-cultured with nonC-TL-specific T cell populations. After 20 h, T-cell reactivity was assessed by 4-1BB expression on CD8⁺ cells. *B cells from the same donor used for the IVS were not electroporated. ¥TCL naturally expressing HLA-A*11:01. n.a, not assessed.

2.5.3. The immunogenic nonC-TL identified arise from the translation of 5'UTR regions and ncRNA

Next, we focused on the genomic origin of the three immunogenic nonC-TL identified through IVS. Two peptides mapped to 5'UTR regions of the canonical genes HOXC13 (5'U-HOXC13) and ZKSCAN1 (5'U-ZKSCAN1), and one peptide mapped to a non-coding spliced variant of C5orf22C gene (nc-C5orf22) (Figure 45A). The nonC-TL derived from 5'U-HOXC13 and nc-C5orf22 were located close to a 5' in-frame ATG start codon (-3 and -36 NT, respectively) as well as relatively close to an alternative start codon (-21 and -84, respectively). In contrast, no in-frame ATG were found close for 5'U-ZKSCAN1, but two alternative starts at -90 and -96 were detected.

To determine whether the immunogenic nonC-TL could be processed and presented from these non-canonical coding regions, we co-cultured TCR transduced T cells with APC transfected with RNA encoding the longest ORF for each of the nonC-TL and evaluated whether these led to expression, processing and antigen presentation (Figure 44B). As measured by 4-1BB upregulation, APC expressing the nc-C5orf22, 5'U-HOXC13 and 5'U-ZKSCAN1 ORF were recognized at similar levels to APC pulsed with minimal peptides. This demonstrates that these peptides are derived from considerably small ORF, potentially containing only 21, 46, or 49 amino acids at the most.

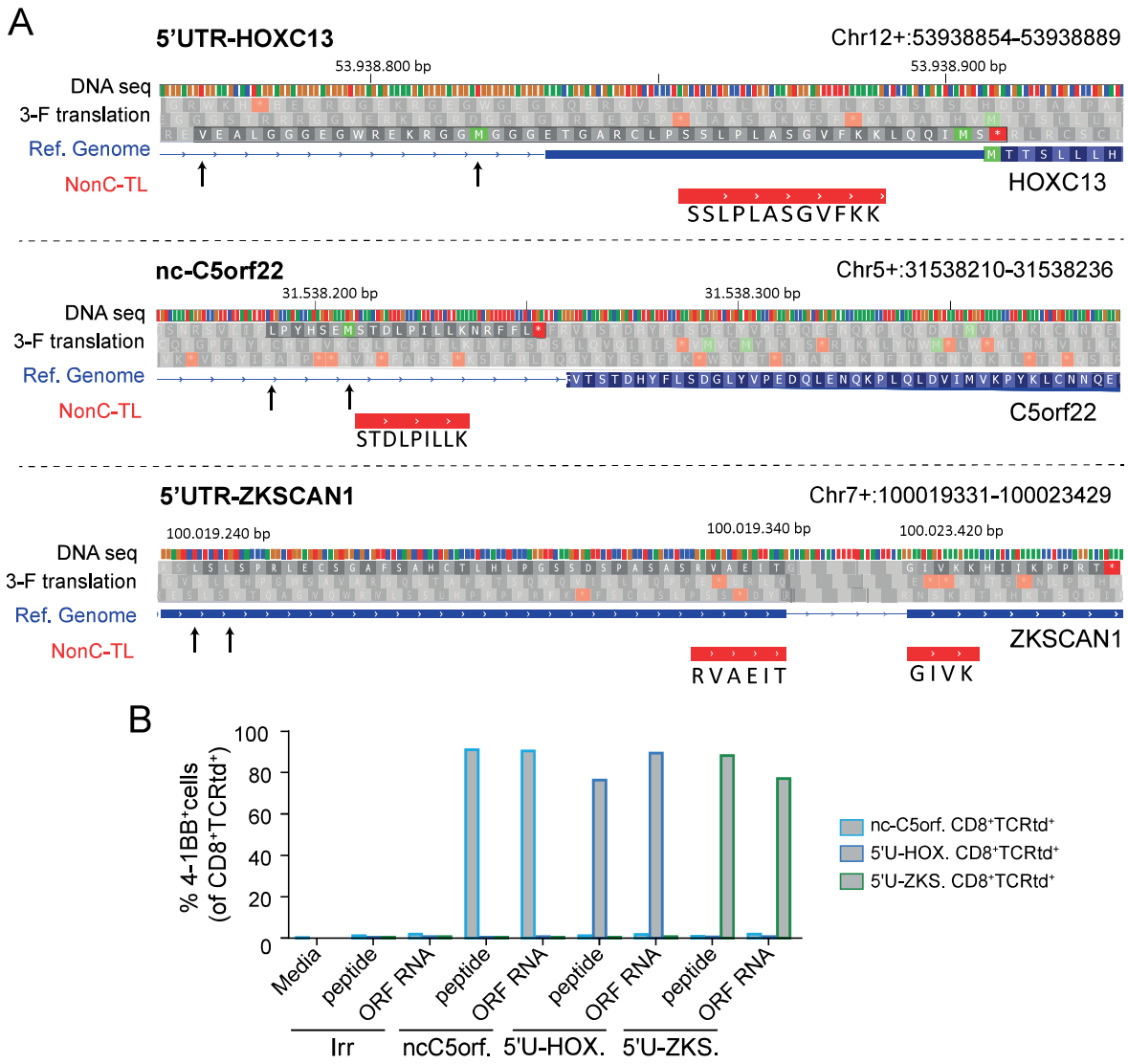


Figure 45. Genomic origin of the three immunogenic nonC-TL identified through IVS. (A) Genomic location of nonC-TL. Annotated translated sequences are depicted in blue and the nonC-TL sequence of interest are shown in red. The possible start codons are marked with an arrow. The predicted ORF is highlighted. Image modified from IGV viewer.(B) Recognition of the predicted ORF encoding the immunogenic nonC-TL by TCR-td cells. The DNA sequences of the predicted open reading frames (ORF) for each immunogenic nonC-TL were cloned into pcDNA3.1 without any modification except for a Kozak sequence upstream, and the RNA was generated by *in vitro* transcription. PBL transduced with the antigen-specific TCR specified were co-cultured with B cells either electroporated with the RNA encoding the predicted or an irrelevant ORF, or pulsed with the nonC-TL or an irrelevant peptide. After 20 h, T-cell activation was evaluated by measuring 4-1BB upregulation by flow-cytometry on live $CD8^+TCRtd^+$ cells.

Furthermore, we attempted to knock out these non-canonical coding regions by CRISPR Cas9 using customized single guide RNAs (sgRNA) specifically targeting the sequences encoding the peptides or truncating the ORF at 5′ region. However, we could only design effective guides for 5′U-HOXC13 (n=5), since the sequences surrounding the genomic location for nc-C5orf22 and 5′U-ZKSCAN1 were not amenable for effective targeted gene knockout. Of note, the dramatic loss of recognition of TCL transduced with Cas9-sgRNA targeting the 5′U-HOXC13 unequivocally confirmed the correct genomic location of this immunogenic nonC-TL (Figure 46).

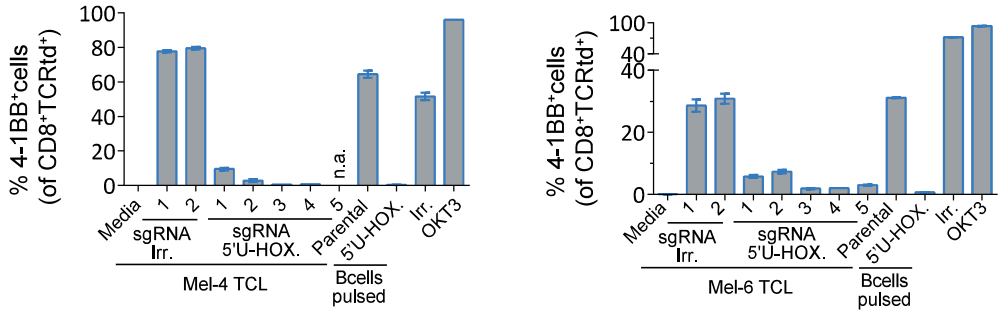


Figure 46. Loss of recognition of TCL transfected with Cas9-sgRNA targeting predicted 5′U-HOXC13. Patient-derived tumor cell lines Mel-4 (right) and Mel-6 (left) were transfected with lentiviruses encoding for Cas9 together with an irrelevant sgRNA or different sgRNA specifically targeting the genomic location of the immunogenic nonC-TL peptide 5′U-HOXC13. Following puromycin selection, TCL were co-cultured with PBL transduced with the 5′U-HOXC13-specific TCR-transduced cells and the activation was measured after 20 h by flow cytometry. Plotted cells were gated on live CD8+TCRtd+ T cells. The experiment was performed with technical duplicates, mean and SD are plotted.

2.5.4. Tumor-specificity of the canonical gene encoding the immunogenic nonC-TL

A potential concern regarding the therapeutic use of nonC-TL is whether they are actually tumor-specific. To address this question, we first analyzed the RNA expression of the canonical genes encoding the three immunogenic nonC-TL in several solid tumors and matched healthy tissues from publicly available data. Their expression pattern was compared to PMEL and MLANA as examples of melanoma-associated antigens, and MAGEA3 and MAGEC2 as representative cancer-germline antigens (Figure 47). Whereas C5orf22 and ZKSCAN1 canonical genes displayed variable, but ubiquitous expression among tissues, the expression of the canonical HOXC13 gene in healthy tissues appeared to be restricted to melanocytes, resembling the expression pattern of MLANA. (Figure 47).

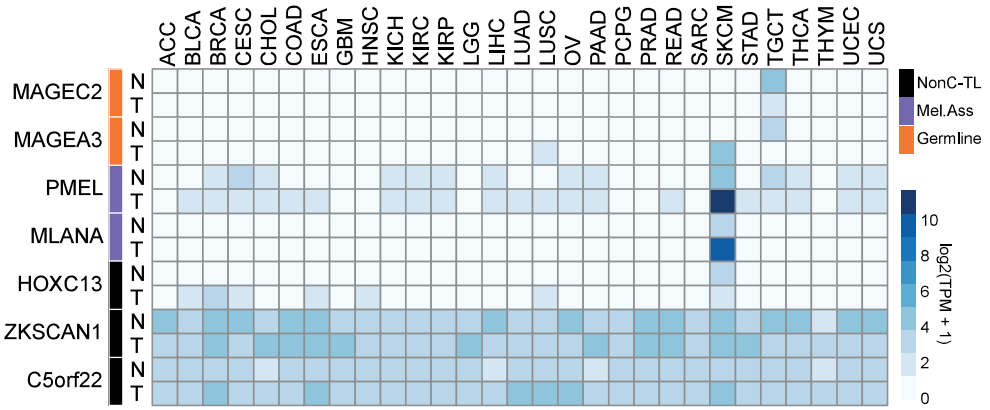


Figure 47. Tumor specificity of the canonical transcripts encoding immunogenic nonC-TL. RNA expression analysis in tumors (T) and matched healthy tissues (N) of the canonical genes for the three immunogenic nonC-TL compared to melanoma-associated antigens (PMEL and MLANA) and cancer germline antigens (MAGEA3 and MAGEC2). TCGA and GTEX data were obtained from GEPIA. RNA levels are expressed as Log2 (TPM+1), the density of color in each block represents the median expression value of a gene in a given tissue, normalized by the maximum median expression value across all blocks. Abbreviation of tumor tumor types found in Appendix.

These results suggested that the identified immunogenic nonC-TL might not be tumor-specific. However, gene expression analysis measures the level of mRNA transcript and is not able to distinguish whether nonC translation events occur. On the contrary, the healthy

immunopeptidome data that was employed to select for nonC-TL derived from ORF absent in non-malignant cells suggested that those peptides might be tumor-specific. Although LC-MS/MS is less sensitive, this technique evaluates the translation since the amino acid sequence is directly identified, as opposed to RNAseq data.

2.5.5. Empirical evaluation of the tumor specificity of nonC-TL using human healthy cells

To overcome the limitations above mentioned and gain insights into the applicability of using nonC-TL as targets for cancer immunotherapy, we empirically evaluated the expression and translation of the immunogenic nonC-TL in several human healthy cell types by exploiting the ability of antigen-specific T cells to detect peptides presented on HLA with high sensitivity. Additionally, because of the expression pattern observed for the canonical HOXC13 gene and as an example of a TCR tested in clinical trials, we included the HLA-A*02:01-restricted TCR DMF5. This TCR recognizes the melanoma-associated antigen MART-1_{27-35} (encoded by MLANA gene) and has shown objective tumor regression as well autoimmunity due to on target off tumor recognition of melanocytes in the skin, eyes and ears (Johnson et al. 2009).

To this end, we first transduced PBL with retroviruses encoding nonC-TL-specific TCR as well as the MART-1-specific TCR DMF5. Because the nonC-TL specific TCR were CD8 dependent as shown in section 2.5.1 as opposed to MART-1 (not shown, published elsewhere), the CD8$^+$TCRtd$^+$ were sorted as mentioned before, to obtain a homogeneous and functional population. Lastly, the four antigen-specific CD8$^+$TCRtd$^+$ cells were co-cultured with human healthy melanocytes, cardiac myocytes, renal epithelial, and fibroblast, as well as some of the melanoma cell lines tested in section 2.5.3. Similarly to the experiment previously shown where we evaluated the presence of nonC-TL in patient-derived TCL (Figure 48), the HLA-I alleles were artificially expressed in target cells and the recognition was evaluated by measuring 4-1BB upregulation in CD8$^+$TCRtd$^+$ cells by flow cytometry (Figure 48).

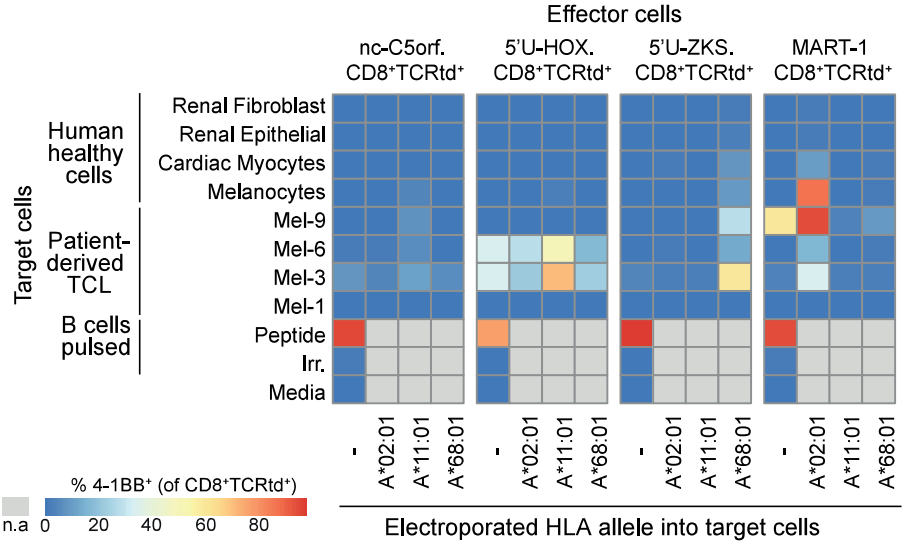


Figure 48. Tumor specific of the three immunogenic nonC-TL. Expression and translation of nonC-TL in healthy human and selected tumor cells was evaluated indirectly by co-culturing control target cells or target cells electroporated with RNA encoding the specified HLA-I alleles with nonC-TL CD8$^+$TCRtd$^+$ T cells After 20 h, T-cell activation was assessed by 4-1BB expression on CD8$^+$ cells. *B cells were not electroporated. n.a, non-assessed.

As expected, MART-1 TCRtd$^+$ cells strongly recognized human healthy melanocytes as well as most of the patient-derived melanoma cell lines tested, when electroporated with the restriction allele HLA-A*02:01. Surprisingly, cardiac myocytes were also recognized by MART-1 TCRtd$^+$ cells, albeit the reactivity was substantially lower. Similar results were observed for 5'U-ZKSCAN1 TCRtd$^+$ cells when target cells were electroporated with the relevant allele HLA-A*68:01, although the recognition was generally lower. On the contrary, 5'U-HOXC13 and nc-C5orf22 TCRtd$^+$ cells did not recognize cardiac myocytes and barely recognized melanocytes when electroporated with the corresponding allele HLA-A*11:01. These results revealed that the nonC 5'U-HOXC13 peptide was the antigen displaying the highest tumor-specificity followed by nc-C5orf22, 5'U-ZKSCAN1, and MART-1. Although we cannot rule out the possibility that they could be expressed and translated in other cells, our findings suggest that the aberrant translation giving rise to the nonC-TL studied occurs preferentially in tumor cells rather than in normal cells.

Overall, our results indicate that these nonC-TL, particularly 5'U-HOXC13, are promising alternative tumor antigens to neoantigens, cancer-germline antigens or melanoma-associated antigens not only because they can be naturally presented, but also because they are immunogenic and broadly expressed across diverse tumor types but not, or at very low levels, in healthy cells.

Discussion

The identification of tumor antigens is critical to better understand the immune response to tumors and for the development of more effective cancer immunotherapies. Recent evidence suggests that peptides derived from nonC proteins are specifically presented on HLA-I by tumor cells and contribute to tumor immune surveillance.

In this thesis, we aimed to determine the personalized repertoire of tumor antigens recognized by T cells in nine patients with distinct histological cancer types, including canonical tumor antigens, but also expanding the scope to non-canonical proteins.

We found that, although nonC-TL identified by proteogenomics represents the most abundant source of candidate tumor antigens, as compared to neoantigens, cancer-germline, and melanoma-associated antigens, naturally occurring tumor-reactive lymphocytes consistently display preferential recognition of neoantigens. Nonetheless, we showed that nonC-TL can elicit *de novo* T-cell responses via *in vitro* sensitization of donor PBL. We identified TCR specific to three nonC-TL, two of which derived from 5' UTR regions and one derived from a non-coding RNA spliced variant. Importantly, we found that these immunogenic nonC-TL display therapeutic potential, as they were expressed across diverse tumor types, but were barely detected in healthy cells.

However, despite the extensive immunological screening including nonC-TL, neoantigens, cancer-germline, and melanoma-associated antigens, the specificity of many tumor-reactive lymphocytes remains unknown and warrants further investigation.

HLA-I immunopeptidomics narrows down the list of candidates and captures immunogenic tumor antigens

In the first part of the thesis, we evaluated whether tumor-reactive lymphocytes recognize conventional tumor antigens. To this end, we used three different methods to identify tumor antigens: personalized neoantigen screen with TMG, *in silico* minimal epitope prediction, and HLA-I immunopeptidomics.

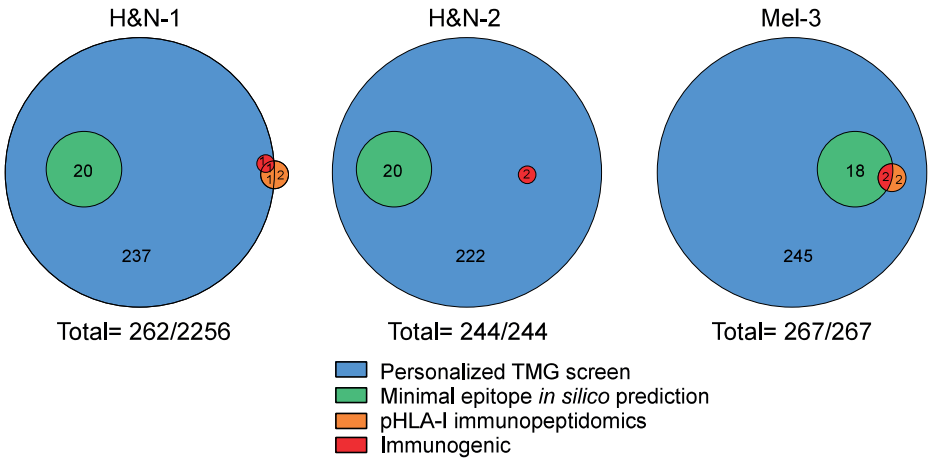


Figure 49. Candidate and immunogenic NSM identified by currently used methods. The number of NSM screened by patient is represented in circles and depicted in different colors depending on the method used. The number of immunogenic NSM identified are depicted in red regardless of the method used. Circles within larger circles indicate that the NSM was investigated through different methods. The total number of NSM tested and identified by WES are shown on the bottom.

