Dissertation

submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto-Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Jonas Philipp Becker, M.Sc.
born in Hermeskeil, Germany

Oral examination: July 21st, 2021

Pharmacological modification of RNA surveillance in microsatellite-unstable colorectal cancer

Referees: Prof. Dr. Frauke Melchior

Prof. Dr. Andreas Kulozik, Ph.D.

“*Wissen ersetzt Denken nicht. Weiß ich nicht, denke ich.*“

OK Kid

The work described in this thesis was carried out in the Molecular Medicine Partnership Unit (MMPU) at the Medical Faculty of the University of Heidelberg and the European Molecular Biology Laboratory (EMBL) Heidelberg between October 2016 and May 2021 under the supervision of Prof. Dr. med. Andreas E. Kulozik, Ph.D., and Prof. Dr. med. Matthias W. Hentze.

The content of this work is partially published in the following peer-reviewed article:

Becker, J.P., Helm, D., Rettel, M., Stein, F., Hernandez-Sanchez, A., Urban, K., Gebert, J., Kloor, M., Neu-Yilik, G., von Knebel Doeberitz, M., Hentze, M.W., Kulozik, A.E. (2021) NMD inhibition by 5-azacytidine augments presentation of immunogenic frameshift-derived neoepitopes. iScience. 10.1016/j.isci.2021.102389.

Abstract

Microsatellite-unstable colorectal cancers are characterized by the accumulation of somatic insertion/deletion mutations in repetitive DNA stretches termed microsatellites. Two-thirds of these mutations cause a shift of the reading frame thus encoding for tumor-specific protein sequences. HLA class I-presented peptides derived from such frameshifted protein sequences, termed InDel neoepitopes, allow patrolling $CD8^+$ T cells to identify and target the presenting tumor cell. The expression of frameshifted protein sequences, and therefore the HLA class I-mediated presentation of InDel neoepitopes, is, however, limited by nonsense-mediated RNA decay (NMD), a translation-dependent quality control pathway that recognizes and degrades mRNAs with premature termination codons introduced by nonsense or frameshift mutations. In the context of microsatellite-unstable cancers, pharmacological NMD inhibition is therefore hypothesized to provide clinical benefit by increasing the abundance of HLA class I-presented InDel neoepitopes on cancer cells and thus their immune recognition. In this thesis, a mass spectrometry-based approach was established to identify neoepitopes and investigate the effect of pharmacological NMD inhibition on HLA class I-presented peptides and neoepitopes.

For the identification of HLA class I-presented peptides, a previously published high-throughput protocol was extended and tailored for the identification of frameshift-derived InDel neoepitopes. To this end, standard mass spectrometry fragmentation and precursor selection methods were combined with advanced methodologies providing deep coverage of the immunopeptidome and additional sequence information on the identified peptides. Together with cell line-specific protein databases based, *e.g.*, on publicly available sequencing data, this enabled the unbiased, mass spectrometry-based identification of InDel neoepitopes.

Using a label-free quantification approach integrating untargeted and targeted MS2 data, it was shown that pharmacological NMD inhibition by the approved drug 5-azacytidine augments the HLA class I-mediated presentation of peptides, including InDel neoepitopes, derived from NMD-targeted transcripts. Moreover, the mutation underlying one of the previously unknown InDel neoepitopes exhibits a high frequency in microsatellite-unstable cancer cell lines and colorectal cancer patients, and *in vivo* immunization with the InDel neoepitope induces strong $CD8^+$ T cell responses in an HLA-humanized mouse model.

Taken together, the presented work provides an extendable immunopeptidomics platform for neoepitope discovery utilizing sequencing data and advanced mass spectrometry. The use of this platform allowed for the first time the unbiased, mass spectrometry-based identification of previously unknown, immunogenic InDel neoepitopes. Moreover, it was shown that NMD inhibition by 5-azacytidine augments HLA class I-mediated InDel neoepitope presentation suggesting that future immunotherapeutic strategies for microsatellite-unstable cancers may benefit from a combination with NMD inhibition, thereby turning cancer cells into easily identifiable targets for tumor-specific T cells.

Zusammenfassung

Mikrosatelliten-instabile kolorektale Karzinome sind durch die Anhäufung somatischer Insertions-/Deletionsmutationen in als Mikrosatelliten bezeichneten repetitiven DNA-Abschnitten gekennzeichnet. Zwei Drittel dieser Mutationen verursachen eine Verschiebung des Leserasters und kodieren daher für tumorspezifische Proteinsequenzen. HLA-Klasse-I-präsentierte Peptide, die von solchen leserasterverschobenen Proteinsequenzen stammen und als InDel-Neoepitope bezeichnet werden, ermöglichen es patrouillierenden $CD8^+$ T-Zellen, die präsentierende Tumorzelle zu identifizieren und anzugreifen. Die Expression von leserasterverschobenen Proteinsequenzen und damit die HLA-Klasse I-vermittelte Präsentation von InDel-Neoepitopen wird jedoch durch den Nonsens-vermittelten RNA-Abbau (NMD) begrenzt. NMD ist ein translationsabhängiger Qualitätskontrollmechanismus, der mRNAs mit vorzeitigen Terminationskodons, die durch Nonsens- oder Leserastermutationen eingeführt wurden, erkennt und abbaut. Eine pharmakologische NMD-Inhibition könnte daher bei Mikrosatelliten-instabilen Krebserkrankungen einen klinischen Nutzen bringen, indem sie die Menge der von HLA-Klasse I präsentierten InDel-Neoepitope auf Krebszellen erhöht und damit deren Immunerkennung verbessert. In dieser Arbeit wurde zunächst eine Massenspektrometrie-basierte Methode zur Identifizierung von Neoepitopen etabliert und verwendet, um den Effekt von pharmakologischer NMD-Inhibition auf HLA-Klasse-I-präsentierte Peptide und Neoepitope zu untersuchen.

Für die Identifizierung von HLA Klasse I-präsentierten Peptiden wurde zunächst ein zuvor veröffentlichtes Hochdurchsatz-Protokoll erweitert und für die Identifizierung von InDel-Neoepitopen, welche aus Leserastermutationen entstehen, optimiert. Zu diesem Zweck wurden standardmäßige massenspektrometrische Fragmentierungs- und Precursorselektionsmethoden mit neuen Methoden kombiniert, die eine tiefgehende Abdeckung des Immunopeptidoms und zusätzliche Sequenzinformationen zu den identifizierten Peptiden liefern. Zusammen mit zelllinienspezifischen Proteindatenbanken, die zum Beispiel auf öffentlich verfügbaren Sequenzierungsdaten basieren, wurde hiermit die unvoreingenommene, Massenspektrometrie-basierte Identifizierung von InDel-Neoepitopen ermöglicht.

Unter Verwendung eines markierungsfreien Quantifizierungsansatzes, der ungezielt und gezielt gewonnene MS2-Daten integriert, konnte gezeigt werden, dass pharmakologische NMD-Inhibition durch das bereits zugelassene Medikament 5-Azacytidin die HLA-Klasse I-vermittelte Präsentation von Peptiden und InDel-Neoepitopen erhöht, welche von Transkripten stammen die normalerweise von NMD abgebaut werden. Eine der Mutationen, die einem bisher unbekannten InDel-Neoepitop zugrunde liegt, zeigt ein hochfrequentes Auftreten in Mikrosatelliten-instabilen Krebszelllinien und Darmkrebspatienten. Zudem konnte gezeigt werden, dass eine *in vivo* Immunisierung mit diesem InDel-Neoepitop starke $CD8^+$ T-Zell-Antworten in einem HLA-humanisierten Mausmodell induziert.

Die vorgestellte Arbeit beschreibt eine erweiterbare Immunopeptidomics-Plattform für die Entdeckung von Neoepitopen unter Verwendung von Sequenzierungsdaten und moderner Massenspektrometrie. Der Einsatz dieser Plattform ermöglichte zum ersten Mal die unvoreingenommene, Massenspektrometrie-basierte Identifizierung von bisher unbekannten, immunogenen InDel-Neoepitopen. Darüber hinaus konnte gezeigt werden, dass NMD-Inhibition durch 5-Azacytidin die HLA-

Klasse-I-vermittelte Präsentation von InDel-Neoepitopen steigert. Dies deutet darauf hin, dass zukünftige immuntherapeutische Strategien für Mikrosatelliten-instabile Krebsarten von einer Kombination mit NMD-Inhibition profitieren könnten, indem Krebszellen in leicht identifizierbare Ziele für tumorspezifische T-Zellen verwandelt werden.

Acknowledgments

I am grateful to everyone who helped, supported, and sometimes distracted me during my Ph.D. journey. In particular, these are:

Andreas Kulozik and **Matthias Hentze** for the opportunity to conduct my Ph.D. research in your research group and the inspiring environment of the MMPU, and for all your advice, support, and guidance in the last years.

Monika Fastenrath-Vedder and **Gerhard Fastenrath** for supporting my research with a Ph.D. scholarship of the **Fastenrath-Stiftung**.

Gabriele Neu-Yilik for all the discussions about science and everything else, for sharing your immense knowledge on molecular biology in general and NMD in particular, and for proofreading basically every piece of text I produced.

Beate Amthor for being the most cheerful seatmate, for always offering a helping hand and keeping the lab running, and for accepting “my” music on some days.

Michael Backlund, **Annika Brosig**, **Manouk Gerritsen**, **Claudia Gruber**, **Daria Lavysh**, and **Yang Zhou** for making the lab a place I liked to go to every day, for numerous breaks with good conversations, for unforgettable nights out, and for celebrating the successful days as well as the unsuccessful ones!

All other current and former members of the Kulozik and Muckenthaler labs for fun times, useful advice, and lots of cake!

Mandy Rettel, **Dominic Helm**, and **Frank Stein**, and all other members of the EMBL Proteomics Core Facility for sparking my interest in mass spectrometry, for taking your time to answer all my questions, for sacrificing countless LC columns for HLA samples, but not giving up, and for providing me a second home in the PCF every time I was “up the hill”.

Vera Fuchs, Katharina Urban, Alejandro Hernandez-Sanchez, Jonathan Dörre, Johannes Gebert, **Matthias Kloor**, and **Magnus von Knebel Doeberitz** of the Department of Applied Tumor Biology for your expertise and experimental support on the immunological part of my Ph.D. project.

Frauke Melchior and **Mikhail Savitsiki** for your scientific advice and for sometimes playing the devil's advocate to make the most out of my project as members of my Thesis Advisory Committee.

Frauke Melchior, **Michael Brunner,** and **Viktor Umansky** for being members of my Thesis Defense Committee.

My friends inside and outside of Heidelberg for freeing my mind from science during dinners and BBQs, concerts, football games, vacations, handball sessions, and other fun activities. You kept me sane!

My whole family, first and foremost my parents **Petra** and **Hartmut** and my sister **Eva,** for always supporting me, for motivating me, and for the fact that I can always count on you.

Elena for always being there for me and believing in me, and for your endless support in all situations of life.

Table of contents

Abbreviations

µg	micrograms
µl	microliters
µm	micrometer
µM	micromolar
3' UTR	3' untranslated region
5AZA	5-azacytidine
Å	angstrom
aa	amino acids
ac	N-terminal acetylation
ACN	acetonitrile
AGC	automatic gain control
APC	antigen-presenting cell
B2M	β-2-microglobulin
BCR	B cell receptor
bp	basepair
cam	carbamidomethylation of cysteine
CAR	chimeric antigen receptor
CCLE	Cancer Cell Line Encyclopedia
CHX	cycloheximide
CID	collision-induced dissociation
CLIP	class II-associated invariant chain peptide
cm^2	square centimeters
COSMIC	Catalogue Of Somatic Mutations In Cancer
CRC	colorectal cancer
CTA	cancer/testis antigen
CTL	cytotoxic T lymphocyte
CTLA-4	cytotoxic T-lymphocyte-associated protein 4
Da	Dalton
DAMP	damage-associated molecular pattern
dAZA	5-aza-2'-deoxycytidine
DB	database
ddH2O	double-distilled water
DMP	dimethyl pimelimidate
DMSO	dimethyl sulfoxide
DRiPs	defective ribosomal products
EDTA	ethylenediaminetetraacetic acid
EJC	exon junction complex
ER	endoplasmatic reticulum
ESI	electrospray ionization
ETciD	electron-transfer/collisional-induced dissociation
ETD	electron transfer dissociation
EThcD	electron-transfer/higher-energy collisional dissociation
EtOH	ethanol
FA	formic acid

FBS	fetal bovine serum
FC	fold change
FDR	false discovery rate
fmol	femtomol
FT	flowthrough
g	grams
h	hours
HBB	human hemoglobin subunit beta
HCD	higher-energy collisional dissociation
HI	hydrophobicity index
HLA	human leukocyte antigen
HLAp	HLA class I-presented peptides
IFNγ	interferon-γ
Ig	immunoglobulin
Ii	invariant chain
InDel	insertion/deletion
IP	immunoprecipitation
iTRAQ	isobaric tags for relative and absolute quantitation
kDa	Kilodalton
kV	kilovolt
l	liters
LC-MS	liquid chromatography-mass spectrometry
LC-MS/MS	liquid chromatography-tandem mass spectrometry
M	molar
m/z	mass-to-charge ratio
m1	one-base deletion
m2	two-base deletion
mA	milliampere
MAE	mild acid elution
MALDI	matrix-assisted laser desorption/ionization
MBL	manose-binding lectin
MeOH	methanol
min	minutes
ml	milliliters
mM	millimolar
MMIIC	MHC class II compartment
MMR	mismatch repair
MNR	mononucleotide repeat
MS	mass spectrometry
MSI	microsatellite instability
MSI CRC	microsatellite-unstable colorectal cancer
ng	nanograms
NK cell	natural killer cell
nm	nanometer
NMD	nonsense-mediated RNA decay
nt	nucleotides

ox	oxidation
p1	one-base insertion
p2	two-base insertion
PAMP	pathogen-associated molecular pattern
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PD-1	programmed cell death protein 1
PD-L1	programmed cell death 1 ligand 1
PLC	peptide loading complex
ppm	parts per million
PRR	pattern recognition receptor
PSM	peptide spectrum match
PTC	premature termination codon
PVDF	polyvinylidene difluoride
qPCR	quantitative real-time PCR
RPLC	reverse-phase liquid chromatography
RT	room temperature
RT	reverse transcriptase
s	seconds
SD	standard deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	standard error of mean
SILAC	Stable isotope labeling with amino acids in cell culture
SNP	single nucleotide polymorphism
TAA	tumor-associated antigen
TBS-T	Tris-buffered saline supplemented with Tween 20
Tc cell	cytotoxic T cell
TCLP	TRON Cell Line Portal
TCR	T cell receptor
TFA	trifluoroacetic acid
Th cell	T helper cell
TLR	Toll-like recpetor
TMT	tandem mass tag
TSA	tumor-specific antigen
uORF	upstream open reading frame
V	volt
wt	wild-type
z	charge

List of figures

Several figures within the presented thesis have been published in Becker *et al.* (2021). In particular, these are: Figure 14A, Figure 15C, Figure 31, Figure 34, Figure 35, Figure 36A, Figure 40, Figure 41, Figure 44, Figure 45, Figure 49, Figure 51B, Figure 52A and C, Figure 55A, Figure 57, Figure 58A, Figure 59, Figure 60, Figure 63, and Figure 65.

List of tables

1 Introduction

1.1 Microsatellite-unstable colorectal cancer

Colorectal cancer (CRC) is the third most common cancer worldwide and has the second-highest, cancer-related mortality rate (Bray *et al.*, 2018). According to their genetic and epigenetic properties, CRCs can be divided into several subtypes. While most CRCs follow the adenoma-carcinoma pathway and exhibit chromosomal instability, about 15 % of all CRCs display microsatellite instability (MSI; Kloor and von Knebel Doeberitz, 2016; Dekker *et al.*, 2019). Of these, approximately 12 % are sporadic MSI CRCs, while 3 % are caused by the Lynch syndrome, the most common inherited cancer predisposition (Boland and Goel, 2010). Common to all cancers with MSI is their high mutational burden due to a defective DNA mismatch repair (MMR) system.

1.1.1 The human DNA mismatch repair system – a guardian of genome stability

The human DNA MMR system is responsible for correcting small insertion/deletion (InDel) mutations and single nucleotide mismatches produced during DNA synthesis (Ruiz-Banobre and Goel, 2019). The key players in this system are MSH2, MSH6, MSH3, MLH1, MLH3, and PSM2 which form specific heterodimers to execute their function (Boland and Goel, 2010). MSH2 pairs either with MSH6 (MutSα complex) to recognize single-nucleotide mismatches and small InDels or with MSH3 (MutSβ) to recognize larger InDels (Jiricny, 2006). The MutS heterodimers then recruit a second heterodimer, either MutLα (MLH1-PMS2) or MutLγ (MLH1-MLH3), which possess an intrinsic endonuclease domain to break open the DNA strand. After excision of the mispaired DNA stretch by exonuclease 1 (EXO1) or proliferating cell nuclear antigen (PCNA), the resulting gap is filled up by the DNA polymerase Polδ (Reyes *et al.*, 2015). The DNA MMR system increases the replication fidelity by 10-fold to 100,000-fold and, together with high nucleotide selectivity and proofreading capacity of DNA polymerases, guarantees (almost) error-free DNA synthesis (Arana and Kunkel, 2010). In sporadic MSI CRC, DNA MMR is disabled epigenetically by hypermethylation of the *MLH1* gene promotor due to the CpG island methylator phenotype (CIMP). In hereditary Lynch syndrome MSI CRCs monoallelic germline mutations in any of the previously introduced DNA MMR key players in combination with a second, somatic mutation leads to the inactivation of the DNA MMR system (Kloor and von Knebel Doeberitz, 2016). Defects in the DNA MMR machinery lead to the accumulation of somatic mutations, microsatellite instability (MSI) and ultimately cancer.

1.1.2 Microsatellite instability – causes and consequences

Microsatellites are short (1-6 bp), repetitive DNA stretches which are particularly prone to errors made by the DNA polymerase during replication. Mutations in microsatellites arise as a consequence of strand slippage (Figure 1). In this process, pausing of the DNA polymerase causes a transient dissociation of the replicating strand. Subsequent misaligned rehybridization of the two DNA strands and further extension by the DNA polymerase leads to either shortening (deletion mutation) or lengthening (insertion mutation) of the nascent strand relative to the template strand (Levinson and Gutman, 1987; Ellegren, 2004). These slippage events are frequently overlooked by the proofreading function of DNA polymerase

(Arana and Kunkel, 2010), but the resulting insertion/deletion loops protruding from the DNA helix, are recognized and repaired by the DNA MMR system (Jiricny, 2006). However, in MSI CRC these replication errors are not corrected due to the defective DNA MMR system and the erroneous microsatellite is passed on to daughter cells at each replication round, thus manifesting the InDel mutation.

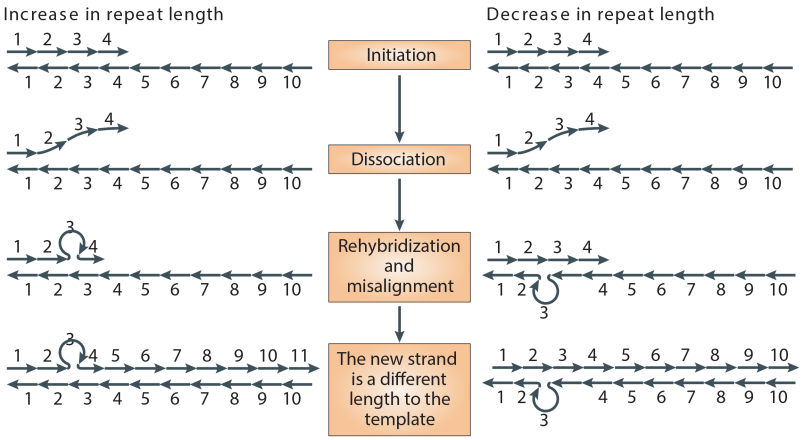


Figure 1: Strand slippage can cause InDel mutations. During replication of a repetitive sequence, the DNA strands might dissociate and rehybridize improperly. Misalignment in the nascent strand will cause an insertion mutation (left side), while loop formation in the template strand will cause a deletion mutation (right side). Reprinted and adapted from Ellegren (2004) with permission from Springer Nature under License Number 5050741464923.

According to the “Bethesda guidelines” proposed by the US National Cancer Institute, a tumor is considered microsatellite-unstable if two out of five markers tested in the MSI panel (mononucleotide markers BAT25, and BAT26; dinucleotide markers D2S123, D5S346, and D17S250) exhibit size variability, *i.e.* InDel mutations (Boland *et al.*, 1998; Umar *et al.*, 2004). Cancers with MSI rapidly acquire hundreds to thousands of InDel mutations. Those in protein-coding regions of the genome eventually lead to the inactivation of tumor suppressor genes or (very rarely) the activation of oncogenes by introducing translational frameshifts. Frequently mutated genes belong to a variety of pathways related to crucial hallmarks of cancer (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011) such as transcriptional regulation (e.g. *TAF1B*, *TCF-4*), signal transduction (e.g. *TGFβR2*, *ACVR2*), apoptosis (e.g. *BAX*, *CASP5*), and immune surveillance (e.g. *β2M*, *TAP1*, *TAP2*). Their inactivation provides a selective growth advantage that allows malignant transformation (Woerner *et al.*, 2006; Imai and Yamamoto, 2008; Kim *et al.*, 2013a; Jonchere *et al.*, 2018). Besides the inactivation of tumor suppressor genes, translational frameshift-inducing InDel mutations have the potential to generate tumor-specific protein sequences which have been associated with an extensive interaction between MSI CRCs and the immune system as well as with increased patient survival (Tougeron *et al.*, 2009; Giannakis *et al.*, 2016; Kloor and von Knebel Doeberitz, 2016; Mlecnik *et al.*, 2016).

1.2 The human immune system

The immune system protects the human body from intruders such as bacteria, fungi, parasites, or viruses and can detect infected or cancerous cells using its remarkable ability to discriminate between "self" and "foreign". Structures recognized by the immune system are termed antigens while specific parts of antigens recognized by immune receptors are termed epitopes. Classically, the immune system can be divided into two branches: the innate and the adaptive immune system. The innate immune system is highly conserved in all species and represents the immediate first-line defense, while the adaptive immune system represents the slightly delayed but highly sophisticated and specific branch that developed in vertebrates roughly 500 million years ago (Flajnik and Kasahara, 2010; Roth-Walter *et al.*, 2014). Moreover, the adaptive immune system is capable of creating an "immunological memory", which allows a faster response upon a second encounter with a threat and is the basis for vaccination (Murphy and Weaver, 2017). Immune responses of both branches can be further subclassified into humoral immune responses (*i.e.* responses mediated by soluble components of the immune system) and cellular immune responses.

1.2.1 The innate immune system

1.2.1.1 Humoral part – barriers and cascades

Breaching the host's anatomic barriers (*i.e.* epithelial surfaces such as the skin, the oral mucosa, or the intestine) by a pathogen marks the start of an infection and triggers an immediate response of some mechanisms of the innate immune system. Pathogens first encounter antimicrobial proteins and peptides which possess the ability to specifically disrupt cell membranes of microorganisms (Murphy and Weaver, 2017).

Pathogens that overcome these first lines of defense subsequently face the humoral immune response of the complement system. The complement system consists of more than 25 proteins, mostly proteases in the form of inactive pro-enzymes. During the proteolytic activation and amplification cascade, the proteins sequentially cleave each other to trigger various effector pathways. The complement system can be activated by three different mechanisms: the lectin pathway, the alternative pathway, and the classical pathway. Mannose-binding lectins (MBLs) and ficolins are capable to distinguish between the carbohydrates present on the surface of microbes and vertebrate cells. Pathogen-bound MBLs and ficolins then recruit specific proteases that activate the complement cascade by generating a C3 convertase. In the alternative pathway, spontaneous hydrolysis of C3 to C3b and subsequent deposition at cell surfaces constantly activates the complement system at a low level. C3b deposited on host cells is rapidly inactivated by regulatory proteins which further cleave it to an inactive form. In contrast, C3b deposited on pathogens lacking these regulatory proteins is stabilized by properdin (factor P) leading to full activation of the complement system by a C3 convertase. The classical pathway is activated by either direct binding of the C1 complex to pathogenic surfaces or antigen-engaged antibodies. Subsequent cleavage of C4 and C2 generates an active C3 convertase. Since antibodies belong to the adaptive branch of the immune system, the classical pathway represents the evolutionarily youngest activation mechanism of the complement system and is only found in vertebrates (Roth-Walter *et al.*, 2014; Merle *et al.*, 2015a; Murphy and Weaver, 2017). All three activation mechanisms result in the formation of a

C3 convertase which subsequently triggers three different effector pathways. Small fragments of complement proteins such as C3a and C5a act as chemotaxins to promote local inflammation and direct phagocytic cells to the site of infection. The attracted phagocytes express complement receptors that can detect C3b and C4b. Pathogens coated with these two factors are ingested in a process termed opsonization by, *e.g.*, phagocytic antigen-presenting cells (APCs) and thereby influence adaptive immunity by enhancing the presentation of pathogen-derived antigens to T cells. Finally, completion of the complement cascade results in the formation of the membrane attack complex leading to the disruption of the cell membrane and subsequent lysis of the target cell (Merle *et al.*, 2015b; Barnum, 2017; Murphy and Weaver, 2017).

1.2.1.2 Cellular part – it's all about patterns

The cellular effectors of the innate immune system include several cell types that contribute to the host's defense by mechanisms such as phagocytosis, secretion of cytokines and chemokines as well as cell-mediated cytotoxicity.

Cells capable of phagocytosis such as neutrophils, macrophages, dendritic cells, and eosinophils express pattern recognition receptors (PRRs) that sense danger by binding to pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs) as a result of cellular infection, stress, transformation, or damage. PRRs can be subclassified in five PRR families (Toll-like receptors (TLRs), RIG-1-like receptors, C-type lectin receptors, NOD-like receptors, and AIM2-like receptors) that detect both intracellular and extracellular PAMPs and DAMPs. TLRs are the best-studied superfamily and are expressed at the cell surface and in endosomal compartments. They recognize PAMPs such as LPS of Gram-negative bacteria (TLR4) or unmethylated CpG DNA of viruses (TLR9) and DAMPs such as heat shock proteins released from apoptotic cells (TLR2/4). TLR signaling via NFκb results in the production of pro-inflammatory cytokines (Mogensen, 2009; Newton and Dixit, 2012; Schaefer, 2014; Brubaker *et al.*, 2015; Hato and Dagher, 2015; Murphy and Weaver, 2017). Today, many vaccines contain TLR ligands as potent immunostimulatory adjuvants (Weeratna *et al.*, 2005; Ho *et al.*, 2018). Importantly, PRR-mediated recognition and subsequent phagocytosis of pathogens and cellular debris and presentation of fragments of these to cells of the adaptive immune system by macrophages and dendritic cells bridge the two branches of the immune response.

Neutrophils, eosinophils, basophils, and mast cells constitute the group of granulocytes. These cells carry granules loaded with microbial enzymes, toxic proteins, free radicals, histamine, and cytokines which are released upon activation (Roth-Walter *et al.*, 2014).

Finally, the host defense by cytotoxicity is mediated by natural killer (NK) cells that act non-specifically against aberrant virus-infected or malignant cells. The fate of their target cells is determined by the integration of both activating and inhibitory signals. Activating receptors can detect dysregulated or stress-induced cell surface proteins, while inhibitory receptors recognize highly expressed cell surface proteins on potential target cells. Lack of inhibitory signals leads to the release of the NK cell's granules which contain granzymes and perforins resulting in apoptosis of the target cell (Roth-Walter *et al.*, 2014; Murphy and Weaver, 2017; Choucair *et al.*, 2019).

1.2.2 The adaptive immune system

The adaptive branch of the immune system takes longer to respond, but is highly specific and can mediate long-term protection. The main effectors of the adaptive immune system are B and T cells, which are special types of lymphocytes, derived from hematopoietic stem cells in the bone marrow. Similar to PPRs of the innate immune system, both humoral and cellular components of the adaptive immune system are able to recognize and distinguish "self" from "foreign". In contrast to the germline-encoded PPRs, the receptors of B cells (immunoglobulin (Ig), the membrane-bound formed is termed BCR, the soluble form is termed antibody) and T cells (TCRs) show a high diversity and are expressed clonally allowing them to recognize virtually all possible antigens. The high diversity of BCRs and TCRs is generated by V(D)J recombination. In this gene rearrangement process, variable, diverse, and joining gene segments are combined by chance to generate varying antigen-binding regions. Diversity is further enhanced by random addition of up to 20 nucleotides at the site of recombination. It is estimated that there are ~$5x10^{13}$ possible BCR combinations and ~10^{18} possible TCR combinations (Boehm, 2011; Roth-Walter *et al.*, 2014). While BCRs and antibodies recognize intact, soluble antigens as well as antigens at the surface of pathogens, TCRs mainly recognize peptides presented by the major histocompatibility complex (MHC; termed human leukocyte antigen (HLA) in humans) at the surface of host cells.

1.2.2.1 B cells – humoral part of the adaptive immune response

After V(D)J recombination B cells remain in the bone marrow and are tested for autoreactivity. B cells with affinity to "self" antigens undergo further editing of their BCR and only successful editing averts apoptosis of B cells carrying autoreactive BCRs thereby establishing central B cell tolerance. Next, B cells move to the secondary lymphoid tissues and acquire the ability to co-express different immunoglobulin isotypes (IgM and IgD). These now mature B cells are still considered "naïve" since they have not yet encountered their cognate antigen. Encountering their cognate antigen in the periphery represents the second stage of B cell development which can be either T cell-independent or T cell-dependent. T cell-independent B cell activation is mediated by proliferation-promoting mitogens or polymerized proteins and polysaccharides. Activated B cells further differentiate into short-lived plasma cells secreting low-affinity IgM antibodies. T cell-dependent B cell activation requires three signals. The first signal is the binding of an antigen by the BCR which leads to internalization and subsequent cell-surface presentation by HLA class II molecules. Peptides presented by HLA class II molecules, in turn, can be recognized by the TCR of T helper (Th) cells delivering the costimulatory second signal. The third signal is mediated by cytokines released by the Th cell in response to the interaction with the B cell and binding to their respective receptors on the B cell. Importantly, B cells activated with the help of T cells can undergo further optimization of the antigen-recognizing part of the BCR by somatic hypermutation and affinity maturation. Additionally, in a process called "isotype switching" activated B cells can change the constant part of the BCR enabling them to interact with different effector molecules. The B cells activated in that way further differentiate either into long-lived plasma cells that secrete high levels of high-affinity antibodies or into memory B cells enabling a faster and more efficient response upon a second encounter with their cognate antigen (Bonilla and Oettgen, 2010; Mak *et al.*, 2014a; Roth-Walter *et al.*, 2014).

1.2.2.2 T cells – cellular part of the adaptive immune response

In contrast to B cells, T cells mature in the thymus rather than in the bone marrow. Depending on the expression of the CD4 and CD8 co-receptors, T cell maturation can be divided into three phases. In the double-negative phase, neither of the two co-receptors is expressed and the TCRβ chain undergoes V(D)J recombination. Successful rearrangement of the TCRβ chain leads to a survival/differentiation signal and late pre-T cells start to express both, CD4 and CD8 ($CD4^+/CD8^+$). In the second phase, called the double-positive phase, the TCRα chain is subjected to V(D)J recombination and the TCRαβ heterodimer (*i.e.* the TCR) undergoes thymic selection. The central T cell tolerance is conveyed by three different selection processes. T cells that do not bind to the host's HLA molecules are non-selected ("death by neglect"). T cells that possess a high affinity to HLA complexed with self-peptides are negatively selected. Only 1-2 % of T cells expressing TCRs with neither too weak nor too strong binding to HLA presenting self-peptides receive a positive selection signal. In the last phase of T cell development, double-positive T cells lose the expression of either CD4 or CD8 co-receptors depending on the interaction partner during thymic selection. T cells that interacted with an HLA class I:peptide complex become $CD8^+$ cytotoxic T (Tc) cells, while T cells that interacted with HLA class II:peptide complexes retain CD4 expression and become $CD4^+$ Th cells. Both single-positive, naïve T cell species leave the thymus after completion of maturation and travel to secondary lymphoid tissues via blood vessels. Similar to B cells, the activation of naïve T cells requires three signals. Interaction between the TCR and an HLA molecule presenting a foreign peptide represents the first signal. Naïve Th cells are activated by the interaction with peptides presented by HLA class II molecules of dendritic cells, while naïve Tc cells are activated by the interaction with peptides presented by HLA class I molecules. After the initial binding of a TCR with an HLA molecule, clustering of multiple TCRs and HLAs leads to the formation of the so-called immunological synapse which allows sustained signaling. The second signal is delivered by the interaction of co-stimulatory molecules B7 (dendritic cell) and CD28 (T cells). Furthermore, the interaction of a naïve Th cell with a dendritic cell induces the expression of B7 molecules, thereby "licensing" the dendritic cell to activate naïve Tc cells. Upon receiving the first two activation signals, both naïve T cell subsets upregulate cytokine, chemokine, and growth factor receptors which allow them to receive the third signal needed for full activation mediated mainly by interleukin-2. While naïve Th cells can produce sufficient quantities of interleukin-2 to provide the third signal by autocrine signaling, naïve Tc cells rely on activated Th cells providing the third signal. Activated Th cells further differentiate into different effector subtypes characterized by different cytokine secretion profiles. Effector Th cells can be activated more easily and more rapidly and induce antibody production by B cells or activate phagocytosis by macrophages and neutrophils. Activated Tc cells proliferate and leave the lymph node as pre-CTLs to migrate in the direction of the inflammatory site where they further differentiate into mature CTLs upon stimulation with cytokines released by macrophages and dendritic cells. During this differentiation process, CTLs become "armed" with cytotoxic granules allowing lytic action upon recognition of their target. Armed CTLs establishing high-affinity binding to their cognate HLA class I:peptide complex induce apoptosis of the target cell by the release of cytotoxic granules containing perforin and granzymes, the interaction of Fas ligand with Fas expressed by the target cell, or by the release of cytokines such as TNF. Armed CTLs dissociate from their target cell after cell death induction and, after re-synthesis of cytotoxic granules, proceed to kill other target cells. After the complete elimination of a threat, most effector T cells undergo apoptosis. However, a small subset

generates memory T cells, which can react faster and stronger upon a second encounter (Bonilla and Oettgen, 2010; Boehm, 2011; Koch and Radtke, 2011; Mak *et al.*, 2014e; d; Roth-Walter *et al.*, 2014; Murphy and Weaver, 2017).

1.2.2.3 Antigen presentation by the HLA system – present yourself!

The peptides presented by the HLA class I and II molecules are the crucial link between professional APCs of the innate immune system and T cells of the adaptive immune system. While HLA class II expression and peptide presentation are restricted to APCs and are necessary for the activation of $CD4^+$ Th cells, HLA class I molecules are expressed by virtually all nucleated cells and allow $CD8^+$ Tc cells to identify infected or malignant target cells. The expression of both, HLA class I and II molecules is induced during an immune response by cytokines such as IFNγ (Sen and Ransohoff, 1993).

The HLA gene locus is located on chromosome 6 in humans and is thought to be the most polymorphic region in the entire genome (Aguado *et al.*, 1999). To date, more than 27,000 HLA alleles have been reported (Robinson *et al.*, 2020)[1]. Furthermore, the HLA genes are polygenic, which means that in each individual there are several functional copies derived from gene duplications during evolution. For HLA class I, there are three different loci per allele (HLA-A, -B, and -C) encoding the HLA class I α chain, while for HLA class II there are HLA-DR, -DQ, and -DP loci on each allele. Each of the three HLA class II loci encodes one or more HLA class II α and β chains adding further complexity. In total, every individual expresses three to six different HLA class I molecules and three to twelve different HLA class II molecules on their cells. Thus, polygeny and polymorphism ensure that a wide range of peptides can be displayed to T cells (Mak *et al.*, 2014b; Rock *et al.*, 2016; Murphy and Weaver, 2017).