We detected tumor-reactive T cells recognizing neoantigens derived from NSM in all three patients. However, the overlap of the identified immunogenic variants between the three methods strikingly varied among patients (Figure 49). Whereas for Mel-3 all approaches led to the detection of the same two neoantigens, for H&N-2 we could only detect immunogenic variants through the personalized neoantigen screen with TMG. In fact, no candidate neoantigens presented on HLA-I were detected through immunopeptidomics in H&N-2 TCL. For H&N-1, one of the two neoantigens detected through TMG was also captured by TCL immunopeptidomics, but none of them were included in the top 20 mutated minimal epitope selected by *in silico* prediction.

The lack of identification of some immunogenic peptides using immunopeptidomics can be explained by the fact that T cells are far more sensitive than current MS instruments, as it has been postulated that a single to a few hundred pHLA-I per cell suffice to trigger an immune response (Becker & Riemer, 2022). IIn addition, it is known that some peptides are not well detected by MS methods due to inefficient ionization and fragmentation which depends on the amino acid composition (Faridi et al., 2018). Alternatively, some of the tumor cell lines have mutations in HLA or could harbor alterations in the antigen presentation machinery that could limit antigen presentation on HLA-I. Nonetheless, the advantage of immunopeptidomics is that it substantially narrows down the list of candidates to screen to those that are presented on HLA-I and the number of false-positives is reduced On the contrary, despite HLA-I binding prediction has substantially improved over the last years, they typically lead to an long list of candidates, as shown in Figure 21B.

Despite the extensive immunological screening including conventional tumor antigens, we could not determine the specific targets recognized by the majority of tumor-reactive lymphocytes. Nonetheless, it is important to mention that we mainly focused on neoantigens and cancer-germline or tissue differentiation antigens which are frequently detected in cancer patients, but we did not include other self-antigens that might also play a role in cancer immune surveillance. In the second part of the thesis, we sought to explore non-canonical HLA-I ligands as an alternative source of tumor antigens. Because including non-canonical transcription and translation massively expands the proteome leading to an overwhelming number of candidates to screen, we decided to exploit immunopeptidomics to select the peptides presented on HLA-I, even at the expense of probably missing some immunogenic tumor antigens.

NonC-TL are frequently identified in patient-derived TCL through Peptide-PRISM

In the second part of the thesis, we used Peptide-PRISM to interrogate the peptides eluted from pHLA-I in the 9 patient-derived TCL. This proteogenomics pipeline developed by F.Erhard and colleagues in 2020 (Erhard et al., 2020), is independent of RNA-seq and Ribo-seq, but enables the detection of HLA-I ligands potentially originating from any region of the genome, including CDS, UTR, off-frame, ncRNA, intronic and intergenic regions (Figure 31). Similar to previous immunopeptidomics-based studies (Chong et al., n.d.; Erhard et al., 2020;

Laumont et al., 2016; Ouspenskaia et al., 2022; Ruiz Cuevas et al., 2021; Scull et al., 2021; Xiang et al., 2022; Zhao et al., 2020), we observed that nonC HLA-I ligands were frequently found in the 9 patient-derived TCL studied, ranging from 0.5% to 5.4% of the total peptides presented (Figure 32A).

Because we were interested in tumor-specific candidates we used healthy immunopeptidome data to exclude peptides derived from ORF present in non-malignant cells as an alternative approach to the above-mentioned studies. As a result, 61.5% of the nonC HLA-I-ligands detected were considered enriched in tumors were referred to as nonC-TL (Figure 32B).

In accordance with previously described immunopeptidomics-based studies coupled with Ribo-seq data (Chong et al., 2020; Ouspenskaia et al., 2022; Ruiz Cuevas et al., 2021), we observed that nonC-TL mainly originated from upstream regions of canonical ORF. More specifically, 5'UTR and off-frame were the major sources, but ncRNA also represented a substantial part (Figures 33A). This differs from other immunopeptidomics-based studies where RNA-seq data from healthy tissues was used to select tumor antigen candidates (Cleyle et al., 2022; Zhao et al., 2020). This discrepancy could be attributed to the fact that transcriptomic data does not distinguish whether aberrant translation events occur from a canonical transcript and, therefore, the contribution of this source of nonC HLA-I ligands might be underestimated.

One limitation of our study lies in the nature of the proteogenomics approach used combined with the stringent and uniform 1% FDR threshold set to select the tumor antigen candidates. Although this pipeline potentially detects peptides originating from any region of the genome, the FDR is calculated in a stratified mixture model that is probably detrimental for the categories with an increased search space, for example intergenic regions. Moreover, because we did not focus on a particular type of nonC source, but rather aimed to study the repertoire of peptides presented in an unbiased manner, some described nonC sources were not properly or not interrogated at all. For instance, we detected few peptides derived from endogenous retroelements (ERE) compared to reported immunopeptidomics-based studies focused on the detection of this particular source (Bonté et al., 2022; Kong et al., 2019; Larouche et al., 2020). Nonetheless, we detected at least one peptide (SSFSTLASLDK) derived from an annotated transcript overlapping with long-terminal repeat elements (LINC02099-201) previously described (Attig et al., 2019). In this study focused on LTR retroelements, the RNA expression of LINC02099-201 was considered cancer-specific and the peptide was found to be naturally presented in melanoma biopsies, although they did not evaluate its immunogenicity. We did not detect T cells targeting this peptide, neither pre-existing nor induced by IVS. Moreover, RNA editing (Zhang et al., 2018) or peptide splicing (Liepe J et al., 2016) were not considered in this study. Although the contribution to the immunopeptidome of the latter is controversial and has generated a lot of debate in the field (Mylonas et al., 2018), (see introduction 4.1.4).

NonC ligands are preferentially presented on HLA-A*11:01 and HLA-A*03:01

In agreement with previous immunopeptidomics-based studies (Erhard et al., 2020; Laumont et al., 2016), we observed that the presentation of nonC peptides is biased toward HLA-A*11:01 and HLA-A*03:01, alleles that display a similar motif containing basic residues at p9 (Figures 33B-C). In fact, we noticed that patients harboring any of those alleles contained a higher percentage of nonC peptides regardless of the tumor histology (Appendix).

F.Erhard et al. suggested that this might be attributed to a differential processing mechanism of nonC, such as a protease with basic amino acid specificity (Erhard et al., 2020). They compared the percentage of HLA-I peptides processed or not processed at C termini (presenting a stop codon immediately downstream of the peptide), derived from canonical or canonical protein. However, although it cannot be completely ruled out, their analyses did not support that hypothesis as they did not find significant differences between the canonical and non-canonical peptides. Laumont et al. hypothesized that processing of nonC HLA-I ligands may be proteasome-independent since nonC ORF were shorter and the aminoacid usage around the C termini (Ct) was different from that of canonical HLA-I ligands (Laumont et al., 2016). Although this is a possibility, no further analysis nor experiments supporting this hypothesis were shown.

We wondered whether this bias could be the result of an MS-identification artifact derived from the "tryptic" characteristics (i.e., K/R at Ct, Figure 50A) of peptides bound to these alleles. On one hand, tryptic-like peptides are generally more prone to the LC MS/MS detection (Faridi, Purcell, et al., 2018). On the other hand, it could be expected that nonC peptides

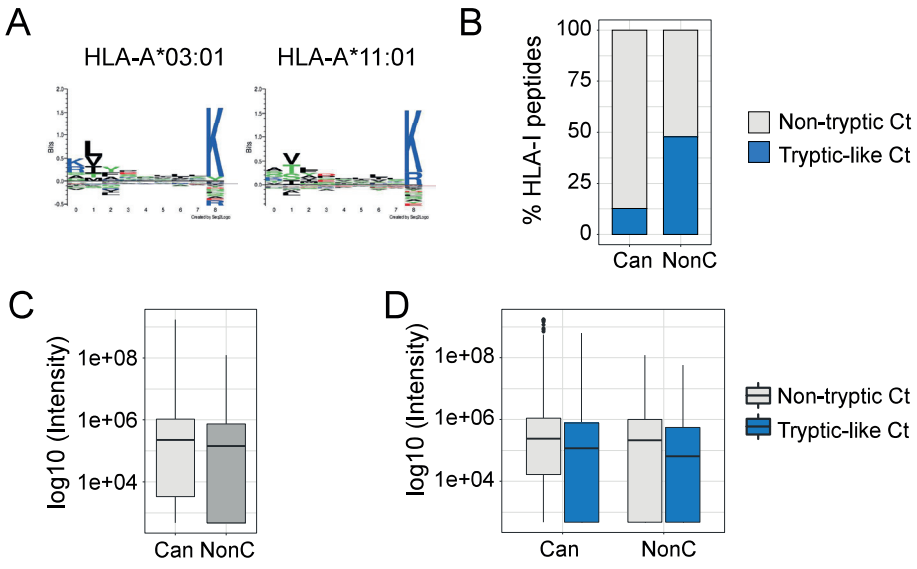


Figure 50. Peptides displaying tryptic-like C-termini are identified from lower intensity signals in both canonical and non-canonical groups. (A) Peptide binding motif of the alleles HLA-A*11:01 and A*03:01 according to NetMHCPan4.0. (B) Percentatge of HLA-I ligands containing K or R at C-termini (tryptic-like Ct) or any other amino acid (non-tryptic Ct) by group. (C) Mean intensity of HLA-I ligands derived from canonical or non-canonical proteins represented in log10 scale. Peptides with intensity values=0 were replaced by the minimal intensity observed >0. (D) Same analysis as in C grouping by tryptic-like or non-tryptic Ct characteristics. Canonical (Can), no-canonical (NonC).

are found at a lower abundance compared to peptides derived from canonical proteins. Taken together, this could result on an over representation of tryptic-like peptides in the nonC peptides observed.

Although the percentage of peptides compatible with tryptic Ct traits was clearly higher for peptides derived from nonC proteins (Figure 50B), we could not draw a clear conclusion from the analysis performed. We observed an overall slightly lower peptide intensity of nonC peptides compared to peptides derived from canonical proteins (Figure 50C). However, when the peptides were grouped according to the Ct characteristics, we also observed differences within the canonical group. As shown in Figure 50D, peptides displaying tryptic-like Ct are identified from lower intensity signals, in both the canonical and non-canonical groups. This data indicates that HLA-I peptides matching tryptic characteristics at the Ct, regardless of the canonical or non-canonical origin, require lower amounts to be detected by LC-MS as compared to non-tryptic Ct, in agreement with their expected higher MS-detectability. Although this data is compatible with our hypothesis, it does not imply that the preferential presentation of nonC HLA-I ligands on HLA-A*03:01 and HLA-A*11:01 can be explained solely by this technical reason and does not rule out any of the other proposed explanations described above.

It has been shown that some polymorphisms in HLA-I class I molecules influence their interactions with components of the antigen processing and presentation pathway (Aflalo & Boyle, 2021). For instance, there is a significant variability in the dependency of TAP, tapasin or ERAP depending on the HLA-I allotype, which is particularly relevant for some HLA-B alleles. TAP independency is associated to the thermal stability of the pHLA-I complex, as well as the ability to present peptides derived from transmembrane domains that do not require TAP to enter the ER (Geng et al., 2018). Similarly, tapasin-independent allotypes exhibit reduced aggregation and bind peptides more readily when refolded *in vitro* compared with tapasin-dependent allotypes (Rizvi et al., 2014). One study investigated the contribution of proteasome inhibition to the generation of HLA-I-associated peptides and found striking differences depending on the HLA allotype(Luckey et al., 2001). They showed that those HLA alleles with a preference for basic residues including HLA-A3.1, HLA-A11, and HLA-68 were less proteasome-dependent. Interestingly, in our study, nonC peptides were preferentially presented on HLA-A*03:01 and HLA-A*11:01. Although this data perfectly supports the hypothesis proposed by Laumont et al. and it seems a plausible explanation, more experiments should be performed to determine whether this allele preference is mainly due to a proteasome-independent route or a yet unknown biological mechanism.

Only one mutated nonC-TL was detected through Peptide-PRISM

Due to the high tumor specificity and immunogenicity but generally private nature of neoantigens derived from NSM (Parkhurst et al., 2019), we also wanted to explore the existence of canonical as well as non-canonical neoantigens arising from tumor-specific mutations in the 9 patient-derived TCL using Peptide Prism.

Even if we did not interrogate all the mutations potentially encoded in the entire genome, since we used WES to identify the tumor-specific NSM, we reasoned that at least a fraction could be detected, including NSM in the off-frame translation of coding regions, some ncRNA, and part of the UTR to a lesser extent. However, most of the identified mutated HLA-I ligands mapped to canonical or annotated coding regions (n= 33, Figures 34C) and only one nonC peptide derived from a 5'UTR containing a mutation was detected for Gyn-3 (Figure 33A). In a recent study, whole-genome sequencing (WGS) of tumor and matched healthy tissue was used in combination with Ribo-seq profiling to estimate the potential contribution of novel ORF with somatic mutations to the neoantigen repertoire (Ouspenskaia et al., 2022). Although 22% of the potentially translated NSM were exclusively found in newly described ORF according to Ribo-seq data, none of them were confirmed to be naturally presented on HLA-I.

Nonetheless, we believe that non-exonic mutations could still be an important source of neoantigens as it has been shown that the vast majority of somatic mutations in cancer accumulate in non-coding regions (Piraino & Furney, 2016). Most of them are passenger mutations, but driver non-coding mutations have also been recently described (Rheinbay et al., 2020). Although the identification and, more importantly, the selection of relevant mutations in non-coding regions is complex since the number of candidates drastically increases, we believe that WGS combined with immunopeptidomics to detect mutated nonC-TL warrants further examination.

NonC-TL were the main source of candidate tumor antigens but pre-existing T-cell responses to NonC-TL were not detected

In this work, we studied in detail the repertoire of presented tumor antigen candidates in 9 patient-derived TCL. Consistent with previous immunopeptidomics studies (Cleyle et al., 2022; Laumont et al., 2018; Zhao et al., 2020), we showed that nonC-TL (n=507) outnumber the HLA-I ligands derived from conventional tumor antigens such as mutations (n=33), cancer-germline antigens (n=36) and melanoma-associated antigens (n=24) (Figure 34A). As mentioned before, the number of HLA-I ligands derived from NSM was relatively low compared to the NSM identified by WES (Figure 34C). In this regard, our results are in line with previous immunopeptidomics-based studies where few presented mutations are typically detected (Bassani-Sternberg et al., 2016; Kalaora et al., 2016, 2018; Newey et al., 2019). Nonetheless, our data adds a considerable number of mutated HLA-I peptides to those identified through immunopeptidomics so far, underscoring the great performance of Peptide-PRISM at detecting neoantigen candidates.

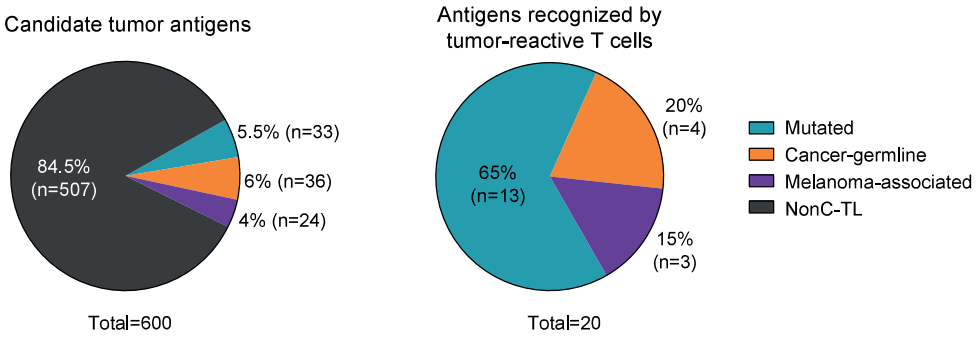


Figure 51. NonC-TL are the main source of candidate tumor antigens identified by proteogenomics but tumor-reactive T cells preferentially recognize neoantigens derived from NSM. The percentage and absolute number of candidate tumor antigens presented on HLA-I identified through proteogenomics in the 9 patient-derived TCL are depicted by antigen category (left). The absolute number and percentage of HLA-I ligands recognized by tumor-reactive lymphocytes are depicted by antigen category (right)

Our results show that although HLA-I ligands derived from nonC proteins were frequently presented in patient-derived TCL and were, apparently, the main source of candidate tumor antigens, pre-existing T-cell responses targeting nonC-TL were not detected in the 9 patients studied (Figure 51). In contrast, nearly 65% of the tumor antigens recognized by T cells were derived from mutations (n=13), 20% from cancer-germline (n=4) and 15% from melanoma-associated antigens (n=3) (Figure 39 and 51). Moreover, the isolation and characterization of the reactive T cells showed that these antigen-specific T cells recognize the autologous tumor cell line (Appendix), which ultimately demonstrates that those peptides are genuine tumor antigens.

It is worth mentioning that in the second part of the thesis we probably underestimated the number of tumor antigen candidates derived from cancer-germline antigens, since we were very stringent selecting genes whose expression is truly restricted to testis according to the human protein atlas. Nonetheless, in the first part of the thesis, we were less restrictive as we selected all genes considered cancer-testis antigens according to the CTPedia database (Almeida et al., 2009) and yet, no responses against any peptide derived from those antigens were detected in H&N-1, H&N-2, or Mel-3 either. Similarly, we only selected defined melanoma-associated antigens since spontaneous T-cell responses to differentiation antigens have been well documented only in patients with melanoma (Coulie et al., 2014). However, we cannot rule out the possibility that other interesting tissue-specific antigens for Gyn or H&N cancer patients might have been overlooked.

Reported immunogenic nonC antigens

To our knowledge, the existence and frequency of naturally occurring T cells targeting nonC-TL compared to clinically relevant antigens such as neoantigens, cancer-germline, or melanoma-associated antigens has not previously been investigated in such detail. Most previous studies solely rely on the identification of nonC peptides presented on HLA-I but did not investigate the immunogenicity in patients (Erhard et al., 2020; Ouspenskaia et al., 2022; Ruiz Cuevas et al., 2021; Scull et al., 2021; Wang et al., 2021; Xiang et al., 2022; Zhao et al.,

2020). The identification of recall T-cell responses against nonC peptides have been generally described in case reports by using cDNA libraries (Garcia-Garijo et al., 2019) or immunopeptidomics-based studies focused on a particular type of nonC source (Zhang et al., 2018). Moreover, in most immunopeptidomics studies the immunogenicity has been typically studied through *in vitro* sensitization of PBL from healthy donors which detects naïve rather than antigen-experienced T cells (Bartok et al., 2021; Bonté et al., 2022; Laumont et al., 2016; Nelde et al., 2022), or immunized mouse models (Laumont et al., 2018). There are, however, two studies in particular, where T-cell responses against nonC peptides were investigated in parallel to classic tumor antigens in cancer patients.

In 2012, Andersen and colleagues (Andersen et al., 2012) sought to dissect the T-cell antigen specificity of TIL in a cohort of 19 melanoma patients by using a "universal" pMHC multimer library of melanoma-associated antigens. They used a combination of available databases to include T-cell epitopes described in the literature with a defined minimal epitope sequence, HLA restriction, and evidence that the peptide was processed and presented on tumor cells. Similar to us, they divided the antigen candidates (n=230) into 4 different categories: differentiation antigens (n=17), cancer-germline (n=50), mutation antigens (n=39), and overexpressed antigens (n=119). Importantly, although they did not categorize nonC as an independent group, 17 peptides derived from nonC proteins, previously shown to be immunogenic, were included. In total, they detected T-cell responses against 18 different melanoma-associated T-cell epitopes and found that differentiation (n=8) and cancer-germline antigens (n=5) dominated the response. On the contrary, only one neoantigen derived from a mutation in a driver gene, and 4 out of the 119 overexpressed antigens were recognized by T cells. In our study, we did not interrogate the category of overexpressed antigens since we consider that they are very often not tumor-specific, highly variable depending on the tumor type and it is overall very difficult to select relevant candidates. Indeed, although they did detect responses against overexpressed antigens (4 out of the 119 tested), the percentage of reactive vs. tested indicates that they are not as immunogenic as differentiation or cancer-germline antigens in patients. Interestingly, 3 out of the 4 responses to overexpressed antigens and 2 out of the 5 specific for cancer-germline antigens were derived from alternative ORF. These findings suggested that nonC peptides might be more immunogenic than peptides derived from canonical proteins of the same group. However, taking into account that most immunogenic mutations arise from private NSM (Parkhurst et al., 2019), these were not properly interrogated in this study as their identification was not personalized. According to our results, neoantigens derived from NSM are preferentially recognized by tumor-reactive TIL (Figure 39 and 51), hence we believe that this study missed a highly relevant source of tumor antigens. Moreover, this study focused on detecting reactivity to nonC peptides previously identified through screening of tumor cDNA libraries, but they did not systematically interrogate the personalized nonC HLA-I ligandome.

Only one proteogenomics study evaluated T-cell responses against nonC HLA-I ligands identified from patient-derived TCL and tumor samples of melanoma and lung cancer patients (Chong et al., 2020). They used RNA-seq to identify expressed non-coding genes, such as

lncRNAs, pseudogenes as well as transposable elements, and other non-protein-coding genes. All three forward ORF were subsequently translated *in silico* and concatenated to personalized canonical proteome references containing allelic variant information from patient tumor exome data. Moreover, for one patient-derived TCL, they additionally performed Ribo-seq to generate a translatome-based database for MS search. This technique reduced the size of the search space and was used as an independent discovery method to identify additional nonC HLA-I ligands, such as UTR or alternative ORF. Importantly, the authors investigated T-cell recognition to the nonC HLA-I ligands identified (n=571) as well as other relevant tumor antigens derived from melanoma-associated antigens and cancer germline antigens (n=215). They detected functional T-cell responses to 5 peptides derived from melanocyte differentiation antigens (PMEL, TYR, and TYRP) in autologous TIL and IVS PBL. In addition, they also reported reactivity against one nonC peptide, although the recognition was weak compared to the above-mentioned differentiation antigens and T cells were not isolated to further confirm and characterize the tumor recognition. This immunogenic nonC HLA-I ligand was derived from a downstream ORF of a canonical gene (ABCB5) that was identified through Ribo-seq. Unlike them, we did not detect reactivity against any of the 507 nonC-TL candidates tested. Nonetheless, we believe that our results are still in accordance with their findings since the percentage of reactive peptides clearly favored melanocyte differentiation antigens, suggesting that nonC HLA-I ligands are not as immunogenic.

NonC peptides capable of mounting T-cell responses in cancer patients have been identified through cDNA libraries. Of note, many of them were derived from genomic regions coding for cancer-germline antigens (Aarnoudse et al., 1999; Topalian et al., 1998)or well-defined melanoma-associated antigens (Lupetti et al., 1998; P. Robbins et al., 1997; Rong-Fu Wang et al., 1996). With the exception of one reported nonC peptide derived from intron retention containing a mutation (Coulie et al., 1995), the remaining reported T-cell responses targeted genes displaying a pattern of overexpression in cancer and, with few exceptions (Guilloux et al., 1996), they were also identified in normal cells. (Harada et al., 2001; Poullion et al., 1999; S. A. Rosenberg et al., 2002; van den Eynde et al., 1999). Altogether this suggests that non-canonical transcription or translation *per se* does not frequently generate relevant immunogenic antigens *in vivo*, but rather their immunogenicity seems to be influenced by the gene of origin and/or level of RNA expression.

NonC-TL can elicit *de novo* T-cell responses via *in vitro* sensitization

The human naïve T-cell repertoire is exceptionally diverse because it contains a large number of distinct T-cell clones (10^9-10^{11}), each represented by only a few cells (Goldrath & Bevan, 1999). Upon antigenic stimulation, infrequent antigen-specific naïve T cells undergo clonal expansion and differentiate into effector and memory cells. Thus, the memory T-cell repertoire contains a collection of expanded T-cell clones that reflect the antigenic experience of the individual (Sallusto et al., 2004). Due to the higher frequency and lower activation threshold, the direct detection of T-cell responses from TIL or PBL through functional immunogenic assays without previous stimulation identifies antigen-experienced cells (Geiger et al., 2009).

In contrast, T cells detected through IVS typically originate from the naïve rather than the antigen-experienced T-cell compartment, since very infrequent T cells can be clonally expanded upon several rounds of stimulation *ex vivo*. Due to the extraordinarily diverse naïve repertoire, T cells targeting any virtual protein can be potentially identified using IVS, since central tolerance is imperfect and the TCR are relatively promiscuous (Morris & Allen, 2012) . Therefore, the detection of antigen-specific T cells through this approach indicates that the peptide can be recognized by a T-cell clone, and thus can be immunogenic, but it does not prove whether the antigen is capable of eliciting a relevant T-cell response *in vivo.* In contrast, the detection of naturally occurring anti-tumor T-cell responses in cancer patients suggests that the cognate antigen was cross-presented at some point by DC and presented on the patient's tumor and was able to stimulate a T-cell response *in vivo*.

Since we can only screen a limited fraction of the broad patient's TCR repertoire, hopefully representative yet incomplete, the existence of antigen-experienced T cells targeting nonC-TL cannot be rule out. Moreover, T cells targeting tumor antigens have also been pursued by reverse immunology through mouse immunization (Parkhurst et al., 2009) or IVS of PBL derived from cancer patients (Bethune et al., 2018) and healthy donors (Gerdemann et al., 2011). Of note, some of the TCR identified by these strategies have been used in clinical trials (Johnson et al., 2009; Parkhurst et al., 2009).

Even if we could not detect recall T-cell responses to nonC-TL, we addressed whether they could still be immunogenic. Through IVS of non-autologous PBL expressing the allele of interest (A*11:01), we were able to isolate and expand T cells specifically targeting three nonC-TL, which enabled us to explore those antigens in more detail. By co-culturing nonC-TL-specific-T cells with a panel of 24 patient-derived TCL, we evaluated the expression and translation of these targets in distinct tumor types. We found that two immunogenic nonC-TL derived from the aberrant translation of 5'UTR of HOXC13 and ZKSCAN genes were frequently detected in patient-derived melanoma TCL as well as other less immunogenic tumor types such as GI or Gyn. In addition, an immunogenic nonC-TL derived from a non-coding spliced variant of C5orf22 gene was also detected in several melanoma TCL. These results revealed that nonC-TL can be immunogenic and shared across tumor types, thus representing attractive targets for off-the-shelf vaccines or T-cell therapies.

Immunogenic nonC-TL are preferentially presented in TCL over healthy cells

A potential concern regarding the therapeutic use of nonC-TL is whether they are actually tumor-specific. We used a healthy immunopeptidome dataset to select nonC HLA-I ligands absent in non-malignant cells, despite we were aware that this filter might not be perfect, in part due to the limited sensitivity of LC/MS-MS methods. However, as opposed to RNA-seq analysis, its advantage is that it can provide evidence that an a priori normal transcript can be aberrantly translated. Recently, Ribo-seq has emerged as a potent tool to predict non-annotated ORF and, hence, can potentially be used to select tumor-specific nonC ORF. However, this

technique is relatively new, the sensitivity is still limited and little data from healthy tissue is currently available.

To gain additional insight into the tumor specificity of the three immunogenic NonC identified, we empirically evaluated their expression and translation and HLA-I presentation in healthy human cells derived from several vital organs as well as melanocytes by using lymphocytes transduced with antigen-specific TCR. Our results revealed that 5'U-HOXC13 nonC-TL showed the highest tumor specificity, since TCR-transduced T cells targeting this antigen strongly recognized several TCL, but barely detected melanocytes or any other healthy human cells tested. Similar results were observed for nc-C5orf22 TCR transduced T cells although the recognition of TCL was generally lower, suggesting a poor transcription/translation of this nonC-TL. Importantly, cells transduced with DMF5 TCR which recognizes MART-1 and has been used to treat melanoma patients (Johnson et al., 2009), displayed far higher recognition of melanocytes (>90%), as compared to the TCR targeting nonC antigens. Unexpectedly, Mart-1 specific T cells also recognized cardiac myocytes at a similar level as TCR transduced T cells targeting the nonC-TL 5'U-ZKSCAN1 (10-20%). DMF5 TCR has shown objective tumor responses in patients but has also shown severe on-target toxicities in skin, eyes and ears due to the high expression of MART-1 in melanocytes. However, neither off-target cardiac toxicities nor adverse effects have been reported in any other organ.

Our findings suggest that the low level of expression of these nonC-TL in healthy tissues may restrict the development of severe toxicities and that patients whose tumors express these antigens could benefit from T cell-based therapies targeting them. Nonetheless, these data must be interpreted with caution, since we cannot rule out the possibility that these immunogenic nonC-TL are expressed and translated in other healthy cells. More experiments should be performed to continue assessing the safety of these targets, especially for 5'U-HOXC13 nonC-TL, which is the most promising.

Why were T-cell responses to nonC-TL not detected in cancer patients?

Altogether these results revealed that nonC-TL can be immunogenic, frequently presented on HLA-I, shared across tumor types, and not or poorly presented in non-malignant cells. However, pre-existing T-cell responses in cancer patients were not detected, contrary to neoantigens derived from NSM, cancer-germline, or melanoma-associated antigens.

Different mechanisms can explain the lack of recognition. For instance, we were only able to screen long-term expanded lymphocytes or lymphocytes enriched based on TCL recognition in some patients which can bias the TCR repertoire to more oligoclonal populations leading to the depletion of some tumor-reactive clones (Figure 52A) (Poschke 2020). In fact, patient H&N-1 is an example of such a mechanism as we could not detect neoantigen-specific T cells in the second screening, even though we previously identified T cells targeting the neoantigen RPL14p.H20Y using less expanded TIL (Table 5). Nonetheless, except for H&N-1, in the second part, all candidates tested including conventional tumor antigens and nonC-TL were interrogated in parallel using the same expanded populations and therefore the comparison was fair.

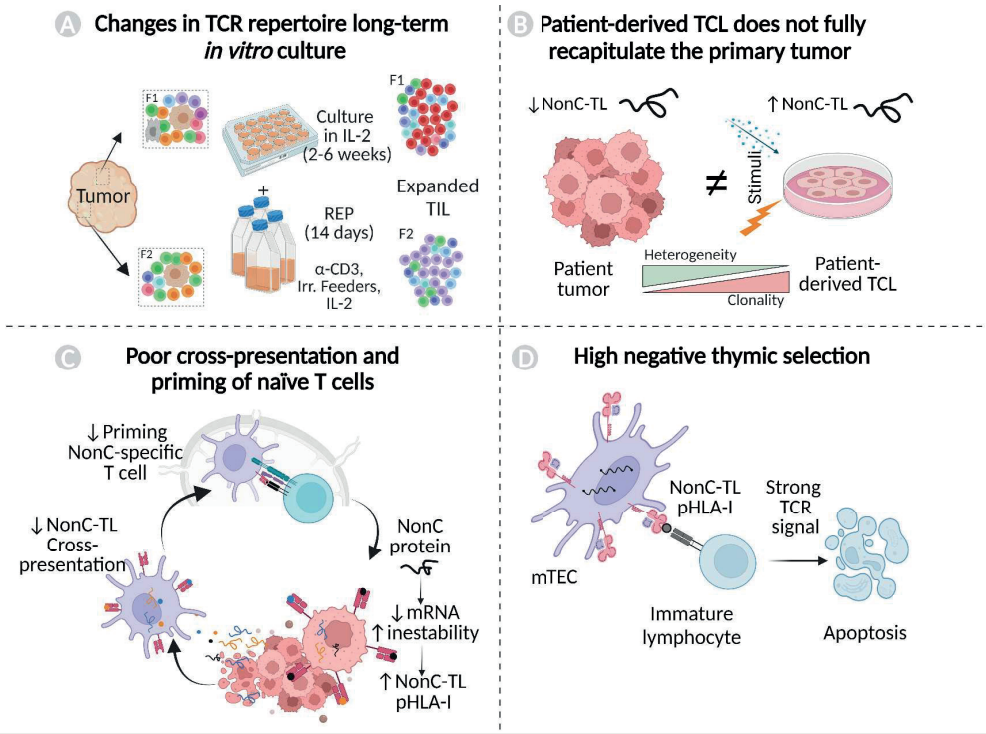


Figure 52. Potential mechanisms explaining the lack of recognition of nonC-TL by T cells in cancer patients. (A) TIL are expanded from tumor biopsies in high doses of IL-2 and/or using a rapid expansion protocol (REP). changing the TCR repertoire and leading to more oligoclonal populations. (B) Whilst nonC-TL are infrequently presented on the original patient's tumor, the established patient-derived TCL increase their presentation as a result of in vitro culture or overgrowth of clones expressing those nonC antigens. (C) NonC-TL derived from low abundance and unstable proteins are efficiently presented on HLA-I of tumor cells but poorly cross-presented hampering naïve T-cell priming in the draining lymph nodes. (D) mTEC cells present nonC-TL on HLA-I and T cells targeting nonC-TL are eliminated during negative thymic selection.

A possible explanation for this might be that patient-derived TCL do not entirely recapitulate the tumors *in vivo* (Figure 52B). On one hand, it is plausible to speculate that tumor heterogeneity can be lost in established TCL, as some clones may overgrow others due to a selective growth advantage, and this may be aggravated for late passages. On the other hand, we cannot exclude the possibility that the nonC expression and/or translation can be induced or exacerbated by *in vitro* culture. To address whether nonC-TL are presented solely as a result of *in vitro* culture, it would be ideal to perform HLA-I immunopeptidomics on matched patient tumors. Unfortunately, we did not have these samples. Instead, we explored whether our nonC-TL candidates are naturally presented by analyzing publicly available immunopeptidomics datasets from human tumor biopsies (Bassani-Sternberg et al., 2016; Kalaora et al., 2021). We detected at least 76 of the nonC-TL we identified using a stringent 1% FDR in uncultured melanoma samples. Considering the limited sensitivity of this technique in addition to the HLA-I diversity, these finding suggests that the nonC translation and, consequently, the nonC-TL observed are not an artifact of *in vitro* cultured cells and that nonC-TL can be naturally presented *in vivo*.

Another possible explanation might be that the level of expression of nonC-TL is sufficient to be presented on tumor HLA-I, but inadequate for efficient cross-presentation *in vivo* (Figure 52C). In line with this hypothesis, nonC peptides are thought to be less abundant and largely originated from disordered or unstable proteins with shorter half-lives compared to

functionally annotated proteins (Dersh et al., 2021; Ruiz Cuevas et al., 2021). In addition, some nonC ORF are so small that they do not require processing (Ouspenskaia et al., 2022). Similar to DRiPs, these characteristics can facilitate the accessibility of nonC peptides into the antigen processing machinery and, therefore, their HLA-I presentation is highly efficient (Dersh et al., 2021; Yewdell, 2011). By contrast, some studies have demonstrated that rapidly degraded proteins and minimal epitopes are unable to provoke cross-priming by APC as opposed to stable, full-length antigens (Norbury et al., 2004). Furthermore, in a model of colorectal cancer, for example, low expression of neoantigens prevented productive cross-priming and drives immediate T-cell dysfunction (Westcott et al., 2021). Altogether, we think that it is possible that the native characteristics of the aberrant translation events given rise to nonC-TL (i.e. low abundance and instability) account, at least in part, for the lack of recognition in cancer patients.

An additional alternative explanation for our findings is that nonC-TL could also be presented by mTEC cells (Figure 52D). Hence, T cells targeting those antigens might have undergone incomplete, but higher, central tolerance limiting the abundance of T cells in the periphery. This, combined with the low level of expression and/or poor priming may hamper the detection of antigen-experienced T cells targeting nonC-TL in cancer patients.

Therapeutic potential of nonC tumor antigens

Even though pre-existing T-cell responses in patients were not detected, we believe that nonC-TL may still be valuable targets for cancer immunotherapies. For instance, the hypothetic defective priming could be rescued by using vaccines, eventually rendering T cells capable of recognizing low-abundance/unstable antigens derived from nonC proteins on tumor cells. The adoptive transfer of TCR-engineered T cells would also overcome this limited natural priming, and the fact that nonC-TL can be shared, as we demonstrated, would facilitate the development of off-the-shelf T-cell therapies.

Although treatment with immunotherapies, including ACT, can mediate objective and durable responses in some patients, many patients show tumor progression following an initial objective response. This can sometimes be explained by the acquisition of resistance mechanisms presumably caused by tumor immunoediting. Two independent reports of patients treated with ACT of either TIL (Tran et al., 2016) or TCR-engineered T cells (Kim et al., 2022) targeting driver mutations evidenced such resistance mechanisms. The study of the lesions after progression revealed that a portion of chromosome 6 was lost resulting in HLA loss of heterozygosity which included the alleles targeted by the infused T cells, thus providing a direct mechanism of tumor immune evasion. Although driver mutations are optimal targets for ACT because they are essential for tumor cell survival and hence, are largely clonal, the immune pressure exerted on such neoantigens is expected to be strong. Further, given that T cells targeting these driver mutations were, in fact, detected in the patient's own tumors, it is plausible that a small fraction of tumor cells found a mechanism to escape, via reducing the expression of the specific antigen or altering its presentation. Such pre-existing tumor het-

erogeneity could contribute to tumor resistance and progression following treatment with T cells targeting such neoantigens. In this respect, the lack or defective T-cell response to nonC tumor antigens in cancer patients could be advantageous, since these antigens and the HLAs presenting them are less exposed to immune selective pressure and, consequently, tumors are less likely to present antigen escape variant heterogeneity.

Given that our work shows that the immunogenic nonC-TL that we identified are potentially interesting targets for off-the-shelf therapies, one of the remaining questions is how to select patients for such therapeutic interventions. An obvious test would be to select HLA-A*11:01 or A*03:01 patients and determine whether these nonC-TL are naturally presented on HLA-I of tumor tissue through immunopeptidomics using targeted methods to increase the sensitivity. However, the limited size of resectable tumors and the percentage of tumor vs. stroma, which is often difficult to control, can negatively affect the quality and success of such a complex technique. Another potential strategy that we envision in a near future would be to perform Ribo-seq on the patient's tumor, although the sensitivity of this technique is currently limited and technically challenging.

As an alternative, we wondered whether a straightforward method such as RNA-sequencing could guide the selection of patients that potentially would benefit from this therapy. We retrospectively analyzed whether the level of expression of the canonical genes correlated with the recognition of the nonC-TL-specific cells tested using available RNA-seq data from some of the TCL and healthy cell lines tested (Figure 53). Whilst no detection of the transcript implied no recognition, the opposite did not stand true. In other words, when the canonical transcript was not detected by RNA-seq, the encoded nonC-TL could not be aberrantly translated. Consequently, the recognition by either the isolated antigen-specific T cells, (Figure 53 left) or TCR-transduced (Figure 53 right) as measured by 4-1BB upregulation (x-axis) was prevented. However, detectable levels of the transcripts did not always result in the presentation of aberrantly translated nonC-TL.

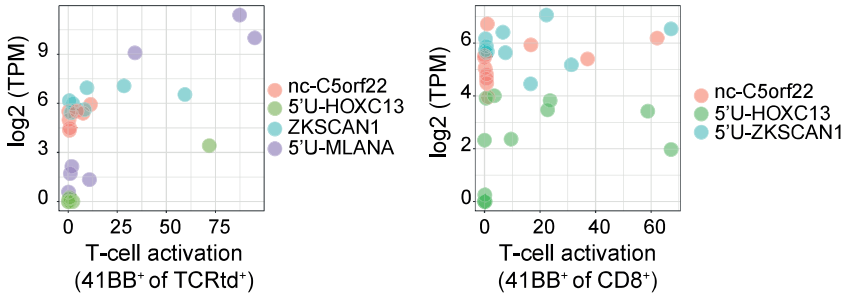


Figure 53. Gene expression does not correlate with the translation of nonC-TL indirectly measured by recognition of nonC-TL-specific T cells. Correlation between antigen-specific T-cell activation and expression of the targeted gene. The expression was retrospectively analyzed from RNA-seq data of TCL and healthy cell lines that were tested for specific T-cell recognition in experiments previously described in Figure 48 (left) or Figure 44 (right). The y-axis shows the gene expression measured by log2 (TPM) and the x-axis the percentage of T-cell activation after co-culturing antigen-specific T cells with target cells expressing the relevant allele (A*11:01 for nc-C5orf22 and 5'U-HOXC13, and A*68:01 for 5'U-ZKSCAN1).