HLA class I and II molecules share substantial structure similarity. Both are membrane-anchored heterodimers composed of an α chain and β-2-microglobulin (B2M) in the case of HLA class I and an α chain and a β chain in the case of HLA class II. Peptides presented by both HLA class I and II molecules are accommodated in a peptide-binding groove formed by a β-sheet and two α-helices. HLA class II molecules possess an open binding groove allowing them to present peptides with 13-25 amino acids (aa) in length. In contrast, the size of the HLA class I binding groove is limited by two conserved tyrosine residues at both ends restricting the length of peptides it can bind (Adams and Luoma, 2013; Wieczorek *et al.*, 2017). Usually, HLA class I molecules present peptides of 8-10 aa, however, there have been reports of longer peptides (up to 15 aa) due to terminally extended or bulged peptides (Tynan *et al.*, 2005; Gfeller *et al.*, 2018; Guillaume *et al.*, 2018). Peptides are held in the binding groove of both HLA class I and II molecules by hydrogen bonds between the anchor positions at the peptide backbone and binding pockets formed by aa side chains of HLA. The previously described polymorphism directly affects the peptide repertoire bound by a specific HLA molecule by forming different binding pockets (Wieczorek *et al.*, 2017). Conserved aa at the anchor position of peptides bound to HLA molecules allow sequence clustering to identify HLA binding motifs (Falk *et al.*, 1991; Andreatta *et al.*, 2017a; Racle *et al.*, 2019) which can be used to define HLA supertypes (Sette and Sidney, 1999; Lund *et al.*, 2004;

[1] https://www.ebi.ac.uk/ipd/imgt/hla/

Sidney *et al.*, 2008; Greenbaum *et al.*, 2011; Saha *et al.*, 2013) or for peptide binding prediction algorithms (Rammensee *et al.*, 1999; Jurtz *et al.*, 2017; O'Donnell *et al.*, 2018; Reynisson *et al.*, 2020).

The entirety of peptides presented by HLA class I and II at the cell surface is termed immunopeptidome and enables immune surveillance, *e.g.* the constant monitoring of intra- and extracellular protein expression. HLA class II molecules mainly present peptides derived from exogenous antigens to $CD4^+$ Th cells, but can also present intracellular peptides captured by autophagy. HLA class II molecules are inserted into the ER membrane during their synthesis together with a protein called invariant chain (Ii). Ii serves as a stabilizer of the HLA class II molecule and prevents the binding of endogenous peptides in the ER. Moreover, Ii contains a motif targeting HLA class II molecules to the endosomal pathway. After trafficking to endolysosomes, the Ii is cleaved by proteases such as cathepsin L and S, leaving a shorter fragment of the Ii called class II-associated invariant chain peptides (CLIP) in the peptide-binding groove of the HLA class II molecule. Pathogens and macromolecules taken up by the cell follow a similar way where their cargo ultimately is degraded to peptides of 10-30 aa in length in endolysosomes. Fusion of these endolysosomes with endolysosomes containing HLA class II:CLIP complexes forms the MIIC (MHC class II compartment). In the MIIC, the low-affinity CLIP is exchanged by higher-affinity peptides with the help of the chaperone HLA-DM by opening the binding groove. After binding of a higher-affinity peptide, HLA-DM dissociates from the complex, and the peptide-loaded HLA class II molecule is transported to the cell surface and stably inserted into the cell membrane (Vyas *et al.*, 2008; Neefjes *et al.*, 2011; Mak *et al.*, 2014c; Roche and Furuta, 2015; Rock *et al.*, 2016).

In contrast to HLA class II molecules, HLA class I molecules mainly present peptides of endogenous origin to $CD8^+$ Tc cells, but can also present peptides of extracellular origin by a process termed cross-presentation (Mak *et al.*, 2014c). The peptides presented by HLA class I molecules are produced from ubiquitinated proteins by proteasome-mediated degradation in the cytosol. The proteasome exists in three different forms producing slightly different peptides due to altered active sites. Besides the constitutive proteasome, there is the immunoproteasome which is induced by interferons and produces peptides that are preferentially processed by the HLA class I processing and presentation machinery (Gaczynska *et al.*, 1993). The thymoproteasome is another specialized form expressed exclusively in cortical thymic epithelial cells and thus is involved in positive and negative selection of developing $CD8^+$ Tc cells (Murata *et al.*, 2007; Kincaid *et al.*, 2016). Peptides produced by the proteasome are released into the cytosol where they can be further trimmed by cytosolic peptidases. Peptides surviving the hostile, cytosolic environment are transported into the ER via a heterodimeric ATP-binding cassette transporter formed from transporters associated with antigen presentation-1 and -2 (TAP1, TAP2). Inside the ER, empty HLA class I molecules, together with the chaperones tapasin, calreticulin, and ERp57, and the TAP form the peptide loading complex (PLC) which ensures efficient peptide loading. Peptides may be further trimmed by ER-resident peptidases to fit the binding groove of HLA class I molecules. Peptide-loaded HLA class I molecules are released from the PLC and transported to the cell membrane (Neefjes *et al.*, 2011; Mak *et al.*, 2014c; Rock *et al.*, 2016; Murphy and Weaver, 2017). Cross-presentation refers to the ability of APCs to present peptides of extracellular origin via HLA class I molecules and allows priming of $CD8^+$ Tc cells for example in the setting of cancer vaccines (Fehres *et al.*, 2014). Antigens taken up by phagocytosis can enter the HLA class I pathway either if the content of the phagosome is released in the cytosol or by fusion of the phagosome and endosomes with empty,

recycling HLA class I molecules (Kovacsovics-Bankowski and Rock, 1995; Grommé *et al.*, 1999). Figure 2 provides a graphical overview of the HLA class I and class II antigen presentation pathways.

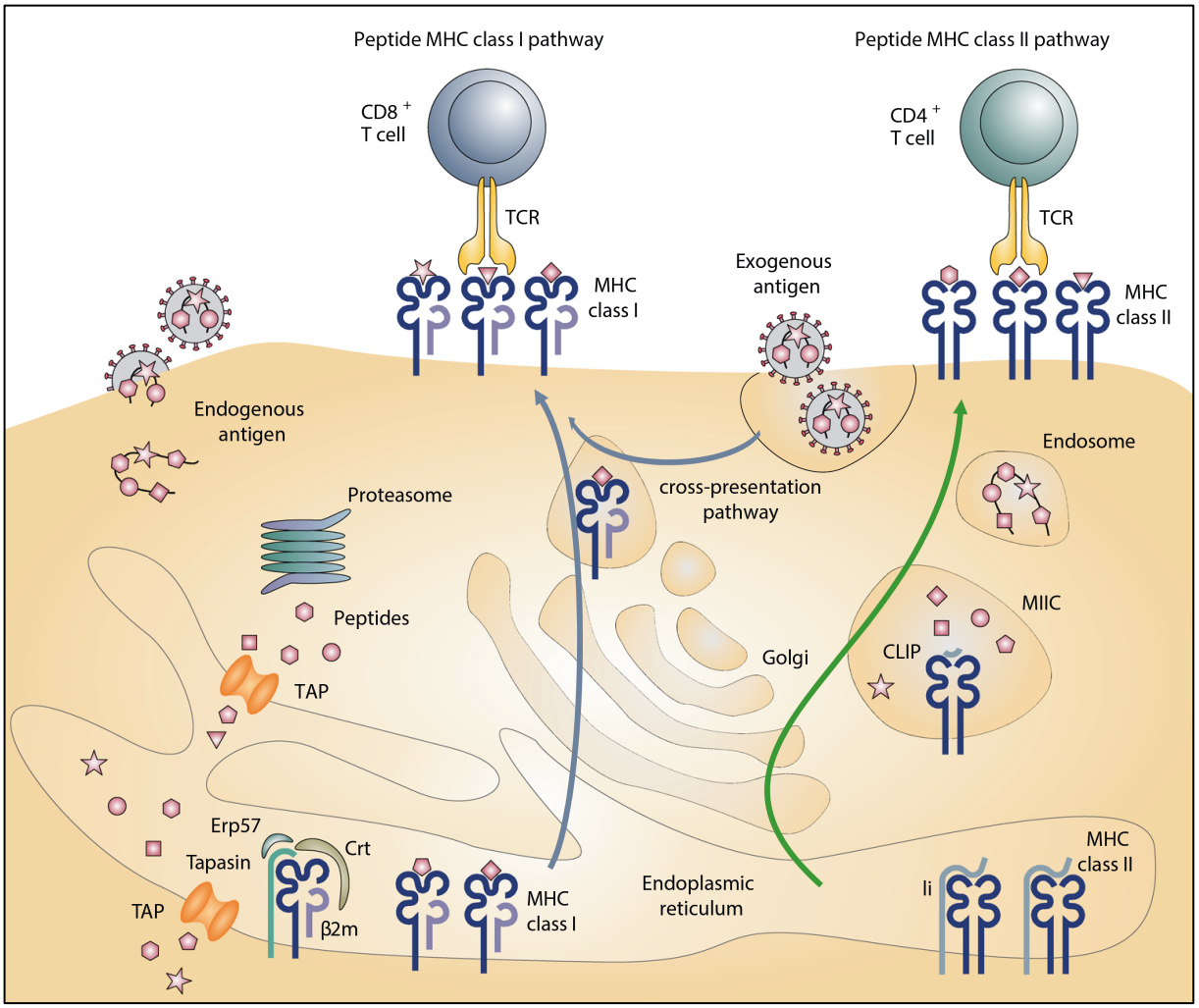


Figure 2: HLA class I and class II antigen presentation pathways. Peptides presented by HLA class I are typically derived from endogenous proteins degraded by the proteasome. After further trimming by aminopeptidases and import into the endoplasmatic reticulum via TAP, peptides are loaded onto HLA class I molecules stabilized by chaperones. After peptide binding, complexes are stabilized and transported to the cell surface. Recycled HLA class I molecules can be equipped with exogenously derived peptides in a process termed cross-presentation. HLA class II-presented peptides are derived from exogenous sources provided by endosomes. Immature HLA class II molecules are stabilized by the invariant chain (Ii) which is further trimmed to a short peptide sequence (CLIP). In the MHC class II compartments (MIIC), CLIP is exchanged with peptides from the endosome and the mature HLA class II molecules are transported to the cell surface. Reprinted and adapted from Purcell *et al.* (2019) with permission from Springer Nature under License Number 5050741330703.

1.2.3 Cancer and the immune system

Cancer development follows an evolutionary, Darwinian selection process with iterative rounds of clonal expansion, (epi)genetic modification, and ultimately clonal selection and “survival of the fittest” (Greaves and Maley, 2012). During this process, cancer cells acquire hallmark capabilities such as sustained proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis. More recently, the hallmark capabilities of deregulating cellular energy metabolism and avoiding detection and destruction by the immune system have been added. Genomic instability and mutation, and more recently tumor-promoting

inflammation have been proposed as the two enabling characteristics of cancer cells and their environment to acquire the described hallmark capabilities (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011).

1.2.3.1 Cancer immunoediting – elimination, equilibrium, escape

The hypothesis that the immune system can suppress tumors dates back to the beginning of the 20th century (Ehrlich, 1909) and was later coined "cancer immunosurveillance" and attributed to tumor-specific antigens (Burnet, 1957; Thomas, 1959; Burnet, 1964; Klein, 1966; Burnet, 1970). At the beginning of the 21st century, the hypothesis was further refined and extended. "Cancer immunoediting" describes both the host-protective and cancer-modeling effects of the immune system and can be divided into three stages: elimination, equilibrium, and escape (Dunn *et al.*, 2002; Dunn *et al.*, 2004). In the elimination phase immune cells of both, the innate and adaptive branches recognize and destroy transformed cells to prevent the outgrowth of tumors. The innate immune response to developing tumors is triggered by the sensing of danger signals as a result of local tissue disruption. Macrophages, NK cells, and other innate immune cells are recruited to the site of the developing tumor, secrete cytokines such as INFγ, and recognize transformed cells via DAMPs eventually killing them (Dunn *et al.*, 2004). Subsequently, tumor antigens released from dead cells are taken up from dendritic cells to engage cells of the adaptive immune system such as tumor-specific $CD4^+$ Th cells and tumor-specific $CD8^+$ Tc cells (via cross-presentation) in lymph nodes. Activated, tumor-specific $CD4^+$ Th and $CD8^+$ Tc cells migrate from the lymph node to the site of the developing tumor and to assist the innate immune system in the elimination of transformed cells (Dunn *et al.*, 2004; Vesely *et al.*, 2011). However, elimination is not always complete and some tumor cell variants enter the equilibrium phase in which immunoediting takes place. During the equilibrium phase, net tumor outgrowth is contained but the selective pressure applied by the immune system constantly leads to the generation of new tumor variants due to genetic instability mechanisms, *e.g.* MSI, with improved fitness (Dunn *et al.*, 2004; Mittal *et al.*, 2014). In the last phase, edited tumor cells escape the surveillance of the immune system through different mechanisms. The release of immunosuppressive cytokines by the tumor cells recruit myeloid suppressor cells which in turn inhibit tumor-specific $CD8^+$ Tc cell function as well as maturation and function of dendritic cells. Tryptophan degradation by indoleamine 2,3-dioxygenase and IFNγ secretion by dendritic cells as well as increased production of TGFβ by myeloid suppressor cells impairs $CD8^+$ Tc cell proliferation and function. Upregulation of immune checkpoint molecules on tumor cells such as CTLA-4 and PD-L1 provide inhibitory signals to $CD8^+$ Tc cells thus attenuating and terminating T cell responses. Finally, alterations of the antigen processing and presentation or the epitopes presented by it can render a tumor "invisible" to the immune system (Zitvogel *et al.*, 2006; Sharma and Allison, 2015). The latter two mechanisms can also be increasingly observed in MSI CRC. MSI CRCs have been shown to upregulate various immune checkpoint molecules including PD-L1 to counterbalance their high mutational burden and high immune cell infiltrates (Llosa *et al.*, 2015). Presentation of peptides by HLA class I molecules can be impaired by defects in the antigen-processing machinery (e.g. mutations in *TAP1/TAP2*) or HLA class I expression can be completely lost due to mutations or loss of *β2m*. Moreover, tumors can exhibit selective loss of an HLA haplotype (by loss of heterozygosity on chromosome 6), locus (by defective transcriptional regulation), or allele (by point mutations or gene deletions) leading to the presentation of

a less diverse peptide repertoire (Khong and Restifo, 2002; McGranahan *et al.*, 2017). Immunoselection of escaping tumors is further exemplified by the fact that oncogenic SNP mutations producing peptides poorly presented by both, HLA class I and II molecules are positively selected (Marty *et al.*, 2017; Marty Pyke *et al.*, 2018). Moreover, the expression of neoantigens can be suppressed transcriptionally (Rosenthal *et al.*, 2019), by translational repression (You *et al.*, 2007), or by the selective degradation of transcripts harboring premature termination codons (PTCs) by the nonsense-mediated RNA decay (NMD) pathway (Bokhari *et al.*, 2018; Jonchere *et al.*, 2018). MSI CRCs recurrently harbor mutations in *β2m*, HLA class I genes, and other genes involved in the antigen processing and presenting pathway (Kloor *et al.*, 2007; Grasso *et al.*, 2018). Moreover, InDel mutations frequently cause PTCs that activate the NMD pathway and MSI CRCs have been shown to express increased amounts of key players of the NMD pathway to limit the expression of frameshifted and potentially antigenic protein sequences and evade the detection by the immune system (El-Bchiri *et al.*, 2008; Bokhari *et al.*, 2018; Lindeboom *et al.*, 2019; Litchfield *et al.*, 2020).

1.3 Nonsense-mediated RNA decay

NMD is a highly conserved, translation-dependent quality control pathway that degrades mRNAs with PTCs introduced by either nonsense or frameshift mutations, transcription errors, incorrect splicing, or unprogrammed ribosomal frameshifting and thus limits the synthesis of C-terminally truncated and potentially harmful proteins (Kurosaki *et al.*, 2019). PTCs have been estimated to underlie one-third of all inherited and acquired diseases including cancer. For heterozygous carriers of nonsense mutations, NMD can be beneficial, because the degradation of the faulty mRNA prevents dominant-negative effects of truncated proteins, *e.g.* of the hemoglobin subunit beta in β-thalassemia, or in the case of dominant-negative forms of tumor suppressor proteins such as E-cadherin (stomach cancer), BRCA1 (breast and ovarian cancers), BRCA2 (ovarian cancer), p53 (breast cancer) or WT1 (kidney cancer; Holbrook *et al.*, 2004; Bhuvanagiri *et al.*, 2010; Popp and Maquat, 2018). In other cases, NMD activity can be detrimental, because it inhibits the expression of (partially) functional proteins, e.g. in Duchenne muscular dystrophy or cystic fibrosis (Holbrook *et al.*, 2004; Bhuvanagiri *et al.*, 2010) or limits anti-tumor immunity by degrading mRNAs coding for potentially antigenic protein sequences, *e.g.* in MSI CRC (El-Bchiri *et al.*, 2008; Litchfield *et al.*, 2020). Besides, NMD regulates the expression of error-free mRNAs (so-called endogenous targets) with upstream open reading frames (uORFs), exceptionally long 3' UTRs, splice junctions in the 3' UTR due to alternative splicing, or programmed ribosomal frameshifting and thereby influences physiological processes such as embryogenesis, development and stress responses (Ottens and Gehring, 2016).

1.3.1 Components of the NMD machinery

The NMD effector proteins include proteins of the up-frameshift (UPF) family (UPF1, UPF2, UPF3A, and UPF3B), proteins of the suppressor with morphogenetic effect on genitalia (SMG) proteins (SMG1, SMG5, SMG6, SMG7, SMG8, and SMG9), the components of the exon junction complex (EJC) as well as several other peripheral components such as NAG/NBAS, ICE1, SRSF1, DHX34, and MOV10 (Bhuvanagiri *et al.*, 2010; Lavysh and Neu-Yilik, 2020).

Both, UPF1 and SMG families were first identified and shown to be involved in mRNA turnover using genetic screens in *Saccharomyces cerevisiae* and *Caenorhabditis elegans*, respectively (Leeds *et al.*, 1992; Pulak and Anderson, 1993). UPF1 represents the key player of the NMD pathway and is an ATP-dependent DNA/RNA 5'-3' helicase with S/T-Q phosphorylation motifs at both, the N- and C-termini. UPF2 and UPF3B stimulate the phosphorylation of UPF1 to activate its ATPase and helicase functions and recruit mRNA degradation enzymes (Chamieh *et al.*, 2008; Ivanov *et al.*, 2008). More recently, UPF3B has been shown to directly interact with the release factors eRF1 and eRF3 and to dissociate terminating ribosomes (Neu-Yilik *et al.*, 2017). Its paralog UPF3A competes with UPF3B for UPF2 binding thus inhibiting NMD (Shum *et al.*, 2016). The phosphorylation cycle of UPF1 is mediated by the proteins of the SMG family. The kinase SMG1 forms a complex with SMG8/9 and is responsible for UPF1 phosphorylation, while SMG5/6/7 promote UPF1 dephosphorylation by the recruitment of phosphatases (Chakrabarti *et al.*, 2014). Furthermore, the SMG5/7 heterodimer recruits the CCR4-NOT deadenylation complex and licenses SMG6 endonuclease activity (Loh *et al.*, 2013; Boehm *et al.*, 2020). The EJC consists of the four core proteins Y14, MAGOH, eIF3A4, and BTZ (Gehring *et al.*, 2005). It is

deposited 20-24 nucleotides (nt) upstream of an exon-exon junction during splicing by the splicesome and serves as a binding platform for other proteins, *e.g.* UPF2 and UPF3B (Le Hir *et al.*, 2000; Le Hir *et al.*, 2001; Gehring *et al.*, 2005).

Importantly, only UPF1 and a translation termination event are essential to target an mRNA to the NMD pathway. All other components are promoting NMD, but none of them is indispensable which is also exemplified by the existence of various NMD branches, *e.g.* the UPF2-independent branch, the UPF3B-independent branch, and the EJC-independent branch (Lavysh and Neu-Yilik, 2020).

1.3.2 Target identification and decay – telling right from wrong

The mechanism by which the NMD machinery distinguishes a PTC from a normal termination codon, *i.e.* the difference between an aberrant and a normal termination process, has not yet been fully deciphered. However, two paradigmatic models have emerged for the termination phase of NMD to explain the degradation of different target mRNAs (Figure 3). In the “downstream marker” model an EJC persisting 3’ to a termination codon serves as a signal for aberrant translation termination (Kashima *et al.*, 2006), while in the “*faux* 3’ UTR” model the absence of the interaction between the stop codon-engaged ribosome and the termination-enhancing Poly(A) binding protein (PABP) triggers NMD (Amrani *et al.*, 2004). More recently, a unified model combined aspects of both, the “downstream marker” and “*faux* 3’ UTR” model and can, therefore, explain the degradation of mRNAs with spliced and unspliced 3’UTRs, respectively (Ivanov *et al.*, 2008; Mühlemann, 2008; Brogna and Wen, 2009).

Independent of the model, the phosphorylation of UPF1 represents the mechanistic endpoint of the aberrant translation phase. UPF1 promiscuously binds to mRNAs and is, together with the EJCs, usually removed by elongating ribosomes (Hogg and Goff, 2010; Kurosaki and Maquat, 2013). According to traditional models, premature termination of translation is thought to allow the interaction of UPF1 with the terminating ribosome as well as the SMG1/8/9 complex to form the so-called SURF complex (Gehring *et al.*, 2005; Kashima *et al.*, 2006; Neu-Yilik *et al.*, 2017). The “faux 3’ UTR” model postulates that the lacking interaction of the termination-enhancing PABP with the terminating ribosome instead allows the interaction with UPF1 with the terminating ribosome to trigger NMD (Singh *et al.*, 2008). According to the “downstream marker” model, an EJC more than 50-55 nt downstream of the PTC provides a suitable protein-binding platform for the NMD factors UPF2 and UPF3B which stabilize the interaction between UPF1 and the terminating ribosome (Thermann *et al.*, 1998; Gehring *et al.*, 2005). Subsequently, ribosomes are disassembled by either UPF3B or ABCE1 (Neu-Yilik *et al.*, 2017; Annibaldis *et al.*, 2019) and phosphorylation of UPF1 by SMG1/8/9 leads to translational repression and initiates the decay phase (Kashima *et al.*, 2006; Isken *et al.*, 2008; Yamashita *et al.*, 2009). Here, target mRNAs are degraded endo- and exonucleolytically. Hyperphosphorylated UPF1 provides the binding sites for the SMG5/7 heterodimer which recruits the decapping complex DCP1/2 and the exonuclease XRN1 via CCR4-NOT (Yamashita *et al.*, 2005; Okada-Katsuhata *et al.*, 2011; Loh *et al.*, 2013; Chakrabarti *et al.*, 2014). Moreover, SMG5/7 authorizes the endonucleolytic activity of SMG6 which cleaves target mRNAs directly at or close to the PTC, and SMG5/7 as well as SMG6 are involved in UPF1 dephosphorylation and recycling (Chiu *et al.*, 2003; Boehm *et al.*, 2014; Boehm *et al.*, 2020).

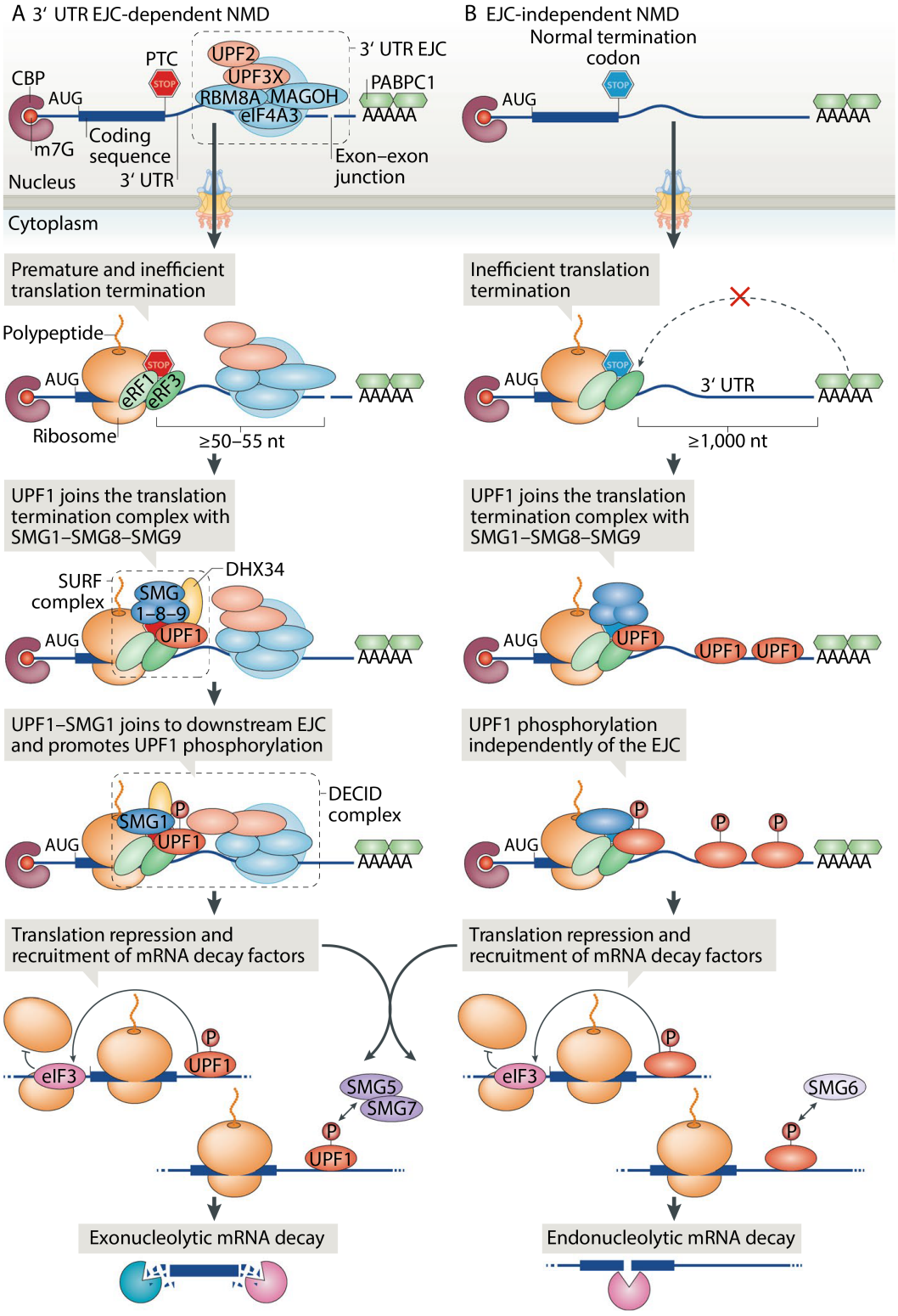


Figure 3: NMD target identification and decay. EJC-dependent NMD occurs when the ribosome encounters a premature termination codon located more than 50-55 nucleotides upstream of the last exon junction complex. UPF1 and SMG-1-8-9 join the terminating ribosome to form the SURF complex, and UPF1 phosphorylation by SMG1 ultimately leads to mRNA degradation by exonucleolytic and endonucleolytic decay (panel A). EJC-independent NMD occurs on mRNAs with long 3' UTRs. Inefficient translation termination due to missing interaction between PABPC1 and the terminating ribosome allows the interaction of UPF1 instead. Subsequent UPF1 phosphorylation and recruitment of decay factors lead to mRNA degradation (panel B). Reprinted and adapted from Kurosaki *et al.* (2019) with permission from Springer Nature under License Number 5050740821512.

1.3.3 Pharmacological modification of NMD

Pharmacological modification of NMD has emerged as a promising strategy for the treatment of various genetic diseases as well as some forms of cancer. To date, several small molecules have been described to inhibit NMD and three major mechanisms are known: general inhibition of translation, direct inhibition of components of the NMD machinery, and translational readthrough.

Since NMD is a translation-dependent quality control mechanism, translation inhibitors such as cycloheximide, emetine, puromycine, or anisomycin interfere with NMD (Carter *et al.*, 1995). However, their general effect impairs both the translation of normal transcripts as well as the expression of partially functional proteins from NMD-targeted transcripts. Caffeine and Wortmannin inhibit SMG1 thus preventing UPF1 phosphorylation and subsequently NMD activation but their known toxicity makes clinical use impossible (Usuki *et al.*, 2004). The natural product Pateamine A directly interacts with eIF4A3 and inhibits NMD by increasing the stability of the interaction between UPF1 and the EJC (Dang *et al.*, 2009). More recently, two other inhibitors of the SMG family have been identified (Durand *et al.*, 2007; Martin *et al.*, 2014). NMDI1 prevents the interaction between UPF1 and SMG5 leading to the stabilization of hyperphosphorylated UPF1 and NMD inhibition (Durand *et al.*, 2007). NMDI14 was identified by virtual ligand screening and disrupts the interaction between UPF1 and SMG7 and stabilizes NMD targets (Martin *et al.*, 2014). Amlexanox also stabilizes NMD targets by a yet unknown mechanism and additionally leads to the synthesis of functional full-length proteins by promoting translational readthrough (Gonzalez-Hilarion *et al.*, 2012). Other readthrough-promoting agents are aminoglycosides such as G418 and gentamycin which cause the insertion of a near-cognate AA at the site of the PTC (Howard *et al.*, 1996; Keeling *et al.*, 2012). A major drawback of aminoglycosides is their toxicity at concentrations needed for NMD inhibition (Bhuvanagiri *et al.*, 2010). 2,6-diaminopurine was also shown to promote readthrough which is, however, limited to UGA stop codons (Trzaska *et al.*, 2020). PTC124 (Ataluren) has been suggested as an alternative promoting selective readthrough at PTCs and with a more favorable safety profile (Welch *et al.*, 2007; Roy *et al.*, 2016). However, the NMD inhibitory effect of PTC124 has been a matter of debate and the initial discovery has been attributed to false-positive results in flawed reporter gene assays (Auld *et al.*, 2009; Auld *et al.*, 2010; McElroy *et al.*, 2013).

Recently, 5-azacytidine (5AZA) has been identified as a novel NMD inhibitor. While its exact mechanism of NMD inhibition remains unknown, it has been shown that the effect depends on the expression of MYC (Bhuvanagiri *et al.*, 2014). 5AZA is a pyrimidine analog of the nucleoside cytidine approved for the treatment of myelodysplastic syndromes and acute myeloid leukemia (Keating, 2012; Müller and Florek, 2014). 5AZA is incorporated into both RNA and DNA and causes their demethylation by trapping and thus inhibiting methyltransferases (Stresemann and Lyko, 2008; Schaefer *et al.*, 2009; Hagemann *et al.*, 2011). Interestingly, the deoxy derivate 5-aza-2'-deoxycytidine (dAZA), which is incorporated in DNA but not RNA, does not affect NMD efficiency (Bhuvanagiri *et al.*, 2014). Therefore, it has been hypothesized, that 5AZA could exert its NMD-inhibitory effect by altering the RNA interactome. The NMD-inhibitory effect is achieved at concentrations similar or lower than those used in anti-leukemic therapy (Stresemann and Lyko, 2008), which renders 5AZA a strong candidate with clinical potential to increase the expression of proteins from PTC-bearing transcripts. In the setting of MSI CRC, inhibition of NMD has been hypothesized to increase anti-tumor immunity and subsequently the effectiveness of

cancer immunotherapy by augmenting the expression of tumor-specific antigens (El-Bchiri *et al.*, 2008; Pastor *et al.*, 2010; Kim *et al.*, 2013b; Lindeboom *et al.*, 2019; Pawlicka *et al.*, 2020).

1.4 Cancer immunotherapy

Only recently cancer immunotherapy has been established as the fourth pillar of cancer treatment, however, the idea of harnessing the immune system to treat cancer dates back to the 19th century when it was discovered that infection with *Streptococcus pyogenes* caused tumor shrinkage (Busch, 1868; Fehleisen, 1882). William B. Coley, who is today known as the "Father of Immunotherapy", further expanded these observations and developed the first commercially available immune-based treatment for cancer, *Coley's toxins* (Coley, 1891; Dobosz and Dzieciatkowski, 2019). This strategy was applied again almost 100 years later with the introduction of the tuberculosis vaccine Bacille Calmette-Guérin (BCG) as a treatment for bladder cancer in 1976 and the introduction of vaccines preventing cancer-causing HBV and HPV infections marked another milestone in the development of cancer immunotherapy (Morales *et al.*, 1976; Finn, 2018). Moreover, the identification and functional characterization of CTLA-4 and PD-1 (and its ligands PD-L1/PD-L2) led to the development of monoclonal antibodies targeting these crucial immune checkpoint molecules and immune checkpoint blockade is now approved for several cancer types (Brunet *et al.*, 1987; Ishida *et al.*, 1992; Krummel and Allison, 1995; Leach *et al.*, 1996; Bardhan *et al.*, 2016; Dobosz and Dzieciatkowski, 2019). In 2013, cancer immunotherapy was named *Science*'s "Breakthrough of the Year" and in 2019, James P. Allison and Tasuku Honjo were awarded the Nobel Prize in Physiology and Medicine for their discovery of the immune checkpoint molecules (Couzin-Frankel, 2013). Today, the major categories of immunotherapy are immune checkpoint inhibitors, cancer vaccines, and adoptive cell transfer which all rely on the expression of tumor (neo)antigens and their recognition by cells of the immune system (Dobosz and Dzieciatkowski, 2019; Zhang and Zhang, 2020).

1.4.1 Tumor antigens

Generally, tumor antigens can be classified into three different subcategories, tumor-specific antigens (TSAs), tumor-associated antigens (TAAs), and cancer/testis antigens (CTAs). CTAs are absent on healthy, adult cells except in reproductive tissues and are frequently re-expressed in several tumor types. Well-studied CTAs include the MAGE (melanoma-associated antigen) family and cancer-testis antigen 1 (NY-ESO-1). TAAs show low expression levels on healthy host cells, but a significantly higher expression on cancer cells. Examples of overexpressed TAAs include HER2 (epidermal growth factor receptor 2) in breast cancer or CD19 in B cell malignancies (Yarchoan *et al.*, 2017b; Hu *et al.*, 2018). Importantly, immune responses against both CTAs and TAAs are limited by central tolerance mechanisms and lack of complete specificity to cancer cells. Moreover, meta-analyses of clinical studies using CTA- or TAA-based vaccines reported only a limited success rate concluding that there is no clinical effectiveness (Rosenberg *et al.*, 2004; Tran *et al.*, 2017). In contrast, TSAs represent optimal targets for vaccine development and cancer immunotherapy since they are expressed only in cancer cells but not in healthy host cells and therefore not subject to central tolerance. TSAs can be either oncogenic viral antigens (such as the HPV oncoproteins E6 and E7) or are generated due to somatic mutations in cancer cells and termed neoantigens (Yarchoan *et al.*, 2017b; Hu *et al.*, 2018; Schoenberger, 2018).

1.4.2 Adoptive cell transfer – engineered immunity

For adoptive cell transfer, lymphocytes (*i.e.* B and T cells) are isolated from peripheral blood or tumor tissue of the patient, activated and expanded in vitro, and eventually re-infused after lymphodepletion. Importantly, the molecular targets of these lymphocytes are exonic mutations expressed by cancer cells (Robbins *et al.*, 2013; Lu *et al.*, 2014; Tran *et al.*, 2014). Patients with melanoma and cervical cancer have been successfully treated using this approach (Rosenberg *et al.*, 1988; Stevanović *et al.*, 2015). Moreover, cells can be enriched for those recognizing tumor-specific epitopes presented via autologous HLA by APCs or engineered to recognize tumor-specific antigens or epitopes on tumor cells before re-infusion. Currently, two kinds of engineered cells are used in adoptive cell transfers: chimeric antigen receptor (CAR) T cells, and TCR-engineered T cells (Rosenberg and Restifo, 2015). CAR T cells express receptors composed of a variable antibody domain to recognize non-HLA (tumor-specific) antigens expressed at the cell surface linked to CD3-zeta for the intracellular signal transmission and other costimulatory domains for T cell activation (Gross *et al.*, 1989; Maher *et al.*, 2002). Most CAR T cell therapies target CD19 and thus are clinically used to treat B cell malignancies such as chronic lymphocytic leukemia and acute lymphocytic leukemia (Kochenderfer *et al.*, 2010; Kalos *et al.*, 2011; Brentjens *et al.*, 2013; Maude *et al.*, 2018). TCR-engineered T cells recognize HLA-restricted, tumor-associated epitopes originating from *e.g.* MART-1 or NY-ESO and have successfully been used in the treatment of melanoma and synovial sarcoma (Clay *et al.*, 1999; Morgan *et al.*, 2006; Robbins *et al.*, 2011).