According to these results, whereas the RNA expression of MLANA gene was associated with the recognition of MART-1-specific T cells (Figure 43 left), the recognition of

nonC-TL-specific T cells was poorly associated with the level of their corresponding canonical transcript. This shows that transcriptome analysis is not a useful tool to stratify the patients, which was in part expected, as RNA analysis is not able to determine whether UTR regions are translated.

Thus, although immunopeptidomics appears to be the best strategy to select patients at present, gaining insights into the mechanism driving aberrant translation from 5′ UTR or the nonC RNA may prove critical to select patients with greater sensitivity.

What do tumor-reactive T cells recognize?

Despite the extensive immunological screening using different methodologies, including conventional tumor antigens but also nonC -TL, the specificity of most tumor-reactive lymphocytes remains unknown.

Although we included a large number of tumor antigen candidates originated from presumably any region of the genome, we certainly did not interrogate all of them. Firstly, we used immunopeptidomics to prioritize the candidates presented by HLA-I, but we know from our own experience that some immunogenic antigens can be missed. Secondly, Peptide-PRISM is not able to efficiently identify all the existing sources of non-canonical antigens. For instance, this pipeline can detect non-coding transcripts and spliced variants as long as they are annotated in *Ensembl*. Indeed, one of the immunogenic nonC-TL identified was derived from an annotated non-coding RNA resulting from alternative splicing due to intron retention of C5orf22 gene. However, to efficiently detect tumor-specific aberrant splicing, a personalized approach using RNA-seq would be needed. As mentioned before, although HERVs and other transposable elements can potentially be detected, the identification of sequences from repetitive regions is challenging, and Peptide-PRISM does not seem to be well-suited to do so as reflected by the few peptides detected from this source. Because we used a proteomics method, we could have tried to detect post-translational modifications such as phosphorylation or acetylation, for instance. However, this would substantially expand the search space (already huge) and, because the samples were not processed for this purpose, although possible, their identification is not ideal. In addition, some potential nonC sources were completely ignored. For instance, RNA editing (Zhang et al., 2018a), protein decoding (Pataskar et al., 2022), or peptide splicing (Liepe J et al., 2016) were not addressed by this approach. Third, although neoantigens arising from NSM identified by WES were included, still, some immunogenic neoantigens derived from genomic alterations could have been missed. A feasible improvement to detect more neoantigens would be to exploit alternative methods of NGS such as WGS as previously discussed, but also other technologies. Short-read sequencing is the most common NGS approach used, however, it has some technical limitations than can be overcome by long-read sequencing (Pollard et al., 2018). For instance, this technology allows to accurately identify repetitive regions as well as structural variants such as segmental duplications, gene loss, and fusion events. Last, we did not interrogate self-antigens overexpressed by tumor cells. A high central tolerance against them is expected as they are derived from normal proteins, and some studies have suggested

that they are not as immunogenic as tissue-differentiation or cancer-germline antigens (Andersen et al., 2012). However, we cannot rule out the possibility that some tumor-reactive T cells might actually recognize them.

However, screening of such a massive list of new potential candidates, must be accompanied by the development of new high-throughput tools to test the immunogenicity with high sensitivity and accuracy while reducing the number of T cells needed.. As mentioned before, long-term expansion of TIL frequently leads to more oligoclonal populations and some tumor-reactive clones can be lost. A currently available option would be to clone the TCR from non-expanded TIL or any population of interest using single-cell sequencing (Zheng et al., 2022). Consequently, an endless source of T cells representative of the initial TCR repertoire is generated. Although feasible, the cost and human effort is currently still cumbersome. New antigen discovery platforms that allow the interrogation of a large list of candidates have been recently developed (Gee et al., 2018; Joglekar et al., 2019; Kisielow et al., 2019; Kula et al., 2019; G. Li et al., 2019). However, the systematic identification of personalized antigens is challenging as these systems typically use libraries that have to be customized for each patient and the APC are genetically engineered to express the allele of interest. Moreover, the T cells investigated are generally clonal or oligoclonal populations since the detection of infrequent T-cell responses is limited.

The remarkable success of cancer immunotherapy has reached an important inflection point in the history of cancer treatment, however, not all the patients benefit from these therapies. Broadening our understanding of the immune response to tumors is essential to develop more effective cancer immunotherapies and long-lasting responses in patients. In addition to conventional antigens, in this thesis we investigated whether non-canonical antigens from which there is limited knowledge and no clinical experience could contribute to cancer immune surveillance. Although we did not detect pre-existing T-cell responses to nonC-TL in cancer patients, we identified TCR targeting these antigens that were capable of recognizing patient-derived TCL. Hence, we believe that they may still be valuable targets for cancer immunotherapies. However, despite the extensive efforts made, we could not determine the specific recognition of many tumor-reactive lymphocytes. So much still remains to be discovered!

Conclusions

Tumor-reactive lymphocytes are frequently detected in cancer patients and consistently display preferential recognition of neoantigens as well as cancer-germline or tissue differentiation antigens. Although non-canonical HLA-I tumor ligands identified by proteogenomics are frequently presented in patient-derived tumor cell lines and can be immunogenic, pre-existing T cell responses could not be detected in cancer patients.

1. Tumor-reactive lymphocytes can frequently be detected in TIL and blood of cancer patients irrespective of the tumor origin.

2. The three independent methods used to identify tumor antigens led to the detection of lymphocytes targeting neoantigens (n=6) and tissue-diferentiation antigens (n=2) in the three patients studied.

3. Despite the extensive immunological screening performed, the specific antigen/s targeted by the majority of tumor-reactive lymphocytes could not be explained by conventional tumor antigens identified through multiple methods.

4. Non-canonical HLA-I tumor ligands identified using Peptide-PRISM are frequently presented on patient-derived tumor cell lines, are mainly derived from 5'UTR regions, and preferentially bind to HLA-A*11:01 and HLA-A*03:01.

5. Non-canonical HLA-I tumor ligands represent the most abundant source of candidate tumor antigens, as compared to neoantigens, cancer-germline, and melanoma-associated antigens.

6. Pre-existing T cell responses targeting non-canonical HLA-I tumor ligands were not detected in the 9 patients studied. Conversely, tumor-reactive lymphocytes consistently display preferential recognition of neoantigens as well as cancer-germline or tissue differentiation antigens.

7. Non-canonical HLA-I tumor ligands can elicit *de novo* T-cell responses via ***in vitro*** sensitization of donor peripheral blood lymphocytes.

8. We identified TCR specific to three non-canonical HLA-I tumor ligands, two of which mapped to the 5' UTR regions of the HOXC13 and ZKSCAN1 genes, and one mapping to a non-coding spliced variant of C5orf22 gene.

9. The 3 immunogenic non-canonical HLA-I tumor ligands detected, particularly the one derived from 5'U-HOXC13, display therapeutic potential, as they are naturally presented on HLA-I, expressed across diverse tumor types, but are barely detected in healthy cells.

10. Nonetheless, the specific antigens recognized tumor-reactive TIL cannot be explained by recognition of non-canonical HLA-I tumor ligands. Still, the specificity of many tumor-reactive TIL remains unknown and warrants further investigation.

References

Aarnoudse, C. A., van den Doel, P. B., Heemskerk, B., & Schrier, P. I. (1999). INTERLEUKIN-2-INDUCED, MELANOMA-SPECIFIC T CELLS RECOGNIZE CAMEL, AN UNEXPECTED TRANSLATION PRODUCT OF LAGE-1. In *J. Cancer* (Vol. 82). Wiley-Liss, Inc.

Aflalo, A., & Boyle, L. H. (2021). Polymorphisms in MHC class I molecules influence their interactions with components of the antigen processing and presentation pathway. In *International Journal of Immunogenetics* (Vol. 48, Issue 4, pp. 317–325). John Wiley and Sons Inc. https://doi.org/10.1111/iji.12546

Akers, S. N., Odunsi, K., & Karpf, A. R. (2010). Regulation of cancer germline antigen gene expression: Implications for cancer immunotherapy. In *Future Oncology* (Vol. 6, Issue 5, pp. 717–732). https://doi.org/10.2217/fon.10.36

Albert, B. (2015). *Molecular Biology of the Cell* (6th ed.).

Aleksic, M., Liddy, N., Molloy, P. E., Pumphrey, N., Vuidepot, A., Chang, K. M., & Jakobsen, B. K. (2012). Different affinity windows for virus and cancer-specific T-cell receptors: Implications for therapeutic strategies. *European Journal of Immunology*, *42*(12), 3174–3179. https://doi.org/10.1002/eji.201242606

Almeida, L. G., Sakabe, N. J., de Oliveira, A. R., Silva, M. C. 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. G., Old, L. J., Camargo, A. A., & Vasconcelos, A. T. R. (2009). CTdatabase: A knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Research*, *37*(SUPPL. 1). https://doi.org/10.1093/nar/gkn673

Altman, J. D., Moss, P. A. H., Goulder, P. J. R., Barouch, D. H., Mcheyzer-Williams, M. G., Bell, J. I., Mcmichael, A. J., & Davis, M. M. (1996). Phenotypic Analysis of Antigen-Specific T Lymphocytes. In *New Series* (Vol. 274, Issue 5284).

Andersen, R. S., Thrue, C. A., Junker, N., Lyngaa, R., Donia, M., Ellebæk, E., Svane, I. M., Schumacher, T. N., Straten, P. T., & Hadrup, S. R. (2012). Dissection of T-cell antigen specificity in human melanoma. *Cancer Research*, *72*(7), 1642–1650. https://doi.org/10.1158/0008-5472.CAN-11-2614

Attermann, A. S., Bjerregaard, A. M., Saini, S. K., Grønbaek, K., & Hadrup, S. R. (2018). *HUMAN ENDOGENOUS RETROVIRUSES AND THEIR IMPLICATION FOR IMMUNOTHERAPEUTICS OF CANCER*. https://doi.org/10.1093/annonc/mdy413/5100881

Attig, J., Young, G. R., Hosie, L., Perkins, D., Encheva-Yokoya, V., Stoye, J. P., Snijders, A. P., Ternette, N., & Kassiotis, G. (2019). LTR retroelement expansion of the human cancer transcriptome and immunopeptidome revealed by de novo transcript assembly. *Genome Research*, *29*(10), 1578–1590. https://doi.org/10.1101/gr.248922.119

Barber, D. L., Wherry, E. J., Masopust, D., Zhu, B., Allison, J. P., Sharpe, A. H., Freeman, G. J., & Ahmed, R. (2006). Restoring function in exhausted CD8 T cells during chronic viral infection. *Nature*, *439*(7077), 682–687. https://doi.org/10.1038/nature04444

Bartok, O., Pataskar, A., Nagel, R., Laos, M., Goldfarb, E., Hayoun, D., Levy, R., Körner, P. R., Kreuger, I. Z. M., Champagne, J., Zaal, E. A., Bleijerveld, O. B., Huang, X., Kenski, J., Wargo, J., Brandis, A., Levin, Y., Mizrahi, O., Alon, M., … Agami, R. (2021). Anti-tumour immunity induces aberrant peptide presentation in melanoma. *Nature*, *590*(7845), 332–337. https://doi.org/10.1038/s41586-020-03054-1

Basler, M., Kirk, C. J., & Groettrup, M. (2013). The immunoproteasome in antigen processing and other immunological functions. In *Current Opinion in Immunology* (Vol. 25, Issue 1, pp. 74–80). Elsevier Ltd. https://doi.org/10.1016/j.coi.2012.11.004

Bassani-Sternberg, M., Bräunlein, 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., Busch, D. H., 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. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms13404

Bassani-Sternberg, M., & Coukos, G. (2016). Mass spectrometry-based antigen discovery for cancer immunotherapy. In *Current Opinion in Immunology* (Vol. 41, pp. 9–17). Elsevier Ltd. https://doi.org/10.1016/j.coi.2016.04.005

Becker, J. P., & Riemer, A. B. (2022). The Importance of Being Presented: Target Validation by Immunopeptidomics for Epitope-Specific Immunotherapies. In *Frontiers in Immunology* (Vol. 13). Frontiers Media S.A. https://doi.org/10.3389/fimmu.2022.883989

Bentzen, A. K., & Hadrup, S. R. (2017). Evolution of MHC-based technologies used for detection of antigen-responsive T cells. In *Cancer Immunology, Immunotherapy* (Vol. 66, Issue 5, pp. 657–666). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s00262-017-1971-5

Bethune, M. T., Li, X. H., Yu, J., McLaughlin, J., Cheng, D., Mathis, C., Moreno, B. H., Woods, K., Knights, A. J., Garcia-DIaz, A., Wong, S., Hu-Lieskovan, S., Puig-Saus, C., Cebon, J., Ribas, A., Yang, L., Witte, O. N., & Baltimore, D. (2018). Isolation and characterization of NY-ESO-1-specific T cell receptors restricted on various MHC molecules. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(45), E10702–E10711. https://doi.org/10.1073/pnas.1810653115

Bigot, J., Lalanne, A. I., Lucibello, F., Gueguen, P., Houy, A., Dayot, S., Ganier, O., Gilet, J., Tosello, J., Nemati, F., Pierron, G., Waterfall, J. J., Barnhill, R., Gardrat, S., Piperno-Neumann, S., Popova, T., Masson, V., Loew, D., Mariani, P., … Lantz, O. (2021). Splicing patterns in sf3b1mutated uveal melanoma generate shared immunogenic tumor-specific neoepitopes. *Cancer Discovery*, *11*(8), 1938–1951. https://doi.org/10.1158/2159-8290.CD-20-0555

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennet, W. S., Strominger, J. L., & Wiley, D. C. (1987). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature*, *329*, 506–512.

Blass, E., & Ott, P. A. (2021). Advances in the development of personalized neoantigen-based therapeutic cancer vaccines. In *Nature Reviews Clinical Oncology* (Vol. 18, Issue 4, pp. 215–229). Nature Research. https://doi.org/10.1038/s41571-020-00460-2

Blum, J. S., Wearsch, P. A., & Cresswell, P. (2013). Pathways of antigen processing. In *Annual Review of Immunology* (Vol. 31, pp. 443–473). https://doi.org/10.1146/annurev-immunol-032712-095910

Bobisse, S., Genolet, R., Roberti, A., Tanyi, J. L., Racle, J., Stevenson, B. J., Iseli, C., Michel, A., le Bitoux, M. A., Guillaume, P., Schmidt, J., Bianchi, V., Dangaj, D., Fenwick, C., Derré, L., Xenarios, I., Michielin, O., Romero, P., Monos, D. S., … Harari, A. (2018). Sensitive and frequent identification of high avidity neo-epitope specific CD8 + T cells in immunotherapy-naïve ovarian cancer. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03301-0

Bonté, P.-E., Arribas, Y. A., Merlotti, A., Carrascal, M., Zhang, J. V., Zueva, E., Binder, Z. A., Alanio, C., Goudot, C., & Amigorena, S. (2022). Single-cell RNA-seq-based proteogenomics identifies glioblastoma-specific transposable elements encoding HLA-I-presented peptides. *Cell Reports*, *39*(10), 110916. https://doi.org/10.1016/j.celrep.2022.110916

Boshoff, C., & Weiss, R. (2002). AIDS-related malignancies. In *Nature Reviews Cancer* (Vol. 2, Issue 5, pp. 373–382). European Association for Cardio-Thoracic Surgery. https://doi.org/10.1038/nrc797

Bourdetsky, D., Schmelzer, C. E. H., & Admon, A. (2014). The nature and extent of contributions by defective ribosome products to the HLA peptidome. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(16). https://doi.org/10.1073/pnas.1321902111

Brichard, V., Pel, A. van, W6lfel, T., W61fel, C., de Plaen, E., Leth6, B., Coulie, P., & Boon, T. (1993). *The Tyrosinase Gene Codes for an Antigen Recognized by Autologous Cytolytic T Lymphocytes on HLA-A2 Melanomas*. http://rupress.org/jem/article-pdf/178/2/489/1267835/489.pdf

Burnet, F. M. (1970). The Concept of Immunological Surveillance. *Progr. Exp. Tumor Res*, *13*, 1–27.

Chandran, S. S., & Klebanoff, C. A. (2019). T cell receptor-based cancer immunotherapy: Emerging efficacy and pathways of resistance. In *Immunological Reviews* (Vol. 290, Issue 1, pp. 127–147). Blackwell Publishing Ltd. https://doi.org/10.1111/imr.12772

Chen, D. S., & Mellman, I. (2013a). Oncology meets immunology: The cancer-immunity cycle. In *Immunity* (Vol. 39, Issue 1, pp. 1–10). https://doi.org/10.1016/j.immuni.2013.07.012

Chen, D. S., & Mellman, I. (2013b). Oncology meets immunology: The cancer-immunity cycle. In *Immunity* (Vol. 39, Issue 1, pp. 1–10). https://doi.org/10.1016/j.immuni.2013.07.012

Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M.,

Stevenson, B. J., Michaux, J., Bilic, I., Hirsekorn, A., Calviello, L., Simó-Riudalbas, L., Planet, E., Lubiński, J., Bryśkiewicz, M., Wiznerowicz, M., … Bassani-Sternberg, M. (n.d.). *Integrated Proteogenomic Deep Sequencing and Analytics Accurately Identify Non-Canonical Peptides in Tumor Immunopeptidomes*. https://doi.org/10.1101/758680

Chong, C., Müller, M., Pak, H. S., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B. J., Michaux, J., Bilic, I., Hirsekorn, A., Calviello, L., Simó-Riudalbas, L., Planet, E., Lubiński, J., Bryśkiewicz, M., Wiznerowicz, M., … Bassani-Sternberg, M. (2020). Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-14968-9

Cleyle, J., Hardy, M.-P., Minati, R., Courcelles, M., Durette, C., Lanoix, J., Laverdure, J.-P., Vincent, K., Perreault, C., & Thibault, P. (2022). Immunopeptidomic analyses of colorectal cancers with and without microsatellite instability. *Molecular & Cellular Proteomics*, 100228. https://doi.org/10.1016/j.mcpro.2022.100228

Correale, P., Walmsley, K., Nieroda, C., Zaremba, S., Zhu, M., Schlom, J., & Tsang, K. Y. (1997). *In Vitro Generation of Human Cytotoxic T Lymphocytes Specific for Peptides Derived From Prostate-Specific Antigen*. http://jnci.oxfordjournals.org/

Coulie, P. G., Brichard, V., Pel, A. van, W61fel, T., Schneider, J., Traversari, C., Mattei, S., de Plaen, E., Lurquin, C., Szikora, J.-P., Renauld, J.-C., & Boon, T. (n.d.). *A New Gene Coding for a Differentiation Antigen Recognized by Autologous Cytolytic T Lymphocytes on HLA-A2 Melanomas*. http://rupress.org/jem/article-pdf/180/1/35/1105507/35.pdf

Coulie, P. G., Lehmann, F., Lethe, B., Herman, J., Lurquin, C., Andrawiss, M., & Boon, T. (1995). A mutated intron sequence codes for an antigenic peptide recognized by cytolytic T lymphocytes on a human melanoma. In *Immunology* (Vol. 92).

Coulie, P. G., van den Eynde, B. J., van der Bruggen, P., & Boon, T. (2014). Tumour antigens recognized by T lymphocytes: At the core of cancer immunotherapy. In *Nature Reviews Cancer* (Vol. 14, Issue 2, pp. 135–146). https://doi.org/10.1038/nrc3670

Cox, A. L., Skipper, J., Chen, Y., Henderson, R. A., Darrow, T. L., Shabanowitz, J., Engelhard, V. H., Hunt, D. F., & Slingluff, C. L. (1994). Identification of a Peptide Recognized by Five Melanoma-Specific Human Cytotoxic T Cell. In *Source: Science, New Series* (Vol. 264, Issue 5159).

Cruz, F. M., Colbert, J. D., Merino, E., Kriegsman, B. A., & Rock, K. L. (2017). The biology and underlying mechanisms of cross-presentation of exogenous antigens on MHC-I molecules. In *Annual Review of Immunology* (Vol. 35, pp. 149–176). Annual Reviews Inc. https://doi.org/10.1146/annurev-immunol-041015-055254

Dalet, A., Robbins, P. F., Stroobant, V., Vigneron, N., Li, Y. F., El-Gamil, M., Hanada, K. I., Yang, J. C., Rosenberg, S. A., & van den Eyndea, B. J. (2011). An antigenic peptide produced by reverse splicing and double asparagine deamidation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(29). https://doi.org/10.1073/pnas.1101892108

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. (n.d.). *Identification of tumor-associated, MHC class II-restricted phosphopeptides as targets for immunotherapy*. www.pnas.org/cgi/content/full/

Dersh, D., Hollý, J., & Yewdell, J. W. (2021). A few good peptides: MHC class I-based cancer immunosurveillance and immunoevasion. In *Nature Reviews Immunology* (Vol. 21, Issue 2, pp. 116–128). Nature Research. https://doi.org/10.1038/s41577-020-0390-6

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101–108. https://doi.org/10.1038/nature11233

Duhen, T., Duhen, R., Montler, R., Moses, J., Moudgil, T., de Miranda, N. F., Goodall, C. P., Blair, T. C., Fox, B. A., McDermott, J. E., Chang, S. C., Grunkemeier, G., Leidner, R., Bell, R. B., & Weinberg,

A. D. (2018). Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05072-0

Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J., & Schreiber, R. D. (2002). *Cancer immunoediting: from immuno-surveillance to tumor escape*. http://www.nature.com/natureimmunology

Dunn, G. P., Old, L. J., & Schreiber, R. D. (2004a). The Immunobiology of Cancer Immunosurveillance and Immunoediting. *Immunity*, *21*, 137–148.

Dunn, G. P., Old, L. J., & Schreiber, R. D. (2004b). The three Es of cancer immunoediting. In *Annual Review of Immunology* (Vol. 22, pp. 329–360). https://doi.org/10.1146/annurev.immunol.22.012703.104803

Eberhardt, C. S., Kissick, H. T., Patel, M. R., Cardenas, M. A., Prokhnevska, N., Obeng, R. C., Nasti, T. H., Griffith, C. C., Im, S. J., Wang, X., Shin, D. M., Carrington, M., Chen, Z. G., Sidney, J., Sette, A., Saba, N. F., Wieland, A., & Ahmed, R. (2021a). Functional HPV-specific PD-1+ stem-like CD8 T cells in head and neck cancer. *Nature*, *597*(7875), 279–284. https://doi.org/10.1038/s41586-021-03862-z

Eberhardt, C. S., Kissick, H. T., Patel, M. R., Cardenas, M. A., Prokhnevska, N., Obeng, R. C., Nasti, T. H., Griffith, C. C., Im, S. J., Wang, X., Shin, D. M., Carrington, M., Chen, Z. G., Sidney, J., Sette, A., Saba, N. F., Wieland, A., & Ahmed, R. (2021b). Functional HPV-specific PD-1+ stem-like CD8 T cells in head and neck cancer. *Nature*, *597*(7875), 279–284. https://doi.org/10.1038/s41586-021-03862-z

Eisenberg, E., & Levanon, E. Y. (2018). A-to-I RNA editing - Immune protector and transcriptome diversifier. In *Nature Reviews Genetics* (Vol. 19, Issue 8, pp. 473–490). Nature Publishing Group. https://doi.org/10.1038/s41576-018-0006-1

Erhard, F., Dölken, L., Schilling, B., & Schlosser, A. (2020). Identification of the cryptic HLA-I immunopeptidome. *Cancer Immunology Research*, *8*(8), 1018–1026. https://doi.org/10.1158/2326-6066.CIR-19-0886

Erhlich.P. (1909). Ueber den jetzigen Stand der Karzinomforschung. *Ned. Tijdschr. Geneeskd*, *5*, 273–290.

Faridi, P., Li, C., Ramarathinam, S. H., Vivian, J. P., Illing, P. T., Mifsud, N. A., Ayala, R., Song, J., Gearing, L. J., Hertzog, P. J., Ternette, N., Rossjohn, J., Croft, N. P., & Purcell, A. W. (2018). *A subset of HLA-I peptides are not genomically templated: Evidence for cis-and trans-spliced peptide ligands*. http://immunology.sciencemag.org/

Faridi, P., Purcell, A. W., & Croft, N. P. (2018). In Immunopeptidomics We Need a Sniper Instead of a Shotgun. *Proteomics*, *18*(12). https://doi.org/10.1002/pmic.201700464

Fisk, B., Blevins, T. L., Wharton, J. T., & Ioannides, C. G. (1995). Identification of an Immunodominant Peptide of HER-2/neu Protooncogene Recognized by Ovarian Tumor-specific Cytotoxic T Lymphocyte Lines. *Journal of Experimental Medicine*, *181*, 2109–2117. http://rupress.org/jem/article-pdf/181/6/2109/1106716/2109.pdf

Frankiw, L., Baltimore, D., & Li, G. (2019). Alternative mRNA splicing in cancer immunotherapy. In *Nature Reviews Immunology* (Vol. 19, Issue 11, pp. 675–687). Nature Publishing Group. https://doi.org/10.1038/s41577-019-0195-7

Freudenmann, L. K., Marcu, A., & Stevanović, S. (2018). Mapping the tumour human leukocyte antigen (HLA) ligandome by mass spectrometry. In *Immunology* (Vol. 154, Issue 3, pp. 331–345). Blackwell Publishing Ltd. https://doi.org/10.1111/imm.12936

Garcia-Garijo, A., Fajardo, C. A., & Gros, A. (2019). Determinants for neoantigen identification. In *Frontiers in Immunology* (Vol. 10, Issue JUN). Frontiers Media S.A. https://doi.org/10.3389/fimmu.2019.01392

Gatti, R. A., & Good, R. A. (1971). OCCURRENCE OF MALIGNANCY IN IMMUNODEFICIENCY DISEASES A Literature Review. *Cancer*, *28*, 89–98.

Gee, M. H., Han, A., Lofgren, S. M., Beausang, J. F., Mendoza, J. L., Birnbaum, M. E., Bethune, M. T., Fischer, S., Yang, X., Gomez-Eerland, R., Bingham, D. B., Sibener, L. v., Fernandes, R. A., Velasco, A., Baltimore, D., Schumacher, T. N., Khatri, P., Quake, S. R., Davis, M. M., & Garcia, K. C. (2018). Antigen Identification for Orphan T Cell Receptors Expressed on Tumor-Infiltrating Lymphocytes. *Cell*, *172*(3),

549-563.e16. https://doi.org/10.1016/j.cell.2017.11.043

Geiger, R., Duhen, T., Lanzavecchia, A., & Sallusto, F. (2009). Human naïve and memory CD4+ T cell repertoires specific for naturally processed antigens analyzed using libraries of amplified T cells. *Journal of Experimental Medicine*, *206*(7), 1525–1534. https://doi.org/10.1084/jem.20090504

Geng, J., Zaitouna, A. J., & Raghavan, M. (2018). Selected HLA-B allotypes are resistant to inhibition or deficiency of the transporter associated with antigen processing (TAP). *PLoS Pathogens*, *14*(7). https://doi.org/10.1371/journal.ppat.1007171

Gerdemann, U., Katari, U., Christin, A. S., Cruz, C. R., Tripic, T., Rousseau, A., Gottschalk, S. M., Savoldo, B., Vera, J. F., Heslop, H. E., Brenner, M. K., Bollard, C. M., Rooney, C. M., & Leen, A. M. (2011). Cytotoxic T lymphocytes simultaneously targeting multiple tumor-associated antigens to treat ebv negative lymphoma. *Molecular Therapy*, *19*(12), 2258–2268. https://doi.org/10.1038/mt.2011.167

Godet, Y., Moreau-Aubry, A., Guilloux, Y., Vignard, V., Khammari, A., Dreno, B., Jotereau, F., & Labarriere, N. (2008). MELOE-1 is a new antigen overexpressed in melanomas and involved in adoptive T cell transfer efficiency. *Journal of Experimental Medicine*, *205*(11), 2673–2682. https://doi.org/10.1084/jem.20081356

Goldrath, A. W., & Bevan, M. J. (1999). *Selecting and maintaining a diverse T-cell repertoire*.

Golstein, P., & Griffiths, G. M. (2018). An early history of T cell-mediated cytotoxicity. *Nature Reviews Immunology*, *18*(8), 527–535. https://doi.org/10.1038/s41577-018-0009-3

Gros, A., Parkhurst, M. R., Tran, E., Pasetto, A., Robbins, P. F., Ilyas, S., Prickett, T. D., Gartner, J. J., Crystal, J. S., Roberts, I. M., Trebska-Mcgowan, K., Wunderlich, J. R., Yang, J. C., & Rosenberg, S. A. (2016). Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nature Medicine*, *22*(4), 433–438. https://doi.org/10.1038/nm.4051

Gros, A., Robbins, P. F., Yao, X., Li, Y. F., Turcotte, S., Tran, E., Wunderlich, J. R., Mixon, A., Farid, S., Dudley, M. E., Hanada, K. I., Almeida, J. R., Darko, S., Douek, D. C., Yang, J. C., & Rosenberg, S. A. (2014). PD-1 identifies the patient-specific CD8+ tumor-reactive repertoire infiltrating human tumors. *Journal of Clinical Investigation*, *124*(5), 2246–2259. https://doi.org/10.1172/JCI73639

Gros, A., Tran, E., Parkhurst, M. R., Ilyas, S., Pasetto, A., Groh, E. M., Robbins, P. F., Yossef, R., Garcia-Garijo, A., Fajardo, C. A., Prickett, T. D., Jia, L., Gartner, J. J., Ray, S., Ngo, L., Wunderllich, J. R., Yang, J. C., & Rosenberg, S. A. (2019). Recognition of human gastrointestinal cancer neoantigens by circulating PD-1+ lymphocytes. *Journal of Clinical Investigation*, *129*(11), 4992–5004. https://doi.org/10.1172/JCI127967

Gubin, M. M., & Vesely, M. D. (2022). Cancer Immunoediting in the Era of Immuno-oncology. *Clinical Cancer Research*, OF1–OF12. https://doi.org/10.1158/1078-0432.ccr-21-1804

Guilloux, Y., Lucas, S., Brichard, V. G., van Pel, A., Viret, C., de Plaen, E., Brasseur, F., Lethé, B., Jotereau, F., & Boon, T. (1996). A peptide recognized by human cytolytic T lymphocytes on HLA-A2 melanomas is encoded by an intron sequence of the N-acetylglucosaminyltransferase V gene. *Journal of Experimental Medicine*, *183*(3), 1173–1183. https://doi.org/10.1084/jem.183.3.1173

H Bakker, A., S, M. W., B, A. J., A, G. J., & F, C. G. (1995). IDENTIFICATION OF A NOVEL PEPTIDE DERIVED FROM THE MELANOCYTE-SPECIFIC GP100 ANTIGEN AS THE DOMINANT EPITOPE RECOGNIZED BY AN HLA-A2.1-RESTRICTED ANTI-MELANOMA CTL LINE. *Int. J. Cancer*, *62*, 97–102.

Hanada, K.-I., Yewdell, J. W., & Yang, J. C. (2004). *Immune recognition of a human renal cancer antigen through post-translational protein splicing*. www.nature.com/nature.

Hanada, K.-I., Zhao, C., Gil-Hoyos, R., Gartner, J. J., Chow-Parmer, C., Lowery, F. J., Krishna, S., Prickett, T. D., Kivitz, S., Parkhurst, M. R., Wong, N., Rae, Z., Kelly, M. C., Goff, S. L., Robbins, P. F., Rosenberg, S. A., & Yang, J. C. (2022). A phenotypic signature that identifies neoantigen-reactive T cells in fresh human lung cancers. *Cancer Cell*. https://doi.org/10.1016/j.ccell.2022.03.012

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. In *Cancer Discovery* (Vol. 12, Issue 1, pp. 31–46). American Association for Cancer Research Inc. https://doi.org/10.1158/2159-8290.CD-21-

1059

Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer Review evolve progressively from normalcy via a series of pre. In *Cell* (Vol. 100).

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. In *Cell* (Vol. 144, Issue 5, pp. 646–674). https://doi.org/10.1016/j.cell.2011.02.013

Harada, M., Li, Y. F., El-Gamil, M., Ohnmacht, G. A., Rosenberg, S. A., & Robbins, P. F. (2001). *Melanoma-Reactive CD8 + T Cells Recognize a Novel Tumor Antigen Expressed in a Wide Variety of Tumor Types*.

Im, S. J., Hashimoto, M., Gerner, M. Y., Lee, J., Kissick, H. T., Burger, M. C., Shan, Q., Hale, J. S., Lee, J., Nasti, T. H., Sharpe, A. H., Freeman, G. J., Germain, R. N., Nakaya, H. I., Xue, H. H., & Ahmed, R. (2016). Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature*, *537*(7620), 417–421. https://doi.org/10.1038/nature19330

Irla, M. (2022). *Annual Review of Immunology Instructive Cues of Thymic T Cell Selection*. https://doi.org/10.1146/annurev-immunol-101320

Iyer, J. G., Afanasiev, O. K., McClurkan, C., Paulson, K., Nagase, K., Jing, L., Marshak, J. O., Dong, L., Carter, J., Lai, I., Farrar, E., Byrd, D., Galloway, D., Yee, C., Koelle, D. M., & Nghiem, P. (2011). Merkel cell polyomavirus-specific CD8+and CD4+ T-cell responses identified in Merkel cell carcinomas and blood. *Clinical Cancer Research*, *17*(21), 6671–6680. https://doi.org/10.1158/1078-0432.CCR-11-1513

Jayasinghe, R. G., Cao, S., Gao, Q., Wendl, M. C., Vo, N. S., Reynolds, S. M., Zhao, Y., Climente-González, H., Chai, S., Wang, F., Varghese, R., Huang, M., Liang, W. W., Wyczalkowski, M. A., Sengupta, S., Li, Z., Payne, S. H., Fenyö, D., Miner, J. H., … Ding, L. (2018). Systematic Analysis of Splice-Site-Creating Mutations in Cancer. *Cell Reports*, *23*(1), 270-281.e3. https://doi.org/10.1016/j.celrep.2018.03.052

Joglekar, A. v., Leonard, M. T., Jeppson, J. D., Swift, M., Li, G., Wong, S., Peng, S., Zaretsky, J. M., Heath, J. R., Ribas, A., Bethune, M. T., & Baltimore, D. (2019). T cell antigen discovery via signaling and antigen-presenting bifunctional receptors. *Nature Methods*, *16*(2), 191–198. https://doi.org/10.1038/s41592-018-0304-8

Johnson, L. A., Morgan, R. A., Dudley, M. E., Cassard, L., Yang, J. C., Hughes, M. S., Kammula, U. S., Royal, R. E., Sherry, R. M., Wunderlich, J. R., Lee, C. C. R., Restifo, N. P., Schwarz, S. L., Cogdill, A. P., Bishop, R. J., Kim, H., Brewer, C. C., Rudy, S. F., VanWaes, C., … Rosenberg, S. A. (2009). Gene therapy with human and mouse T-cell receptors mediates cancer regression and targets normal tissues expressing cognate antigen. *Blood*, *114*(3), 535–546. https://doi.org/10.1182/blood-2009-03-211714

Kahles, A., Lehmann, K. van, Toussaint, N. C., Hüser, M., Stark, S. G., Sachsenberg, T., Stegle, O., Kohlbacher, O., Sander, C., Caesar-Johnson, S. J., Demchok, J. A., Felau, I., Kasapi, M., Ferguson, M. L., Hutter, C. M., Sofia, H. J., Tarnuzzer, R., Wang, Z., Yang, L., … Rätsch, G. (2018). Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell*, *34*(2), 211-224.e6. https://doi.org/10.1016/j.ccell.2018.07.001

Kalaora, S., Barnea, E., Merhavi-Shoham, E., Qutob, N., Teer, J. K., Shimony, N., Schachter, J., Rosenberg, S. A., Besser, M. J., Admon, A., & Samuels, Y. (n.d.-a). *Use of HLA peptidomics and whole exome sequencing to identify human immunogenic neo-antigens* (Vol. 7, Issue 5). www.impactjournals.com/oncotarget

Kalaora, S., Barnea, E., Merhavi-Shoham, E., Qutob, N., Teer, J. K., Shimony, N., Schachter, J., Rosenberg, S. A., Besser, M. J., Admon, A., & Samuels, Y. (n.d.-b). *Use of HLA peptidomics and whole exome sequencing to identify human immunogenic neo-antigens* (Vol. 7, Issue 5). www.impactjournals.com/oncotarget

Kalaora, S., Lee, J. S., Barnea, E., Levy, R., Greenberg, P., Alon, M., Yagel, G., Bar Eli, G., Oren, R., Peri, A., Patkar, S., Bitton, L., Rosenberg, S. A., Lotem, M., Levin, Y., Admon, A., Ruppin, E., & Samuels, Y. (2020). Immunoproteasome expression is associated with better prognosis and response to checkpoint therapies in melanoma. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-14639-9

Kalaora, S., Nagler, A., Nejman, D., Alon, M., Barbolin, C., Barnea, E., Ketelaars, S. L. C., Cheng, K., Vervier, K., Shental, N., Bussi, Y., Rotkopf, R., Levy, R., Benedek, G., Trabish, S., Dadosh, T., Levin-Zaidman, S., Geller, L. T., Wang, K., … Samuels, Y. (2021). Identification of bacteria-derived HLA-bound peptides in melanoma. *Nature*, *592*(7852), 138–143. https://doi.org/10.1038/s41586-021-03368-8

Kalaora, S., Wolf, Y., Feferman, T., Barnea, E., Greenstein, E., Reshef, D., Tirosh, I., Reuben, A., Patkar, S., Levy, R., Quinkhardt, J., Omokoko, T., Qutob, N., Golani, O., Zhang, J., Mao, X., Song, X., Bernatchez, C., Haymaker, C., … Samuels, Y. (2018). Combined analysis of antigen presentation and T-cell recognition reveals restricted immune responses in melanoma. *Cancer Discovery*, *8*(11), 1366–1375. https://doi.org/10.1158/2159-8290.CD-17-1418

Kamphorst, A. O., Wieland, A., Nasti, T., Yang, S., Zhang, R., Barber, D. L., Konieczny, B. T., Daugherty, C. Z., Koenig, L., Yu, K., Sica, G. L., Sharpe, A. H., Freeman, G. J., Blazar, B. R., Turka, L. A., Owonikoko, T. K., Pillai, R. N., Ramalingam, S. S., Araki, K., & Ahmed, R. (2017). Rescue of exhausted CD8 T cells by PD-1-targeted therapies is CD28-dependent. *Science*, *355*(6332), 1423–1427. https://doi.org/10.1126/science.aaf0683

Kaplan, D. H., Shankaran, V., Dighe, A. S., Stockert, E., Aguet, M., Old, L. J., & Schreiber, R. D. (1998). Demonstration of an interferon-dependent tumor surveillance system in immunocompetent mice. In *Immunology* (Vol. 95). www.pnas.org.