1.4.3 Immune checkpoint inhibition – releasing the break

Immune checkpoint molecules such as CTLA-4 and PD-1 (and its ligands PD-L1/PD-L2) are involved in the regulation of T cell activation. CTLA-4 acts as a co-inhibitory molecule and competes with the co-stimulatory molecule CD28 for the binding of B7 molecules on dendritic cells. T cells missing the co-stimulatory second signal are not activated leading to tumor escape (Pardoll, 2012; Chen and Flies, 2013). PD-1 interacts with PD-L1 to induce T cell exhaustion and suppress proliferation (Freeman *et al.*, 2000; Keir *et al.*, 2008; Pardoll, 2012). PD-L1 is overexpressed by cancer cells to escape immunosurveillance (Iwai *et al.*, 2002). Consequently, disruption of the CTLA-4:B7 or PD-1:PD-L1 interaction by monoclonal antibodies releases the break of T cell activation and various antibodies have been approved for the treatment of an ever-growing list of cancers including melanoma, lymphoma, lung cancers, renal cell cancer, head and neck squamous cell cancer, bladder cancer, liver cancer gastroesophageal cancer as well as MSI CRCs (Hargadon *et al.*, 2018; Vaddepally *et al.*, 2020). In a first of its kind approval, the FDA approved the PD-1 inhibitor Pembrolizumab for the treatment of both adult and pediatric patients independently of the tumor site and only based on the prognostic biomarker of MSI-high or MMR deficiency (Marcus *et al.*, 2019). Moreover, it has been shown that T cells, re-activated by immune checkpoint inhibition target tumor-specific epitopes explaining the extraordinary success of these strategies in cancers with high mutational burden (van Rooij *et al.*, 2013; Gubin *et al.*, 2014; McGranahan *et al.*, 2016).

1.4.4 Therapeutic cancer vaccines – steering T cells

In contrast to preventive cancer vaccines against HBV or HPV, therapeutic cancer vaccines target existing cancers by inducing T cell-mediated immune responses against tumor (neo)antigens. As previously described, earlier clinical vaccine trials based on CTAs and TAAs had very limited success. However, advances in both next-generation sequencing and more recently mass spectrometry improved the identification of tumor-specific neoantigens and vaccines targeting mutant neoepitopes in glioma, melanoma, and MSI CRC have been reported (Schumacher *et al.*, 2014; Ott *et al.*, 2017; Sahin *et al.*, 2017; Bassani-Sternberg *et al.*, 2019; Hilf *et al.*, 2019; Kloor *et al.*, 2020b; Sahin *et al.*, 2020; Platten *et al.*, 2021).

1.5 Identification of tumor-specific neoepitopes

Recent breakthroughs in sensitivity and reproducibility of mass spectrometry (MS) enable the unbiased exploration of the global immunopeptidome presented by the HLA system, including neoepitopes (Bassani-Sternberg *et al.*, 2015; Bassani-Sternberg *et al.*, 2016; Chong *et al.*, 2018; Chong *et al.*, 2019; Bilich *et al.*, 2020; Chen *et al.*, 2020b). Previously utilized methodologies in the search for neoantigens such as HLA binding predictions and indirect immunological read-outs are laborious and time-consuming or have been afflicted with a high rate of false discoveries (Bassani-Sternberg and Coukos, 2016; Schmidt *et al.*, 2017). MS-based immunopeptidomics approaches overcome these limitations by the direct identification of actually HLA-presented (neo)epitopes and thus represent a promising technology for the development of personalized cancer immunotherapies.

The following chapter aims to convey basic knowledge in the field of MS-based proteomics and immunopeptidomics and introduces the terms and concepts underlying the presented thesis.

1.5.1 Fundamentals of mass spectrometry for proteomics and peptidomics

A mass spectrometer is capable of measuring the mass to charge (*m/z*) ratios and intensities of ions. The resulting diagram is termed mass spectrum and allows conclusions about the composition of the sample, *i.e.* the contained analytes and their characteristics.

Both in proteomics and peptidomics, these analytes are peptides. While in proteomics sample preparation, peptides are generated by (tryptic) digestion of proteins contained in the sample, in peptidomics enzymatic digestion is not required. Before the peptides enter the mass spectrometer, they are often pre-separated by liquid chromatography, gas chromatography, and/or capillary electrophoresis to reduce sample complexity. The most commonly used separation technique in modern proteomics/peptidomics is reverse-phase liquid chromatography (RPLC) which separates peptides according to their hydrophobicity. The stationary phase in RPLC usually consists of silica particles covered with long alkyl groups (*e.g.* C_{18}) allowing adsorption and retention of hydrophobic analytes. The adsorbed peptides are subsequently eluted by an aqueous mobile phase with an increasing gradient of organic solvent (Aguilar and Hearn, 1996; Sandra *et al.*, 2008; Rozenbrand and van Bennekom, 2011).

The ionization of the peptides is mandatory in order to be detected and analyzed by the mass spectrometer. The two most commonly used techniques are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI; Karas and Hillenkamp, 1988; Tanaka *et al.*, 1988; Fenn *et al.*, 1989). While today MALDI is mostly applied for specialized applications such as MS imaging, ESI is widely used in both proteomics and peptidomics (Mann, 2019). ESI can be performed at atmospheric pressure, is well suited for small biomolecules such as peptides, and is considered as a “soft” ionization form conserving molecular ions. Moreover, it is performed in solution and thus allows the direct (“on-line”) coupling of RPLC to the mass spectrometer (Aebersold and Mann, 2003; Cramer, 2009). ESI is performed by applying a high voltage to the analyte solution causing the enrichment of positive ions near its meniscus. This leads to the destabilization of the meniscus and the formation of the so-called Taylor cone and jet. The jet eventually splits into highly positive charged droplets and evaporation of the analyte solution further increases the charge density and destabilizes the droplets. For the transmission of ions into the gas phase, two models have been proposed: Either droplets continue to disintegrate until they

contain only a single ion and the liquid completely evaporates or droplets shrink until they reach a critical size and then emit ions into the gas phase (Fenn *et al.*, 1989; Kebarle and Verkerk, 2009; Wilm, 2011; Konermann *et al.*, 2013). Subsequently, the ions, *i.e.* charged peptides, enter the mass spectrometer and can be measured by different mass analyzers.

Mass analyzers determine the *m/z* ratio of either intact ions (termed “precursor ions”) or fragments thereof (termed “fragment ions”). For this purpose, they can operate either in “scanning” mode recording all ions within a specific *m/z* range or in “selecting” mode isolating ions based on a specific *m/z*. Different types of mass analyzers such as quadrupole, ion trap, time-of-flight, and Fourier transform analyzers exist and are used independently or in combination (Noor *et al.*, 2020). In the presented thesis, an Orbitrap Fusion Lumos tribrid mass spectrometer (Figure 4) was used which combines a quadrupole, a linear ion trap, and an orbitrap mass analyzer (Martins *et al.*, 2016; Hecht *et al.*, 2019). Quadrupole analyzers or mass filters consist of four parallel rods forming a tube. A direct current and a radiofrequency current are applied at opposing rods to create an electric field. Depending on the electric field applied, only ions with a specific *m/z* are allowed to pass the tube and can be detected. Quadrupoles can also be used for ion transmission within the mass spectrometer by applying only a radiofrequency allowing all ions to pass through. Linear ion traps have a similar architecture as quadrupoles with the addition of two cap electrodes at both ends of the tube. By applying a potential field at the cap electrodes ions can be trapped within the quadrupole. Moreover, linear ion traps can accumulate ions over time thereby improving the sensitivity and can be used for ion acceleration and fragmentation (Niessen and Falck, 2015; Haag, 2016). The Orbitrap Fusion Lumos tribrid mass spectrometer uses a dual-pressure linear ion trap combining a high-pressure cell to capture, select and fragment ion and a low-pressure cell with a fast scanning capability (Olsen *et al.*, 2009; Second *et al.*, 2009). Orbitrap mass analyzers are composed of an inner spindle electrode and an outer hollow electrode. A voltage potential between the inner and outer electrodes causes ions injected into the orbitrap to spin around the inner electrode. Ions with the same *m/z* have the same frequency of oscillation along the spindle electrode. These oscillations are recorded by the outer electrode and can be transformed into a mass spectrum via Fourier transformation (Makarov, 2000; Niessen and Falck, 2015; Haag, 2016). After the mass analyzer, the ion abundance is measured by a detector that converts ion incidence into an electrical current. While for quadrupole, ion trap, and time-of-flight analyzers the signal is amplified and detected separately, orbitrap mass analyzers possess built-in image current detectors (Makarov, 2000; Noor *et al.*, 2020).

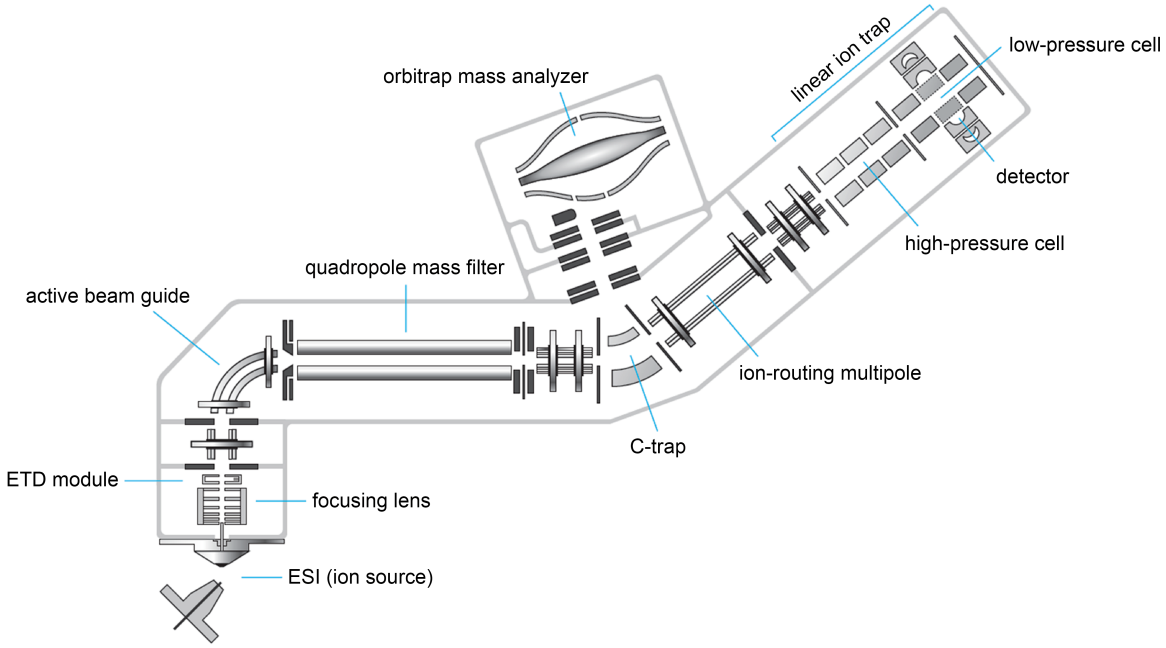


Figure 4: Schematic illustration of an Orbitrap Fusion Lumos tribrid mass spectrometer used in this thesis. Adapted from https://planetorbitrap.com/orbitrap-fusion-lumos#tab:schematic.

Modern proteomics and peptidomics workflows make use of multiple analyzers to record mass spectra of precursor and fragment ions. First, *m/z* ratios of the injected peptides are recorded (MS1/full scan) and based on this information precursor ions can be selected and fragmented. Precursor ions can either be selected based on their intensity (untargeted approach) or based on a specific, pre-defined *m/z* ratio (targeted approach). Next, the *m/z* ratios of the corresponding fragment ions are obtained (MS2). Finally, the combined information can be used to determine the peptide identity, *i.e.* the aa sequence of the analyzed precursor. Generally, fragment ions are generated by the addition of energy to selected and isolated precursor ions, and two commonly used fragmentation techniques are collision-induced dissociation and electron transfer dissociation (Noor *et al.*, 2020). For collision-induced dissociation (CID) selected precursor ions are accelerated and collided with a target gas converting translational ion energy to internal ion energy and as a consequence, precursor ions are decomposed into fragment ions (Niessen and Falck, 2015). Higher-energy collisional dissociation (HCD) is an advancement of CID improving the detection of ions with low molecular masses (Olsen *et al.*, 2007). In electron transfer dissociation (ETD), an electron is transferred from a radical anion (usually generated from ionized fluoranthene) to the positively charged precursor ion. This generates an unstable radical precursor ion which is decomposed into fragment ions. The principles of the described fragmentation techniques are fundamentally different and generate different fragment ions: CID and HCD cause cleavage of the peptide bond resulting in b- and y-ions, while ETD causes cleavage between the α-carbon and the nitrogen atom of the peptide backbone resulting in c- and z-ions (Roepstorff and Fohlman, 1984; Biemann, 1988; Syka *et al.*, 2004; Kim and Pandey, 2012). Therefore, spectra obtained with the HCD and ETD techniques provide complementary information. Recently, a combination of the two techniques termed electron-transfer/higher-energy collisional dissociation (EThcD) has been introduced. Here, the unreacted precursor ions after ETD are further fragmented using HCD to generate information-rich mass spectra with b-, c-, y-, and z-ions (Frese *et al.*, 2012).

The obtained spectrum data can be used for the identification of peptides and their corresponding aa sequence using different strategies such as spectral matching, database search, or *de novo* sequencing. Spectral matching is performed by comparing previously recorded MS2 spectra with correct identification of peptides with newly obtained data (Noor *et al.*, 2020). In contrast, database searches compare theoretical MS2 spectra of all precursors with a similar *m/z* ratio to a given recorded MS2 spectrum (Eng *et al.*, 1994). For each of these comparisons, the search engine calculates a score, and the highest-scoring peptide-spectrum match (PSM) is reported. Moreover, recorded spectra are matched against a decoy database in order to separate correct and false identifications based on a false discovery rate (FDR; Elias and Gygi, 2007; 2010; Zhang *et al.*, 2012). Peptide *de novo* sequencing performs peptide identification solely based on spectrum information and without the need for a database. The aa sequence (or a sequence tag for spectra with incomplete ion series) is deduced from the mass differences of spectrum peaks (Ma *et al.*, 2003; Sinitcyn *et al.*, 2018; Chen *et al.*, 2020a). Additionally, hybrid approaches have been introduced combining peptide *de novo* sequencing and database search approaches (Taylor and Johnson, 1997; Zhang *et al.*, 2012).

Finally, quantification of analytes can be performed either label-based or label-free. The two most commonly used approaches for label-based quantification are SILAC and TMT/iTRAQ labeling (Bantscheff *et al.*, 2012). Stable isotope labeling with amino acids in cell culture (SILAC) is a metabolic labeling strategy in which isotopically labeled amino acids are incorporated into proteins *in vivo* and differentially labeled samples are quantified based on their MS1 isotope patterns (Ong *et al.*, 2002). In experiments using tandem mass tags (TMT) and isobaric tags for relative and absolute quantitation (iTRAQ) for quantification, peptides from different samples are labeled using isobaric tags. The isobaric tags have the same mass (*i.e.* the same peptides with different tags have an identical *m/z* ratio in MS1), but can be distinguished and quantified at the MS2 level after fragmentation (Thompson *et al.*, 2003; Ross *et al.*, 2004; Wiese *et al.*, 2007; Sinitcyn *et al.*, 2018). Label-free quantification is based on the integration of MS1 ion intensities of a peptide over its elution profile and compared between different MS runs. In contrast to label-based quantification strategies, the number of conditions compared is not limited by the availability of distinguishable labels (Bantscheff *et al.*, 2012).

1.5.2 Immunopeptidomics

Immunopeptidomics is the MS-based investigation of HLA-presented peptides and to date, represents the only unbiased method to identify tumor-specific neoepitopes. The field of immunopeptidomics first emerged in the early 1990s when peptides presented by HLA molecules were characterized by sequencing via Edman degradation and mass spectrometry (Rötzschke *et al.*, 1990; Falk *et al.*, 1991; Hunt *et al.*, 1992). More recently, the field gained new momentum with the introduction of high-sensitivity MS instruments as well as the advent of cancer immunotherapy and the finding that T cells, re-activated by immune checkpoint blockade, recognize tumor-specific neoepitopes (van Rooij *et al.*, 2013; Gubin *et al.*, 2014; Ramarathinam *et al.*, 2018). In contrast to these technical advancements, the principles used for immunopeptidomics sample preparation have remained largely the same and involve either mild acid elution (MAE) or immunoprecipitation (IP). MAE is performed by washing cells with low pH for dissociating non-covalently bound B2M and peptides from HLA molecules. The major advantages of MAE are the fast and detergent-free protocol as well as the reduced loss of low-affinity peptides

compared to IP. However, the unspecific dissociation of other peptides at the cell membrane frequently causes contaminations of up to 60 % in MAE samples (Freudenmann *et al.*, 2018; Chen *et al.*, 2019). These contaminations can be drastically reduced by using a specific antibody for IP in the sample preparation protocol. For IP, cells or tissues are lysed with non-denaturing detergents to conserve HLA:peptide complexes which can be purified by immunoaffinity chromatography. The antibodies commonly used are either the well-described pan-HLA class I antibody W6/32 or antibodies specific to a defined HLA allotype (Barnstable *et al.*, 1978; Freudenmann *et al.*, 2018). The antibodies can be coupled covalently or non-covalently to Protein A/G beads, or sepharose and agarose resins. After the capture of the HLA:peptide complex, the beads or resins are washed to remove unspecific binders, and finally, HLA:peptide complexes are eluted with a strong acid. Isolation of HLA-presented peptides by IP represents the currently preferred method because of its high specificity and significantly higher yield compared to MAE (Lanoix *et al.*, 2018). Additionally, a recently published high-throughput immunoprecipitation protocol reduces sample handling and thus peptide losses generating highly reproducible data to investigate clinical samples or the effects of drug treatments on the immunopeptidome (Chong *et al.*, 2018). The peptides can be further separated from the much larger HLA and B2M molecules using ultrafiltration or C_{18} fractionation before being subjected to MS analysis (Freudenmann *et al.*, 2018; Purcell *et al.*, 2019). Both MS acquisition and subsequent data analysis have been developed for the analysis of tryptic peptides and are therefore inherently more difficult to apply to immunopeptidomics. Recent developments in the field to overcome these limitations are the use of dual fragmentation methods such as EThcD to generate information-rich spectra and the incorporation of peptide *de novo* sequencing tools for reliable identification of HLA-presented peptides (Mommen *et al.*, 2014; Faridi *et al.*, 2018; Narayan *et al.*, 2019).

1.6 Aims of the thesis

CRC is the fourth most common malignancy diagnosed and caused more than 500,000 cancer-related deaths worldwide in 2018. MSI CRCs are a common and therefore paradigmatic hereditary malignancy in developed countries caused by the inactivation of the DNA MMR system and account for about 15 % of all CRCs. Importantly, the inactivation of the DNA MMR system leads to the accumulation of somatic InDel mutations generating tumor-specific protein sequences which, if presented by the HLA system, represent an attractive target for personalized cancer immunotherapy approaches. Traditionally used workflows for the identification of tumor-specific neoepitopes such as HLA binding prediction and indirect cellular assays are laborious and time-consuming. Moreover, their indirect nature often identifies candidate peptides not necessarily presented by tumor cells *in vivo*. However, recent advances in mass spectrometry enable the unbiased exploration of the global immunopeptidome presented by HLA molecules and thus the direct identification of tumor-specific epitopes that may serve as targets for immunotherapy.

It was the first aim of this thesis to establish an immunopeptidomics pipeline allowing the comprehensive, MS-based identification of HLA class I-presented peptides including tumor-specific neoepitopes originating from InDel mutations.

Besides generating tumor-specific protein sequences, InDel mutations frequently introduce PTCs into the open reading frame. PTC-bearing transcripts are eliminated by NMD limiting their expression and consequently, their HLA class I-mediated presentation and recognition of tumor-specific InDel neoepitopes.

It was the second aim of this thesis to test if NMD inhibition by the clinically approved drug 5AZA could increase HLA class I-mediated presentation of InDel neoepitopes and thus, potentially, the immune recognition of cancer cells by the host’s immune system.

2 Materials and Methods

All general reagents, chemicals, and materials were purchased from Carl Roth (Karlsruhe, Germany), Sigma-Aldrich (St. Louis, MO, USA), Merck Millipore (Darmstadt, Germany), or Thermo Fisher Scientific (Waltham, MA, USA) unless otherwise stated. Table 9 - Table 16 in Section 2.8 provide the key resources essential to reproduce the results presented in this thesis.

2.1 Cell culture

Cell culture work was carried out under sterile conditions in a Safe 2000 Class II Biological Safety Cabinet (Thermo Fisher Scientific; Waltham, MA, USA). All standard plasticware and consumables used for cell culture were obtained from Sarstedt (Nümbrecht, Germany).

2.1.1 Cell lines and general culture conditions

HCT-116, RKO, and LoVo cell lines were provided by the Department of Applied Tumor Biology at the Heidelberg University Hospital's Institute of Pathology (Heidelberg, Germany). All three cell lines used have been established from human colon carcinoma (Drewinko *et al.*, 1976; Brattain *et al.*, 1981; Brattain *et al.*, 1984). HLA class I types of HCT-116 were previously determined as HLA-A*01:01, HLA-A*02:01, HLA-B*18:01, HLA-B*45:01, HLA-C*05:01, and HLA-C*07:01. HLA class I types of RKO were previously determined as HLA-A*03:01, HLA-B*18:01, HLA-B*35:08, and HLA-C*07:01. HLA class I types for LoVo cells were previously determined as HLA-A*01:01, HLA-A*32:01, HLA-B*57:01, HLA-B*27:08, and HLA-C*06:02 (Scholtalbers *et al.*, 2015). HLA class I types for HCT-116 and RKO were confirmed by sequencing in the DKMS Life Science Lab GmbH (Dresden, Germany). HCT-116 and LoVo cells were maintained in Gibco RPMI 1640 [+] L-Glutamine (Thermo Fisher Scientific; Waltham, MA, USA) supplemented with 10 % FBS (Thermo Fisher Scientific; Waltham, MA, USA) and 1 % penicillin-streptomycin (Sigma-Aldrich; St. Louis, MO, USA). RKO cells were maintained in DMEM/F-12 (1:1) [+] L-Glutamine (Thermo Fisher Scientific; Waltham, MA, USA) supplemented with 10 % FBS and 1 % penicillin-streptomycin. Growth was monitored by visual inspection with a CKX41 inverted microscope (Olympus; Tokyo, Japan). Depending on their density, all cell lines were passaged two to three times per week. For this purpose, cells were washed once with pre-warmed phosphate-buffered saline (PBS) to remove dead cells. The remaining adherent cells were detached using Trypsin/EDTA (Biochrom; Berlin, Germany), diluted in fresh complete medium (medium containing FBS and penicillin-streptomycin), and then usually split 1:5 (LoVo), 1:10 (HCT-116), and 1:20 (RKO), respectively. The HB-95 cell line was a generous gift by Michal Bassani-Sternberg (University of Lausanne; Lausanne, Schwitzerland). The HB-95 cell line is an antibody-secreting hybridoma cell line generated by the fusion of murine myeloma cells and murine spleen cells. Murine spleen cells were extracted from mice immunized with membrane from human tonsil lymphocytes (Barnstable *et al.*, 1978). The hybridoma cell line produces the monoclonal W6/32 antibody directed against HLA-A, HLA-B, and HLA-C (Parham *et al.*, 1979; Brodsky and Parham, 1982). HB-95 cells were cultured in small-scale bioreactors as described in detail in section 2.1.3. All cells were cultivated in a Binder CB Series CO_2 incubator (Tuttlingen, Germany) at 37 °C and 5 % CO_2.

2.1.2 Mycoplasma testing

Mycoplasma infection of cell cultures can significantly alter gene and protein expression profiles and thus confound experimental results (Miller *et al.*, 2003; Hoff *et al.*, 2018). Cell lines used in this study were regularly tested for mycoplasma contamination by PCR. To this end, either 100 µl supernatant of a dense cell culture or $1x10^5$ cells were used as samples. Both were incubated at 95 °C for 10-15 min and centrifuged at 13,000 x g for 5 min to pellet cellular debris. Supernatants were transferred to fresh tubes and directly used in the PCR reaction according to Table 3 and Table 4. The primers used (forward: 5'-CGGATAACGCTTGCGACCTATG-3'; reverse: 5'-GGCGAATGGGTGAGTAACACG-3') are specific to the 16S rRNA of mycoplasma and amplify a 450 bp PCR product. PCR reactions were analyzed by agarose gel electrophoresis.

2.1.3 Hybridoma cell culture

HB-95 cells were cultivated in CELLine CL1000 bioreactors (Sigma-Aldrich; St. Louis, MO, USA). These small-scale bioreactors consist of two compartments separated by a semi-permeable membrane. The large upper compartment contains the growth medium, while the smaller, lower compartment is used for cell cultivation. The semi-permeable membrane allows the exchange of small molecules for continuous nutrient supply and waste elimination, however retains the produced antibody in the cell cultivation compartment. An additional gas membrane at the bottom of the cell cultivation compartment ensures optimal oxygenation and gas exchange.

Pre-cultures of HB-95 cells were maintained in Gibco RPMI 1640 [+] L-Glutamine supplemented with 10 % FBS and 1 % penicillin-streptomycin and kept in standard TC flasks. Growth was monitored daily by visual inspection and cell numbers and viability was determined by trypan blue staining and counting of the cells using a TC20 automated cell counter (Bio-Rad Laboratories; Hercules, CA, USA). When pre-cultures contained at least $2x10^7$-$4x10^7$ viable cells, ideally in the log growth phase to prevent an initial lag phase, they were transferred to the bioreactors. For this, the membrane was first equilibrated by the addition of 50 ml of culture medium to the medium compartment. Next, $2x10^7$-$4x10^7$ viable cells of the pre-culture were diluted in 16 ml of pre-warmed culture medium and added to the cell cultivation compartment. Finally, an additional 950 ml of culture medium were added to the medium compartment and the bioreactor was incubated at 37 °C and 5 % CO_2. The growth of hybridoma cells was monitored every third day by determining cell numbers and viability. After the cell number reached a maximum of $4x10^8$ cells ($2.6x10^7$ cells/ml), the cells were harvested. To this end, the culture medium was removed from the medium compartment and cell suspension was removed from the cell cultivation compartment. $4x10^7$ viable cells in 15 ml fresh culture medium were put back to the cell cultivation compartment and 1 l of fresh culture medium was placed in the medium compartment. Typically, harvesting was performed every 72 h. The remaining cell suspension was cleared by centrifugation with 1,660 x g for 15 min, filtered through a 0.45 µm syringe filter, and stored at -20 °C until usage for antibody purification (described in section 2.3.1).

2.1.4 Plasmid transfection

To assess the NMD efficiency of the used cell lines, dual-luciferase NMD reporter plasmids (Boelz *et al.*, 2006) were transiently transfected using JetPRIME transfection reagent (Polyplus transfection; Illkirch-Graffenstaden, France). Briefly, $1x10^5$ cells/well in 2 ml medium supplemented with 10 % FBS and 1 % penicillin-streptomycin were seeded in 6-well plates and incubated for 24 h at 37 °C and 5 % CO_2. The next day, the culture medium was changed to fresh medium for transfection. The transfection mix for one well of a 6-well plate was prepared by combining 0.5 µg of either *Renilla*-HBB (wild-type) or *Renilla*-HBB (NS39) plasmid, 0.175 µg firefly plasmid, and 0.1 µg of a plasmid coding for eYFP in 200 µl 1X JetPRIME buffer. After vortexing for 10 s, 1.55 µl JetPRIME reagent was added, vortexed for 10 s again, and incubated for 10-15 min at RT. Then 200 µl of the transfection mix per well were added dropwise and cells were incubated for 24 h at 37 °C and 5 % CO_2. After 24 h, transfection efficiency was controlled visually by the eYFP expression. Luciferase assay was performed as described in section 2.2.3.

2.1.5 siRNA transfection

siRNA-mediated knockdown of the key NMD factor UPF1 was performed to validate identified frameshift-bearing transcripts as *bona fide* NMD targets. The siRNA targeting UPF1 (Gehring *et al.*, 2003), as well as the AllStars Negative Control siRNA, were purchased from Qiagen (Hilden, Germany). Lipofectamine RNAiMAX (Invitrogen; Carlsbad, CA, USA) or INTERFERin (Polyplus; Illkirch-Graffenstaden, France) were used as transfection reagents for siRNA delivery. Briefly, $1x10^5$ cells/well in 2 ml medium supplemented with 10 % FBS and 1 % penicillin-streptomycin were seeded in 6-well plates and incubated for 24 h at 37 °C and 5 % CO_2. Before transfection, the medium was replaced with fresh medium. For transfection with Lipofectamine, the transfection mix per well was prepared by first diluting the siRNA in 150 µl Gibco OPTI-MEM (Thermo Fisher Scientific; Waltham, MA, USA). For a final concentration of 10 nM, 1.5 µl of 20 µM siRNA stock solution were used, while 1.5 µl of a 2 µM stock solution were used for a final concentration of 1 nM. 9 µl of RNAiMAX reagent were diluted in 150 µl OPTI-MEM. 150 µl of the transfection reagent solution were then mixed with 150 µl of the siRNA solution and incubated for 5 min at RT. 250 µl per well of this mix were added dropwise to the cells. For transfection with INTERFERin, the siRNA was diluted in 200 µl OPTI-MEM and 4 µl of INTERFERin transfection reagent was added. For a final concentration of 10 nM, 1.1 µl of 20 µM siRNA stock solution were used, while 1.1 µl of a 2 µM stock solution were used for a final concentration of 1 nM. The transfection mix was vortexed for 10 s, incubated for 10 min at RT, and added dropwise to the cells. After incubation for 48 h at 37 °C and 5 % CO_2, cells were lysed using RIPA buffer (50 mM Tris (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 % NP-40, 0.5 % sodium deoxycholate, 0.1 % SDS; 1 cOmplete Protease Inhibitor Cocktail tablet (Roche; Basel, Schwitzerland) per 10 ml of lysis buffer) and knockdown was assessed by quantitative real-time PCR and western blot analysis.

2.1.6 Preparation of drugs used for *in vitro* treatment

In this study, treatments with 5AZA, 5-aza-2'-deoxy-cytidin (dAZA), and cycloheximide (CHX) were performed. 5AZA and dAZA were purchased from Sigma-Aldrich (St. Louis, MO, USA), CHX was

purchased from Merck (Darmstadt, Germany). 5AZA and dAZA stock solutions with a concentration of 100 µM were prepared in CryoMACS DMSO (Miltenyi Biotech; Bergisch Gladbach, Germany), CHX stock solutions (5 µg/µl) were prepared in Aqua *ad iniectabilia* (Braun; Melsungen, Germany). Given their low stability, further dilutions of 5AZA and dAZA in Aqua *ad iniectabilia* were prepared freshly before each experimental usage. Negative controls were treated with DMSO corresponding to the amount of DMSO in the highest drug concentration used.

2.1.7 Cryopreservation and resuscitation of frozen cells

Low passage cells were used to prepare cryopreserved stocks of cells used in this study. For the preparation of cryopreserved stocks, suspension cells or trypsinized adherent cells, respectively, were pelleted for 5 min at 500 x g. After removal of the medium, the pelleted cells were resuspended in FBS supplemented with 10 % CryoMACS DMSO to a final concentration of $1x10^7$-$3x10^7$ cells/ml. Aliquots of 1 ml of the cell suspension were transferred to CryoTubes (Nunc; Roskilde, Denmark). Stocks were slowly frozen using a Mr. Frosty container (Thermo Fisher Scientific; Waltham, MA, USA) at -80 °C for at least 24 h before being transferred to a liquid nitrogen storage tank.

HCT-116, RKO, and LoVo cells were typically used for 25-30 passages before being replaced with freshly resuscitated cells. HB-95 cells were used for up to 3 months before freshly resuscitated cells were used to inoculate a new small-scale bioreactor. To resuscitate frozen cells, they were thawed at 37 °C in a water bath until only a small portion was frozen and then transferred to a TC flask containing pre-warmed culture medium supplemented with 10 % FBS and 1 % penicillin-streptomycin. Both adherent and suspension cells were allowed to recover for 8-12 h before the culture medium was replaced. Adherent cells were split at least two times before being used in experiments.

2.2 Biochemical and molecular biology methods

2.2.1 Bacterial transformation and plasmid purification

Plasmid DNAs used in the luciferase assays were amplified in *Eschericha coli* XL1-Blue competent cells (Stratagene; San Diego, CA, USA). Briefly, 200 ng plasmid DNA were added to 50 µl of competent cells and incubated on ice for 30 min. After a heat-shock for 2 min at 42 °C, cells were again incubated for 5 min on ice. Next, 450 µl LB medium (10 g/l NaCl, 10 g/l tryptone, 5 g/l yeast extract; in ddH_2O, pH 7.0) were added and the cell suspension was incubated for 1 h at 37 °C with constant shaking. After 1 h, cells were pelleted, resuspended in 40 µl LB medium, and spread on an LB agar plate (LB medium supplemented with 15 g/l agar) with 100 mg/l ampicillin. Plates were incubated overnight at 37 °C with 5 % CO_2. Clones picked from the plates were used to inoculate 100 ml LB medium supplemented with 0.1 mg/ml ampicillin. Bacterial cultures were incubated overnight at 37 °C and 5 % CO_2. Plasmids were purified from the overnight cultures using the NucleoBond Xtra Midi kit according to the manufacturer's protocol for high-copy plasmids. Purified plasmid DNA was reconstituted in 1:10 diluted TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) and DNA concentration was determined by measuring the absorbance at 260 nm using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific; Waltham, MA, USA).

2.2.2 Determination of protein concentrations

Protein concentrations were determined using either colorimetric or spectrophotometric methods. The protein concentrations of mixtures, *e.g.* cell lysates, were determined using colorimetric assays such as the Bradford or DC assay. The Bradford assay (Bradford, 1976) is based on the observation that the dye Coomassie Brilliant Blue G-250 forms complexes with proteins in acidic solution. While the unbound dye has an absorption maximum at 470 nm, the absorption maximum of the complexed dye is at 595 nm. The increase in absorption at 595 nm is linear for protein concentrations up to 2 mg/ml and can be monitored to determine the protein concentration. The DC assay (Lowry *et al.*, 1951; Peterson, 1979) involves a two-step reaction in which first peptide bonds react with copper atoms. In the so-called "biuret reaction" an electron is released which reduces the added Folin reagent in a second reaction to produce an intense blue color. The absorbance of the color behaves linearly to the protein concentration up to 1.5 mg/ml and can be measured at 750 nm. Concentrations of single proteins, *e.g.* purified antibodies, were determined spectrophotometrically at 280 nm (Desjardins *et al.*, 2009).

2.2.2.1 Bradford protein assay

The Bradford assay was performed using the Bio-Rad Protein Assay reagents (Hercules, CA, USA). Briefly, the Protein Assay Dye Reagent concentrate was diluted 1:5 with ddH_2O and 1 ml of this solution was mixed with 2 µl of the sample with unknown concentration. A sample containing only the lysis buffer of the sample was used as blank. Absorbance at 595 nm was measured in cuvettes using a BioPhotometer Plus (Eppendorf; Hamburg, Germany) to determine a relative protein concentration.

2.2.2.2 DC protein assay

The DC assay was performed using the Bio-Rad DC Protein Assay reagents (Hercules, CA, USA). Purified BSA (New England Biolabs; Ipswich, MA, USA) was used to generate a standard curve for absolute protein quantification. BSA standards with concentrations from 0-1.5 mg/ml were prepared in the sample lysis buffer and diluted 1:4 with ddH_2O. Samples with unknown concentrations were also diluted 1:4 in ddH_2O. Reagent A' was prepared by mixing 1 ml of Reagent A with 25 µl of Reagent S. 5 µl of both standards and samples were added in triplicates to a clear, flat-bottomed 96-well polystyrene microplate (Corning; New York, NY, USA). Next, 25 µl of Reagent A' and 200 µl of Reagent B were added per well and incubated for 15-60 min in the dark. Absorbance at 750 nm was measured using a SpectraMax M2 microplate reader (Molecular Devices; San Jose, CA, USA) and absolute protein concentrations were determined using the generated standard curve.

2.2.2.3 Determination of IgG concentration

Given its defined extinction coefficient, the concentration of purified IgG antibodies was determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific; Waltham, MA, USA). 300 µl of 1 M Tris-HCl (pH 8.0) diluted in 1 ml of 0.1 N acetic acid (pH 3.0) served as blank.

2.2.3 Luciferase assay

The luciferase assay was utilized to determine the NMD efficiency of MSI CRC cell lines using a previously published dual-luciferase reporter system (Boelz *et al.*, 2006). The system consists of plasmids containing the human hemoglobin subunit beta (*HBB*) gene without (wild-type) or with an NMD-triggering nonsense mutation at position 39 (NS39) fused in-frame to the 3' end of the *Renilla* luciferase gene. In the luciferase assay, the signal of a co-expressed firefly luciferase is used to normalize the *Renilla* luciferase signal and the NMD efficiency is calculated by comparing the *Renilla* luciferase signals obtained from wild-type reporter constructs with those obtained from the NS39.

For the luciferase assay, $1x10^5$ cells in 2 ml medium per well were seeded in 6-well plates. After 24 h, cells were transfected using JetPrime reagent (Polyplus Transfection; Illkirch-Graffenstaden, France) and incubated at 37 °C and 5 % CO_2. After 24 h, cells were washed two times with ice-cold PBS. Next, cells were lysed for 10 min using 1X Passive Lysis Buffer (Promega; Madison, WI, USA). Lysates were transferred to Eppendorf tubes and cleared by centrifugation for 10 min at 4 °C and 21,130 x g. Supernatants were transferred to a new tube and luciferase activity was measured using the Dual-Luciferase Reporter Assay System kit (Promega; Madison, WI, USA) and a Centro LB 960 microplate luminometer (Berthold Technologies; Bad Wildbad, Germany). For this, 5 µl of either diluted (1:5-1:25) or undiluted supernatant were loaded to white, flat-bottomed microtiter plates (Berthold Technologies; Bad Wildbad, Germany) in triplicates. First, 30 µl of firefly luciferase substrate per well were injected and after 4 s luminescence with a wavelength of 562 nm was measured. Next, 30 µl of the Stop & Glo Buffer per well were injected, which quenches the firefly luminescence and contains the Renilla luciferase substrate. Luminescence with a wavelength of 480 nm was measured after 4 s. Signals obtained from *Renilla* luciferase were first normalized by signals obtained from firefly luciferase. Normalized signals

obtained from wild-type-transfected cells were set as 100 % and signals obtained from NS39-transfected cells were set in relation to this to determine NMD efficiency.

2.2.4 Proliferation assay

Proliferation assays were performed to evaluate the effect of 5AZA treatment on cell growth of MSI CRC cell lines. $3x10^4$ cells per well in 100 µl medium supplemented with 10 % FBS and 1 % penicillin/streptomycin were seeded in clear, flat-bottomed 96-well polystyrene microplates (Corning; New York, NY, USA). Eight wells contained only the supplemented medium and served as control. After incubation for 24 h at 37 °C and 5 % CO_2, culture medium was replaced with supplemented medium containing either 0.001-5 µM 5AZA or DMSO (negative control). Cells were incubated again at 37 °C and 5 % CO_2 for either 24, 48, or 72 h before proliferation was assessed by crystal violet staining. For crystal violet staining, culture medium was removed and cells were washed once with 200 µl PBS per well. Then, 50 µl crystal violet staining solution (5 g/l crystal violet, 20 % MeOH in ddH_2O) were added to each well and incubated for 15 min at RT on a plate shaker. After 15 min, the staining solution was removed and wells were washed three times with 200 µl ddH_2O. After the last wash step, ddH_2O was removed as completely as possible and plates were air-dried overnight. Crystal violet was resolved by adding 200 µl methanol per well and absorbance at 555 nm was measured using a SpectraMax M2 microplate reader. Relative proliferation was calculated by subtracting the mean absorbance of control wells from absorbance values of wells with cells. The proliferation of untreated controls was defined as 100 % and 5AZA-treated samples were set in relation.

2.2.5 SDS-PAGE analysis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was utilized to separate proteins based on their molecular weight. Proteins in solution were denatured by addition of SDS-containing loading buffer and incubation at 95 °C for 5 min. Additionally, treatment with SDS provides a uniformly negative charge to the denatured linearized proteins which is approximately proportional to the proteins' molecular weight. In SDS-PAGE, proteins migrate through the polyacrylamide matrix when a current is applied depending on their size: Smaller proteins will travel faster, while larger proteins will be retained by the matrix and migrate slower. SDS-PAGE was performed either using in-house casted gels (10 % acrylamide; see Table 1) or pre-casted NuPAGE 4-12 % Bis-Tris protein gels (Invitrogen; Carlsbad, CA, USA). PageRuler Prestained Protein Ladder (10-180 kDa) or PageRuler Plus Prestained Protein Ladder (10-250 kDa; both from Thermo Fisher Scientific; Waltham, MA, USA) were used as size standards. In-house casted gels were run at 35 mA per gel for 60 min using a Biometra Multigel chamber (Analytik Jena; Jena, Germany) and 1X Laemmli running buffer (4X: 768 mM glycine, 100 mM Tris-HCl, 0.4 % SDS). Pre-casted gels were run at 200 V for 40-55 min using an XCell SureLock Electrophoresis System (Thermo Fisher Scientific; Waltham, MA, USA) and the NuPAGE MOPS SDS running buffer (Thermo Fisher Scientific; Waltham, MA, USA). 6X SDS loading dye (7 ml 4X Tris-HCl/SDS (0.5 M Tris-HCl, 0.4 % SDS, pH 6.8), 3.6 ml glycerol, 1 g DTT, 1 mg bromophenol blue) was added to diluted protein samples before loading them to the gels. Proteins separated by SDS-PAGE were either visualized by silver staining or transferred to a membrane for further analysis by immunoblotting.

Table 1: Composition of stacking gel and resolving gel (10 % acrylamide) used for SDS-PAGE.

stacking gel	
30 % acrylamide/bisacrylamide (37.5:1)	0.75 ml
4X Tris-HCl/SDS (pH 6.8)	0.5 ml
ddH_2O	1.72 ml
10 % APS	30 µl
TEMED	3 µl

resolving gel	
30 % acrylamide/bisacrylamide (37.5:1)	2.3 ml
4X Tris-HCl/SDS (pH 8.8)	1.75 ml
ddH_2O	2.9 ml
10 % APS	70 µl
TEMED	4 µl

2.2.6 Silver staining of polyacrylamide gels

The silver staining procedure is based on the selective reduction of silver ions to metallic silver on protein bands (Kumar, 2018). Here, silver staining was utilized to analyze antibody purifications and antibody crosslinking procedures using the SilverQuest Silver Staining Kit (Invitrogen; Carlsbad, CA, USA). Briefly, the SDS-PAGE gel was rinsed with ddH_2O and then placed in 50 ml fixative (40 % EtOH, 10 % acetic acid in ddH_2O) for at least 20 min. Next, the gel was washed with 50 ml of 30 % EtOH in ddH_2O for 10 min and then incubated with 50 ml of sensitizing solution (30 % EtOH, 10 % Sensitizer in ddH_2O) for 10 min. The gel was washed again with 50 ml of 30 % EtOH in ddH_2O for 10 min and with 50 ml of ddH_2O for 10 min. Next, the gel was incubated for 15 min with 50 ml of staining solution (1 % Stainer in ddH_2O) and then rinsed with ddH_2O. For development, the gel was incubated with 50 ml of developing solution (10 % Developer, 1 drop developer enhancer in ddH_2O) until the desired band intensity was reached. Development was stopped by the addition of 5 ml of Stopper and further incubation for 10 min. Finally, the gel was washed for at least 10 min with ddH_2O and then digitalized using a Fusion-FX7 imaging platform (Vilber Lourmat; Eberhardzell, Germany). All incubation steps were carried on a shaker to guarantee uniformity.

2.2.7 Western blot analysis

Detection of specific proteins by immunoblotting was carried out to confirm the efficiency of immunoprecipitation and siRNA knockdown experiments. Proteins were transferred from SDS-PAGE gels to 0.2 µm Hybond-P polyvinylidene difluoride membranes (PVDF; GE Healthcare; Amersham, UK) under semi-dry conditions. To this end, PVDF membranes were activated by incubation for at least 1 min in methanol and placed on top of three Whatman blotting papers (GE Healthcare; Amersham, UK) soaked in anode buffer (25 mM Tris, 20 % MeOH). The gel was rinsed once with cathode buffer (25 mM Tris, 40 mM 6-aminohexanoic acid, 0.01 % SDS, 20 % MeOH) and placed on the PVDF membrane. On top of the gel, three Whatman blotting papers soaked in cathode buffer were placed. Proteins were transferred from the gel to the PVDF membrane at a constant current of 1-1.5 mA per cm^2 of membrane for 90 min using a Biometra Fastblot blotting chamber (Analytik Jena; Jena, Germany). Membranes were blocked for 1 h at RT with 5 % milk in Tris-buffered saline supplemented with Tween 20 (TBS-T; 20 mM Tris-HCl (pH 7.5), 125 mM NaCl, 0.125 % Tween 20) to prevent subsequent unspecific binding of antibodies. Specific proteins were probed by incubation of primary antibodies diluted in 5 % milk powder (w/v) in TBS-T (see Table 2 for antibody-specific dilutions and incubation times) under constant shaking. Membranes were rinsed three times with TBS-T and then washed three times for 10 min with TBS-T to remove unspecifically bound and excess antibody before incubation with a horseradish

peroxidase (HRP)-coupled secondary antibody for 60 min at RT. Again, membranes were rinsed three times with TBS-T and then washed three times for 10 min with TBS-T. For evaluation of co-eluting W6/32 antibody during immunoprecipitation experiments, corresponding membranes were incubated with an anti-mouse HRP-coupled secondary antibody before incubation with primary antibodies to specific proteins. Membranes were developed using the Western Lightning Plus-ECL Enhanced Luminol Reagent and Oxidizing Reagent (Perkin Elmer; Waltham, MA, USA) and a Fusion-FX7 imaging platform.

Table 2: Antibodies used for immunoblotting in this thesis. All antibodies were diluted in 5 % milk in TBS-T.

target	source	working dilution	incubation time	supplier
α-tubulin	mouse	1:10,000	1 h	Sigma-Aldrich (St. Louis, MO, USA)
B2M	rabbit	1:1,000	overnight	Abcam (Cambridge, UK)
pan-HLA	mouse	1:1,000	overnight	in-house produced
UPF1	goat	1:1,000	overnight	Bethyl Laboratories (Montgomery, TX, USA)
anti-goat	rabbit	1:10,000	1 h	Sigma-Aldrich (St. Louis, MO, USA)
anti-mouse	rabbit	1:10,000	1 h	Sigma-Aldrich (St. Louis, MO, USA)
anti-rabbit	goat	1:10,000	1 h	Sigma-Aldrich (St. Louis, MO, USA)

2.2.8 Mutation analysis

Underlying frameshift mutations of InDel neoepitopes and single nucleotide polymorphisms of SNP neoepitopes were first validated by Sanger sequencing in the corresponding MSI CRC cell line. The frequency of validated mutations was further assessed in MSI CRC cell line and patient samples using gel capillary electrophoresis.

2.2.8.1 Cell line samples

All human CRC cell lines tested have been previously described (Woerner *et al.*, 2001; Woerner *et al.*, 2007; Michalak *et al.*, 2020). Genomic DNA was isolated and provided by the Department of Applied Tumor Biology at the Heidelberg University Hospital's Institute of Pathology (Heidelberg, Germany).

2.2.8.2 Human tissue samples

Human tissue samples were obtained from the local tissue bank established within the German Collaborative Group on HNPCC. Informed consent was obtained from all patients and the study protocol was approved by the local Ethics Committee (S-582/2016). Genomic tumor DNA was isolated in the Department of Applied Tumor Biology at the Heidelberg University Hospital's Institute of Pathology (Heidelberg, Germany) and MSI status was determined based on the National Cancer Institute/ICGHNPCC reference marker panel (Boland *et al.*, 1998) and CAT35 as an additional mononucleotide marker (Findeisen *et al.*, 2005). MSI is defined by instability in at least 30 % of tested markers.

2.2.8.3 Polymerase chain reaction

Polymerase chain reaction (PCR) was utilized for the amplification of DNA fragments for mutation analysis. Target-specific primers (Table 17 - Table 19) were designed using Primer-BLAST (Ye *et al.*, 2012) and ordered from Sigma-Aldrich (Steinheim, Germany). Standard primers were ordered desalted, while FAM-labeled primers used for gel capillary electrophoresis were purchased HPLC-purified. Lyophilized primers were reconstituted in Aqua *ad iniectabilia* to a final concentration of 100 mM and further diluted to 10 mM for PCRs. DFS-Taq polymerase and accompanying buffers were purchased from BIORON Diagnostics (Römerberg, Germany), dNTPs were purchased from Carl Roth (Karlsruhe, Germany). PCR mixtures were prepared according to Table 3. For hard-to-amplify fragments, DMSO and betaine were added to the reaction mixture. Standard PCRs in tubes were performed using a Biometra T3000 thermocycler (Analytik Jena; Jena, Germany), small-scale PCRs in 96-well plates (Steinbrenner Laborsysteme; Wiesenbach, Germany) were performed using a PTC-100 thermal cycler (MJ Research; Waltham, MA, USA) according to the PCR protocol in Table 4. Elongation times were set to at least 30 s and adjusted for the size of the amplified fragment with roughly 1 min elongation time per kilobase. Fragments were analyzed by agarose gel electrophoresis. Depending on the fragment size, agarose concentration was between 1 and 3 %. Agarose was dissolved in 1X TBE buffer (8X: 0.88 M Tris, 0.8 M boric acid, 20 mM EDTA) and mixed with 5 µl ethidium bromide (Carl Roth; Karlsruhe, Germany) per 50 ml of agarose solution. GeneRuler 50 bp DNA Ladder or GeneRuler 100 bp DNA Ladder (both from Thermo Fisher Scientific; Waltham, MA, USA) were used as size standards. DNA fragments were separated using a PerfectBlue gel system (PEQLAB Biotechnologie; Erlangen, Germany) at 140-170 V. PCR products were analyzed visually using a Herolab gel documentation system (Wiesloch, Germany) with UV light.

Table 3: Composition of PCR reactions performed in standard and small scale format. Standard format was used for further analysis by Sanger sequencing, small scale format was used for further analysis by gel capillary electrophoresis.

reagent	standard PCR (1 sample)	small-scale PCR (1 sample)
10X DFS Taq buffer	5 µl	1 µl
dNTPs [10 mM]	1 µl	0.2 µl
forward primer [10 mM]	3 µl	0.6 µl
reverse primers [10 mM]	3 µl	0.6 µl
DFS Taq polymerase [5 U/µl]	0.25 µl	0.05 µl
DNA template	1 µl (~50-200 ng/µl)	2 µl (~5 ng/µl)
DMSO (optional)	2.5 µl	5 µl
betaine [5 M] (optional)	3.75 µl	0.75 µl
Aqua *ad iniectabilia*	*ad* 50 µl	*ad* 10 µl

Table 4: PCR program used for the amplification from genomic DNA.

step	temperature	time	
initial denaturation	95 °C	5 min	
denaturation	95 °C	30 s	35-40 X
hybridization	60 °C	30 s	
elongation	72 °C	30-90 s	
final elongation	72 °C	10 min	

2.2.8.4 Sample preparation for further analysis

PCR products that were further analyzed by Sanger sequencing were purified from agarose gels and/or PCR reactions after visual inspection using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel; Düren, Germany). For agarose gel extraction, the corresponding bands were excised using a clean scalpel and the gel slice was dissolved in 200 µl NTI buffer per 100 mg agarose gel at 50 °C for 5-10 min. For DNA extraction from PCR mixtures, 2 volumes of NTI buffer were added per volume of PCR mixtures. Next, the sample was loaded onto a NucleoSpin column and centrifuged for 30 s at 11,000 x g. The column was washed twice with 700 µl NT3 buffer and centrifugation for 30 s at 11,000 x g and the silica membrane was dried by centrifugation at 11,000 x g for 1 min. Finally, DNA was eluted by 25 µl NE buffer diluted 1:10 in ddH_2O to the column, incubation for 1 min at RT, and centrifugation at 11,000 x g for 1 min. DNA concentration was measured using a NanoDrop 2000 and adjusted to 50-150 ng/µl if necessary.

2.2.8.5 Sanger sequencing

Sanger sequencing was performed either at Eurofins Genomics (Köln, Germany) or Microsynth Seqlab (Göttingen, Germany). Chromatogram files were manually analyzed using the Indigo webtool (Rausch *et al.*, 2020)[2].

2.2.8.6 Gel capillary electrophoresis

Mutation analysis of MSI CRC cell lines and patient samples was performed by gel capillary electrophoresis. Gel capillary electrophoresis allows the separation of DNA molecules at a resolution of one nucleotide and thus is suited for the detection of small insertion and deletion mutations caused by microsatellite instability (Butler *et al.*, 2004). DNA fragments for analysis by gel capillary electrophoresis were amplified using fluorescently labeled primers designed using Primer-BLAST (Table 19). PCR products were diluted in ddH_2O based on their band intensity and mixed 1:13 with GeneScan 500 ROX dye size standard/Hi-Di formamid mix (1:81; both from Applied Biosystems; Foster City, CA, USA). Fragments were visualized on an ABI3130xl (Applied Biosystem; Foster City, CA, USA) as described previously (Findeisen *et al.*, 2005). Data were analyzed using the GeneMapper software (version 5; Applied Biosystems; Foster City, CA, USA) and the regression-based frameshift quantification algorithm (ReFrame; Ballhausen *et al.*, 2020).

[2] https://www.gear-genomics.com/indigo/

2.2.9 Real-time quantitative PCR

Relative quantification of mRNA expression was performed using real-time quantitative PCR. Here, RNA is first isolated from cell lysates and then reverse-transcribed into complementary DNA (cDNA) using oligo(dT) primers which hybridize to the poly(A) tail of mRNAs. In a second reaction, the previously synthesized cDNA serves as template for a PCR with target-specific primers. During the PCR, the dye SYBR Green 1 intercalates into the synthesized double-stranded DNA and thereupon emits a strong fluorescent signal which is used for the quantification of amplified DNA fragments after each elongation step (Arya *et al.*, 2005).

2.2.9.1 RNA isolation

Total RNA isolation was performed using TRI Reagent (Sigma-Aldrich; St. Louis, MO, USA) and is based on the single-step method by acid guanidinium thiocyanate-phenol-chloroform extraction (Chomczynski and Sacchi, 2006). Briefly, biological samples are lysed in TRI reagent which contains phenol and guanidinium thiocyanate. The latter efficiently denatures proteins, including RNases, in the sample thus protecting mRNAs from degradation. After the addition of chloroform and subsequent centrifugation, an aqueous phase, interphase, and an organic phase form. Under acidic conditions, total RNA remains in the aqueous phase while proteins and DNA are located in the interphase and organic phase. Total RNA can be precipitated from the aqueous phase using isopropanol.

For RNA isolation from cleared cell lysates, 750 µl TRI Reagent were added per 250 µl of lysate and incubated for 5 min at RT. Next, 200 µl chloroform were added, samples were vortexed for 30 s and incubated for 15 min at RT. Phase separation was performed by centrifugation at 12,000 x g and 4 °C for 15 min and the upper aqueous phase was transferred to a new tube. After the addition of 200 µl isopropanol and 1 µl glycogen (20 mg/ml; Roche Diagnostics; Mannheim, Germany), samples were mixed by vortexing and incubated for 10 min at RT. Precipitated RNA was pelleted by centrifugation at 12,000 x g and 4 °C for 15 min. The supernatant was discarded and the RNA pellet was washed once with 1 ml ice-cold 70 % EtOH. After centrifugation for 5 min at 12,000 x g and 4 °C, the supernatant was removed as completely as possible, RNA was dissolved in 11 µl Aqua *ad iniectabilia* and incubated at 50 °C for 5 min. RNA concentration was measured using a NanoDrop 2000 and RNA samples were stored at -20 °C (short term) or -80 °C (long term).

2.2.9.2 cDNA synthesis

cDNA synthesis was performed using the RevertAid H Minus Reverse Transcriptase (Thermo Fisher Scientific; Waltham, MA, USA). Briefly, 0.5-5 µg isolated RNA were diluted in 9.5 µl Aqua *ad iniectabilia* and 1 µl oligo(dT)$_{18}$ primers (biomers.net; Ulm, Germany). Samples were incubated at 65 °C for 5 min and then cooled on ice to allow hybridization of primers. Next, 9.5 µl of RT mix (Table 5) containing revere transcriptase and RiboLock RNase inhibitor (Thermo Fisher Scientific; Waltham, MA, USA) were added, and samples were incubated for 60 min at 42 °C. After 1 h, reverse transcriptase was inactivated by incubation at 70 °C for 10 min. cDNA samples were diluted 1:10 with Aqua *ad iniectabilia* and stored at -20 °C until usage in qPCR.

Table 5: Composition of reverse transcriptase reaction mix used for cDNA synthesis.

reagent	RT mix (1 sample)
5X RevertAid H Minus Reverse Transcriptase buffer	4 µl
dNTPs [10 mM]	2 µl
DTT [0.1 M]	2 µl
RiboLock RNase inhibitor [40 U/µl]	0.5 µl
RevertAid H Minus Reverse Transcriptase [200 U/µl]	1 µl
Aqua *ad iniectabilia*	*ad* 9.5 µl

2.2.9.3 qPCR

qPCRs were performed in technical triplicates on a StepOnePlus system (Applied Biosystems; Foster City, CA, USA) using primaQuant CYBR qPCR Master Mix (Steinbrenner Laborsysteme; Wiesenbach, Germany) according to Table 6 and Table 7. Target-specific primers for *ATF3*, *ATF4*, *SC35A*, *SC35C*, *SC35D*, *UPP1*, and the housekeeping gene *HPRT1* have been reported elsewhere (Gehring *et al.*, 2003; Viegas *et al.*, 2007; Bhuvanagiri *et al.*, 2014; Cheruiyot *et al.*, 2018). qPCR primers for transcripts of neoepitopes have been designed using Primer-BLAST and are reported in Table 20. Melting curves were inspected manually to verify the specificity of the PCR reaction. Thresholds for Ct values were set manually to the start of the exponential phase of target amplification in the StepOne Software (version 2.3; Applied Biosystems; Foster City, CA, USA). Ct values were exported and further analyzed using Excel (Microsoft; Redmond, WA, USA) using the "delta-delta Ct" method (Livak and Schmittgen, 2001). Briefly, Ct values of target genes were normalized with the corresponding Ct value of the housekeeping gene *HPRT1* (ΔCt). Normalized Ct values were compared between treated and untreated conditions (ΔΔCt) and fold changes in gene expression were calculated as $2^{-\Delta\Delta Ct}$.

Table 6: Composition of qPCR reaction mix.

reagent	RT mix (1 well)
Aqua *ad iniectabilia*	6 µl
primaQuant CYBR qPCR Master Mix	10 µl
forward primer [10 µM]	1 µl
reverse primer [10 µM]	1 µl
cDNA (1:10 diluted)	2 µl

Table 7: PCR program used for analysis by qPCR. Fluorescence measurements were taken at the end of each elongation step and after each temperature increment during melting curve analysis.

step	temperature	time	
initial denaturation	95 °C	10 min 10 s	
denaturation	95 °C	3 s	40 X
hybridization	60 °C	10 s	
elongation	72 °C	20 s	
melting curve	95 °C	10 s	
	60 °C	60 s	
	60 °C → 95 °C (0.3 °C increment)		
	95 °C	15 s	

2.3 Immunopeptidomics sample preparation

Samples for the mass spectrometry-based analysis of HLA class I-presented peptides were generated by immunoprecipitation of HLA class I:peptide complexes using the pan-HLA class I-reactive monoclonal antibody W6/32 and subsequent separation of HLA class I-presented peptides using solid-phase extraction.

2.3.1 Antibody purification

HB-95 hybridoma cells were utilized for the production of the pan-HLA class I-reactive monoclonal antibody W6/32. The antibody was purified from cell culture medium using Sepharose-Protein A beads as described previously (Bassani-Sternberg, 2018). Briefly, empty gravity-flow columns (9 ml; Bio-Rad; Hercules, CA, USA) were washed once with 9 ml of a 1 % SDS solution and 40 ml of ddH_2O before the first usage. Next, 4 ml of a 50 % Sepharose-Protein A conjugate 4B beads solution (Invitrogen; Carlsbad, CA, USA) were filled into the column, preservation buffer was allowed to drain and beads were washed with 9 ml of 100 mM Tris-HCl (pH 8.0). After the washing buffer had drained off, the column tip was closed and up to 8 ml of cleared cell culture supernatant was added to the column. Next, the column top was closed and the cell culture supernatant and beads were incubated at RT for 10 min with gentle rotation. After 10 min, the cell culture supernatant was allowed to drain and columns were first washed with 9 ml of 100 mM Tris-HCl (pH 8.0) and then with 9 ml of 20 mM Tris-HCl (pH 8.0). The W6/32 antibodies were eluted with 6X 1 ml of 0.1 N acetic acid (pH 3.0) into tubes containing 300 µl of 1 M Tris-HCl (pH 8.0) to neutralize the solution. Antibody elutions were mixed by gentle vortexing, pH was verified using pH test strips and the IgG concentration was determined using a NanoDrop 2000. Fractions with high antibody concentrations were pooled and purified antibodies were stored at -20 °C until further usage. Gravity flow columns and beads were washed with 9 ml of 100 mM Tris-HCl (pH 8.0) and 2 ml of PBS supplemented with 0.02 % NaN_3. Finally, the column tip was closed, beads were covered with 2 ml of PBS supplemented with 0.02 % NaN_3, and stored at 4 °C until further usage.

2.3.2 Antibody crosslinking

For the immunoprecipitation of HLA class I:peptide complexes, W6/32 antibodies were covalently coupled to Sepharose-Protein A beads using the chemical crosslinker dimethyl pimelimidate (DMP) in order to reduce co-elution and potentially clogging of columns during the immunoprecipitation procedure. Crosslinking was performed in gravity-flow columns as described previously (Bassani-Sternberg, 2018). Briefly, empty gravity-flow columns (9 ml) were washed once with 9 ml of a 1 % SDS solution and 40 ml of ddH_2O before the first usage. Next, 4 ml of a 50 % Sepharose-Protein A conjugate 4B beads solution were filled into the column, preservation buffer was allowed to drain and beads were washed with 9 ml of 100 mM Tris-HCl (pH 8.0). After the washing buffer had drained off, the column tip was closed and 10 mg of W6/32 antibody in 9 ml of 100 mM Tris-HCl (pH 8.0) were added to the beads. Next, the column top was closed and the beads and the antibody solution were incubated at RT for 30 min with gentle rotation. After 30 min, the liquid was allowed to drain, beads were washed with 9 ml of 0.2 M sodium borate buffer (pH 9.0) and the liquid was allowed to drain again. A 40 mM DMP solution was freshly prepared for crosslinking and 2 ml of this solution were added to the beads to achieve a final

DMP concentration of 20 mM. The column top was closed and the beads and the DMP solution were incubated at RT for 30 min with gentle rotation for covalent crosslinking. After 30 min, the liquid was allowed to drain and the crosslinking reaction was quenched by washing the beads with 5 ml of 0.2 M ethanolamine (pH 8.0). Quenching solution was allowed to drain, column tip was closed and beads were resuspended in 5 ml of 0.2 M ethanolamine (pH 8.0). The column top was closed and the crosslinking reaction was further quenched by incubation of beads in the quenching solution at RT for 2 h with gentle rotation. After 2 h, the liquid was allowed to drain and beads were washed once with 9 ml of PBS supplemented with 0.02 % NaN_3 and then stored at 4 °C in 2 ml of PBS supplemented with 0.02 % NaN_3 until usage in immunoprecipitation experiments.

2.3.3 Immunoprecipitation of HLA class I:peptide complexes

HLA class I:peptide complexes were isolated by immunoprecipitation using the pan-HLA class I-reactive monoclonal antibody W6/32. Here, cells are first lysed using the non-ionic detergent N-octyl-β-D-glucopyranoside which forms micelles that include membrane lipids and proteins thus enhancing the isolation of cell surface proteins. Next, the lysate is cleared by centrifugation, and the supernatant including the micelles is used for immunoprecipitation. For immunoprecipitation, the cleared lysate is incubated with W6/32 antibodies immobilized to a solid matrix. Unspecifically bound constituents of the lysate are removed by washing the beads with both high salt and low salt buffers. Finally, HLA class I:peptide complexes are eluted and peptides are separated from the much larger HLA class I molecules by solid-phase extraction (Kuznetsov *et al.*, 2020). The immunoprecipitation was performed either in a batch format or in a high-throughput format.

2.3.3.1 Immunoprecipitation sample preparation

For immunoprecipitation sample preparation, $7.3x10^6$ HCT-116 cells or $5.5x10^6$ RKO cells in 20 ml of complete growth medium were seeded in 150 mm dish plates and incubated at 37 °C and 5 % CO_2. After 72 h, cells were washed once with ice-cold PBS and harvested by scraping in 1 ml of ice-cold PBS. RKO cells were passed through a 40 µm cell strainer (Corning; Corning, NY, USA) to obtain single cells. Cells were counted using the TC20 automated cell counter and distributed in aliquots of $1x10^8$ cells. Cells were pelleted by centrifugation at 1,660 x g for 5 min, the supernatant was removed. The dry pellet was snap-frozen in liquid nitrogen and stored at -20 °C until usage in the immunoprecipitation.

To evaluate the effect of 5AZA on the immunopeptidome, samples were prepared as described above. Additionally, after 48 h of incubation at 37 °C and 5 % CO_2, cells were treated either with 5 µM AZA or the solvent control DMSO. 24 h after the treatment, cells were harvested as described above. Analogous to the samples prepared for immunoprecipitation, samples of $5x10^6$ cells were prepared and used for treatment validation by real-time quantitative PCR.

2.3.3.2 Batch immunoprecipitation procedure

The immunoprecipitation in batch format was carried out according to the protocol published by Bassani-Sternberg and colleagues (Bassani-Sternberg *et al.*, 2015; Bassani-Sternberg *et al.*, 2016). Briefly,

Sepharose-Protein A-W6/32 columns with 200 µl of crosslinked beads were prepared as described above in a gravity-flow column (9 ml). Cell lysis was performed directly before the immunoprecipitation by adding 1 ml of immunoprecipitation lysis buffer (0.25 % sodium deoxycholate, 0.2 M iodoacetamide, 1 mM EDTA, 1 mM PMSF, 1 % octyl-β-D-glucopyranoside in PBS; 1 cOmplete Protease Inhibitor Cocktail tablet per 50 ml of lysis buffer) to snap-frozen dry cell pellets containing $1x10^8$ cells. Once cell pellets were thawed, they were resuspended and incubated for 60 min at 4 °C. Lysates were cleared by centrifugation at 4 °C and 21,130 x g for 30 min. Cleared lysate of $1x10^8$ cells was added to the beads and incubated overnight at 4 °C with gentle rotation. The next day, beads were washed 10X with 400 µl 150 mM NaCl in 20 mM Tris-HCl (pH 8.0). Next, beads were washed 10X with 400 µl 400 mM NaCl in 20 mM Tris-HCl (pH 8.0), again 10X with 400 µl 150 mM NaCl in 20 mM Tris-HCl (pH 8.0), and finally 7X with 400 µl 20 mM Tris-HCl (pH 8.0). All washing steps were performed with ice-cold washing buffers and by gravity flow. HLA class I:peptide complexes were eluted 7X with 50 µl 0.1 N acetic acid into tubes containing 10 µl 1 M Tris-HCl (pH 7.5) to neutralize the solution. Elutions containing HLA class I:peptide complexes were either further processed to separate HLA class I-presented peptides (see section 2.3.3.3) or the volume of the elutions was reduced by vacuum centrifugation and samples were further analyzed by SDS-PAGE and immunoblotting.

2.3.3.3 Separation of HLA class I-presented peptides

Peptides and HLA class I molecules eluted during the batch immunoprecipitation procedure were further separated using OASIS HLB µElution Plate 30 µm cartridges and an Extraction Plate Manifold (both from Waters; Milford, MA, USA). Cartridges were pre-washed with 80 % LC-MS grade acetonitrile (ACN) in 0.05 % LC-MS grade formic acid (FA; both from Fisher Scientific; Schwerte, Germany) and with 0.05 % FA only. Elutions from the batch immunoprecipitation procedure were loaded to the cartridges and the cartridges were washed with 0.05 % FA. The peptides were eluted with 30 % ACN in 0.05 % FA, dried by vacuum centrifugation, and stored at -80 °C until MS analysis.

2.3.3.4 High-throughput immunoprecipitation and peptide separation procedure

The high-throughput immunoprecipitation was performed as previously described (Chong *et al.*, 2018), but omitting the pre-clear and HLA class II plates. The protocol is an updated version of the protocol used for batch immunoprecipitation and utilizes a Positive Pressure-96 Processor (Waters; Milford, MA, USA) allowing the processing of up to 96 samples in parallel. Moreover, the updated protocol reduces sample handling and overall time needed to complete the procedure thereby increasing sensitivity as well as reproducibility (Chong *et al.*, 2018). Cell lysis was performed directly before the immunoprecipitation by adding 1 ml of immunoprecipitation lysis buffer to snap-frozen dry cell pellets containing $1x10^8$ cells. Once cell pellets were thawed, they were resuspended and incubated for 60 min at 4 °C. Lysates were cleared by centrifugation at 4 °C and 21,130 x g for 30 min. Meanwhile, the Sepharose-Protein A-W6/32 plate was prepared. To this end, an empty 96-well filter microplate with 3 µm glass fiber and 10 µm polypropylene membranes (Agilent; Santa Clara, CA, USA) was washed once with 1 ml/well 100 % ACN and once with 1 ml/well 0.1 % trifluoroacetic acid (TFA). Empty plates were equilibrated with 2 ml/well 0.1 M Tris-HCl (pH 8.0). 150 µl of previously crosslinked Sepharose-

Protein A-W6/32 beads were added per well using low-retention ART filter tips (Thermo Scientific; Waltham, MA, USA) and conditioned with 1 ml 0.1 M Tris-HCl (pH 8.0) and 400 µl of immunoprecipitation lysis buffer. For the affinity purification of HLA class I:peptide complexes, 1 ml of cleared lysate was loaded per well and allowed to flow through by gravity at 4 °C. The flow-through fraction was collected for further analysis on a clean 2 ml 96-well collection plate (Agilent; Santa Clara, CA, USA). Next, plates were washed with ice-cold washing buffers using a Positive Pressure-96 Processor with 3-5 psi. First, plates were washed 8X with 1 ml/well 150 mM NaCl in 20 mM Tris-HCl (pH 8.0). The first wash was collected for further analysis on a clean 2 ml 96-well collection plate (Waters; Milford, MA USA). Next, plates were washed 8X with 1 ml/well 400 mM NaCl in 20 mM Tris-HCl (pH 8.0), again 8X with 1 ml/well 150 mM NaCl in 20 mM Tris-HCl (pH 8.0), and finally 4X with 1 ml/well 20 mM Tris-HCl (pH 8.0). Meanwhile, a Sep-Pak tC18 96-well plate with 100 mg sorbent/well (Waters; Milford, MA, USA) was conditioned with 1 ml/well 80 % ACN in 0.1 % TFA and then with 2 ml/well 0.1 % TFA. For the elution of HLA class I:peptide complexes the filter microplate with Sepharose-Protein A-W6/32 beads and bound HLA molecules was stacked on top of the Sep-Pak tC18 96-well plate. HLA class I:peptide complexes were eluted 2X with 500 µl/well 1 % TFA. Next, Sep-Pak tC18 96-well plate was washed 2X with 1 ml/well 0.1 % TFA. Peptides were eluted 2X with 400 µl/well 28 % ACN in 0.1 % TFA into a clean 2 ml 96-well collection plate with 1.5 psi. Eluted peptides were dried by vacuum centrifugation and stored at -80 °C until MS analysis. HLA class I heavy chains and B2M were eluted 2X with 300 µl/well 80 % ACN in 0.1 % TFA in a clean 2 ml 96-well collection plate. Samples were dried by vacuum centrifugation and used for SDS-PAGE and western blot analysis.