Kim, S. P., Vale, N. R., Zacharakis, N., Krishna, S., Yu, Z., Gasmi, B., Gartner, J. J., Sindiri, S., Malekzadeh, P., Deniger, D. C., Lowery, F. J., Parkhurst, M. R., Ngo, L. T., Ray, S., Li, Y. F., Hill, V., Florentin, M., Masi, R. v., Paria, B. C., … Rosenberg, S. A. (2022). Adoptive Cellular Therapy with Autologous Tumor-Infiltrating Lymphocytes and T-cell Receptor–Engineered T Cells Targeting Common p53 Neoantigens in Human Solid Tumors. *Cancer Immunology Research*, OF1–OF15. https://doi.org/10.1158/2326-6066.cir-22-0040

Kisielow, J., Obermair, F. J., & Kopf, M. (2019). Deciphering CD4+ T cell specificity using novel MHC–TCR chimeric receptors. *Nature Immunology*, *20*(5), 652–662. https://doi.org/10.1038/s41590-019-0335-z

Kloetzel, P. (2001). ANTIGEN PROCESSING BY THE PROTEASOME. *Nature Reviews*, *2*(Molecular cell biology), 179–187. www.nature.com/reviews/molcellbio

Kochenderfer, J. N., Wilson, W. H., Janik, J. E., Dudley, M. E., Stetler-Stevenson, M., Feldman, S. A., Maric, I., Raffeld, M., Nathan, D. A. N., Lanier, B. J., Morgan, R. A., & Rosenberg, S. A. (2010). Eradication of B-lineage cells and regression of lymphoma in a patient treated with autologous T cells genetically engineered to recognize CD19. *Blood*, *116*(20), 4099–4102. https://doi.org/10.1182/blood-2010-04-281931

Kong, Y., Rose, C. M., Cass, A. A., Williams, A. G., Darwish, M., Lianoglou, S., Haverty, P. M., Tong, A. J., Blanchette, C., Albert, M. L., Mellman, I., Bourgon, R., Greally, J., Jhunjhunwala, S., & Chen-Harris, H. (2019). Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-13035-2

Kristensen, N. P., Heeke, C., Tvingsholm, S. A., Borch, A., Draghi, A., Crowther, M. D., Carri, I., Munk, K. K., Holm, J. S., Bjerregaard, A. M., Bentzen, A. K., Marquard, A. M., Szallasi, Z., McGranahan, N., Andersen, R., Nielsen, M., Jönsson, G. B., Donia, M., Svane, I. M., & Hadrup, S. R. (2022). Neoantigen-reactive CD8+ T cells affect clinical outcome of adoptive cell therapy with tumor-infiltrating lymphocytes in melanoma. *Journal of Clinical Investigation*, *132*(2). https://doi.org/10.1172/JCI150535

Kubo, R., Sette, A., Grey, H., Appella, E., Sakaguchi, K., Zhu, N., Arnott, D., Sherman, N., Shabanowitz, J., Michel, H., Bodnar, W., Davis, T., & Hunt, D. (1994). *Definition of specific peptide motifs for four major HLA-A alleles*. http://www.jimmunol.org/content/152/8/3913

Kula, T., Dezfulian, M. H., Wang, C. I., Abdelfattah, N. S., Hartman, Z. C., Wucherpfennig, K. W., Lyerly, H. K., & Elledge, S. J. (2019). T-Scan: A Genome-wide Method for the Systematic Discovery of T Cell Epitopes. *Cell*, *178*(4), 1016-1028.e13. https://doi.org/10.1016/j.cell.2019.07.009

Kumar, B. v., Connors, T. J., & Farber, D. L. (2018). Human T Cell Development, Localization, and Function throughout Life. In *Immunity* (Vol. 48, Issue 2, pp. 202–213). Cell Press. https://doi.org/10.1016/j.immuni.2018.01.007

Larouche, J. D., Trofimov, A., Hesnard, L., Ehx, G., Zhao, Q., Vincent, K., Durette, C., Gendron, P., Laverdure, J. P., Bonneil, É., Côté, C., Lemieux, S., Thibault, P., & Perreault, C. (2020). Widespread and tissue-specific expression of endogenous retroelements in human somatic tissues. *Genome Medicine*, *12*(1). https://doi.org/10.1186/s13073-020-00740-7

Laumont, C. M., Daouda, T., Laverdure, J. P., Bonneil, É., Caron-Lizotte, O., Hardy, M. P., Granados, D. P., Durette, C., Lemieux, S., Thibault, P., & Perreault, C. (2016). Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms10238

Laumont, C. M., Vincent, K., Hesnard, L., Audemard, É., Bonneil, É., Laverdure, J. P., Gendron, P., Courcelles, M., Hardy, M. P., Côté, C., Durette, C., St-Pierre, C., Benhammadi, M., Lanoix, J., Vobecky, S., Haddad, E., Lemieux, S., Thibault, P., & Perreault, C. (2018). Noncoding regions are the main source of targetable tumor-specific antigens. *Science Translational Medicine*, *10*(470). https://doi.org/10.1126/scitranslmed.aau5516

Lee, S. P., Chan, A. T. C., Cheung, S.-T., Thomas, W. A., CroomCarter, D., Dawson, C. W., Tsai, C.-H., Leung, S.-F., Johnson, P. J., & Huang, D. P. (2000). CTL Control of EBV in Nasopharyngeal Carcinoma (NPC): EBV-Specific CTL Responses in the Blood and Tumors of NPC Patients and the Antigen-Processing Function of the Tumor Cells. *The Journal of Immunology*, *165*(1), 573–582. https://doi.org/10.4049/jimmunol.165.1.573

Leidner, R., Sanjuan Silva, N., Huang, H., Sprott, D., Zheng, C., Shih, Y.-P., Leung, A., Payne, R., Sutcliffe, K., Cramer, J., Rosenberg, S. A., Fox, B. A., Urba, W. J., & Tran, E. (2022). Neoantigen T-Cell Receptor Gene Therapy in Pancreatic Cancer. *The New England Journal of Medicine*, *386*(22), 2112–2119. https://doi.org/10.1056/NEJMoa2119662

Leister, H., Krause, F. F., Mahdavi, R., Steinhoff, U., & Visekruna, A. (2022). The Role of Immunoproteasomes in Tumor-Immune Cell Interactions in Melanoma and Colon Cancer. In *Archivum Immunologiae et Therapiae Experimentalis* (Vol. 70, Issue 1). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s00005-022-00644-x

Leister, H., Luu, M., Staudenraus, D., Krol Lopez, A., Mollenkopf, H. J., Sharma, A., Schmerer, N., Schulte, L. N., Bertrams, W., Schmeck, B., Bosmann, M., Steinhoff, U., & Visekruna, A. (2021). Pro- And antitumorigenic capacity of immunoproteasomes in shaping the tumor microenvironment. *Cancer Immunology Research*, *9*(6), 682–692. https://doi.org/10.1158/2326-6066.CIR-20-0492

Leko, V., & Rosenberg, S. A. (2020). Identifying and Targeting Human Tumor Antigens for T Cell-Based Immunotherapy of Solid Tumors. In *Cancer Cell* (Vol. 38, Issue 4, pp. 454–472). Cell Press. https://doi.org/10.1016/j.ccell.2020.07.013

Lenschow, D. J., Walunas, T. L., & Bluestone, J. A. (1996). CD28/B7 SYSTEM OF T CELL COSTIMULATION. In *Annu. Rev. Immunol* (Vol. 14). www.annualreviews.org

Li, G., Bethune, M. T., Wong, S., Joglekar, A. v., Leonard, M. T., Wang, J. K., Kim, J. T., Cheng, D., Peng, S., Zaretsky, J. M., Su, Y., Luo, Y., Heath, J. R., Ribas, A., Witte, O. N., & Baltimore, D. (2019). T cell antigen discovery via trogocytosis. *Nature Methods*, *16*(2), 183–190. https://doi.org/10.1038/s41592-018-0305-7

Li, Y., Bleakley, M., & Yee, C. (2005). IL-21 Influences the Frequency, Phenotype, and Affinity of the Antigen-Specific CD8 T Cell Response. *The Journal of Immunology*, *175*(4), 2261–2269. https://doi.org/10.4049/jimmunol.175.4.2261

Liepe, J., Marino, F., Sidney, J., Jeko, A., Bunting, D. E., Sette, A., Kloetzel, P. M., Stumpf, M. P. H., Heck, A. J. R., & Mishto, M. (n.d.). *A large fraction of HLA class I ligands are proteasome-generated spliced peptides*.

Liepe J, Marino F, Sidney J, Jeko A, Bunting D, Sette A, Kloetzel P, Stumpf M, Heck A, & Mishto M. (2016). A large fraction of HLA class Iligands are proteasome-generated spliced peptides. *Science*, *354*(6310), 350–358. https://doi.org/10.1126/science.aag3267

Linette, G. P., Stadtmauer, E. A., Maus, M. v, Rapoport, A. P., Levine, B. L., Emery, L., Litzky, L.,

Bagg, A., Carreno, B. M., Cimino, P. J., Binder-Scholl, G. K., Smethurst, D. P., Gerry, A. B., Pumphrey, N. J., Bennett, A. D., Brewer, J. E., Dukes, J., Harper, J., Tayton-Martin, H. K., … June, C. H. (2013). *Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma*. https://doi.org/10.1182/blood-2013-03

Lu, Y. C., Yao, X., Crystal, J. S., Li, Y. F., El-Gamil, M., Gross, C., Davis, L., Dudley, M. E., Yang, J. C., Samuels, Y., Rosenberg, S. A., & Robbins, P. F. (2014). Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions. *Clinical Cancer Research*, *20*(13), 3401–3410. https://doi.org/10.1158/1078-0432.CCR-14-0433

Luckey, C. J., Marto, J. A., Partridge, M., Hall, E., White, F. M., Lippolis, J. D., Shabanowitz, J., Hunt, D. F., & Engelhard, V. H. (2001). Differences in the Expression of Human Class I MHC Alleles and Their Associated Peptides in the Presence of Proteasome Inhibitors. *The Journal of Immunology*, *167*(3), 1212–1221. https://doi.org/10.4049/jimmunol.167.3.1212

Lupetti, R., Pisarra, P., Verrecchia, A., Farina, C., Nicolini, G., Anichini, A., Bordignon, C., Sensi, M., Parmiani, G., & Traversari, C. (1998). Translation of a Retained Intron in Tyrosinase-related Protein (TRP) 2 mRNA Generates a New Cytotoxic T Lymphocyte (CTL)-defined and Shared Human Melanoma Antigen Not Expressed in Normal Cells of the Melanocytic Lineage. In *J. Exp. Med* (Vol. 188, Issue 6). http://www.jem.org

Malekzadeh, P., Yossef, R., Cafri, G., Paria, B. C., Lowery, F. J., Jafferji, M., Good, M. L., Sachs, A., Copeland, A. R., Kim, S. P., Kivitz, S., Parkhurst, M. R., Robbins, P. F., Ray, S., Xi, L., Raffeld, M., Yu, Z., Restifo, N. P., Somerville, R. P. T., … Deniger, D. C. (2020). Antigen experienced T cells from peripheral blood recognize p53 neoantigens. *Clinical Cancer Research*, *26*(6), 1267–1276. https://doi.org/10.1158/1078-0432.CCR-19-1874

Marcu, A., Bichmann, L., Kuchenbecker, L., Kowalewski, D. J., Freudenmann, L. K., Backert, L., Mühlenbruch, L., Szolek, A., Lübke, M., Wagner, P., Engler, T., Matovina, S., Wang, J., Hauri-Hohl, M., Martin, R., Kapolou, K., Walz, J. S., Velz, J., Moch, H., … Neidert, M. C. (2021). HLA Ligand Atlas: A benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. *Journal for ImmunoTherapy of Cancer*, *9*(4). https://doi.org/10.1136/jitc-2020-002071

Maus, M. v., & June, C. H. (2016). Making better chimeric antigen receptors for adoptive T-cell therapy. *Clinical Cancer Research*, *22*(8), 1875–1884. https://doi.org/10.1158/1078-0432.CCR-15-1433

Mayer, J., Blomberg, J., & Seal, R. L. (2011). A revised nomenclature for transcribed human endogenous retroviral loci. *Mobile DNA*, *2*(1). https://doi.org/10.1186/1759-8753-2-7

Mei, S., Li, F., Leier, A., Marquez-Lago, T. T., Giam, K., Croft, N. P., Akutsu, T., Ian Smith, A., Li, J., Rossjohn, J., Purcell, A. W., & Song, J. (2019). A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction. In *Briefings in Bioinformatics* (Vol. 21, Issue 4, pp. 1119–1135). Oxford University Press. https://doi.org/10.1093/bib/bbz051

Michaux, A., Larrieu, P., Stroobant, V., Fonteneau, J.-F., Jotereau, F., van den Eynde, B. J., Moreau-Aubry, A., & Vigneron, N. (2014). A Spliced Antigenic Peptide Comprising a Single Spliced Amino Acid Is Produced in the Proteasome by Reverse Splicing of a Longer Peptide Fragment followed by Trimming. *The Journal of Immunology*, *192*(4), 1962–1971. https://doi.org/10.4049/jimmunol.1302032

Milner, E., Barnea, E., Beer, I., & Admon, A. (2006). The turnover kinetics of major histocompatibility complex peptides of human cancer cells. *Molecular and Cellular Proteomics*, *5*(2), 357–365. https://doi.org/10.1074/mcp.M500241-MCP200

Morgan, R. A., Dudley, M. E., Wunderlich, J. R., Hughes, M. S., Yang, J. C., Sherry, R. M., Royal, R. E., Topalian, S. L., Kammula, U. S., Restifo, N. P., Zheng, Z., Nahvi, A., de Vries, C. R., Rogers-Freezer, L. J., Mavroukakis, S. A., & Rosenberg, S. A. (2006). Cancer Regression in Patients After Transfer of Genetically Engineered Lymphocytes. *Science*, *314*(5796), 126–129. www.sciencemag.org/cgi/content/full/1129003/DC1

Morris, G. P., & Allen, P. M. (2012). How the TCR balances sensitivity and specificity for the recognition of self and pathogens. In *Nature Immunology* (Vol. 13, Issue 2, pp. 121–128). https://doi.org/10.1038/ni.2190

Mur, P., García-Mulero, S., del Valle, J., Magraner-Pardo, L., Vidal, A., Pineda, M., Cinnirella, G., Martín-Ramos, E., Pons, T., López-Doriga, A., Belhadj, S., Feliubadaló, L., Munoz-Torres, P. M., Navarro, M., Grau, E., Darder, E., Llort, G., Sanz, J., Ramón Cajal, T., … Valle, L. (2020). *Role of POLE and POLD1 in familial cancer*. https://doi.org/10.1038/s41436

Muul, L. M., Spiess, P. J., Director, E. P., & Rosenberg, S. A. (1987). Identification of specific cytolytic immune responses against autologous tumor in humans bearing malignant melanoma. *The Journal of Immunology*, *138*, 989–995. http://www.jimmunol.org/content/138/3/989

Mylonas, R., Beer, I., Iseli, C., Chong, C., Pak, H. S., Gfeller, D., Coukos, G., Xenarios, I., Müller, M., & Bassani-Sternberg, M. (2018). Estimating the contribution of proteasomal spliced peptides to the HLA-I ligandome. *Molecular and Cellular Proteomics*, *17*(12), 2347–2357. https://doi.org/10.1074/mcp.RA118.000877

Neefjes, J., Jongsma, M. L. M., Paul, P., & Bakke, O. (2011). Towards a systems understanding of MHC class i and MHC class II antigen presentation. In *Nature Reviews Immunology* (Vol. 11, Issue 12, pp. 823–836). https://doi.org/10.1038/nri3084

Nelde, A., Flötotto, L., Jürgens, L., Szymik, L., Hubert, E., Bauer, J., Schliemann, C., Kessler, T., Lenz, G., Rammensee, H. G., Walz, J. S., & Wethmar, K. (2022). Upstream open reading frames regulate translation of cancer-associated transcripts and encode HLA-presented immunogenic tumor antigens. *Cellular and Molecular Life Sciences*, *79*(3). https://doi.org/10.1007/s00018-022-04145-0

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. *Journal for ImmunoTherapy of Cancer*, *7*(1). https://doi.org/10.1186/s40425-019-0769-8

Norbury, C. C. (2016). Defining cross presentation for a wider audience. In *Current Opinion in Immunology* (Vol. 40, pp. 110–116). Elsevier Ltd. https://doi.org/10.1016/j.coi.2016.04.003

Norbury, C. C., Basta, S., Donohue, K. B., Tscharke, D. C., Princiotta, M. F., Berglund, P., Gibbs, J., Bennink, J. R., & Yewdell, J. W. (2004). CD8 + T Cell Cross-Priming via Transfer of Proteasome Substrates. In *New Series* (Vol. 304, Issue 5675).

Olson, B. M., Frye, T. P., Johnson, L. E., Fong, L., Knutson, K. L., Disis, M. L., & McNeel, D. G. (2010). HLA-A2-restricted T-cell epitopes specific for prostatic acid phosphatase. *Cancer Immunology, Immunotherapy*, *59*(6), 943–953. https://doi.org/10.1007/s00262-010-0820-6

Ouspenskaia, T., Law, T., Clauser, K. R., Klaeger, S., Sarkizova, S., Aguet, F., Li, B., Christian, E., Knisbacher, B. A., Le, P. M., Hartigan, C. R., Keshishian, H., Apffel, A., Oliveira, G., Zhang, W., Chen, S., Chow, Y. T., Ji, Z., Jungreis, I., … Regev, A. (2022). Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nature Biotechnology*, *40*(2), 209–217. https://doi.org/10.1038/s41587-021-01021-3

Parkhurst, M. R., Joo, J., Riley, J. P., Yu, Z., Li, Y., Robbins, P. F., & Rosenberg, S. A. (2009). Characterization of genetically modified T-cell receptors that recognize the CEA:691-699 peptide in the context of HLA-A2.1 on human colorectal cancer cells. *Clinical Cancer Research*, *15*(1), 169–180. https://doi.org/10.1158/1078-0432.CCR-08-1638

Parkhurst, M. R., Robbins, P. F., Tran, E., Prickett, T. D., Gartner, J. J., Li, J., Ivey, G., Li, Y. F., El-Gamil, M., Lalani, A., Crystal, J. S., Sachs, A., Groh, E., Ray, S., Ngo, L. T., Kivitz, S., Pasetto, A., Yossef, R., Lowery, F. J., … Rosenberg, S. A. (2019). Unique neoantigens arise from somatic mutations in patients with gastrointestinal cancers. *Cancer Discovery*, *9*(8), 1022–1035. https://doi.org/10.1158/2159-8290.CD-18-1494

Parkhurst, M. R., Yang, J. C., Langan, R. C., Dudley, M. E., Nathan, D. A. N., Feldman, S. A., Davis, J. L., Morgan, R. A., Merino, M. J., Sherry, R. M., Hughes, M. S., Kammula, U. S., Phan, G. Q., Lim, R. M., Wank, S. A., Restifo, N. P., Robbins, P. F., Laurencot, C. M., & Rosenberg, S. A. (2011). T cells targeting carcinoembryonic antigen can mediate regression of metastatic colorectal cancer but induce severe tran-

sient colitis. *Molecular Therapy*, *19*(3), 620–626. https://doi.org/10.1038/mt.2010.272

Pataskar, A., Champagne, J., Nagel, R., Kenski, J., Laos, M., Michaux, J., Pak, H. S., Bleijerveld, O. B., Mordente, K., Navarro, J. M., Blommaert, N., Nielsen, M. M., Lovecchio, D., Stone, E., Georgiou, G., de Gooijer, M. C., van Tellingen, O., Altelaar, M., Joosten, R. P., … Agami, R. (2022). Tryptophan depletion results in tryptophan-to-phenylalanine substitutants. *Nature*, *603*(7902), 721–727. https://doi.org/10.1038/s41586-022-04499-2

Penn, I. (1999). Posttransplant Malignancies. *Transplantation Proceedings*, *31*, 1260–1262.