For MS parameter optimization, high-throughput immunoprecipitation was performed in twelve separate wells from a total of $1.2x10^9$ HCT-116 cells. Eluted peptides were pooled together for MS analysis, dried by vacuum centrifugation, and resuspended in 120 µl of 0.1 % LC-MS grade FA. For evaluation of NMD inhibition by 5AZA on the immunopeptidome, the high-throughput immunoprecipitation was performed in technical triplicates of $1x10^8$ cells per biological replicate and condition. Eluted peptides from technical IP replicates were pooled together for MS analysis, dried by vacuum centrifugation, and resuspended in 30 µl 0.1 % LC-MS grade FA.

2.4 Mass spectrometry data acquisition and analysis

2.4.1 Mass spectrometry data acquisition

For LC-MS/MS analysis of HLA class I-presented peptides, 4.5 µl of peptide sample in 0.1 % LC-MS grade FA were used per injection. Lyophilized synthetic peptides (Table 21; purchased from JPT Peptide Technologies; Berlin, Germany) for validation of neoepitopes were diluted to a concentration of 100 fmol/µl in 0.1 % LC-MS grade FA and 3 µl were used per injection. The mass spectrometric analysis was conducted using an UltiMate™ 3000 RSLCnano system directly coupled to an Orbitrap Fusion Lumos with ETD module (both from Thermo Fisher Scientific; Waltham, MA, USA). Peptides were loaded onto the trapping cartridge (Acclaim PepMap 100 C18 µ-Precolumn, 5µm, 300 µm i.d. x 5 mm, 100 Å; Thermo Fisher Scientific; Waltham, MA, USA) for 3 min at 30 µl/min using 0.05 % TFA in Pierce LC-MS grade H_2O (Fisher Scientific; Schwerte, Germany). Peptides were eluted and separated on an analytical column (nanoEase MZ HSS T3 column, 100 Å, 1.8 µm, 75 µm x 250 mm; Waters; Milford, MA, USA) with a constant flow of 0.3 µl/min using solvent A (0.1 % FA in LC-MS grade H_2O) and solvent B (0.1% FA in LC-MS grade ACN. The LC system was coupled on-line to the mass spectrometer using a Nanospray-Flex ion source (Thermo Fisher Scientific; Waltham, MA, USA) and a Pico-Tip Emitter (360 µm OD x 20 µm ID; 10 µm tip; New Objective; Littleton, MA, USA). The mass spectrometer was operated in positive mode and a spray voltage of 2.4 kV was applied for ionization with an ion transfer tube temperature of 275 °C. Full scan MS spectra were acquired in profile mode for a mass range of 300-1,650 m/z at a resolution of 120,000 (RF Lens 30 %, AGC target $4x10^5$ ions, and maximum injection time of 250 ms). The instrument was operated in data-dependent mode for MS/MS acquisition. Peptide fragment spectra were acquired for charge states 1-4 and the quadrupole isolation window was set to 1.2 m/z. For HCD fragmentation, collision energy was set to 35 %, and fragment mass spectra were recorded in the Orbitrap at a resolution of 30,000 for a maximum of $2x10^5$ ions (AGC target) or after 150 ms maximum injection time. For CID fragmentation, collision energy was set to 35 %, activation time to 10 ms, and activation Q to 0.25. Fragment mass spectra were acquired in the Orbitrap with a first mass set to 100 m/z at a resolution of 30,000 for a maximum of $2x10^5$ ions (AGC target) or after 150 ms maximum injection time. The instrument acquired MS/MS spectra for up to 3 s (or 1 s during MS parameter optimization) between MS scans and dynamic exclusion was set to 20 s for both HCD and CID fragmentation methods. Total analysis time for the HCD and CID method was 90 min starting with a 0-2 % solvent B elution step for 5 min and a 2-8 % B elution step for 1 min, proceeded with a gradient containing an 8-25 % solvent B elution step for 69 min, followed by an increase to 40 % solvent B for 5 min, 85 % B for 4 min, and a re-equilibration step to initial conditions for 6 min. Additionally, samples were analyzed using different dual fragmentation/decision tree methods, namely ETciD, EThcD, and lowEThcD. Here, the instrument fragmented precursors with a charge state of +1 using the previously described parameters of either the HCD or CID methods. Precursors with a charge state of 2-7 were fragmented using ETD (Calibrated Charge-Dependent ETD Parameters) with supplemental activation by either CID (35 %) or HCD (30 %). For supplemental activation by CID, first mass was fixed to 120 m/z, AGC target was set to $2x10^5$ ions, a maximum injection time of 200 ms was allowed and the resulting fragment spectra were recorded in the Orbitrap with a resolution of 30,000. For supplemental activation by HCD, AGC target was set to $2x10^5$ ions, a maximum injection time of 200 ms was allowed and the resulting fragment spectra were recorded in the Orbitrap with a resolution of 30,000. For fragment spectra acquisition, either high abundant (EThcD) or low abundant (lowEThcD)

precursors were selected first. For parameter optimization, total analysis time for dual fragmentation methods was 90-210 min. All gradients started with a 0-2 % solvent B elution step for 5 min and a 2-8 % B elution step for 1 min. The 90 min method contained an 8-25 % solvent B elution step for 69 min, followed by an increase to 40 % solvent B for 5 min, 85 % B for 4 min. The 120 min method contained an 8-25 % solvent B elution step for 94 min, followed by an increase to 40 % solvent B for 10 min, 85 % B for 4 min. The 150 min method contained an 8-25 % solvent B elution step for 122 min, followed by an increase to 40 % solvent B for 12 min, 85 % B for 4 min. The 180 min method contained an 8-25 % solvent B elution step for 150 min, followed by an increase to 40% solvent B for 14 min, 85% B for 4 min. The 210 min method contained an 8-25 % solvent B elution step for 185 min, followed by an increase to 40 % solvent B for 9 min, 85 % B for 4 min. For all gradients, a re-equilibration step to initial conditions was performed for 6 min. For analysis of the effect of NMD inhibition by 5AZA on the immunopeptidome, total analysis time for the EThcD and lowEThcD methods was 180 min using the gradient parameters described above. To further validate the presence of the candidates found in the HCT-116 cell line and to obtain reliable quantification data, InDel neoepitope candidates were measured from the same samples that were used for data-dependent acquisition by a targeted MS2 method with the previously described settings. Precursor masses of targeted InDel neoepitope candidates are listed in Table 8 and were fragmented using both the previously described HCD and EThcD parameters with a total analysis time of 90 min. Synthetic peptide pools were analyzed using the previously described parameters for HCD with a total analysis time of 90 min as well as EThcD and lowEThcD with a total analysis time of 180 min.

Table 8: Target list of InDel neoepitopes for the targeted MS2 approach. *m/z* indicate the targeted mass, t start and t stop the retention time window, and z the charge state. ox, oxidation.

peptide	*m/z*	z	t start (min)	t stop (min)	gene	method
SLM(ox)EQIPHL	542.2788	2	60	90	*CKAP2*	HCD, EThcD
SLMEQIPHL	534.2813	2	60	90	*CKAP2*	HCD, EThcD
KRSSTILRL	358.5645	3	27	47	*NFAT5*	HCD, EThcD
REKHSWHEP	302.1507	4	7	27	*PSMC6*	EThcD
REKHSWHEP	402.5319	3	7	27	*PSMC6*	HCD, EThcD
ISERDLLQY	568.8009	2	60	80	*STK38*	HCD, EThcD
GVWEKPRRV	376.2208	3	15	35	*TUBGCP3*	HCD, EThcD

2.4.2 Mass spectrometry data analysis

2.4.2.1 MaxQuant analysis

The Andromeda search engine integrated into the MaxQuant computational proteomics platform (version 1.5.8.3; Cox and Mann, 2008; Cox *et al.*, 2011) was used to search MS raw data against the UniProt/SwissProt database (20,659 entries, February 2019) and a list of common contaminants for the evaluation of different search engines for immunopeptidomics. Enzyme specificity was set as “unspecific”, and N-terminal acetylation (+42.01 Da), oxidation of methionine (+15.99 Da) and carbamidomethylation of cysteine (+57.02 Da) were set as variable modifications. Sequence matches were restricted to 8-15 aa, a maximum mass of 1,500, and a maximum charge state of +3. Precursor ions were allowed to have a maximal mass deviation of 6 ppm and fragment ions were allowed to have a maximal mass deviation of 20 ppm. The “match between runs” option was enabled. The FDR was set

to 1 % at the peptide level while there was no FDR set at the protein level. The “peptide.txt” output file generated by MaxQuant was further filtered excluding hits to the contaminant list and the decoy database.

2.4.2.2 PEAKS analysis

MS raw data were analyzed using PEAKS Studio (version 10.0; Bioinformatics Solutions; Waterloo, ON, Canada). Raw files were subjected to the default data refinement before *de novo* sequencing-assisted database search with precursor mass error tolerance set to 10.0 ppm and fragment mass error tolerance set to 0.02 Da. Fragmentation was set either to “HCD” or “CID” for raw files recorded with single fragmentation methods or to “Mixed” for raw files recorded with dual fragmentation/decision tree methods. All raw files were first searched against the UniProt/SwissProt database (20,659 entries, February 2019) and a database containing standard contaminants. N-terminal acetylation (+42.01 Da), oxidation of methionine (+15.99 Da), and carbamidomethylation of cysteine (+57.02 Da) were set as variable modifications. The enzyme specificity was set to “no enzyme”. “*De novo* only” spectra with an average local confidence of more than 50 % (*i.e.* good spectra not matching a UniProt/SwissProt database entry) were subjected to a multi-round database search using the in-house generated neoepitope databases. HCT-116 data were searched against the cell line-specific CCLE (527 entries) and COSMIC database (705 entries) and the cell line-specific SNP database (1,260 entries). RKO data were searched against the cell line-specific CCLE (235 entries) and COSMIC database (1,019 entries) and the cell line-specific SNP database (1,504 entries). Additionally, both, HCT-116 and RKO data were searched against the ReFrame database (519 entries) and the MNR database (149,907 entries). All peptides identified at a peptide spectrum match FDR of 1 % were exported and further filtered by removing contaminants, decoy hits as well as peptides shorter than 8 aa and longer than 15 aa. Peptides matching entries of the UniProt/SwissProt database were reported as “endogenous”, peptides matching one of the frameshift peptide databases were reported as “InDel neoepitopes” and peptides matching one of the SNP neoepitope databases were reported as “SNP neoepitope”.

2.4.3 Quality control of immunopeptidomics data

All identified peptides were subjected to different *in silico* quality controls. Additionally, the spectra of identified neoepitopes were compared to synthetic counterparts in order to confirm their identity and exact aa sequence.

2.4.3.1 *In silico* quality control

Sequence-specific hydrophobicity indices (HIs) are an orthogonal parameter to validate correct peptide identifications and correlate with experimentally observed retention times (Krokhin, 2006). HIs were calculated for both unmodified as well as modified identified peptides using the web-based tool SSRcalc

(version Q; Krokhin, 2006)[3] with the following parameters: 100Å C18 column, 0.1% Formic Acid (2015 model), no cysteine protection.

Binding prediction of identified peptides in an immunopeptidomics experiment can be used to assess the quality of the data. Here, binding prediction to the HLA allotypes expressed by either HCT-116 cells or RKO cells was performed using the command line version of NetMHCpan (version 4.0a; Jurtz *et al.*, 2017) and GNU Parallel (Tange, 2018). The rank threshold for binders was set to < 0.5 % for strong binders and < 2 % for weak binders. For peptides binding to more than one HLA allotype expressed, only the best-ranked HLA allotype with its corresponding affinity in nM was reported.

Given the similarity of HLA class I-presented peptides due to the allele-specific binding-mediating anchor residues, peptide sequences can be clustered into groups in order to identify or recapitulate HLA binding motifs. Peptide sequences were clustered using the web-based GibbsCluster (version 2.0; Andreatta *et al.*, 2017b)[4] with default parameters for MHC class I ligands. Motifs were visualized using the web-based Seq2Logo (version 2.0; Thomsen and Nielsen, 2012)[5] and compared to literature HLA class I motifs obtained from the NetMHCpan webpage.

2.4.3.2 Synthetic peptide validation

The identity and exact aa sequence of identified neoepitopes were validated by comparison of spectra recorded from synthetic peptides to spectra measured from the samples and used for identification. MS raw data recorded from synthetic peptide pools were analyzed using PEAKS Studio as described above and matched fragment ions (“PSM-ions.txt”) was exported for both synthetic peptides and neoepitope candidates measured in biological samples. All spectra recorded for a given peptide were compared to all spectra recorded from its synthetic counterpart using a custom script in the *R* programming language (ISBN 3-900051-07-0) and partially based on a script by Daniel Zolg (Zolg *et al.*, 2017). To this end, the intensity correlation between the matched fragment ions of the synthetic peptide spectrum and the spectrum of the peptide measured from the biological sample was computed. For visualization of the peptide pair with the highest correlation, intensities were normalized and plotted using the *ggplot* package (ISBN 978-3-319-24277-4). Moreover, retention time differences between the two peptides were calculated and reported.

[3] http://hs2.proteome.ca/SSRCalc/SSRCalcQ.html
[4] http://www.cbs.dtu.dk/services/GibbsCluster/
[5] http://www.cbs.dtu.dk/biotools/Seq2Logo/

2.5 *In vivo* immunizations and ELISpot assays

In order to evaluate the immunogenic potential of the identified and validated InDel neoepitopes, *in vivo* immunizations with the InDel neoepitopes and a TLR9-based adjuvant (van Duikeren *et al.*, 2012) were performed in an "HLA-humanized" mouse model. The utilized mouse model does not express endogenous murine H-2 class I and class II molecules, but instead, expresses human HLA-A2.1 and HLA-DR1 molecules and thus is ideal to investigate HLA-restricted $CD8^+$ and $CD4^+$ T cell responses.

2.5.1 Mouse strain

The HLA-A2.1/HLA-DR1-transgenic H-2 class I-/class II-knockout mice (Pajot *et al.*, 2004) were provided by the Institute Pasteur (Paris, France). All animal procedures followed the institutional laboratory animal research guidelines and were approved by the governmental authorities. The mice were fed a standard chow diet and provided water *ad libitum*. The Animal Care Facilities at German Cancer Research Center (DKFZ; Heidelberg, Germany) have been approved by FELASA and accredited. For the experiments, mice were assigned to age-matched and sex-matched groups.

2.5.2 *In vivo* immunization and organ preparation

HLA-A2.1/HLA-DR1-transgenic H-2 class I-/class II-knockout mice were immunized weekly for three weeks with either a peptide pool consisting of 50 µg of the *CKAP2*-, *NFAT5*-, *PSMC6*-, and *TUBGCP3*-derived InDel neoepitope each or 100 µg of the *CKAP2*- or *TUBGCP3*-derived InDel neoepitopes separately. The known viral HPV epitope E7 11-19 was used as a positive control. All peptides were purchased either from JPT Peptide Technologies (Berlin, Germany) or synthesized by the GMP and T Cell Therapy Unit at the German Cancer Research Center (DKFZ; Heidelberg, Germany) and were HPLC-purified. Peptide pools or single peptides were mixed with 20 µg CpG ODN 1826 (TIB MolBiol; Berlin, Germany) as an adjuvant, suspended in sterile PBS, and injected intradermally. Seven days after the last immunization, mice were sacrificed by CO_2 inhalation. For the analysis of cellular immune responses, the spleen was removed under aseptic conditions and transferred in sterile ice-cold PBS supplemented with 10 % FBS. The tissue was disaggregated by passing it through a 40 µm cell strainer to obtain single cells. Splenocytes were pelleted by gentle centrifugation with 282 x g for 5 min at 4 °C and resuspended in 1 ml of 1X Red Blood Cell (RBC) lysis buffer (Sigma-Aldrich; St. Louis, MO, USA). After lysis of erythrocytes for 5 min at RT, 9 ml Gibco RPMI 1640 [+] L-Glutamine supplemented with 10 % FBS and 1 % penicillin-streptomycin were added, and cells were pelleted by gentle centrifugation with 282 x g for 5 min at 4 °C. The cell pellet was resuspended in Gibco RPMI 1640 [+] L-Glutamine supplemented with 10 % FBS and 1 % penicillin-streptomycin, counted, and adjusted to $1x10^6$ cells/ml. Isolated splenocytes were either directly used in ELIspot assays or $CD8^+$ were isolated from whole splenocytes and used in ELISpot assays.

2.5.3 Isolation of $CD8^+$ T cells

$CD8^+$ T cells were isolated from splenocytes of immunized mice using the murine CD8a+ T Cell Isolation Kit (Miltenyi Biotech; Bergisch Gladbach, Germany). Briefly, 40 µl MACS buffer and 10 µl of Biotin

Antibody Cocktail were added to $1x10^7$ splenocytes and incubated for 5 min at 4 °C. Next, 30 µl of MACS buffer and 20 µl of Anti-Biotin MicroBeads were added and incubated for 10 min at 4 °C. Meanwhile, LS columns (Miltenyi Biotech; Bergisch Gladbach, Germany) were placed on a magnetic separator and rinsed with 3 ml MACS buffer. The cell suspension was loaded onto the columns and washed with 6 ml MACS buffer. Magnetically-labeled cells were retained by the magnetic separator and the flowthrough containing the unlabeled enriched $CD8^+$ T cells was collected. $CD8^+$ T cells were pelleted by gentle centrifugation with 282 x g for 5 min at 4 °C. The cell pellet was resuspended in Gibco RPMI 1640 [+] L-Glutamine supplemented with 10 % FBS and 1 % penicillin-streptomycin, counted, and adjusted to $1x10^6$ $CD8^+$ T cells/ml.

2.5.4 ELISpot assay

IFNγ ELISpot was performed *ex vivo* seven days after the last immunization either with isolated splenocytes or $CD8^+$ T cells as described previously (Ballhausen *et al.*, 2020). Briefly, MultiScreenHTS-IP plates (Merck; Darmstadt, Germany) were activated with 70 µl/well 70% ethanol for 5 min and washed five times with sterile PBS. 100 µl/well of purified rat anti-mouse IFNγ antibody (BD Biosciences; San Jose, CA, USA) diluted 1:200 in PBS were added and incubated overnight at 4 °C. Plates were washed four times with PBS and blocked with 200 µl/well Gibco RPMI 1640 [+] L-Glutamine supplemented with 10 % FBS and 1 % penicillin-streptomycin for 1 h at 37 °C. For the analysis of whole splenocytes, 2 µg of peptide in 100 µl assay medium were directly added to each well, topped up with $1x10^6$ splenocytes in 100 µl assay medium and incubated for 16-20 h at 37 °C. For the analysis of isolated $CD8^+$ cells, plates were coated with $4x10^5$ splenocytes/well from naïve mice as antigen-presenting cells supplemented with 2 µg/well of the corresponding peptide and incubated for 4 h at 37 °C. Next, $1x10^5$ $CD8^+$ T cells/well were added and incubated for 16-20 h at 37 °C. For both the analysis of splenocytes and $CD8^+$ T cells, cells were removed and plates were washed four times with PBS supplemented with 0.01% Tween 20, once with PBS only, and coated with 100 µl/well of biotinylated rat anti-mouse IFNγ antibody (BD Biosciences; San Jose, CA, USA) diluted 1:500 in PBS for 1 h at RT. Next, plates were washed six times with PBS supplemented with 0.01% Tween 20 and once with PBS only. 100 µl/well of AKP Streptavidin (BD Biosciences; San Jose, CA, USA) diluted 1:500 in PBS were added and incubated for 30 min at RT in the dark followed by three wash steps with PBS supplemented with 0.01% Tween 20 and three times wash steps with PBS only. 100 µl/well BCTP/NBT substrate (Thermo Fisher Scientific; Waltham, MA, USA) was added and incubated up to 45 min depending on the color development. The reaction was stopped with ddH_2O, plates were dried overnight and analyzed using the CTL ImmunoSpot Reader and the ImmunoSpot software (version 7.0.24.1; both from CTL Europe; Bonn, Germany). 2 µg/well Concanavalin A (Sigma-Aldrich; St. Louis, MO, USA) was used as an assay high control for IFNγ production in all performed experiments.

2.6 Neoepitope database construction

For the identification of mutation-derived neoepitopes from mass spectrometry raw data, it is mandatory to generate databases that contain the mutated protein sequences. Here, databases with frameshift mutation-derived protein sequences were constructed by using public sequencing information of the Cancer Cell Line Encyclopaedia (CCLE) and the Catalogue of Somatic Mutations in Cancer (COSMIC), information on frequently mutated genes in MSI cancers (ReFrame), and by *in silico* mutation of all coding microsatellites (MNR). Moreover, previously published data on cell-line specific neoepitopes derived from single nucleotide polymorphisms were integrated (SNP). All scripts used are available upon request.

2.6.1 CCLE- and COSMIC-based databases

Custom scripts in the *R* programming language were used to construct protein sequence databases in the FASTA format. For the CCLE-based databases, germline-filtered merged mutation calls for the coding regions ("CCLE_DepMap_18q3_maf_20180718.txt", July 2018) were downloaded from the CCLE webpage[6]. The obtained data frame was filtered for frameshift mutations in the two cell lines of interest using the "Tumor_Sample_Barcode" and the "Variant_Classification" columns. The coding sequence and the 3' UTR sequence for the annotated transcript ("Annotation_Transcript") were obtained from the Ensembl repository using the *biomaRt* package (Durinck *et al.*, 2005; Durinck *et al.*, 2009). For the COSMIC-based databases, HCT-116 and RKO variants data were exported separately from the COSMIC webpage[7]. The obtained data frames were filtered for frameshift mutations in the two cell lines of interest using the "Type" column. The coding sequence and the 3' UTR sequence for the annotated transcript ("Transcript") were obtained from the Ensembl repository using the *biomaRt* package. Next, obtained sequences for the coding part of the transcript as well as the 3' UTR were concatenated and *in silico* mutated using the mutation information provided in the "cDNA_Change" (CCLE) or "CDS Mutation" columns (COSMIC), respectively. The remaining steps were the same for both datasets. Briefly, the mutated sequence was then *in silico* translated using the *Biostrings* package. The resulting protein sequences were trimmed to contain 14 aa of the wild-type protein sequences downstream of the mutation position, followed by the frameshifted protein sequences up to the first stop codon. The 14 wild-type aa were added with the rationale that already one frameshifted aa can alter the protein sequence so that a neoepitope is generated. Providing only the frameshifted protein sequences would exclude these special cases. Finally, the protein sequences were annotated with UniProt identifiers and HGNC symbols of the wild-type protein as well as the corresponding Ensembl description. Mutated protein sequences and corresponding annotations were exported as a FASTA file using the *seqinr* package (Charif and Lobry, 2007).

2.6.2 MNR database

The MNR database contains protein sequences generated by *in silico* mutation of all coding mononucleotide repeats greater or equal to six nt. The length of six nt was chosen with the rationale that

[6] https://portals.broadinstitute.org/ccle/data
[7] https://cancer.sanger.ac.uk/cell_lines

starting from this length the mutation probability is exponentially increasing (Woerner *et al.*, 2010). The human genome was queried for coding mononucleotide repeats using a custom script in the *Perl* programming language (ISBN 978-1-449-30358-7) and partially based on a script by Yan Ping Yuan (European Molecular Biology Laboratory; Heidelberg, Germany; Woerner *et al.*, 2010; Woerner *et al.*, 2015). The *Perl* script uses the Ensembl *Perl* API[8] and *Bioperl* (Stajich *et al.*, 2002) to fetch all identifiers of protein-coding transcripts of all genes on a given chromosome. Next, the transcript IDs are used to obtain the corresponding coding sequences as well as the 3' UTR sequences. The coding sequences were then searched for mononucleotide repeats greater or equal to six nt. The position of found mononucleotide repeats in relation to the coding sequence position together with the type and length of mononucleotide repeat as well as the gene, transcript, and UniProt ID, the Ensembl description, and both coding sequence and 3' UTR sequence were deposited in a separate output file for each queried chromosome. The output files were further processed using a custom script in the *R* programming language to generate a database containing protein sequences in the FASTA format. Briefly, coding sequences were *in silico* mutated by either inserting one or two nt or deleting one or two nt at the previously obtained coding sequence position of the mononucleotide repeat. The mutated sequences were then *in silico* translated using the *Biostrings* package and the resulting protein sequences were trimmed to contain 14 aa of the wild-type protein sequences downstream of the mutation position, followed by the frameshifted protein sequences up to the first stop codon. Mutated protein sequences and corresponding annotations were exported as a FASTA file using the *seqinr* package.

2.6.3 ReFrame and SNP neoepitope databases

The ReFrame database was constructed based on previously published protein sequences originating from recurrent InDel mutations in MSI cancer (Ballhausen *et al.*, 2020). Briefly, "cMNR_peptides.csv" was obtained from GitHub[9] and the *biomaRt* package was used to annotate the protein sequences based on the provided gene symbol. Protein sequences and corresponding annotations were exported as a FASTA file using the *seqinr* package.

The SNP neoepitope database was constructed based on previously published SNP neoepitope data (Scholtalbers *et al.*, 2015). Briefly, cell line-specific neoepitope data were obtained from the TRON Cell Line Portal[10] and filtered for entries for HLA class I molecules. Neoepitope sequences and corresponding annotations were exported as a FASTA file using the *seqinr* package.

[8] https://m.ensembl.org/info/docs/api/core/index.html#api
[9] https://github.com/atb-data/neoantigen-landscape-msi
[10] http://celllines.tron-mainz.de/

2.7 Statistical analysis

2.7.1 General statistical analysis

Luciferase assay and qPCR data are presented as mean ± SD from at least three independent experiments. Proliferation assay data are presented as mean ± SEM from three independent experiments. ELISpot assay data are presented as mean ± SEM scatter dot plots from three to six independent experiments. Statistical analyses were made with GraphPad Prism 6 (version 6.01; GraphPad Software; San Diego, CA, USA) using a two-sided, unpaired t-test with correction for multiple hypothesis testing. $p \leq 0.05$ was considered significant. Results were visualized using the ggplot2 package.

2.7.2 Label-free quantification of HLA class I-presented peptides

Quantification of HLA class I-presented peptides was performed using the raw output (peptides.csv) from PEAKS Studio X and a custom script in the R programming language. The script is a modified version of the standard proteomics quantification pipeline provided by Frank Stein (European Molecular Biology Laboratory; Heidelberg, Germany) which has been used in other publications from our laboratory (Perez-Perri *et al.*, 2018; Backlund *et al.*, 2020). The script was modified to enable the quantification on peptide level rather than on protein level. All steps were performed for each MS method and cell line separately. For quantification of the HCT-116 dataset, intensities measured using the targeted MS2 approach were integrated into the dataset recorded by data-dependent acquisition. First, data were filtered to include only peptides which were measured in at least two out of three replicates per condition and MS method used. Potential batch effects between biological replicates were removed using the *limma* package (Ritchie *et al.*, 2015). Next, the variance stabilization normalization method was performed on the log2 transformed raw data using the *vsn* package (Huber *et al.*, 2002) with individual normalization coefficients for the different MS methods. Missing values were imputed with the *Msnbase* package using nearest neighbor averaging (Gatto and Lilley, 2012). Normalized data were tested for differential HLA class I presentation of peptides between DMSO- and 5AZA-treated samples using the *limma* package. The replicate factor was included in the linear model. Peptides with an FDR ≤ 0.05 and an FC > 2 were defined as “hits” while peptides with an FDR ≤ 0.2 and an FC ≥ 1.5 were defined as “candidates”. GO-term analysis for source proteins of “hit” peptides in the HCT-116 cell line was performed with the PANTHER classification system (Mi et al., 2019)[11] using the following parameters: PANTHER overrepresentation test with “Homo sapiens (all genes in database)” as reference, “GO biological process complete” as annotation dataset, and Fisher’s Exact with FDR correction for multiple testing. All results were visualized using the ggplot2 package.

[11] http://pantherdb.org/webservices/go/overrep.jsp

2.8 Key resources

The following tables list the key resources essential to reproduce the data presented in this thesis including the manufacturers' catalog number and unique Research Resource Identifiers (RRID)[12] if applicable.

Table 9: Antibodies used in this thesis.

name	source	identifier
Goat-anti-rabbit IgG (HRP Conjugate)	Sigma-Aldrich	Cat# A0545; RRID: AB_257896
Goat-anti-RENT1 antibody (anti-UPF1)	Bethyl Laboratories	Cat# A300-038A; RRID: AB_2288326
Mouse monoclonal anti-HLA-ABC antibody (W6/32)	Barnstable *et al.*, 1978; Parham *et al.*, 1979	N/A
Mouse monoclonal anti-α-Tubulin	Sigma-Aldrich	Cat# T5168; RRID: AB_477579
Rabbit monoclonal anti-β-2-microglobulin (EP2978Y)	Abcam	Cat# ab75853; RRID: AB_1523204
Rabbit-anti-goat IgG (HRP Conjugate)	Sigma-Aldrich	Cat# A5420, RRID: AB_258242
Rabbit-anti-mouse IgG (HRP Conjugate)	Sigma-Aldrich	Cat# A9044; RRID: AB_258431
Rat monoclonal anti-IFNγ antibody (R4-6A2)	BD Bioscience	Cat# 551216, RRID: AB_394094
Rat monoclonal anti-IFNγ antibody (XMG1.2, Biotin conjugated)	BD Bioscience	Cat# 554410; RRID: AB_395374

Table 10: Biological samples used in this thesis.

name	source	identifier
Colorectal cancer cell lines (genomic DNA)	Woerner *et al.*, 2001; Woerner *et al.*, 2007; Michalak *et al.*, 2020	N/A
Human tissues (tumor DNA)	Tissue bank of the German Collaborative Group on HNPCC (Heidelberg, Germany)	N/A

[12] https://scicrunch.org/resources

Table 11: Chemical, peptides, recombinant DNA, and oligonucleotides used in this thesis.

name	source	identifier
5-Aza-2′-deoxy-cytidin (dAZA)	Sigma-Aldrich	Cat# A3656
5-Azacytidine (5AZA)	Sigma-Aldrich	Cat# A2385
AllStars Negative Control siRNA	Qiagen	Cat# 1027281
CpG ODN 1826	TIB MolBiol	N/A
Fluorescently labeled primers for mutation analysis by capillary gel electrophoresis; see Table 19	this thesis	N/A
Primers for mutation analysis by Sanger sequencing; see Table 17 and Table 18	this thesis	N/A
Primers for qPCR; see	this thesis; Viegas *et al.*, 2007; Bhuvanagiri *et al.*, 2014; Cheruiyot *et al.*, 2018	N/A
Renilla-HBB WT/*Renilla-HBB* NS39 plasmids, firefly plasmid (NMD efficiency assay)	Boelz *et al.*, 2006	N/A
siRNA targeting UPF1 (AA-GAUGCAGUUCCGCUCCAUU-TT)	Gehring *et al.*, 2003	N/A
synthetic peptides for validation and *in vivo* immunization; see Table 21	JPT Peptide Technologies and GMP & T Cell Therapy Unit at German Cancer Research Center (Heidelberg, Germany)	N/A

Table 12: Experimental models and strains used in this thesis.

name	source	identifier
Eschericha coli: XL1-Blue competent cells	Stratagene	Cat# 200249
HLA-A2.1/HLA-DR1-transgenic H-2 class I-/class II-knockout mice	Institute Pasteur (Paris, France); Pajot *et al.*, 2004	N/A
Homo sapiens: HCT-116 cells	Department of Applied Tumor Biology, Heidelberg University (Heidelberg, Germany)	DSMZ Cat# ACC 581; RRID: CVCL_0291
Homo sapiens: LoVo cells	Department of Applied Tumor Biology, Heidelberg University (Heidelberg, Germany)	ATCC Cat# ACC 350; RRID: CVCL_0399
Homo sapiens: RKO cells	Department of Applied Tumor Biology, Heidelberg University (Heidelberg, Germany)	ATCC Cat# CRL-2577; RRID: CVCL_0504
Mus musculus: HB95 hybridoma cells	Department of Oncology, Ludwig Institute for Cancer Research (Lausanne, Switzerland)	ATCC Cat# HB-95; RRID: CVCL_7872

Table 13: Critical commercial kits used in this thesis.

name	source	identifier
CD8a$^+$ T Cell Isolation Kit	Miltenyi Biotec	Cat# 130-104-075
Dual-Luciferase Reporter Assay System	Promega	Cat# E1910
primaQuant CYBR qPCR Master Mix	Steinbrenner Laborsysteme	Cat# SL-9902

Table 14: Deposited data used and generated in this thesis.

name	source	identifier
Cancer Cell Line Encyclopaedia (CCLE)	Barretina *et al.*, 2012; Ghandi *et al.*, 2019	RRID: SCR_013836
Catalogue of Somatic Mutations in Cancer (COSMIC)	Tate *et al.*, 2018	RRID: SCR_002260
Mass spectrometry raw data and search results	this thesis	PRIDE: PXD021755
ReFrame database of recurrent InDel mutations in MSI CRC	Ballhausen *et al.*, 2020	N/A
TRON Cell Line Portal (database of neoepitopes originating from single nucleotide polymorphisms)	Scholtalbers *et al.*, 2015	N/A
UniProt/SwissProt database	https://www.uniprot.org/	RRID: SCR_002380

Table 15: Software, algorithms, and packages used in this thesis.

name	source	identifier
biomaRt	Durinck *et al.*, 2005; Durinck *et al.*, 2009	RRID: SCR_002987
Biostrings	https://bioconductor.org/packages/release/bioc/html/Biostrings.html	RRID: SCR_016949
Ensembl *Perl* API	https://m.ensembl.org/info/docs/api/core/index.html#api	N/A
ggplot2	https://cran.r-project.org/web/packages/ggplot2/index.html	RRID: SCR_014601
GibbsCluster (version 2.0)	Andreatta *et al.*, 2017b	N/A
GraphPad Prism 6 (version 6.01)	GraphPad Software	RRID: SCR_002798
Indigo webtool for rapid InDel discovery in Sanger chromatograms	Rausch *et al.*, 2020	N/A
limma	Ritchie *et al.*, 2015	RRID: SCR_010943
MaxQuant (version 1.5.8.3)	Cox and Mann, 2008; Cox *et al.*, 2011	RRID: SCR_014485
Msnbase	Gatto and Lilley, 2012	N/A
NetMHCpan (version 4.0a)	Jurtz *et al.*, 2017	RRID: SCR_018182
PANTHER classification system	Mi *et al.*, 2019	RRID: SCR_004869
PEAKS Studio X (version 10.0)	Bioinformatics Solutions	N/A
Perl	https://www.perl.org/	N/A
R	https://www.r-project.org/	N/A
RStudio (version 1.3.1093)	https://rstudio.com/	RRID: SCR_000432
Seq2Logo (version 2.0)	Thomsen and Nielsen, 2012	N/A
seqinr	Charif and Lobry, 2007	N/A
SSRcalc (version Q)	Krokhin, 2006	N/A
vsn	Huber *et al.*, 2002	RRID: SCR_001459

Table 16: Miscellaneous reagents and equipment critical for this thesis.

name	source	identifier
1-Step BCIP/NBT Substrate Solution	Thermo Fisher Scientific	Cat# 34042
96-well filter plate with 3 µm glass fiber and 10 µm polypropylene	Agilent	Cat# 201017-100
ABI3130xl gel capillary electrophoresis system	Applied Biosystems	N/A
AKP Streptavidin	BD Biosciences	Cat# 554065
CELLine CL1000 bioreactor	Sigma-Aldrich	Cat# Z688029
Concanavalin A from *Canavalia ensiformis*	Sigma-Aldrich	Cat# C5275
Extraction Plate Manifold	Waters	Cat# 186001831
INTERFERin (siRNA transfection)	Polyplus transfection	Cat# 409-10
jetPRIME (DNA transfection)	Polyplus transfection	Cat# 114-15
LS Columns	Miltenyi Biotec	Cat# 130-042-401
MultiScreenHTS-IP plates for ELISpot	Merck	Cat# MSIPN4W50
OASIS HLB µElution Plate 30 µm	Waters	Cat# 186001828BA
Orbitrap Fusion with ETD module	Thermo Fisher Scientific	N/A
Passive Lysis 5X Buffer	Promega	Cat# E1941
Poly-Prep Chromatography Columns (9 ml)	Bio-Rad Laboratories	Cat# 7311550
Positive Pressure-96 Processor	Waters	Cat# 186006961
RNAiMAX (siRNA transfection)	Invitrogen	Cat# 13778150
Sepharose-Protein A conjugate 4B beads	Invitrogen	Cat# 101142
Sep-Pak tC18 plate 100 mg sorbent	Waters	Cat# 186002321
StepOnePlus qPCR system	Applied Biosystems	N/A

2.9 Primers and synthetic peptides

2.9.1 Primers

Table 17: Primers used for mutation validation of SNP neoepitopes by Sanger sequencing.

gene	forward primer	reverse primer	cell line
CHMP7	GGTTGCCTTTGCCTTTCCAG	TGGCCCCTTTCTGTACCTCT	HCT-116
RGP1	TTGCCGTGCTAGTCTTGTTCA	TGACTGACTGACCCCGAAAG	HCT-116
PCMT1	GTTTTTCTTTCAGTGGGGGATGG	TCCCAATTCTACTGTTGTCCTAGT	HCT-116
RBB7	AGAGCAGTTAGTTGTCCGTGT	GGAAGCCACTGAACGGTAAGA	HCT-116
NR1D1	GGTGGCGTCATCACCTACAT	GCCACTTGTAGACTCCCAGG	HCT-116
AURKA	GTGAGCTCCAAACCAAGTGTT	TACGTCAAGACAAAACCCACTC	RKO
CNN3	AGTAACTTGTCGGTTGTGCC	GGCTGGCTCCTTTATTAGTGC	RKO
MAFK	ATCAAGCGGGTGACGCAG	CCCAGTGACAGGTACGAGG	RKO
PRMT1	CCTGCAGGAGGTGGACATCTAT	CCATGGTTCCTTCTGCCCGTT	RKO
RHOG	CCTGCAACAGGATGGTGTCA	TACCCAGTGTTCCCAAGCAG	RKO

Table 18: Primers used for mutation validation of InDel neoepitope by Sanger sequencing.

gene	forward primer	reverse primer	cell line
CKAP2	AAGTTTCTCACTTCGGTGAGCTT	TGCATTAGGGCGGCATACAA	HCT-116
NFAT5	TACCAGGTTTGTATCTCATGCTAC	TGGAAGTCACTATGTGGGCAAT	HCT-116
PSMC6	ATGGTTTTACCTAGCATGGAAGTCT	CAATACCAACCTGGTGGTCCAT	HCT-116
STK38	GGTCTAGGCTCTCACGGCTA	GCTTGAGATGTGCTGAAAGGC	HCT-116
TUBGCP3	GGGGAATACGTTTGTGGGTTG	CAGTGCAACGAACATCACCC	HCT-116
DYNC2LI1	CTGCCTCTCATGTAGTGAAACTCT	TCATGCTAAACCAGATTGGTCATC	HCT-116
MSL2	GATGAAACCTTCCTGTAGCTGGT	CTGAGGGTGCTTTCAGACATTG	HCT-116
SIAH2	TTGCTTTGAGCAAAAGTAGTTAAGA	TCGAAAACAAGGCAGTCGCT	HCT-116
EWSR1	AAGGGATTCAATATCCTGATGGTTT	TGCTGCCCATAGGCTGGATA	HCT-116
NMT2	ATTCCTCACCTACATCCTAACACAG	CCACTCATGCACGAGAAACC	HCT-116
RPS2	CGGGTCTAGGCTCCCGAGT	ACCTACCTCCTTATCCTCGGC	HCT-116
AGO2	CCCCCGTTAATGGGTGTTCT	CGCGCAGACCACTTACACAG	RKO
BCAP29	TCTGTAATTTTCACTTGCTACACAA	TGTACACCAATTGTAACATCAACTC	RKO
BTBD10	ATGGACAGGTGAGGTAGGGG	AAGGTCACACATATCCCTCTCA	RKO
C16orf72	GCAGCGGTTATCTGGTCC	GGCTCAGGGCACAATGGG	RKO
CEBPZ	ACCTTATGCTGGCTACTTTGGAT	ATCAGAATGTGGTTCTGGCCTA	RKO
CLEC11A	TGATGGGCTCAGAGACAGGA	GGGTGAGGGCTTCTGGTTAC	RKO
COQ3	GAGGAGGGAGGGGGAAACA	GGCTCAAGAAGGCAGACTGA	RKO
DCAF5	TGTCAGCCTTGCGATGTCAG	ACCAGTAATTCCACAGCTCAATG	RKO
DISP1	GCAGTTGTTGTGCTGCATGA	CGGATAACTCCAGTGAGGGC	RKO
DNAJC16	TGATCAATATGGAGACGCTGGA	AAAGCACCAATCGGAGGTGA	RKO
DNAJC8	AGTTGTTCCTCCTATTGGCCC	AGACAGGGGGTTTGGTAGTTA	RKO
DNTTIP2	TGGCCAAAGGGAGTCATTAAGT	TTAGCCTTTGTGCTATTTAGGAAGT	RKO
DSTN	CAGGAGTGCAAGTAGCTGATGA	GCTGGAACTGGTCTTCCCGA	RKO
EIF5	CGTACAAATGTGAGTGAGTAGCC	TAGCGGTATCACAGAAGCAGG	RKO
FBXO25	ATTACCCTTGTGATTTCAAGCCC	TGAAACGGTTGAGCTGCAAAAAG	RKO
HNRNPH1	GACAGCTGTTGCTCAGGACA	GCCGCACTCCCCCAAG	RKO
LUC7L3	TGCTTCATGATTTAGTAGGATCAGT	TTCCTCAACCCTTGTCCCAC	RKO
MAPKAPK3	TGCCACTAGAAGCGCCAGG	GCACACTTCTGTCCAGTGCG	RKO
MRPL15	GCACTTTGGAGGGAAAAACACTG	GTAGCCAACATGAGACAAAGAGAG	RKO

MRPL3	GGGGAATATGTTTATCTGCGAGG	GGCCATGGTTCATCTTTCAGAG	RKO
NAA35	TCTTGCCCTGAAAGCTACATCC	TGGCTAACAAACTCAACAAACCAC	RKO
NABP1	AGGGCTTGATCCTCACTTCT	AGCTACCAAAGACATTCAGTTTAC	RKO
NFX1	CAGAGTCAAGAAAGCACAGAGTC	CAGCATCCTCCGGTTTTGGA	RKO
PDZD8	TTGGTTTTGTAGAGCACGTGTA	TGCACTGTTCTTCCCGTAACT	RKO
PRRG1	TCCTCACGGGAGAAAAAGCC	AAGACAGTCCAGCACAAGGG	RKO
RBFA-1	GGCTGTGGCGCTCCC	CAGGAACTGGCGAGCTAGG	RKO
RBFA-2	AGCCCCCTATAAACAGTTATGGT	CAATGGCATCCAGTCGCATC	RKO
RBPJ	GTGGAAGTCCCGATTTCTGTC	TAATTGAAGAAACCCCTGCCACG	RKO
RPA2	AGACTCGGGCAGTGTTTGAA	AGATCATTATCGCATCAGTTCACG	RKO
RSRC2	ACAGAGCGAGATGGACTAGC	CCATGCCCAGCCAACAAAC	RKO
SECISBP2	GCCCTCTCCCACAAGAAATGAA	GGACAAATGATGTGAACCTCTAGC	RKO
SH3PXD2A	GCCTCTCTGTCACTGTTGGT	CCCTGCTAGCTCACCTTCTG	RKO
SSBP3-1	CCGTCTCTGCTCCTCTAAGC	TTCCGCCCTTCTCCTCCTT	RKO
SSBP3-2	ACGGGGAATGTGGACGGAC	ATGGCAAACTGAGCTACTTACTTCT	RKO
TAF1B	TGTATTCCAGGCACTGTTTTAGTTA	ACATAGGATAGAAGAACGTATGCC	RKO
TCF19	GAAGCGAACTTAAGCCAGCG	GCGTGGATCCCAGAGATGAG	RKO
TIFA	AGATGACGGTTTACCATCCTGG	TGCTGTCCACGATCAGATTGG	RKO
TRIP12	TGCTGTGATAGTTCCACAACCA	AGCCTTGCTTGTTGATGGTGA	RKO
USP1	TGTTGCCAAAACTGATAACCAAA	TGAGAATCTGTGAAATCCAAAGCTC	RKO
WDR75	CTGTTAACTTTTGTGGGGGAGTG	TTAAATTGAGGTGCAGTGGGGA	RKO
ZNF439	TCCCTCAGTGCATTAGTTTACC	TCCTGGGGTTTTGGTACTCATATT	RKO
ZYX	GGACCGGGACGCAGAGT	ACAGGGCCGAACTTCTTCTG	RKO

Table 19: Primers used for evaluation of mutation frequency in MSI CRC cell lines and patient samples by gel capillary electrophoresis. One primer of each pair was labeled fluorescently with 6-carboxyfluorescein (6FAM).

gene	forward primer	reverse primer
CKAP2	TGTTAACATGTTTTTGAACTCTGGA	[6FAM]CGAGAACGTCTCACTGGTGT
NFAT5	[6FAM]CCTAATGCCCTGATGACTCCAC	ATAGGAGGTTTGTGCACTAGTCAAT
PSMC6	[6FAM]TAGAATTACCTCTTACAAACCCAGAGT	CTGGTGGTCCATATAACAAACAGC
STK38	GAGGAGACTGCTTGAGATGTGC	[6FAM]ATGTTCCCAGTCAACGCCT
TAF1B	[6FAM]ACCCAAATAAAAGCCCTCAAC	CTACTTAAAATTCCATTCCATGTCC
TUBGCP3	[6FAM]GCTGGACTTCAACGAGCATTAC	GTCTAGCAGTGCAACGAACATC

Table 20: Primers used for qPCR analysis of endogenous NMD targets and frameshift-bearing transcripts coding for InDel neoepitopes.

transcript	forward primer	reverse primer	remarks	reference
ATF3	GCCATTGGAGAGCTGTCTTC	GGGCCATCTGGAACATAAGA	endogenous NMD target	Bhuvanagiri *et al.*, 2014
ATF4	ATGTCCCCCTTCGACCA	CCATTTTCTCCAACATCCAATC	endogenous NMD target	Cheruiyot *et al.*, 2018
CKAP2	GAAACGAGGACAAGTTGCTTAAT	CGAGAACGTCTCACTGGTGT	FS-bearing target	
HPRT1	GACCAGTCAACAGGGGACAT	AACACTTCGTGGGGTCCTTTTC	housekeeping gene	Viegas *et al.*, 2007
NFAT5	CCTAATGCCCTGATGACTCCAC	TGAGATGTTTTCTAATGTTTGCTGA	FS-bearing target	
PSMC6	TAGAATTACCTCTTACAAACCCAGAGT	CTGGTGGTCCATATAACAAACAGC	FS-bearing target	
SC35A	CGTGCCTGAAACTGAAACCA	TTGCCAACTGAGGCAAAGC		Bhuvanagiri *et al.*, 2014
SC35C	GGCGTGTATTGGAGCAGATGTA	CTGCTACACAACTGCGCCTTTT	endogenous NMD target	Bhuvanagiri *et al.*, 2014
SC35D	CGGTGTCCTCTTAAGAAAATGATGTA	CTGCTACACAACTGCGCCTTTT	endogenous NMD target	Bhuvanagiri *et al.*, 2014
STK38	TGGAGCTCCTGGAGTTGAGG	TTGTGGCCACTGTTGGCTTA	FS-bearing target	
TUBGCP3	GCTGGACTTCAACGAGCATTAC	GTCTAGCAGTGCAACGAACATC	FS-bearing target	
UPP1	CCAGCCTTGTTTGGAGATGT	ACATGGCATAGCGGTCAGTT	endogenous NMD target	Cheruiyot *et al.*, 2018

2.9.2 Synthetic peptides

Table 21: Synthetic peptides used for sequence validation and *in vivo* immunizations. ac, N-terminal acetylation; cam, carbamidomethylation of cysteine.

gene	peptide sequence	type	used for	cell line
AXIN2	ac-GATAGTPAP	InDel neoepitope	synthetic peptide validation	HCT-116
CDC42EP5	PPRPAAAP	InDel neoepitope	synthetic peptide validation	HCT-116
CHMP7	QTDQMVFNTY	SNP neoepitope	synthetic peptide validation	HCT-116
CKAP2	SLMEQIPHL	InDel neoepitope	synthetic peptide validation, *in vivo* immunization	HCT-116
DBR1	TKSFPTRSG	InDel neoepitope	synthetic peptide validation	HCT-116
DYNC2LI1	MEVKVMEL	InDel neoepitope	synthetic peptide validation	HCT-116
EWSR1	AMDSRVAMV	InDel neoepitope	synthetic peptide validation	HCT-116
LMTK3	ac-VGGGFPPPPP	InDel neoepitope	synthetic peptide validation	HCT-116
LSM1	WSEEKMWSY	InDel neoepitope	synthetic peptide validation	HCT-116
MSL2	NTDYTGTGY	InDel neoepitope	synthetic peptide validation	HCT-116
NFAT5	KRSSTILRL	InDel neoepitope	synthetic peptide validation, *in vivo* immunization	HCT-116
NMT2	STDTSFGTHNRY	InDel neoepitope	synthetic peptide validation	HCT-116
NR1D1	YSDNSNDSF	SNP neoepitope	synthetic peptide validation	HCT-116
PCMT1	AAAPVVPQV	SNP neoepitope	synthetic peptide validation	HCT-116
PHLDA1	DEHISIVF	InDel neoepitope	synthetic peptide validation	HCT-116
PSMC6	REKHSWHEP	InDel neoepitope	synthetic peptide validation, *in vivo* immunization	HCT-116
RBBP7	EERVIDEEY	SNP neoepitope	synthetic peptide validation	HCT-116
RGP1	RLDPGEPKSY	SNP neoepitope	synthetic peptide validation	HCT-116
RPS2	AEAAELAEARP	InDel neoepitope	synthetic peptide validation	HCT-116
SIAH2	AEIDLGGHAP	InDel neoepitope	synthetic peptide validation	HCT-116
STK38	ISERDLLQY	InDel neoepitope	synthetic peptide validation	HCT-116
TGFBR2	RLSSCVPVA	InDel neoepitope	targeted MS2 approach	HCT-116
TUBGCP3	GVWEKPRRV	InDel neoepitope	synthetic peptide validation, *in vivo* immunization	HCT-116
ZNF703	WAAPPPPPP	InDel neoepitope	synthetic peptide validation	HCT-116
AGO2	ISSKWTSPK	InDel neoepitope	synthetic peptide validation	RKO
AGO2	NTWSSTLKH	InDel neoepitope	synthetic peptide validation	RKO
AURKA	TVYREPQKLSK	SNP neoepitope	synthetic peptide validation	RKO
BCAP29	RLSLPLSSY	InDel neoepitope	synthetic peptide validation	RKO
BCORL1	KTSLTIPSQK	InDel neoepitope	synthetic peptide validation	RKO
BTBD10	LLKEELTTPK	InDel neoepitope	synthetic peptide validation	RKO
BTBD10	RVLLKEELTTPK	InDel neoepitope	synthetic peptide validation	RKO
BTBD10	VLLKEELTTPK	InDel neoepitope	synthetic peptide validation	RKO
C16orf72	RTFSLGPLPK	InDel neoepitope	synthetic peptide validation	RKO
CARNMT1	TLYSTDVLK	InDel neoepitope	synthetic peptide validation	RKO
CEBPZ	KVNWKHLFK	InDel neoepitope	synthetic peptide validation	RKO
CLEC11A	RILPELLREK	InDel neoepitope	synthetic peptide validation	RKO
CLOCK	SLHSQMLVK	InDel neoepitope	synthetic peptide validation	RKO
CNN3	SVIGLQTGTNK	SNP neoepitope	synthetic peptide validation	RKO
CNOT11	LRGPVHRLP	InDel neoepitope	synthetic peptide validation	RKO
COQ3	VLYQKVLIH	InDel neoepitope	synthetic peptide validation	RKO
CSGALNACT2	HLFLVLLGK	InDel neoepitope	synthetic peptide validation	RKO

DCAF5	MMSKLSSMMLK	InDel neoepitope	synthetic peptide validation	RKO
DISP1	RIFFRKSIAM	InDel neoepitope	synthetic peptide validation	RKO
DNAJC16	ILNGGTQLTK	InDel neoepitope	synthetic peptide validation	RKO
DNAJC8	MMLTEHKRLLK	InDel neoepitope	synthetic peptide validation	RKO
DNTTIP2	KMKMSLVMK	InDel neoepitope	synthetic peptide validation	RKO
DSTN	VIFVSVQTK	InDel neoepitope	synthetic peptide validation	RKO
EIF5	SVLNVRILK	InDel neoepitope	synthetic peptide validation	RKO
FAM126B	HLYQPSIKLFK	InDel neoepitope	synthetic peptide validation	RKO
FBXO25	KVFIVKNGSMSIK	InDel neoepitope	synthetic peptide validation	RKO
GTPBP3	KLWPRPLRH	InDel neoepitope	synthetic peptide validation	RKO
HAUS3	KTLTGC(cam)LRALK	InDel neoepitope	synthetic peptide validation	RKO
HNRNPH1	RLLLNLNQK	InDel neoepitope	synthetic peptide validation	RKO
LUC7L3	KIYENSMRR	InDel neoepitope	synthetic peptide validation	RKO
MAFK	ATTSIITIVK	SNP neoepitope	synthetic peptide validation	RKO
MAPKAP3	VLRGGGGSPR	InDel neoepitope	synthetic peptide validation	RKO
MRPL15	RILPNFLKH	InDel neoepitope	synthetic peptide validation	RKO
MRPL3	GGMSIFLKK	InDel neoepitope	synthetic peptide validation	RKO
MRPL3	RVVHGGMSIFLKK	InDel neoepitope	synthetic peptide validation	RKO
MRPL3	SLSSWSLMK	InDel neoepitope	synthetic peptide validation	RKO
MRPL3	VVHGGMSIFLK	InDel neoepitope	synthetic peptide validation	RKO
MSANTD2	GLFTGGLGK	InDel neoepitope	synthetic peptide validation	RKO
MSL1	KLLSTLLLK	InDel neoepitope	synthetic peptide validation	RKO
MTHFSD	KVLASSSLKAK	InDel neoepitope	synthetic peptide validation	RKO
NAA35	IQTGWTLPK	InDel neoepitope	synthetic peptide validation	RKO
NABP1	ILYGLFRSAK	InDel neoepitope	synthetic peptide validation	RKO
NFX1	ALYGGRILK	InDel neoepitope	synthetic peptide validation	RKO
NPHP1	KTYSSIDINK	InDel neoepitope	synthetic peptide validation	RKO
PDZD8	SIYNNGHLLK	InDel neoepitope	synthetic peptide validation	RKO
PKMYT1	RTKVDPLFK	InDel neoepitope	synthetic peptide validation	RKO
PRIMPOL	KSYTVLIAQLK	InDel neoepitope	synthetic peptide validation	RKO
PRMT1	KRNDYMHAL	SNP neoepitope	synthetic peptide validation	RKO
PRRG1	RSFGAPTQK	InDel neoepitope	synthetic peptide validation	RKO
PRSS12	RISTTSWRQK	InDel neoepitope	synthetic peptide validation	RKO
RBFA	KVLPWVLTR	InDel neoepitope	synthetic peptide validation	RKO
RBPJ	VVLNKSLNR	InDel neoepitope	synthetic peptide validation	RKO
RHOG	VLNPTPIKH	SNP neoepitope	synthetic peptide validation	RKO
RPA2	RLQPTLFTK	InDel neoepitope	synthetic peptide validation	RKO
RSRC2	KTLQTQIAK	InDel neoepitope	synthetic peptide validation	RKO
SECISBP2	RMISKSLTVK	InDel neoepitope	synthetic peptide validation	RKO
SH3PXD2A	RLDPRMSTLQK	InDel neoepitope	synthetic peptide validation	RKO
SSBP3	EEFLGHSH	InDel neoepitope	synthetic peptide validation	RKO
SSBP3	RVLRGHSPR	InDel neoepitope	synthetic peptide validation	RKO
SUMO1	RIAVRFTSK	InDel neoepitope	synthetic peptide validation	RKO
TAF15	ac-SDMIHSEI	InDel neoepitope	synthetic peptide validation	RKO
TAF1B	KVSSIFFINK	InDel neoepitope	synthetic peptide validation	RKO
TCF19	GVMTGGSAWK	InDel neoepitope	synthetic peptide validation	RKO
TIFA	STAQFSPLK	InDel neoepitope	synthetic peptide validation	RKO

TRIP12	KQISHIVSQR	InDel neoepitope	synthetic peptide validation	RKO
UBE2Q1	GSCPHPLQH	InDel neoepitope	synthetic peptide validation	RKO
UBQLN2	RLLPRLSLK	InDel neoepitope	synthetic peptide validation	RKO
USP1	KLLNIEHLKLIK	InDel neoepitope	synthetic peptide validation	RKO
WDR75	KVYFIIIKK	InDel neoepitope	synthetic peptide validation	RKO
ZNF439	KLKKTVIVEK	InDel neoepitope	synthetic peptide validation	RKO
ZNF559	RVKGRISMK	InDel neoepitope	synthetic peptide validation	RKO
ZYX	ALWWPQSPK	InDel neoepitope	synthetic peptide validation	RKO
E7	YMLDLQPET	HPV16 viral epitope	*in vivo* immunization (positive control)	

3 Results

Several prerequisites are essential in order to establish an immunopeptidomics workflow for the reliable identification of tumor-specific neoepitopes originating from InDel mutations. First, a model system exhibiting a strong NMD phenotype that can be modulated using the pharmacological NMD inhibitor 5AZA is required. Second, databases containing protein sequences arising from genomic mutations have to be constructed which will be used for data analysis of mass spectrometry raw data. Third, a procedure for the enrichment and purification of HLA class I-presented peptides is needed. Finally, parameters for the mass spectrometry-based analysis of these peptides have to be optimized.

After the successful establishment of such an immunopeptidomics workflow, it can be utilized to explore the effect of NMD inhibition by 5AZA on the immunopeptidome.

3.1 Selection of a suitable model system

3.1.1 NMD competence of MSI CRC model cell lines

The NMD competence of the three MSI CRC model cell lines HCT-116, LoVo, and RKO was tested using a previously established dual-luciferase NMD reporter system (Boelz *et al.*, 2006). The NMD reporter consists of the human hemoglobin subunit beta (*HBB*) gene with or without an NMD-triggering nonsense-mutation at position 39 in the second exon fused in-frame to the 3' end of the *Renilla* luciferase gene (*Renilla-HBB* wild-type and *Renilla-HBB* NS39). The *Renilla* luciferase signals were normalized to the signal of a co-transfected firefly luciferase construct in order to account for differences in transfection efficiency.

The three MSI CRC model cell lines tested exhibit a high NMD efficiency demonstrated by the substantially and highly significantly lower *Renilla* luciferase signal in cells transfected with the *Renilla-HBB* NS39 construct compared to *Renilla* luciferase signal of cells transfected with the wild-type control construct (wt; Figure 5). Luciferase activity in cells transfected with the NS39 reporter construct was reduced to 5.5 ± 3.8 % in HCT-116 cells ($p \leq 0.0001$), 17.0 ± 7.0 % in LoVo cells ($p \leq 0.0001$), and 5.2 ± 0.6 % in RKO cells ($p \leq 0.0001$) of the activity in cells transfected with the wt reporter construct.

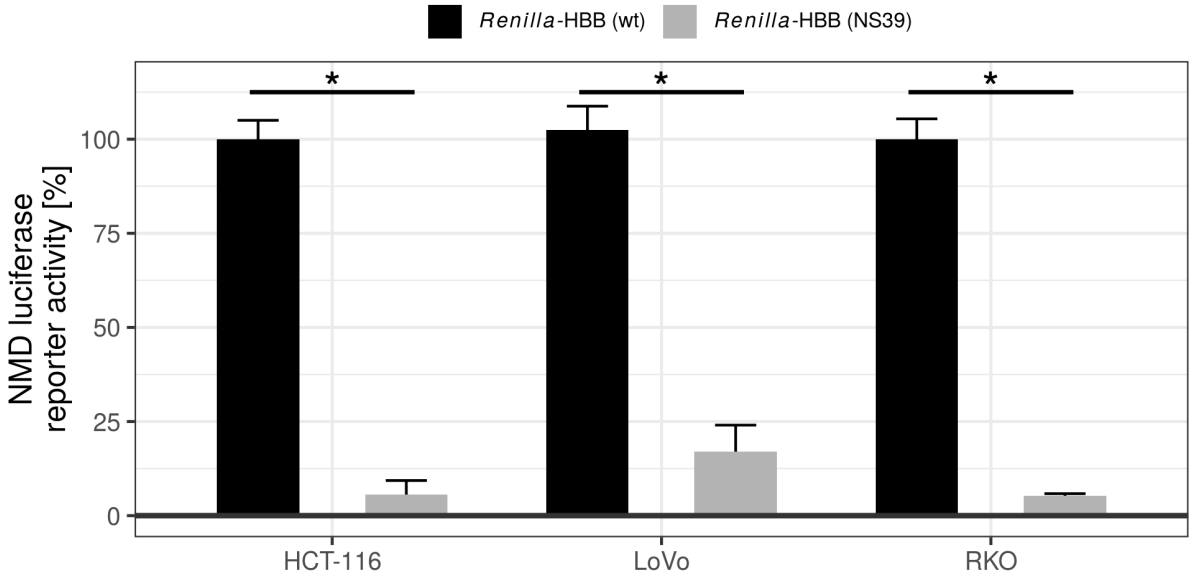


Figure 5: NMD is efficient in MSI CRC cell lines. NMD efficiency determined by significant downregulation of luciferase signal from *Renilla*-HBB (NS39) reporter construct compared to *Renilla*-HBB (wt) reporter construct. Bars represent mean ± SD of three experiments, * $p \leq 0.0001$ (two-sided, unpaired t-test).

3.1.2 Effect of 5AZA on cell proliferation of MSI CRC model cell lines

Cell proliferation assays were used to define conditions that restrict NMD efficiency while at the same time have minimal cytotoxic effects on the cells. First, cell proliferation was assessed for HCT-116, LoVo, and RKO cell lines for treatment with 0.01 µM-5 µM 5AZA for either 24 h, 48 h, or 72 h (Figure 6). Treatment with low 5AZA concentrations slightly increased the proliferation of HCT-116 cells and RKO cells. Treatment with 2.5 µM and 5 µM 5AZA for 24-48 h decreased cell proliferation of HCT-116 cells to about 60 % to 70 % compared to the untreated control. Treatment of HCT-116 cells for 72 h with 2.5 and 5 µM 5AZA led to a slight decrease in cell proliferation when compared to treatment with low 5AZA doses but not with the untreated control. For RKO cells, treatment with 2.5 µM and 5 µM decreased cell proliferation in all treatment durations (60-70 % compared to untreated control for 24 h treatment, 50 % for 48 h treatment, and 50-75 % for 72 h treatment). For LoVo cells, treatment with low 5AZA doses did not have any significant effect on cell proliferation. Higher 5AZA doses decreased proliferation in LoVo cells treated for 24 h (65-75 % compared to untreated control for 2.5 µM and 5 µM 5AZA, respectively) and 48 h (75 % compared to untreated control for 5 µM 5AZA). 5AZA treatment of LoVo cells for 72 h with any of the doses tested did not affect cell proliferation.

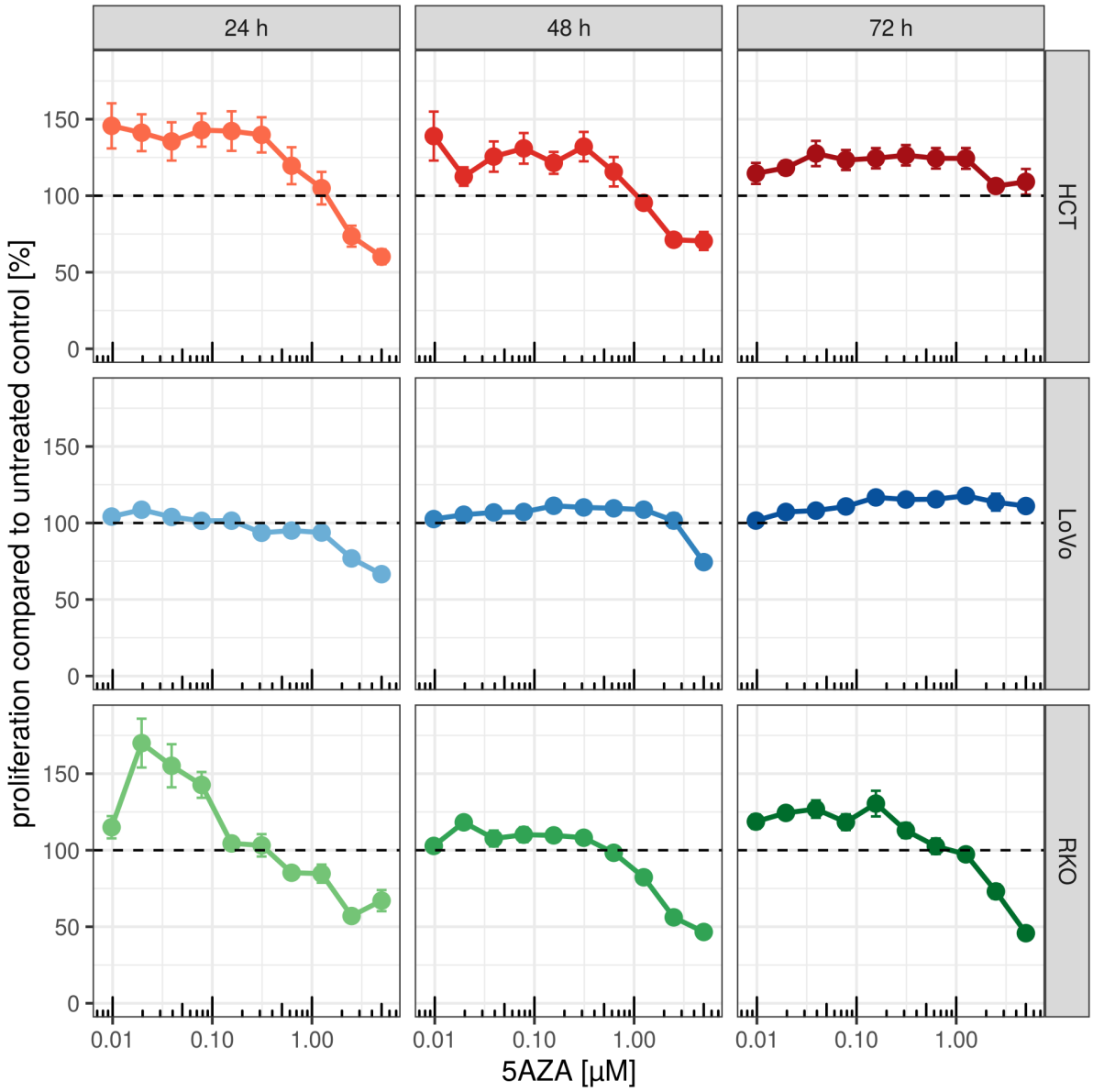


Figure 6: Effect of 5AZA treatment on proliferation of MSI CRC cell lines. MSI CRC cell lines were treated for 24, 48, or 72 h with 0.01-5 µM 5AZA. Data represent mean ± SEM of 3 experiments.

3.1.3 NMD inhibition by 5AZA in MSI CRC model cell lines

NMD inhibition by 5AZA in HCT-116, LoVo, and RKO cell lines was tested by analyzing the mRNA levels of the endogenous NMD target *SC35C*, *SC35D*, and *ATF3* mRNAs as well as the control isoform *SC35A*. The cells were treated for 24 h with either 0.39-3.12 µM 5AZA, 5 µg/ml CHX (positive control), or DMSO solvent control. For HCT-116, 5AZA treatment significantly stabilized *ATF3* (up to 8.9-fold with 3.12 µM 5AZA; $p \leq 0.0001$) and SC35C (2.3-fold with 3.12 µM 5AZA; $p \leq 0.0001$) mRNA levels (Figure 7; left panel). *SC35D* mRNA levels were slightly, but not significantly increased (1.6-fold with 3.12 µM 5AZA). In LoVo cells, mRNA levels of *SC35C* and *SC35D* were significantly stabilized after treatment with 1.56 µM 5AZA (37-fold and 50-fold, respectively, $p \leq 0.0001$; Figure 7, middle panel) and in RKO cells, treatment with 3.12 µM increased mRNA levels of *ATF3* (8.4-fold), *SC35C* (2.5-fold) and *SC35D* (2-fold; Figure 7, right panel). Treatment with the positive control CHX significantly increased the mRNA levels of the endogenous NMD targets in all cell lines except for *ATF3* mRNA levels in LoVo cells.

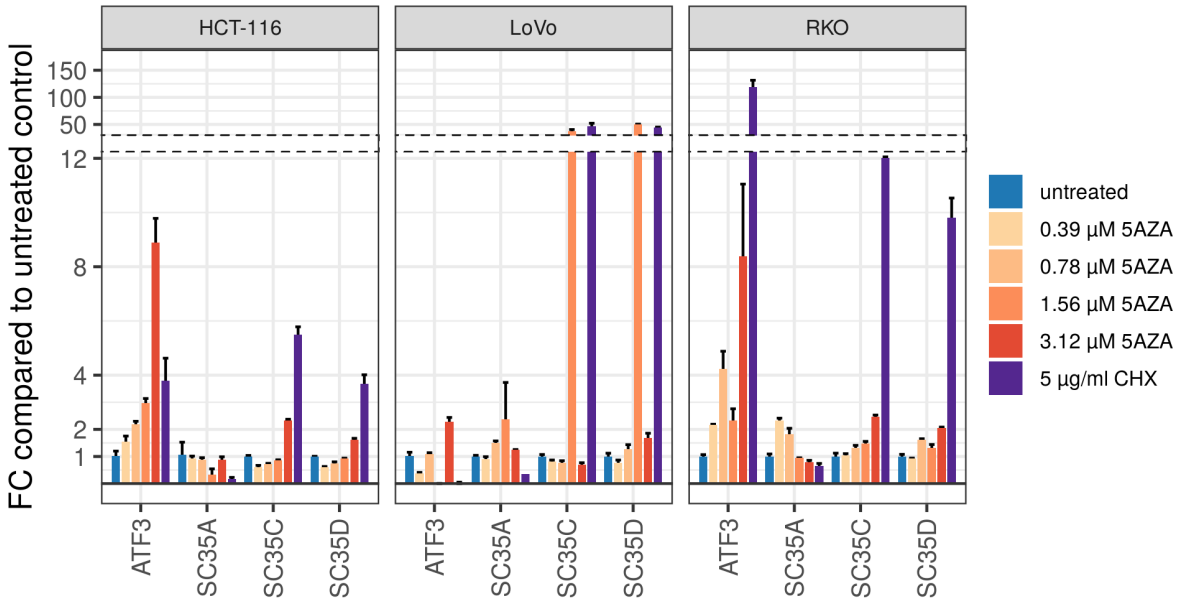


Figure 7: NMD can be inhibited by 5AZA in MSI CRC cell lines. Treatment with 0.39-3.12 µM 5AZA increases mRNA abundance of endogenous NMD targets ATF3, SC35C, and SC35D as determined by qPCR. Each bar represents the mean ± SD of three experiments.