Piraino, S. W., & Furney, S. J. (2016). Beyond the exome: The role of non-coding somatic mutations in cancer. *Annals of Oncology*, *27*(2), 240–248. https://doi.org/10.1093/annonc/mdv561

Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., & Sandhu, M. S. (2018). Long reads: Their purpose and place. In *Human Molecular Genetics* (Vol. 27, Issue R2, pp. R234–R241). Oxford University Press. https://doi.org/10.1093/hmg/ddy177

Poullion, T., Aknouche, N., Gaudin, C., & Ronsin, F. C. (1999). Lymphocytes In Situ Carcinoma-Reactive Tumor-Infiltrating Recognized by Renal Cell Esterase mRNA Generates an Epitope Reading Frame of the Intestinal Carboxyl A Non-AUG-Defined Alternative Open. In *J Immunol References* (Vol. 483). http://www.jimmunol.org/content/163/1/http://www.jimmunol.org/content/163/1/483.full#ref-list-1

Probst-Kepper, M., Stroobant, V., Kridel, R., Gaugler, B., Landry, C., Brasseur, F., Cosyns, J.-P., Weynand, B., Boon, T., & van den Eynde, B. J. (2001). An Alternative Open Reading Frame of the Human Macrophage Colony-stimulating Factor Gene Is Independently Translated and Codes for an Antigenic Peptide of 14 Amino Acids Recognized by Tumor-infiltrating CD8 T Lymphocytes. In *J. Exp. Med* (Vol. 193, Issue 10). http://www.jem.org/cgi/content/full/193/10/1189

Quezada, S. A., & Peggs, K. S. (2019). Lost in translation: Deciphering the mechanism of action of anti-human CTLA-4. In *Clinical Cancer Research* (Vol. 25, Issue 4, pp. 1130–1132). American Association for Cancer Research Inc. https://doi.org/10.1158/1078-0432.CCR-18-2509

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2021). NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, *48*(W1), W449–W454. https://doi.org/10.1093/NAR/GKAA379

Rezvani, K., Brenchley, J. M., Price, D. A., Kilical, Y., Gostick, E., Sewell, A. K., Li, J., Mielke, S., Douek, D. C., & Barrett, A. J. (2005). T-cell responses directed against multiple HLA-A*0201-restricted epitopes derived from Wilms' tumor 1 protein in patients with leukemia and healthy donors: Identification, quantification, and characterization. *Clinical Cancer Research*, *11*(24), 8799–8807. https://doi.org/10.1158/1078-0432.CCR-05-1314

Rheinbay, E., Nielsen, M. M., Abascal, F., Wala, J. A., Shapira, O., Tiao, G., Hornshøj, H., Hess, J. M., Juul, R. I., Lin, Z., Feuerbach, L., Sabarinathan, R., Madsen, T., Kim, J., Mularoni, L., Shuai, S., Lanzós, A., Herrmann, C., Maruvka, Y. E., … Zamora, J. (2020). Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature*, *578*(7793), 102–111. https://doi.org/10.1038/s41586-020-1965-x

Rizvi, S. M., Salam, N., Geng, J., Qi, Y., Bream, J. H., Duggal, P., Hussain, S. K., Martinson, J., Wolinsky, S. M., Carrington, M., & Raghavan, M. (2014). Distinct Assembly Profiles of HLA-B Molecules. *The Journal of Immunology*, *192*(11), 4967–4976. https://doi.org/10.4049/jimmunol.1301670

Robbins, P., El-Gamil, M., Li, Y. F., Fitzgerald, E. B., & Rosenberg SA. (1997). *tumor-infiltrating lymphocytes recognized by melanoma-reactive gp100 gene transcript encodes an epitope The intronic region of an incompletely spliced*. http://www.jimmunol.org/content/159/1/303

Robbins, P. F., Morgan, R. A., Feldman, S. A., Yang, J. C., Sherry, R. M., Dudley, M. E., Wunderlich, J. R., Nahvi, A. v., Helman, L. J., Mackall, C. L., Kammula, U. S., Hughes, M. S., Restifo, N. P., Raffeld, M., Lee, C. C. R., Levy, C. L., Li, Y. F., El-Gamil, M., Schwarz, S. L., … Rosenberg, S. A. (2011). Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. *Journal of Clinical Oncology*, *29*(7), 917–924. https://doi.

org/10.1200/JCO.2010.32.2537

Rong-Fu Wang, B., Kawakami, Y., Robbins, P. F., & Rosenberg, S. A. (1996). *Utilization of an Alternative Open Reading Frame of a Normal Gene in Generating a Novel Human Cancer Antigen*. https://rupress.org/jem/article-pdf/183/3/1131/499993/1131.pdf

Rosenberg, S. A. (2014). IL-2: The First Effective Immunotherapy for Human Cancer. *The Journal of Immunology*, *192*(12), 5451–5458. https://doi.org/10.4049/jimmunol.1490019

Rosenberg, S. A., & Restifo, N. P. (2015). Adoptive cell transfer as personalized immunotherapy for human cancer. In *Science* (Vol. 348, Issue 6230, pp. 62–68). American Association for the Advancement of Science. https://doi.org/10.1126/science.aaa4967

Rosenberg, S. A., Tong-On, P., Li, Y., Riley, J. P., El-Gamil, M., Parkhurst, M. R., & Robbins, P. F. (2002). Identification of BING-4 Cancer Antigen Translated From an Alternative Open Reading Frame of a Gene in the Extended MHC Class II Region Using Lymphocytes From a Patient With a Durable Complete Regression Following Immunotherapy. *The Journal of Immunology*, *168*(5), 2402–2407. https://doi.org/10.4049/jimmunol.168.5.2402

Rosenberg, S. A., Yang, J. C., & Restifo Steven A Rosenberg, N. P. (2004). Cancer immunotherapy: moving beyond current vaccines. *Nature Medicine*, *10(9)*, 909–915.

Rosenberg, S., Lotze, M., Muul, L., Leitman, S., Chang, A., Ettinghausen, S., Matory, Y., Skibber, J., Shiloni, E., & Vetto, J. (1985). Observations on the systemic administration of autologous lymphokine-activated killer cells and rIL2 to patients with metastatic cancer. *New England Journal of Medicine*, 1485–1492.

Rosenberg, S., Packard, B., Aebersold, P., Solomon, D., Topalian, D., Toy, D., Simon, P., Lotze, M., Yang, J., Seipp, C., Steven Bock, C., Schwartzentruber, D., & Wei, J. (1988). Use of Tumor-Infiltrating Lymphocytes and Interleukin-2 in the Immunotherapy of Patients with Metastatic Melanoma. *New England Journal of Medicine*, *319*, 1676–1680.

Ruiz Cuevas, M. V., Hardy, M. P., Hollý, J., Bonneil, É., Durette, C., Courcelles, M., Lanoix, J., Côté, C., Staudt, L. M., Lemieux, S., Thibault, P., Perreault, C., & Yewdell, J. W. (2021). Most non-canonical proteins uniquely populate the proteome or immunopeptidome. *Cell Reports*, *34*(10). https://doi.org/10.1016/j.celrep.2021.108815

Rygaard, J., & Povlsen, C. O. (1974). THE MOUSE MUTANT NUDE DOES NOT DEVELOP SPONTANEOUS TUMOURS. An Argument Against Immunological Surveillance. *Acta Pathologica Microbiologica Scandinavica Section B Microbiology and Immunology*, *82 B*(1), 99–106. https://doi.org/10.1111/j.1699-0463.1974.tb02299.x

Sallusto, F., Geginat, J., & Lanzavecchia, A. (2004). Central memory and effector memory T cell subsets: Function, generation, and maintenance. In *Annual Review of Immunology* (Vol. 22, pp. 745–763). https://doi.org/10.1146/annurev.immunol.22.012703.104702

Samstein, R. M., Lee, C. H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., Barron, D. A., Zehir, A., Jordan, E. J., Omuro, A., Kaley, T. J., Kendall, S. M., Motzer, R. J., Hakimi, A. A., Voss, M. H., Russo, P., Rosenberg, J., Iyer, G., Bochner, B. H., … Morris, L. G. T. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. In *Nature Genetics* (Vol. 51, Issue 2, pp. 202–206). Nature Publishing Group. https://doi.org/10.1038/s41588-018-0312-8

Schmidt, J., Smith, A. R., Magnin, M., Racle, J., Devlin, J. R., Bobisse, S., Cesbron, J., Bonnet, V., Carmona, S. J., Huber, F., Ciriello, G., Speiser, D. E., Bassani-Sternberg, M., Coukos, G., Baker, B. M., Harari, A., & Gfeller, D. (2021). Prediction of neo-epitope immunogenicity reveals TCR recognition determinants and provides insight into immunoediting. *Cell Reports Medicine*, *2*(2). https://doi.org/10.1016/j.xcrm.2021.100194

Scull, K. E., Pandey, K., Ramarathinam, S. H., & Purcell, A. W. (2021). Immunopeptidogenomics: Harnessing RNA-Seq to illuminate the dark immunopeptidome. *Molecular and Cellular Proteomics*, *20*. https://doi.org/10.1016/j.mcpro.2021.100143

Sette, C., & Paronetto, M. P. (2022). Somatic Mutations in Core Spliceosome Components Promote

Tumorigenesis and Generate an Exploitable Vulnerability in Human Cancer. In *Cancers* (Vol. 14, Issue 7). MDPI. https://doi.org/10.3390/cancers14071827

Shankaran, V., Ikeda, H., Bruce, A., White, M., Swanson, P., Old, J., & Schreiber, R. (2001). IFNg and lymphocytes prevent primary tumour development and shape tumor immunogenicity. *Nature*, *410*, 1107–1111.

Sharpe, A. H. (2017). Introduction to checkpoint inhibitors and cancer immunotherapy. In *Immunological Reviews* (Vol. 276, Issue 1, pp. 5–8). Blackwell Publishing Ltd. https://doi.org/10.1111/imr.12531

Skipper, J. A., Hendrickson, R. C., Gulden, P. H., Brichard, V., Pel, A., Shabanowitz, J., Wolfel, T., Slingluff, C. L., Boon, T., Hunt, D. F., & andVictor Engelhard, I. H. (1996). *An HLA-A2-restricted Tyrosinase Antigen on Melanoma Cells Results from Posttranslational Modification and Suggests a Novel Pathway for Processing of Membrane Proteins*.

Smart, A. C., Margolis, C. A., Pimentel, H., He, M. X., Miao, D., Adeegbe, D., Fugmann, T., Wong, K. K., & van Allen, E. M. (2018). Intron retention is a source of neoepitopes in cancer. *Nature Biotechnology*, *36*(11), 1056–1063. https://doi.org/10.1038/nbt.4239

Smith, C. C., Chai, S., Washington, A. R., Lee, S. J., Landoni, E., Field, K., Garness, J., Bixby, L. M., Selitsky, S. R., Parker, J. S., Savoldo, B., Serody, J. S., & Vincent, B. G. (2019). Machine-learning prediction of tumor antigen immunogenicity in the selection of therapeutic epitopes. *Cancer Immunology Research*, *7*(10), 1591–1604. https://doi.org/10.1158/2326-6066.CIR-19-0155

Stevanović, S., Pasetto, A., Helman, S. R., Gartner, J. J., Prickett, T. D., Howie, B., Robins, H. S., Robbins, P. F., Klebanoff, C. A., Rosenberg, S. A., & Hinrichs, C. S. (2017). Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. *Science*, *356*(6334), 200–205. https://doi.org/10.1126/science.aak9510

Strominger, J. L. (2002). Human histocompatibility proteins Past and future: a memoir. In *Immunological Reviews* (Vol. 185).

Thomas, L. (1959). Discussion. In Cellular and Humoral Aspects of the Hypersensitive States. In *Shaftesbury Ave* (Issue 9). http://jama.jamanetwork.com/

Topalian, R., Schwartzentruber, D. J., Rong-Fu, S. A., Wang, S. L., Johnston, G., & Zeng, S. L. (1998). *Reading Frames Peptides Translated from Different Open Antigen: T Cell Responses to Antigenic A Breast and Melanoma-Shared Tumor*. http://www.jimmunol.org/content/161/7/3596

Tran, E., Ahmadzadeh, M., Lu, Y. C., Gros, A., Turcotte, S., Robbins, P. F., Gartner, J. J., Zheng, Z., Li, Y. F., Ray, S., Wunderlich, J. R., Somerville, R. P., & Rosenberg, S. A. (2015). Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science*, *350*(6266), 1387–1390. https://doi.org/10.1126/science.aad1253

Tran, E., Robbins, P. F., Lu, Y.-C., Prickett, T. D., Gartner, J. J., Jia, L., Pasetto, A., Zheng, Z., Ray, S., Groh, E. M., Kriley, I. R., & Rosenberg, S. A. (2016). T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *New England Journal of Medicine*, *375*(23), 2255–2262. https://doi.org/10.1056/nejmoa1609279

Tran, E., Turcotte, S., Gros, A., Robbins, P. F., Lu, Y. C., Dudley, M. E., Wunderlich, J. R., Somerville, R. P., Hogan, K., Hinrichs, C. S., Parkhurst, M. R., Yang, J. C., & Rosenberg, S. A. (2014). Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science*, *344*(6184), 641–645. https://doi.org/10.1126/science.1251102

van den Eynde, B. J., Gaugler, B., Probst-Kepper, M., Michaux, L., Devuyst, O., Lorge, F., Weynants, P., & Boon, T. (1999). A New Antigen Recognized by Cytolytic T Lymphocytes on a Human Kidney Tumor Results from Reverse Strand Transcription. In *J. Exp. Med* (Vol. 190). http://www.jem.org

van der Bruggen, P., Traversari, C., Chomez, P., Lurquin, C., de Plaen, E., van den Eynde, B., Knuth, A., & Boon, T. (1991). A Gene Encoding an Antigen Recognized by Cytolytic T Lymphocytes on a Human. In *New Series* (Vol. 254, Issue 5038).

van Deutekom, H. W. M., & Keşmir, C. (2015). Zooming into the binding groove of HLA molecules: which positions and which substitutions change peptide binding most? *Immunogenetics*, *67*(8), 425–436. https://doi.org/10.1007/s00251-015-0849-y

Verkerk, T., Koomen, S. J. I., Fuchs, K. J., Griffioen, M., & Spaapen, R. M. (2022). *An unexplored angle: T cell antigen discoveries reveal a marginal contribution of proteasome splicing to the immunogenic MHC class I antigen pool*. https://doi.org/10.1073/pnas.2119736119/-/DCSupplemental

Wang, T. Y., Liu, Q., Ren, Y., Alam, S. K., Wang, L., Zhu, Z., Hoeppner, L. H., Dehm, S. M., Cao, Q., & Yang, R. (2021). A pan-cancer transcriptome analysis of exitron splicing identifies novel cancer driver genes and neoepitopes. *Molecular Cell*, *81*(10), 2246-2260.e12. https://doi.org/10.1016/j.molcel.2021.03.028

Warren, E. H., Vigneron, N. J., Gavin, M. A., Coulie, P. G., Stroobant, V., Dalet, A., Tykodi, S. S., Xuereb, S. M., Mito, J. K., Riddell, S. R., & van den Eynde, B. J. (2006). An Antigen Produced by Splicing of Noncontiguous Peptides in the Reverse Order. In *New Series* (Vol. 313, Issue 5792).

Wei, S. C., Duffy, C. R., & Allison, J. P. (2018). Fundamental mechanisms of immune checkpoint blockade therapy. In *Cancer Discovery* (Vol. 8, Issue 9, pp. 1069–1086). American Association for Cancer Research Inc. https://doi.org/10.1158/2159-8290.CD-18-0367

Westcott, P. M. K., Sacks, N. J., Schenkel, J. M., Ely, Z. A., Smith, O., Hauck, H., Jaeger, A. M., Zhang, D., Backlund, C. M., Beytagh, M. C., Patten, J. J., Elbashir, R., Eng, G., Irvine, D. J., Yilmaz, O. H., & Jacks, T. (2021). Low neoantigen expression and poor T-cell priming underlie early immune escape in colorectal cancer. *Nature Cancer*, *2*(10), 1071–1085. https://doi.org/10.1038/s43018-021-00247-z

Xiang, H., Zhang, L., Bu, F., Guan, X., Chen, L., Zhang, H., Zhao, Y., Chen, H., Zhang, W., Li, Y., Lee, L. J., Mei, Z., Rao, Y., Gu, Y., Hou, Y., Mu, F., & Dong, X. (2022). A Novel Proteogenomic Integration Strategy Expands the Breadth of Neo-Epitope Sources. *Cancers*, *14*(12), 3016. https://doi.org/10.3390/cancers14123016

Yarchoan, M., Hopkins, A., & Jaffee, E. M. (2017). Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *New England Journal of Medicine*, *377*(25), 2500–2501. https://doi.org/10.1056/nejmc1713444

Yewdell, J. W. (2011). DRiPs solidify: Progress in understanding endogenous MHC class I antigen processing. In *Trends in Immunology* (Vol. 32, Issue 11, pp. 548–558). https://doi.org/10.1016/j.it.2011.08.001

Zacharakis, N., Chinnasamy, H., Black, M., Xu, H., Lu, Y. C., Zheng, Z., Pasetto, A., Langhan, M., Shelton, T., Prickett, T., Gartner, J., Jia, L., Trebska-Mcgowan, K., Somerville, R. P., Robbins, P. F., Rosenberg, S. A., Goff, S. L., & Feldman, S. A. (2018). Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nature Medicine*, *24*(6), 724–730. https://doi.org/10.1038/s41591-018-0040-8

Zacharakis, N., Lutfi, ;, Huq, M., Seitter, S. J., Kim, S. P., Gartner, J. J., Sindiri, S., Hill, V. K., Li, Y. F., Paria, B. C., Ray, S., Gasmi, B., Chyi-Chia Lee, ;, Todd, ;, Prickett, D., Parkhurst, M. R., Robbins, P. F., Langhan, M. M., Shelton, T. E., … Rosenberg, S. A. (2022). Breast Cancers Are Immunogenic: Immunologic Analyses and a Phase II Pilot Clinical Trial Using Mutation-Reactive Autologous Lymphocytes. *J Clin Oncol*, *40*, 1741–1754. https://doi.org/10.1200/JCO.21

Zhang, M., Fritsche, J., Roszik, J., Williams, L. J., Peng, X., Chiu, Y., Tsou, C. C., Hoffgaard, F., Goldfinger, V., Schoor, O., Talukder, A., Forget, M. A., Haymaker, C., Bernatchez, C., Han, L., Tsang, Y. H., Kong, K., Xu, X., Scott, K. L., … Hwu, P. (2018a). RNA editing derived epitopes function as cancer antigens to elicit immune responses. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-06405-9

Zhang, M., Fritsche, J., Roszik, J., Williams, L. J., Peng, X., Chiu, Y., Tsou, C. C., Hoffgaard, F., Goldfinger, V., Schoor, O., Talukder, A., Forget, M. A., Haymaker, C., Bernatchez, C., Han, L., Tsang, Y. H., Kong, K., Xu, X., Scott, K. L., … Hwu, P. (2018b). RNA editing derived epitopes function as cancer antigens to elicit immune responses. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-06405-9

Zhang, M., Fritsche, J., Roszik, J., Williams, L. J., Peng, X., Chiu, Y., Tsou, C. C., Hoffgaard, F., Goldfinger, V., Schoor, O., Talukder, A., Forget, M. A., Haymaker, C., Bernatchez, C., Han, L., Tsang, Y. H., Kong, K., Xu, X., Scott, K. L., … Hwu, P. (2018c). RNA editing derived epitopes function as cancer antigens to elicit immune responses. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-

06405-9

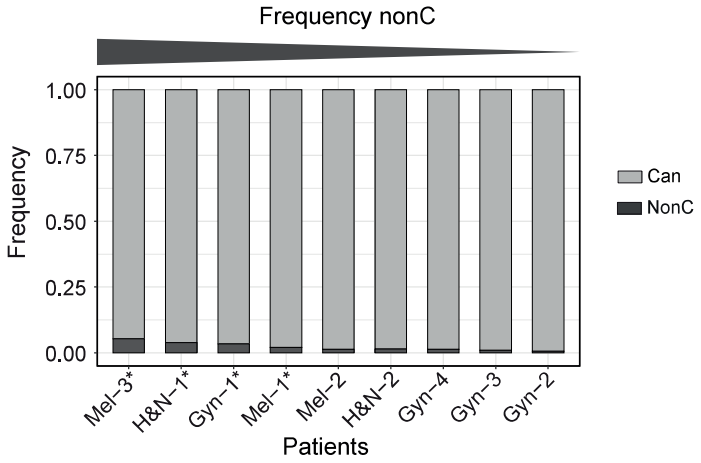
Zhao, Q., Laverdure, J. P., Lanoix, J., Durette, C., Cote, C., Bonneil, E., Laumont, C. M., Gendron, P., Vincent, K., Courcelles, M., Lemieux, S., Millar, D. G., Ohashi, P. S., Thibault, P., & Perreault, C. (2020). Proteogenomics uncovers a vast repertoire of shared tumor-specific antigens in ovarian cancer. *Cancer Immunology Research*, *8*(4), 544–555. https://doi.org/10.1158/2326-6066.CIR-19-0541