NMD inhibition with 5 µM 5AZA in HCT-116 cells further increased mRNA levels of ATF3, SC35C, and SC35D when compared to treatment with 3 µM 5AZA (Figure 8). In line with previous reports, treatment with the chemically closely related 5-aza-2'-deoxycytidine (dAZA) did not affect the mRNA levels of the endogenous NMD targets tested in HCT-116 cells (Bhuvanagiri *et al.*, 2014) indicating that NMD inhibition by 5AZA is specific. Further experiments investigating the response to NMD inhibition at the level of the immunopeptidome were performed with 5 µM 5AZA for 24 h (described in Section 3.5.1).

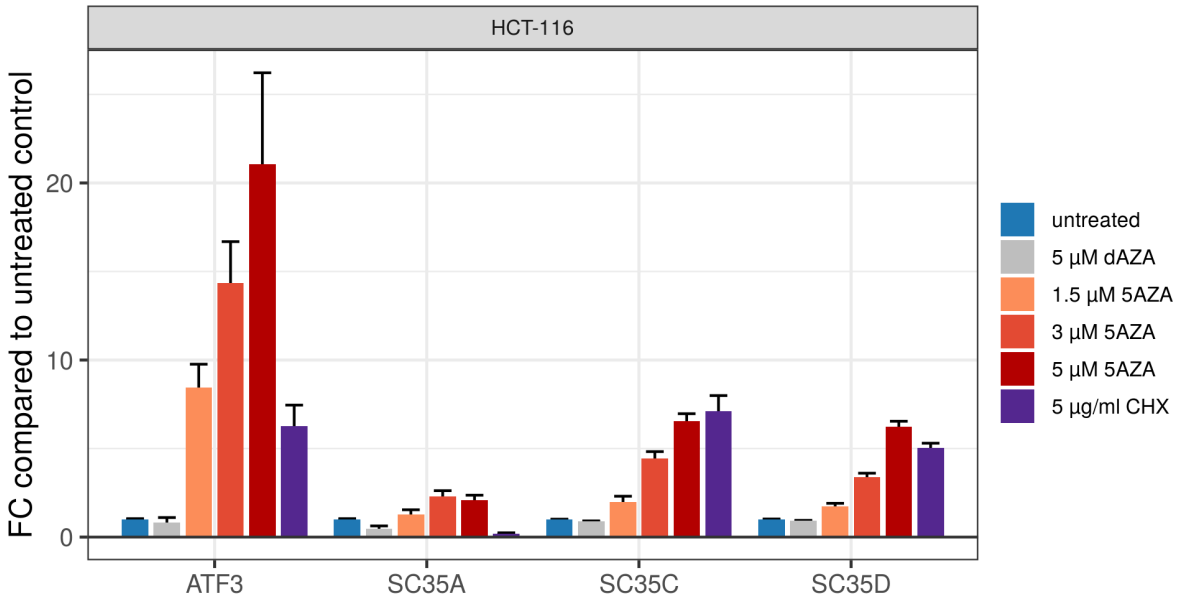


Figure 8: NMD is specifically inhibited by 5AZA in HCT-116 cells. Treatment with 1.5-5 µM 5AZA increases mRNA abundance of endogenous NMD targets ATF3, SC35C, and SC35D in a dose-dependent manner as determined by qPCR. Treatment with the chemically closely related 5-aza-2'deoxycytidine (dAZA) does not affect mRNA abundance. Data represent mean ± SD of three experiments.

3.2 Establishment of an immunoprecipitation-based purification method for HLA class I molecules

The samples used for the mass spectrometry-based analysis of HLA class I immunopeptidomes were prepared using an immunoprecipitation approach. For this purpose, an in-house produced monoclonal antibody against HLA class I molecules (clone W6/32) was chemically crosslinked to sepharose-Protein A beads. The beads were then used in both batch and high-throughput immunoprecipitations.

3.2.1 Quality control of antibody production and crosslinking procedure

Large amounts of W6/32 antibody are needed for both the batch and the high-throughput immunoprecipitation. Therefore, the W6/32 antibody was produced in-house utilizing HB-95 hybridoma cells which were cultivated in a small-scale bioreactor. The antibody was purified from cell culture supernatants using sepharose-Protein A beads. Purity of pooled elutions, as well as elution fractions, was analyzed on SDS-PAGE gels (Figure 9). The in-house produced antibody (lanes 2-9) showed a band pattern comparable to the commercially available W6/32 antibody (lane 1) with prominent bands at approximately 25 kDa corresponding to the IgG light chain and at approximately 55 kDa corresponding to the IgG heavy chain.

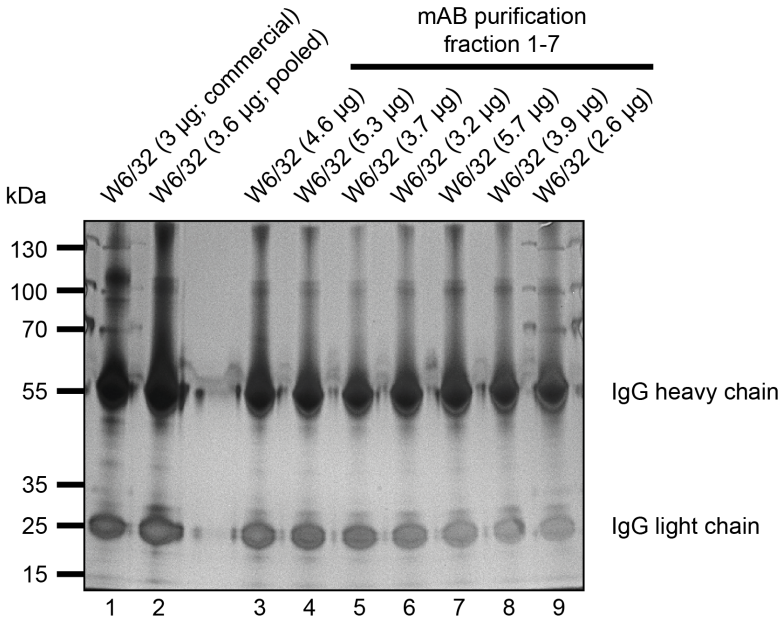


Figure 9: Quality control of in-house purification of the pan-HLA antibody W6/32. The purity of pooled elutions (lane 2) and elution fractions (lanes 3-9) from the antibody purification procedure was compared to the commercially available W6/32 antibody (lane 1) using SDS-PAGE and silver staining.

The specificity of the in-house produced W6/32 antibody was assessed by its ability to detect endogenous HLA class I molecules in membrane-enriched lysates from HCT-116, LoVo, and RKO cells and compared to the performance of a commercially available W6/32 antibody (Figure 10). Both the commercial and the in-house produced W6/32 antibody detect HLA class I molecules at the expected size of 42 kDa in lysates from HCT-116 and RKO cells proofing the functionality of the purified antibody

(Figure 10; lanes 1-6). In the membrane-enriched lysates of LoVo cells, neither the commercial nor the in-house produced antibody was able to detect HLA class I molecules (Figure 10; lanes 7-9) which can be explained by the complete loss of B2M expression due to a genomic deletion in LoVo cells (Browning *et al.*, 1996). Therefore, functional HLA class I molecules (consisting of an α chain and B2M) are not assembled in these cells which consequently lack cell surface expression of HLA class I molecules. For this reason, the LoVo cell line was excluded from further experiments.

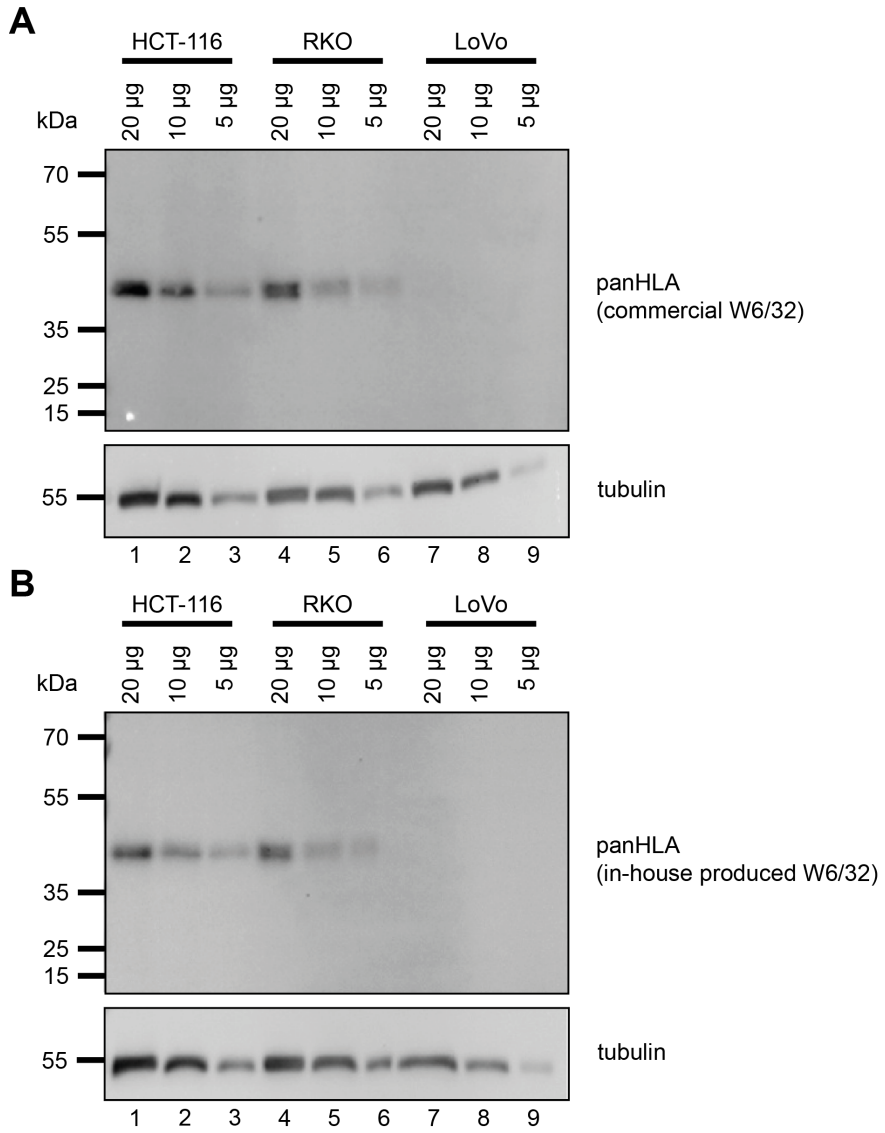


Figure 10: The in-house produced pan-HLA antibody W6/32 specifically detects HLA class I molecules. Specificity of in-house produced pan-HLA antibody W6/32 (panel B) was compared to the commercially available antibody (panel A) by SDS-PAGE and western blot analysis. Tubulin served as the loading control. LoVo cells lack HLA class I surface expression (lanes 7-9).

Besides a pure and functional antibody, efficient crosslinking of the W6/32 antibody to the beads used for the immunoprecipitation of HLA class I molecules is essential to provide high-quality samples for further analysis by mass spectrometry. Here, sepharose beads coated with Protein A were chosen for their strong binding affinity to IgG-type antibodies such as the W6/32 antibody. Additionally, in contrast to uncoated beads, Protein A-coated beads provide the correct orientation of the bound antibodies for

immunoprecipitation procedures as Protein A specifically recognizes the Fc region of antibodies. To prevent leakage of the W6/32 antibody during the sample preparation procedure, the antibodies were covalently linked using the chemical crosslinker DMP. The two imidoester groups of DMP react with primary amines of proteins (*i.e.* Protein A and IgG antibodies) to form a non-cleavable amidine bond. The efficiency of the crosslinking procedure was determined by analyzing samples taken at different steps during the crosslinking procedure. After the incubation of Sepharose-Protein A beads with the antibody solution (Figure 11B and A; lanes 1 and 2), antibody amounts in the supernatant were strongly reduced (< 10 %) due to the binding of IgG antibody to Protein A. Prior to crosslinking bound antibodies are still removable by boiling the beads in SDS sample buffer (Figure 11A; lane 4), while after the crosslinking and quenching reaction only trace amounts of W6/32 antibody are released from the beads (lane 7).

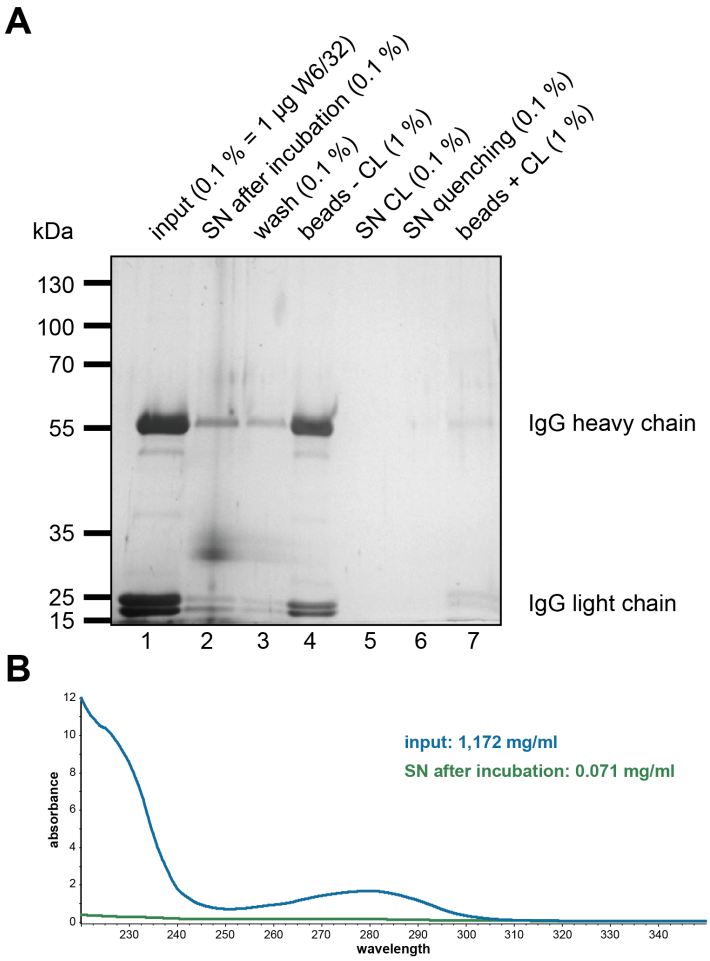


Figure 11: Quality control of chemical crosslinking of W6/32 antibody to Sepharose-Protein A beads. Samples were taken during the different steps of the chemical crosslinking procedure and analyzed by SDS-PAGE and silver staining (panel A). Complete antibody depletion from input by Sepharose-Protein A beads was verified by determining the IgG concentration of inputs before and after incubation (panel B).

3.2.2 Immunoprecipitation of HLA class I:peptide complexes

Immunoprecipitation of HLA class I:peptide complexes was first established in a batch format with the HCT-116 cell line similar to a previously published protocol (Bassani-Sternberg *et al.*, 2015; Bassani-

Sternberg *et al.*, 2016). During the establishment process, various parameters such as the amount of starting material, beads:antibody ratio, beads:sample ratio, washing conditions, as well as number and composition of elutions were tested as exemplified below by the evaluation of the effects of pre-elutions.

After immunoprecipitation of HLA class I:peptide complexes, the bound peptides are separated from the much larger HLA class I molecules based on their hydrophobicity using C_{18} cartridges. Here, a major concern is the overloading of these cartridges by co-eluting antibodies resulting in clogging of the cartridges and peptide loss. For this reason, one of the parameters tested for the immunoprecipitation of HLA class I:peptide complexes was the pre-elution of the W6/32 beads (Figure 12). Comparison of samples taken from an immunoprecipitation experiment performed either with freshly coupled or pre-eluted beads showed that the flowthrough after lysate incubation with pre-eluted W6/32 beads (Figure 12; lane 5) contained slightly more remaining HLA class I:peptide complexes and B2M than the flowthrough depleted with freshly coupled W6/32 beads (lane 4). Both the elutions from freshly coupled (lane 10) and pre-eluted W6/32 beads (lane 11) showed enrichment for HLA class I:peptide complexes albeit slightly stronger for the elution of freshly coupled beads. The specificity of the immunoprecipitation was shown by probing for the control protein tubulin which was only found in the input and flowthrough samples as well as in the sample of the first washing step with buffer A (lanes 3-7). B2M, which is a part of the HLA class I:peptide complex was enriched in both elution samples (lanes 10 and 11) but not completely depleted in the flowthrough samples (lanes 4 and 5). As B2M has additional biological functions, the B2M proportion remaining in the flowthrough probably corresponds to B2M that is not in complex with the HLA α chain and thus not recognized by the W6/32 antibody (Parham *et al.*, 1979).

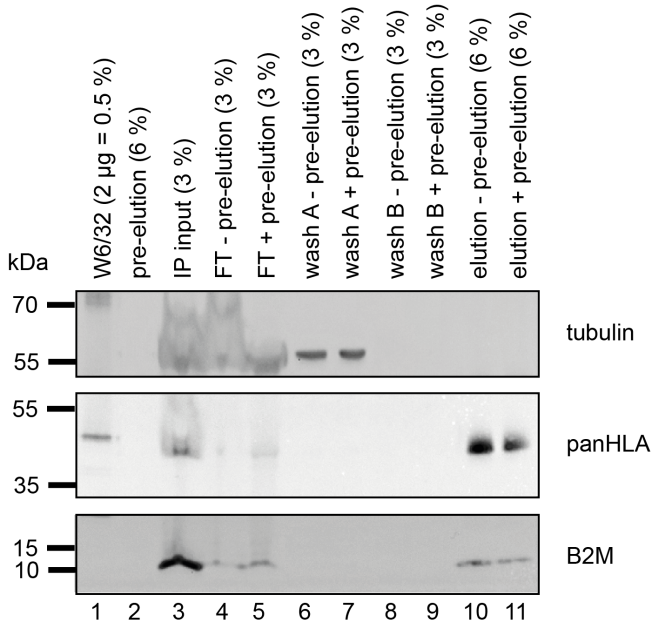


Figure 12: Selective purification of HLA class I molecules by batch immunoprecipitation. Western blot analysis of HLA class I molecules purified by batch immunoprecipitation from HCT-116 cells using the W6/32 antibody. Pre-elution of beads slightly reduced IP yield. FT, flowthrough; B2M, β-2-microglobulin.

To further evaluate the amount of W6/32 antibody eluted during the immunoprecipitation procedure, membranes were additionally probed with an anti-mouse antibody recognizing the W6/32 antibody eluted in the different steps (Figure 13). Here, W6/32 antibody was detected in the pre-elution sample (lane 2) and a similar amount in the elution sample from freshly coupled beads (lane 10). However, in comparison to 2 µg of W6/32 antibody (corresponding to 0.5 % of the total antibody amount used for one immunoprecipitation; lane 1), the amounts of eluted W6/32 antibody are negligible and probably do not interfere with further processing of the samples.

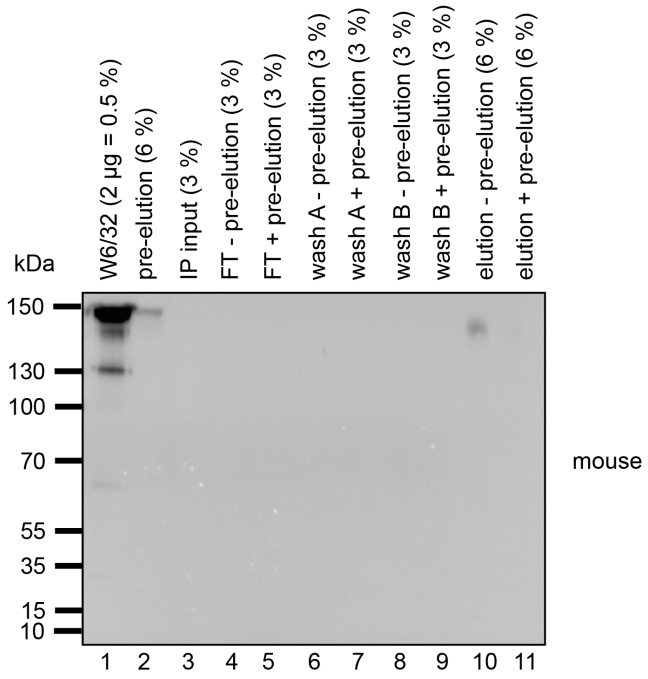


Figure 13: Pre-elution of beads reduces the amount of co-eluted W6/32 antibody during immunoprecipitation. Western blot analysis of samples taken during batch immunoprecipitation procedure. Membrane was incubated with anti-mouse antibody to evaluate the amount of co-eluted W6/32 antibody. FT, flowthrough.

After the successful establishment of HLA class I:peptide complex immunoprecipitation in batch format, an updated high-throughput version of the protocol was published and adopted (Chong *et al.*, 2018). The high-throughput method utilizes a positive pressure microprocessor, thereby reducing both sample handling and peptide loss as well as the time needed to complete the sample preparation protocol. Moreover, it allows the processing of up to 96 samples in parallel thus reducing variability in sample preparation and increasing reproducibility which is a prerequisite for accurate label-free quantification to assess the effect of NMD inhibition on the immunopeptidome.

The parameters established for the batch immunoprecipitation were transferred to the high-throughput method. Different amounts of input material ($1x10^7$-$1x10^8$ cells) were evaluated during the testing of the high-throughput method using the HCT-116 cell line (Figure 14). In elutions from input samples generated from $5x10^7$ and $1x10^8$ cells, respectively, the HLA molecules were detectable (indicated by an arrow; Figure 14A, lanes 6 and 7), while HLA molecules were not detectable in the elution from input samples generated from only $1x10^7$ cells (lane 5). However, B2M was proportionally detectable in all

three elution samples, suggesting that the high-throughput immunoprecipitation can be performed with as little input material as $1x10^7$ cells. The additional bands in the elution samples belong to the co-eluted W6/32 antibody (indicated by an asterisk; Figure 14B). The specificity of the immunoprecipitation was controlled by staining for tubulin which is only found in the input (Figure 14A, lanes 1 and 3) and flowthrough samples (lanes 2 and 4), but not in the elutions. Of note, the signals obtained from HLA molecules for elution samples from the high-throughput method seem to be weaker than from elutions of the batch procedure. This might be explained by the fact that the high-throughput samples are subjected to further steps before being analyzed by immunoblotting. These steps include the separation of HLA molecules and bound peptides at low pH. During this step also the HLA α chain and B2M might be separated causing the HLA molecules to partially lose their three-dimensional structure. However, this three-dimensional structure is essential for the recognition of HLA molecules by the W6/32 antibody used in the immunoprecipitation procedure and for immunoblotting. As this study focuses on the investigation of immunopeptidomes of established MSI CRC cell lines, $1x10^8$ cells were used in all further immunoprecipitations. Downscaling of the input material is not necessary, as these cell lines can be cultivated in virtually unlimited quantities. Moreover, it has been shown that downscaling of input material hampers the identification of low abundant peptides, which might include mutation-derived neoepitopes (Chong *et al.*, 2018).

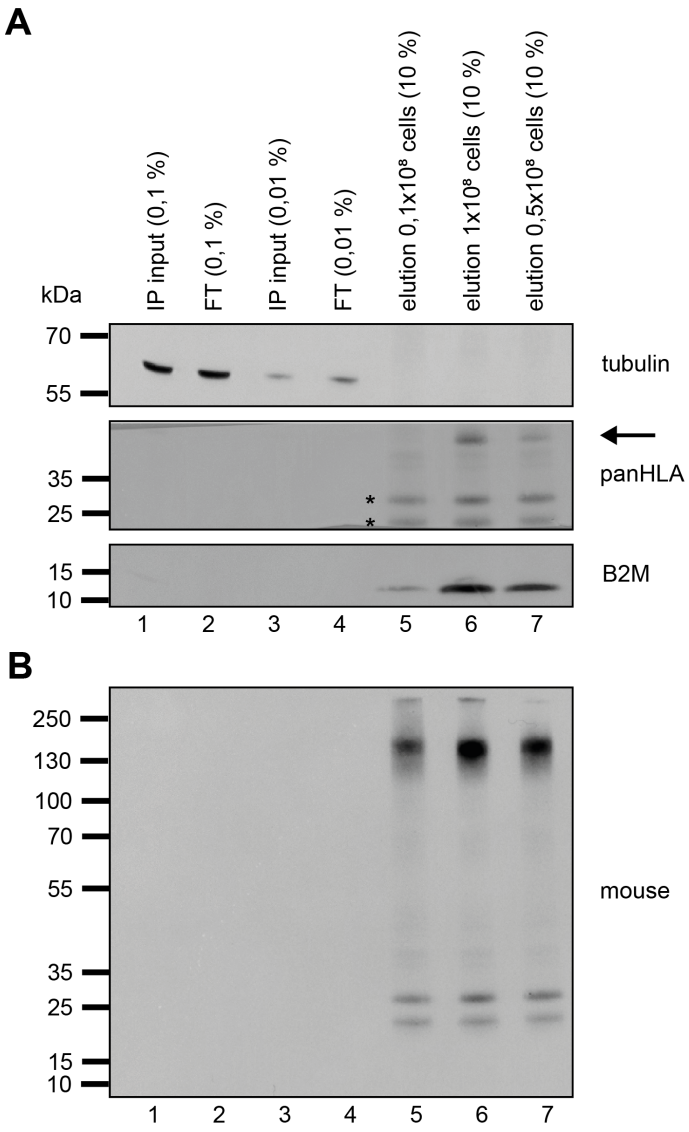


Figure 14: Selective purification of HLA class I molecules by high-throughput immunoprecipitation. Western blot analysis of HLA class I molecules purified by high-throughput immunoprecipitation from 0.1-1x10^8 HCT-116 cells with W6/32 antibody. Arrow in panel A indicates expected band size of HLA class I molecules. Asterisks in panel A indicate background bands of co-eluted W6/32 antibody. Panel B shows membrane incubated with anti-mouse antibody only. FT, flowthrough; B2M, β-2-microglobulin.

3.3 InDel and SNP database construction and *in silico* analysis

In order to identify mutation-derived neoepitopes their amino acid sequences have to be stored in databases that can be matched against MS raw data. Here, different databases were built and HLA class I binding predictions were used to analyze the potential neoepitopes.

3.3.1 InDel neoepitope databases based on public sequencing data

InDel neoepitope databases based on publicly available data were built using sequencing data for the HCT-116 and RKO cell line deposited in the Cancer Cell Line Encyclopedia (CCLE) and the cancer cell line project of the Catalogue Of Somatic Mutations In Cancer (COSMIC). A detailed description for the database construction using in-house developed R scripts can be found in section 2.6.

For the HCT-116 cell line, the CCLE database contains 527 protein-coding InDel mutations in 498 distinct genes, while the COSMIC database contains 705 protein-coding InDel mutations in 645 distinct genes. Of these, 349 mutations are found in both databases. The resulting frameshifted protein sequences have a median length of 31 aa. Of note, the maximum length of frameshift sequences generated is 387 and 393 aa for CCLE and COSMIC, respectively (Figure 15A). The CCLE-based InDel neoepitope database for HCT-116 contains a total of 117525 unique peptides with a length between eight and 15 aa (14680 nonamers). Of these, 4848 peptides (1612 nonamers) are predicted to bind at least one of the six HLA alleles expressed by HCT-116 cells (Figure 15C). The COSMIC-based InDel neoepitope database for HCT-116 contains a total of 166103 unique peptides with a length between eight and 15 aa (20738 nonamers). Of these, 6417 peptides (2161 nonamers) are predicted to bind at least one of the six HLA alleles expressed by HCT-116 cells (Figure 15C). For the RKO cell line, the CCLE database contains 235 protein-coding InDel mutations in 223 distinct genes, while the COSMIC database contains 1019 protein-coding InDel mutations in 920 distinct genes. Of these, 165 mutations are found in both databases. The resulting frameshifted protein sequences have a median length of 31 aa. Of note, the maximum length of frameshift sequences generated is 177 and 520 aa for CCLE and COSMIC, respectively (Figure 15B). The CCLE-based InDel neoepitope database for RKO contains a total of 44992 unique peptides with a length between eight and 15 aa (5619 nonamers). Of these, 2059 peptides (675 nonamers) are predicted to bind at least one of the four HLA alleles expressed by RKO cells (Figure 15D). The COSMIC-based InDel neoepitope database for RKO contains a total of 219719 unique peptides with a length between eight and 15 aa (27430 nonamers). Of these, 8908 peptides (3163 nonamers) are predicted to bind at least one of the four HLA alleles expressed by RKO cells (Figure 15D).

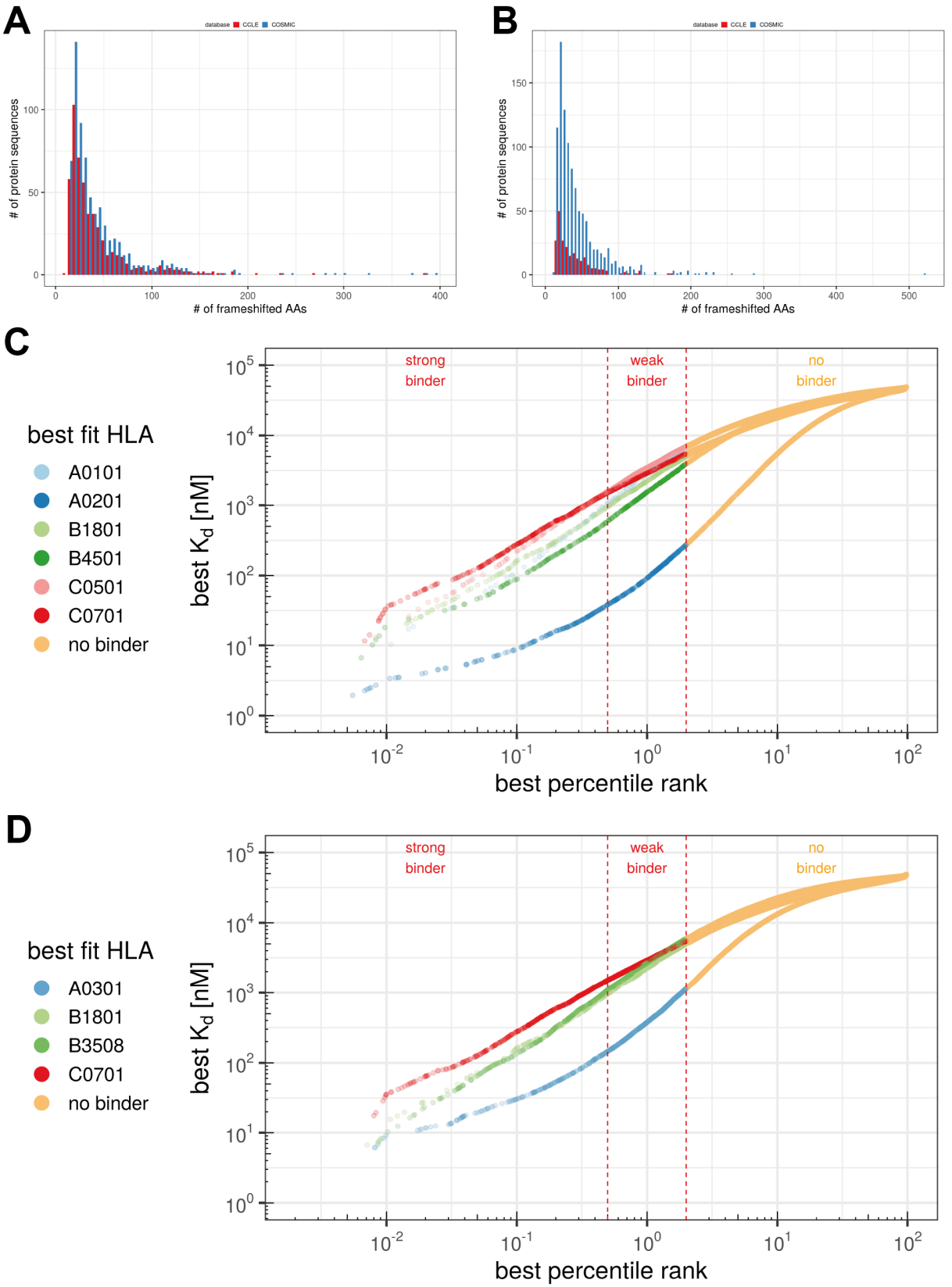


Figure 15: Analysis of frameshifted protein sequences in HCT-116 and RKO cell lines based on CCLE and COSMIC sequencing data. Length distribution of frameshifted protein sequences in HCT-116 (panel A) and RKO (panel B). Frameshifted sequences originating from InDel mutations annotated in CCLE are colored in red, frameshifted sequences originating from InDel mutations annotated in COSMIC are colored in blue. Binding prediction of all potential nonamers resulting from InDel mutations annotated in COSMIC and CLLE databases to HLA alleles expressed by HCT-116 cells (panel C) and RKO cells (panel D). Threshold in panels C and D for strong binders is top 0.5 % ranked, for weak binders top 2 % ranked.

3.3.2 InDel neoepitope database based on recurrent frameshift mutations in MSI CRC

During database generation from publicly available sequencing data and subsequent analysis, it was noticed that the constructed databases lack some of the most commonly found InDel mutations and thus the corresponding frameshifted amino acids sequences, for example, a one-base deletion in AIM2 (reported in Woerner *et al.*, 2007) or a one-base deletion in CASP5 (reported in Woerner *et al.*, 2001). To minimize this potential limitation, a database containing frameshifted amino acid sequences based on recurrent InDel mutations in MSI CRC was constructed (Ballhausen *et al.*, 2020). The so-called ReFrame database contains 519 distinct protein sequences from InDel mutations in 252 different genes. The frameshifted aa sequences have a median length of 20 amino acids with a maximum length of 196 aa (Figure 16A). The ReFrame database contains a total of 61879 unique peptides with a length of eight to 15 aa (7735 nonamers). Of these, 3036 peptides (986 nonamers) are predicted to bind at least one of the six HLA alleles expressed by HCT-116 cells (Figure 16B) and 3125 peptides (1064 nonamers) are predicted to bind to at least one of the four HLA alleles expressed by RKO cells (Figure 16C).