Zheng, C., Fass, J. N., Shih, Y.-P., Gunderson, A. J., Sanjuan Silva, N., Huang, H., Bernard, B. M., Rajamanickam, V., Slagel, J., Bifulco, C. B., Piening, B., Newell, P. H. A., Hansen, P. D., & Tran, E. (2022). Transcriptomic profiles of neoantigen-reactive T cells in human gastrointestinal cancers. *Cancer Cell*, *40*(4), 410-423.e7. https://doi.org/10.1016/j.ccell.2022.03.005

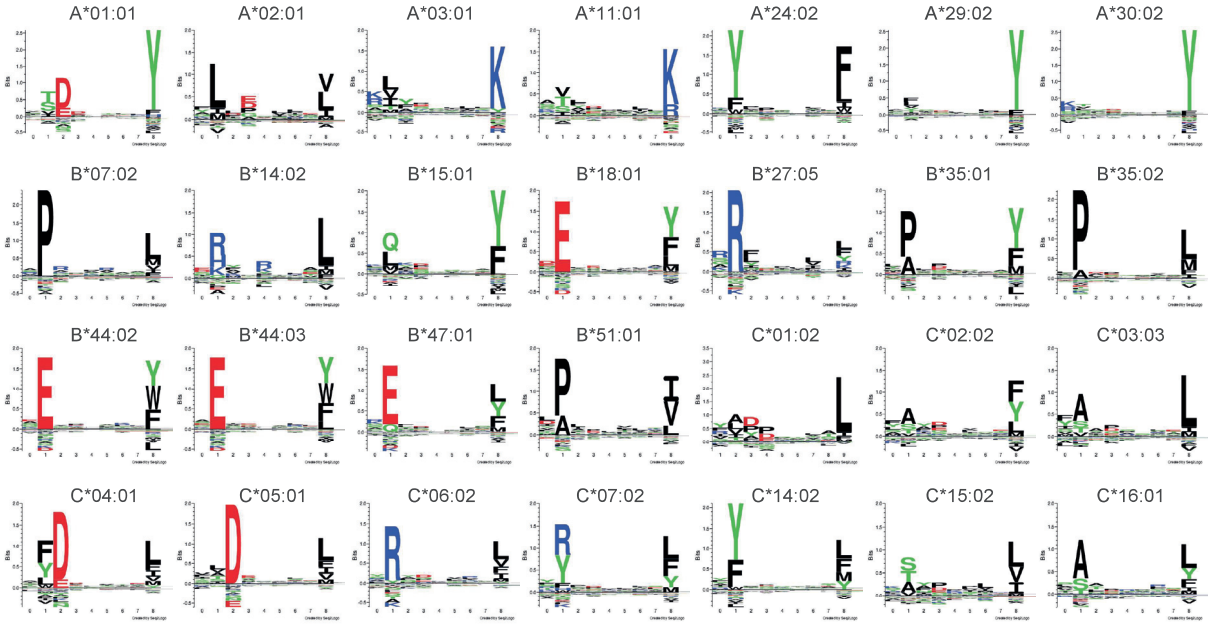
Appendix

Abbreviations tumor type

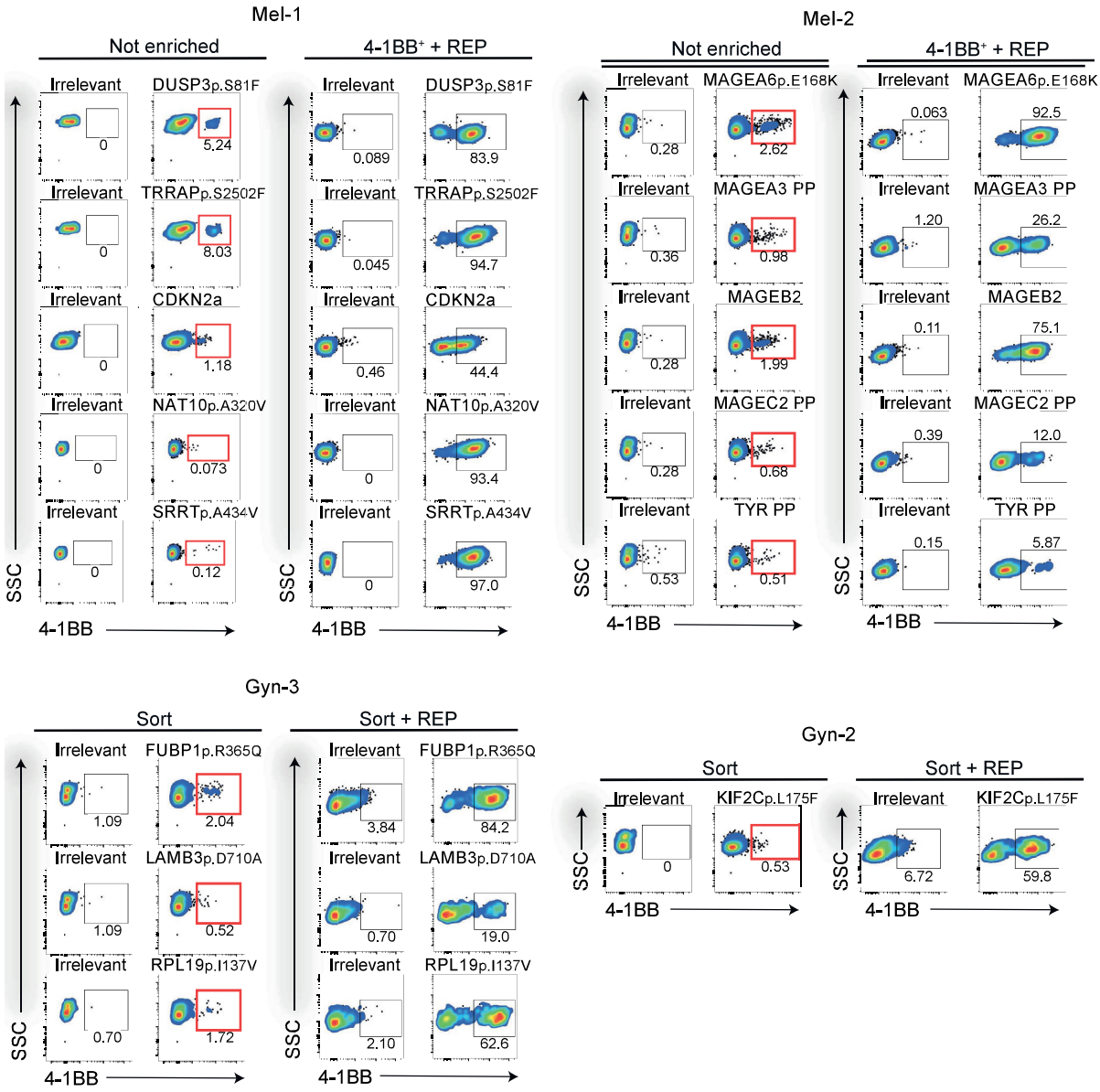
ACC	Adrenocortical carcinoma
BLCA	Bladder Urothelial Carcinoma
BRCA	Breast invasive carcinoma
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL	Cholangio carcinoma
COAD	Colon adenocarcinoma
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
ESCA	Esophageal carcinoma
GBM	Glioblastoma multiforme
HNSC	Head and Neck squamous cell carcinoma
KICH	Kidney Chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LAML	Acute Myeloid Leukemia
LGG	Brain Lower Grade Glioma
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
MESO	Mesothelioma
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PCPG	Pheochromocytoma and Paraganglioma
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
SARC	Sarcoma
SKCM	Skin Cutaneous Melanoma
STAD	Stomach adenocarcinoma
TGCT	Testicular Germ Cell Tumors
THCA	Thyroid carcinoma
THYM	Thymoma
UCEC	Uterine Corpus Endometrial Carcinoma
UCS	Uterine Carcinosarcoma
UVM	Uveal Melanoma



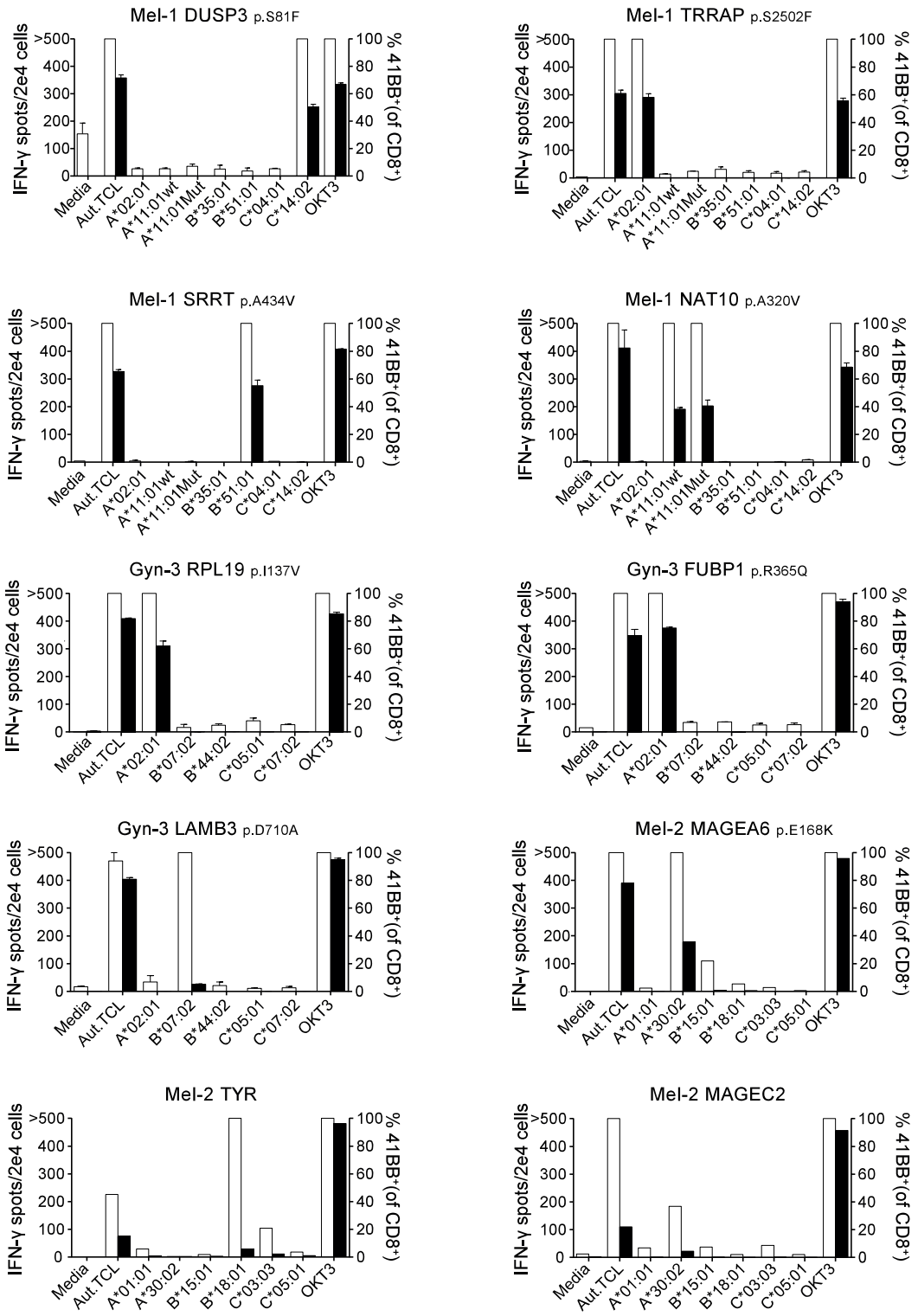
Appendix Figure 1. HLA-I allele binding motif of all patients included in the study. HLA-I binding motifs according to NetMHCpan4.1 are depicted for each allele.



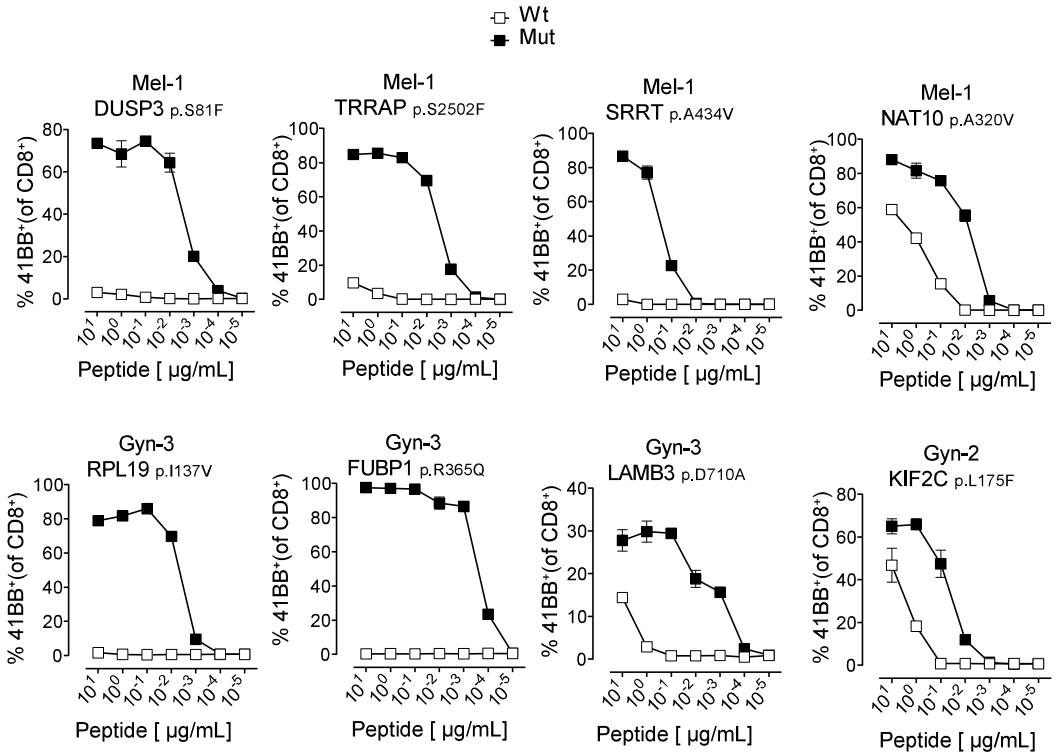
Appendix Figure 2. Frequency of nonC peptides detected in HLA-I. Percentage of peptides derived from canonical or non-canonical proteins for each patient ordered based on the frequency of nonC pHLA-I. Patients harboring HLA-A*11:01 or HLA-A*03:01 are marked with an *.



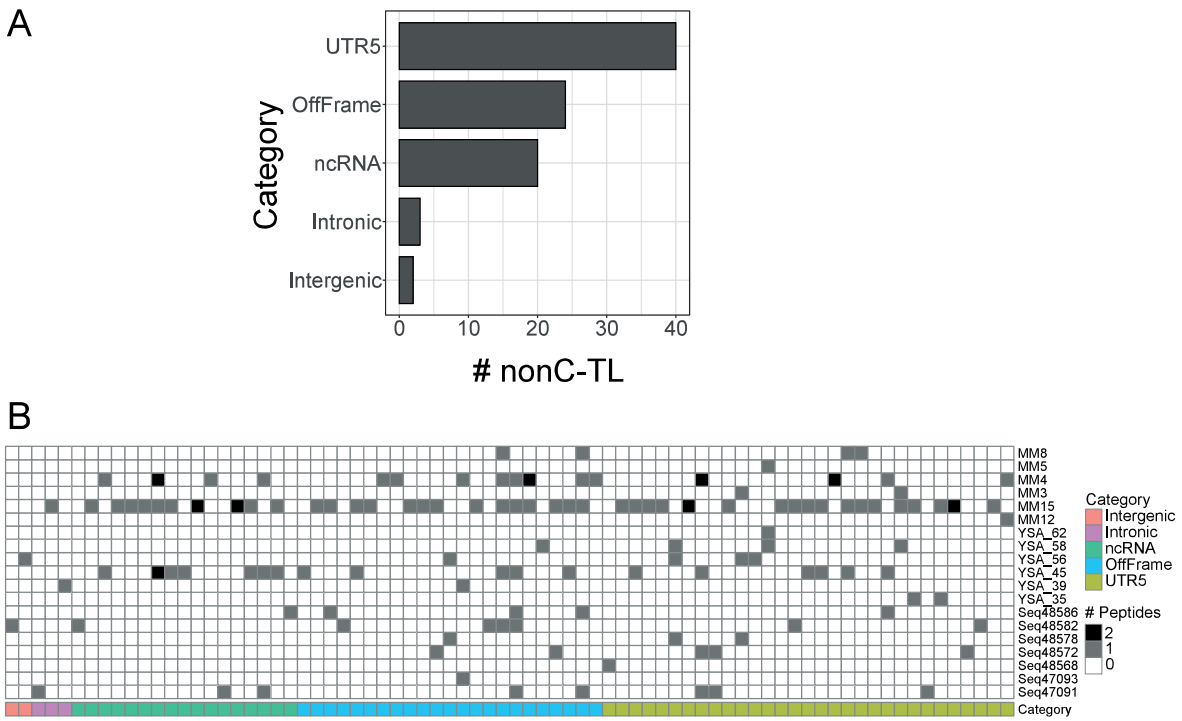
Appendix Figure 3. Enrichment of antigen-specific T cells from cancer patients by flow cytometry-based sorting. Reative T cells from TIL or PBL were enriched by flow cytometry based-sorting based on 4-1BB expression after 20h coculture with autologous APC pulsed with the peptides indicated (red squares). Sorted cells were expanded for 14 days through a rapid expansion protocol (REP). To evaluate the enrichment, 2e4 expanded T cells were cocultured with autologous APC pulsed with the peptides specified and the activation was measured based on 4-1BB upregulation after 20 h by flow cytometry. Gates display the percentage of 4-1BB cells gated on live/ $CD3^+CD8^+$. Autologous APC pulsed with an irrelevant peptide were used as a negative control.



Appendix Figure 4. Evaluation of the restriction element and recognition of the autologous tumor cell lines for the antigen-specific T cells isolated from cancer patients. CD8+ lymphocyte populations recognizing the HLA-I peptides indicated, were enriched by flow cytometry based-sorting based on 4-1BB expression and expanded for 14 days. The recognition of non-pulsed autologous TCL (Aut.TCL) was evaluated by coincubating from 3e4 to 5e4 tumor cells with 2e4 T cells. COS7 cells cotransfected with the indicated individual HLA-I alleles followed by peptide pulsing were used to determine the restriction element. The recognition was evaluated by IFN-γ release (left axis, white bars) and 4-1BB+ upregulation (right axis, black bars). '>' denotes greater than 500 spots/2e4 cells. OKT3 and media were used as positive and negative controls, respectively. Experiments were performed twice with technical duplicates. SD is plotted.



Appendix Figure 5. Functional avidity and specificity of neoantigen-specific T cells isolated from cancer patients. CD8+ lymphocyte populations recognizing the mutated HLA-I peptides indicated were enriched by flow cytometry based-sorting based on 4-1BB expression and expanded for 14 days. To evaluate the avidity and specificity of the antigen-specific sorted populations, 2e4 T cells were incubated with 1-2e5 autologous B cells or CD4+ T cells pulsed with serial dilutions of the wild-type (Wt) or mutant HPLC versions. After 20h, 4-1BB upregulation was measured by flow cytometry. Cells were gated on live/CD3+CD8+ lymphocytes. Experiments were performed twice with technical duplicates. SD is plotted.



Appendix Figure 6. A fraction of nonC-TL were identified in publicly available immunopeptidomics datasets from human tumor biopsies. HLA-I immunopeptidomics raw data of melanoma samples from Bassani-Sternberg et al., 2016; Kalaora et al., 2021 , was downloaded from PRIDE and search against a database containing the proteome reference plus our nonC-TL candidates (n=507) using PEAKS Studio Pro. The analysis shows the identified nonC-TL sequences at 5% FDR. (A) Unique nonC-TL sequences identified by category. (B) NonC-TL identified in each sample. The sample ID are shown on the right, and the nonC category is depicted in different colors at the bottom.