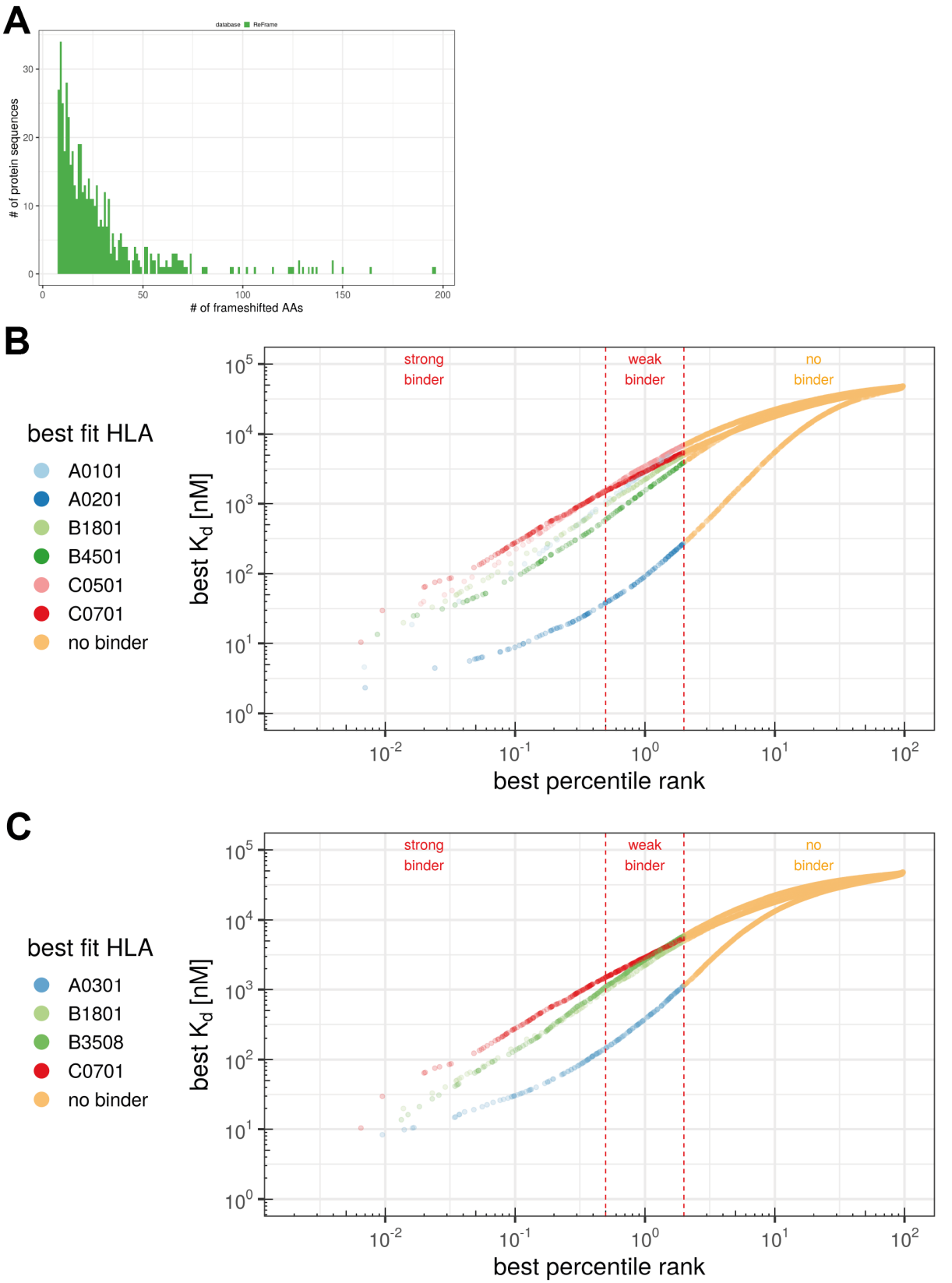


Figure 16: Analysis of frameshifted protein sequences in the ReFrame database. Length distribution of frameshifted protein sequences in the ReFrame database (panel A). Binding prediction of all potential nonamers resulting from frameshifted protein sequences in the ReFrame database to HLA alleles expressed by HCT-116 cells (panel B) and RKO cells (panel C). Threshold in panels C and D for strong binders is top 0.5 % ranked, for weak binders top 2 % ranked.

3.3.3 Mononucleotide repeat neoepitope database

As previously described, sequencing-based databases might be incomplete since variant calling of short InDel mutations remains challenging to date (Treangen and Salzberg, 2011; O'Rawe *et al.*, 2013; Narzisi and Schatz, 2015; Vanderwalde *et al.*, 2018). To this end, a database was constructed which contains the protein sequences resulting from all one-base and two-base InDel mutations of coding microsatellites with a length of six or more nt. The threshold of six nucleotides was chosen because of the increased probability of an InDel mutation starting from this length (Woerner *et al.*, 2010). The MNR database contains 149,907 distinct protein sequences originating from InDel mutations in mononucleotide repeat sequences in 12,692 distinct genes. The median length of the frameshifted protein sequences is 31 aa (Figure 17A). Due to the large size of the database, only nonamers were further analyzed as they represent the preferred binding length of HLA class I molecules. The mononucleotide repeat (MNR) neoepitope database contains 2,141,214 distinct nonamers, nearly twice as many as the entire human proteome (roughly 1,200,000 potential nonamers; Bassani-Sternberg *et al.*, 2015). Of these, 233,199 peptides are predicted to bind at least one of the six HLA alleles expressed by HCT-116 cells (Figure 17B) and 243,909 peptides are predicted to bind to at least one of the four HLA alleles expressed by RKO cells (Figure 17C).

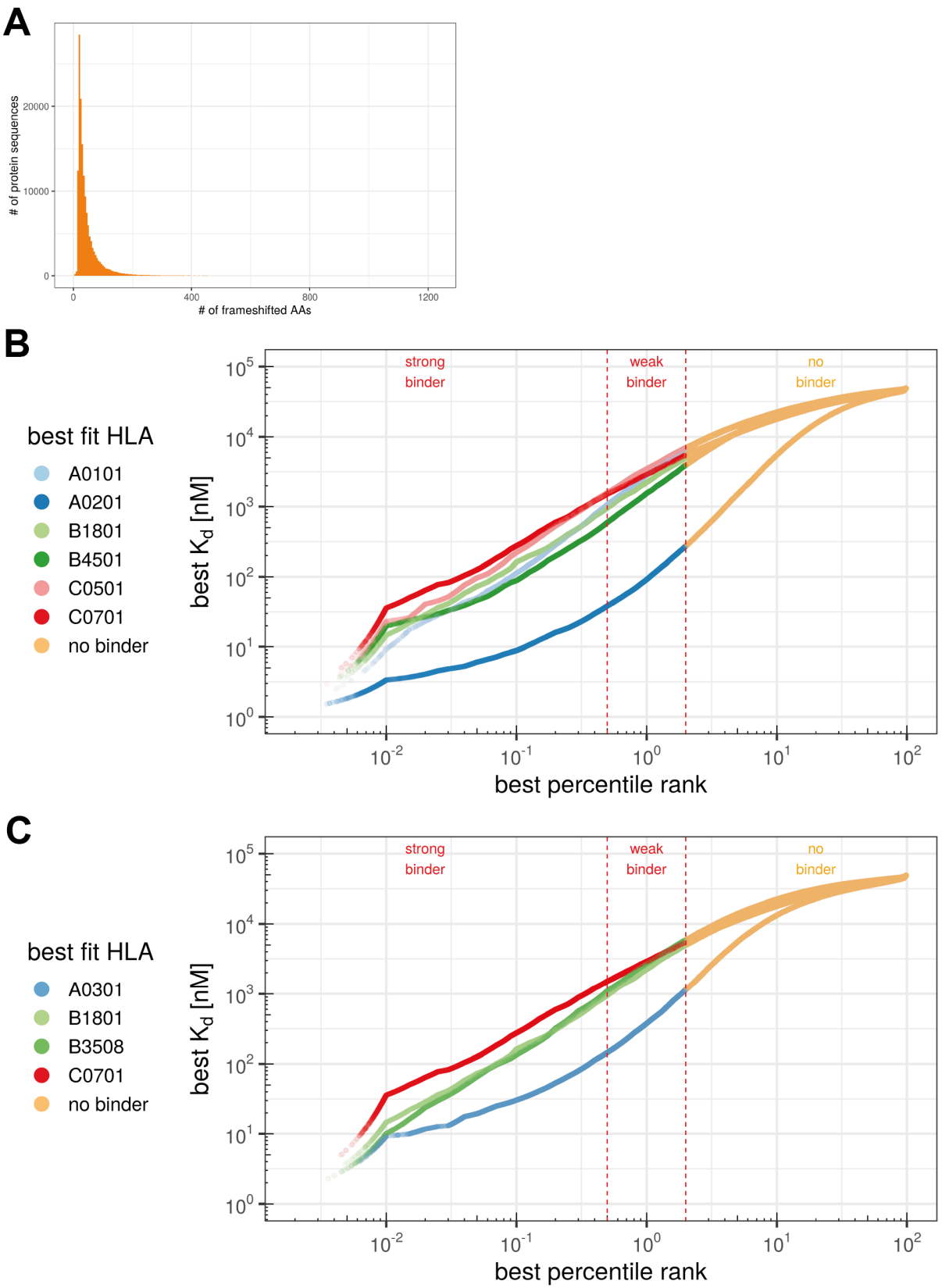


Figure 17: Analysis of frameshifted protein sequences in the MNR database. Length distribution of frameshifted protein sequences in the MNR database (panel A). Binding prediction of all potential nonamers resulting from frameshifted protein sequences in the MNR database to HLA alleles expressed by HCT-116 cells (panel B) and RKO cells (panel C). Threshold in panels C and D for strong binders is top 0.5 % ranked, for weak binders top 2 % ranked.

3.3.4 SNP neoepitope databases

The SNP neoepitope databases were constructed using previously published data stored in the TRON Cell Line Portal (TCLP) database. The database contains peptide sequences of eight to eleven amino acids in length due to single nucleotide polymorphisms (SNPs) leading to the exchange of single amino acids. For the HCT-116 cell line, the TCLP database contains 1260 neoepitope sequences (305 nonamers) originating from 935 different genes. Of these, 773 peptides (263 nonamers) are predicted to bind at least one of the six HLA alleles expressed by HCT-116 cells (Figure 18A). For the RKO cell line, the TCLP database contains 1504 neoepitope sequences (378 nonamers) originating from 1099 different genes. Of these, 864 peptides (329 nonamers) are predicted to bind at least one of the four HLA alleles expressed by RKO cells (Figure 18B). Of note, not all peptides reported in the TCLP database are predicted to bind one of the expressed HLA alleles since the TCLP database lists all SNP neoepitopes with stronger binding affinity than the corresponding wild-type epitope regardless of whether they are classified as binders or not.

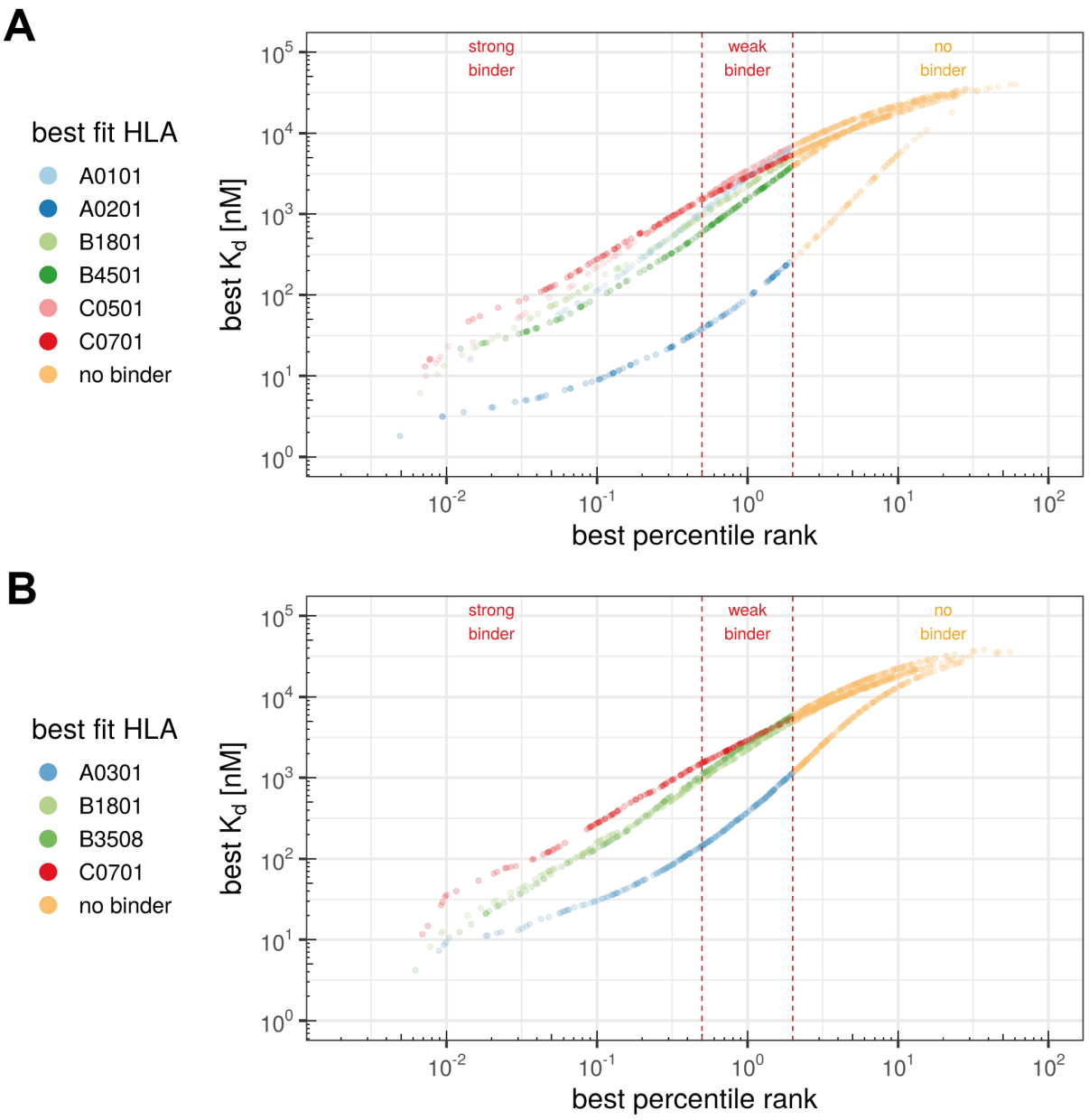


Figure 18: Analysis of SNP neoepitopes for HCT-116 and RKO cell lines based on TRON Cell Line Portal data. Binding prediction of all potential SNP nepepitopes proposed in the TRON Cell Line Portal to HLA alleles expressed by HCT-116 cells (panel A) and RKO cells (panel B). Threshold in panels A and B for strong binders is top 0.5 % ranked, for weak binders top 2 % ranked.

3.4 Establishment of a mass spectrometry-based workflow for immunopeptidome analysis

A mass spectrometry-based workflow was established with the goal to provide the most comprehensive view of HLA class I-presented peptides in MSI CRC model cell lines. To this end, different search engines were tested for their suitability of analyzing challenging immunopeptidomics datasets, and different MS parameters were optimized to provide both reliable identification and quantification data.

3.4.1 Comparison of different search engines for immunopeptidomics

The performance of two search engines on immunopeptidomics raw data was compared: Andromeda, the freely available search engine integrated into the MaxQuant computational proteomics platform (Cox and Mann, 2008; Cox *et al.*, 2011), and the proprietary PEAKS software, which utilizes *de novo* sequencing-assisted database search for peptide identification (Ma *et al.*, 2003; Zhang *et al.*, 2012). First, a previously published immunopeptidomics dataset of the HCT-116 cell line (Bassani-Sternberg *et al.*, 2015) was re-analyzed and the results were compared to the published data (Figure 19). The original report identified 4,182 unique peptides for HCT-116 using MaxQuant (version 1.3.10.15). Re-analysis of the mass spectrometry raw data deposited at the ProteomeXchange Consortium using the Andromeda search engine of a newer MaxQuant version (1.5.8.3), identified 3,228 distinct peptides. Of these, 2,653 peptides were identified in the original report while 575 peptides were only identified using the newer MaxQuant software. 1529 peptides reported in the original publication were not identified using the newer MaxQuant software version. Of note, other publications by Bassani-Sternberg and colleagues used the MaxQuant version 1.5.3.2 for the analysis of immunopeptidomics data. This version was optimized for the analysis of immunopeptidomics datasets but is not publicly available (Bassani-Sternberg, personal communication). Using the *de novo* sequencing-assisted approach implemented in PEAKS, a total of 5,548 unique peptides were identified (32 % more than in the original publication). Of these, 2,682 peptides were identified only using PEAKS while 1526 peptides found in the original report were missed.

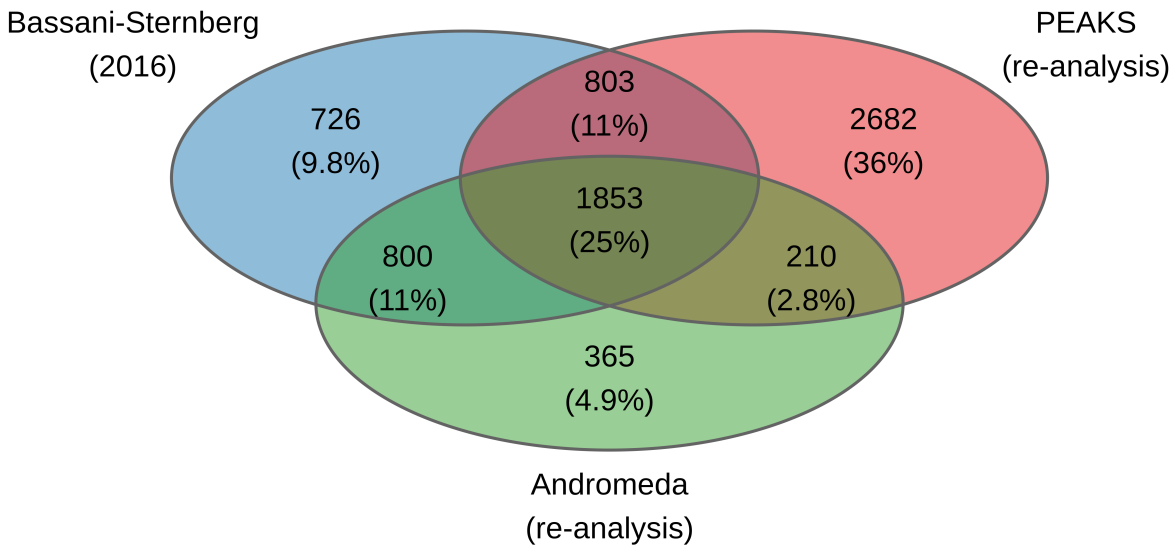


Figure 19: Comparison of different search engines using previously published data. A previously published dataset for the HCT-116 cell line (Bassani-Sternberg et al., 2015) was re-analyzed using the Andromeda search engine (green) and the PEAKS software (red) and identified peptides were compared to the published data (blue).

Next, the two search engines were compared using a test dataset recorded during this study (Figure 20). Again, PEAKS outperformed Andromeda regarding the total number of peptide identifications (3,819 versus 2,048) and only 288 peptides (7 %) were only found when using Andromeda as search engine which is in line with a previous report comparing different MS search engines for immunopeptidomics (Bichmann *et al.*, 2019). An additional benefit of PEAKS is the reporting of so-called "*de novo* only" peptides, which are peptides with a high-scoring *de novo* sequencing result but no database match potentially belonging to modified or novel peptides such as neoepitopes. The "*de novo* only" peptides and the corresponding spectra can be easily matched against custom databases in a subsequent analysis step. Moreover, it was found during search engine testing that analysis using PEAKS is considerably faster when using large databases such as the MNR neoepitope database albeit requiring more user input. Although combining the results of several search engines would potentially further increase both the number of and confidence in peptide identifications (Bichmann *et al.*, 2019), subsequent data analysis was only performed using PEAKS since its *de novo* sequencing capability provides a suitable tool for the identification of InDel mutation-derived neoepitopes.

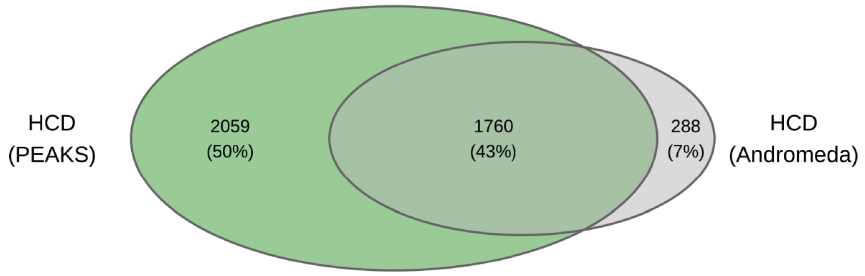


Figure 20: Comparison of search engines using a test dataset. MS raw data was recorded during parameter optimization using HCD fragmentation. Overlap of peptides identified using the PEAKS software (green) and the Andromeda search engine (grey).

3.4.2 Mass spectrometry method optimization

3.4.2.1 Single and dual fragmentation methods for mass spectrometry

The identification of mutation-derived neoepitopes requires an in-depth analysis of peptides presented by HLA class I molecules. In order to gain the most comprehensive view of the immunopeptidome, different single fragmentation (HCD and CID), as well as dual fragmentation (EThcD and ETciD) methods, were assessed and optimized. To this end, the high-throughput immunoprecipitation method was used to generate a peptide pool from $1.2x10^9$ HCT-116 cells thus allowing up to 24 injections of the same sample. First, technical duplicates were measured using either HCD, CID, EThcD, or ETciD fragmentation and the number of peptide identifications was assessed (Figure 21A). In total, 9,284 distinct peptides were identified. Of these, the majority was identified with the single fragmentation methods, namely 7,756 with HCD and 7,809 with CID. 6,136 peptides were identified with both methods, while 1,066 peptides were identified only by HCD and 1,156 peptides only using CID. In samples analyzed using EThcD and ETciD, considerably fewer peptides were identified (3,293 and 2,926 with EThcD and ETciD, respectively). Of these, the majority was also identified with one of the single fragmentation methods (3,280 peptides) while 238 peptides were identified only with one or both of the dual fragmentation methods. For samples analyzed using single fragmentation, the average peptide identification scores were slightly higher for CID compared to HCD, while for dual fragmentation the average peptide identification score was slightly higher in samples analyzed using EThcD (Figure 21B).

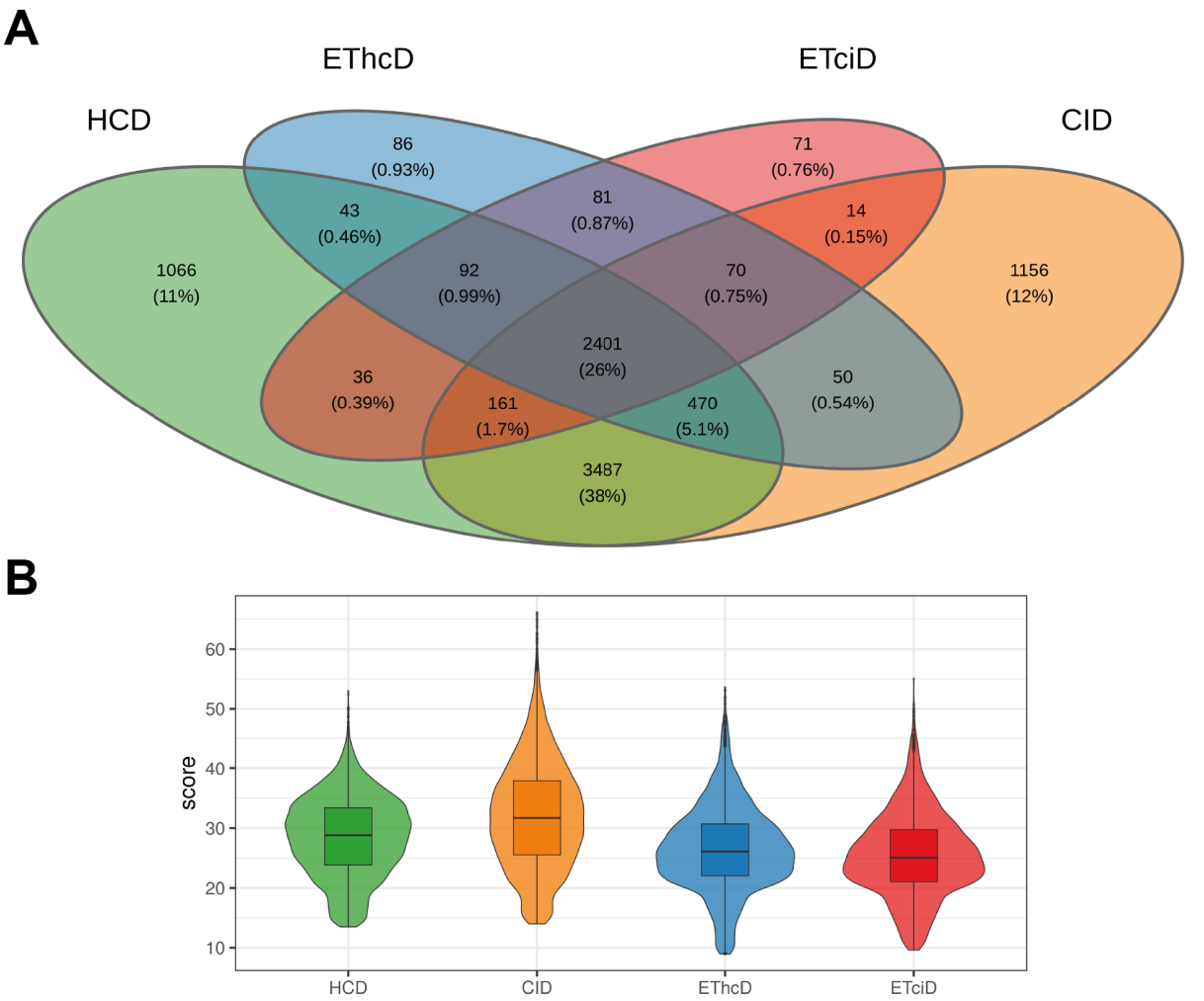


Figure 21: Comparison of single fragmentation and dual fragmentation methods for immunopeptidomics. Overlap of peptides identified using HCD (green), EThcD (blue), ETciD (red), and CID (orange) fragmentation (panel A). Score distribution of peptides identified using HCD, CID, EThcD, and ETciD fragmentation (panel B).

In addition to being analyzed by PEAKS, the data obtained using dual fragmentation methods were also processed using Andromeda. Again, analysis with PEAKS led to substantially more peptide identifications, and most peptides identified with Andromeda were also found using PEAKS (Figure 22).

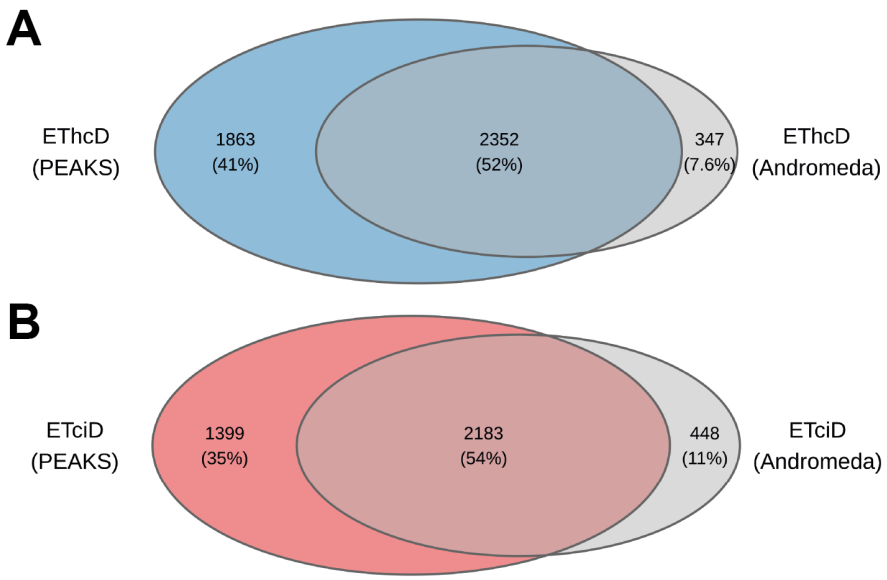


Figure 22: Comparison of search engines for data recorded using dual fragmentation. MS raw data was recorded during parameter optimization using either EThcD or ETciD fragmentation. Overlap of peptides identified using the PEAKS software (blue) and the Andromeda search engine (grey) with EThcD fragmentation (panel A). Overlap of peptides identified using the PEAKS software (red) and the Andromeda search engine (grey) with ETciD fragmentation (panel B).

The technical duplicates of the different fragmentation methods tested showed a high degree of reproducibility in terms of peptide identifications (Figure 23). In the samples analyzed using HCD, 7,756 unique peptides were identified in total with 5,879 peptides (76 %) found in both technical replicates (Figure 23A). In the samples analyzed using CID, 7,809 distinct peptides were identified in total with 5,834 peptides (75 %) found in both technical replicates (Figure 23A). In the samples analyzed using the dual fragmentation method EThcD, 3,293 distinct peptides were identified in total with 2,859 peptides (87 %) found in both technical replicates while in the samples analyzed using ETciD, 2,926 unique peptides were identified in total with 2,590 peptides (89 %) found in both technical replicates (Figure 23B).

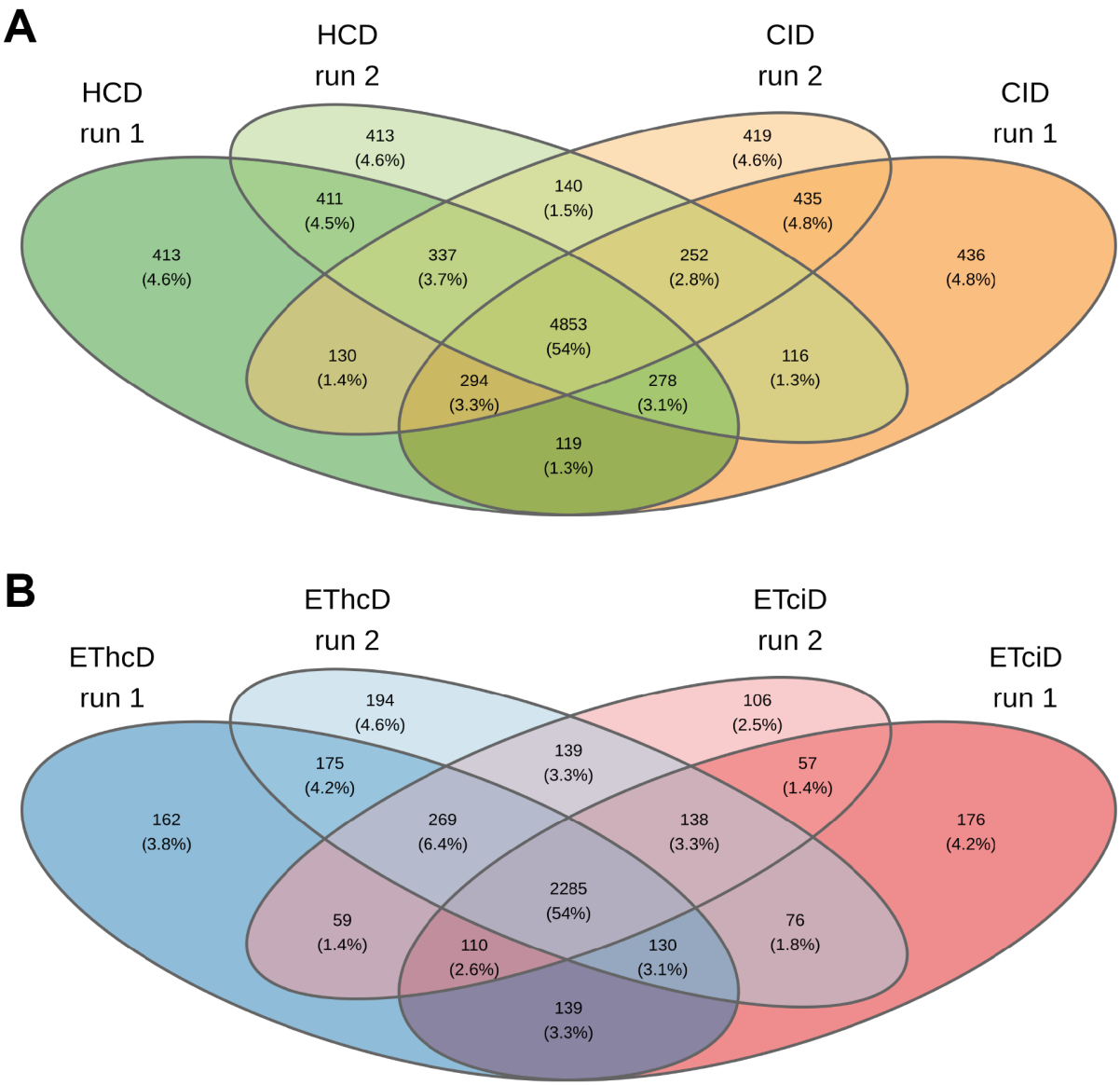


Figure 23: Technical replicates of all fragmentation methods analyzed show a high reproducibility. Overlap of peptides identified in technical replicates using single fragmentation by either HCD (green) or CID (orange; panel A). Overlap of peptides identified in technical replicates using dual fragmentation by either EThcD (blue) or ETciD (red; panel B).

Next, the intensity correlation between the technical duplicates of the samples analyzed using different fragmentation methods was assessed. Both samples analyzed using single fragmentation (Figure 24A and B) as well as samples analyzed using dual fragmentation (Figure 24C and D) showed very good intensity correlations ($R \geq 0.97$) for peptides identified in technical replicates. Concluding from the identification and quantification data presented here, the MS method used is suitable for the label-free quantification of HLA class I-presented peptides.

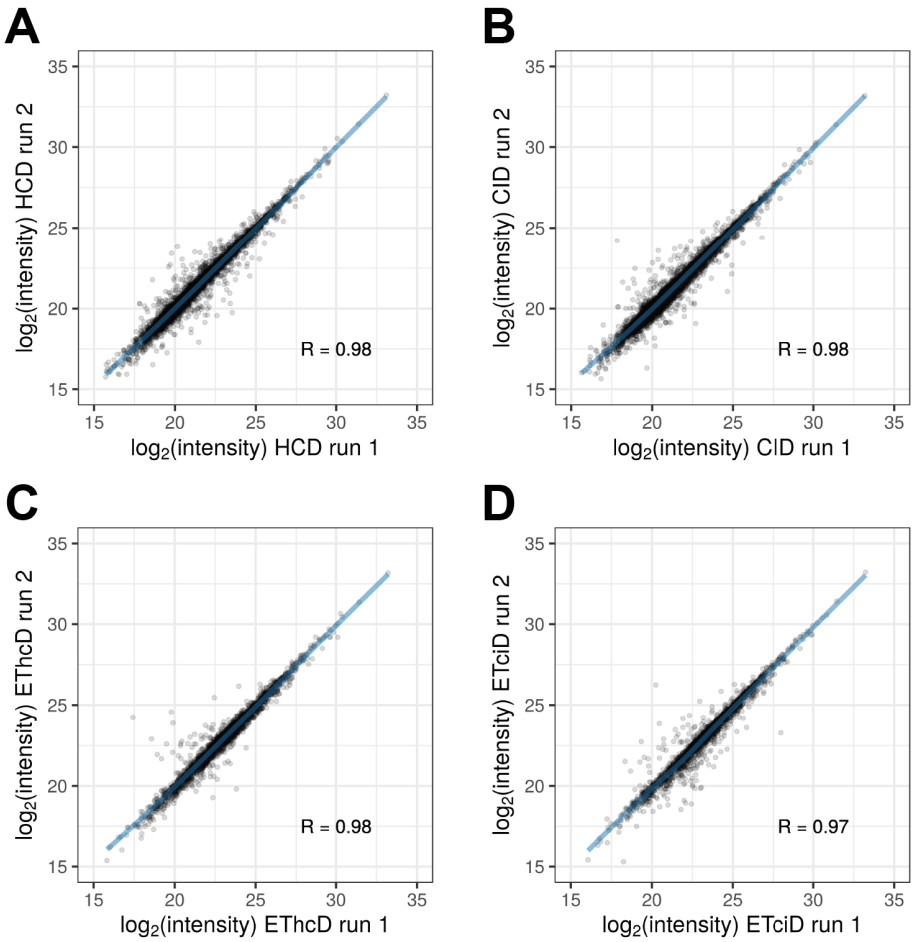


Figure 24: Peptides identified in technical duplicates show a high intensity correlation. Intensity correlations of peptides identified in technical duplicates using single fragmentation by either HCD (panel A) or CID (panel B), or using dual fragmentation by either EThcD (panel C) or ETciD (panel D).

The two major drawbacks of dual fragmentation methods involving ETD are their low speed of measurement and the inability to analyze singly charged peptide precursors. Both are caused by the chemical reaction which is used for electron transfer. During this reaction, an electron is transferred from fluoranthene to the peptide resulting in its fragmentation. The transfer of an electron reduces the charge of the peptide by one, which means that singly charged peptides no longer carry a charge and can therefore no longer be analyzed (Chi *et al.*, 2007). Since singly charged peptides can account for a large proportion of up to 40 % in immunopeptidomics datasets (Bassani-Sternberg *et al.*, 2015), a decision tree method was set up for analyses using dual fragmentation. Here, singly charged peptides were subjected to single fragmentation by either HCD and CID, and only peptides with a charge larger than one were analyzed by either EThcD or ETciD. In the parameter testing dataset, 828 peptides with a single charge were identified by HCD (11 %), while 1,156 peptides with a single charge were identified using the decision tree method for EThcD (35 %). Of these, 567 peptides with a single charge were identified in both samples analyzed with HCD and EThcD. For both datasets, the majority of peptides had a charge state of +2, followed by precursors with a charge state of +3 and only a minor fraction of peptides with a charge state of +4 (Figure 25).

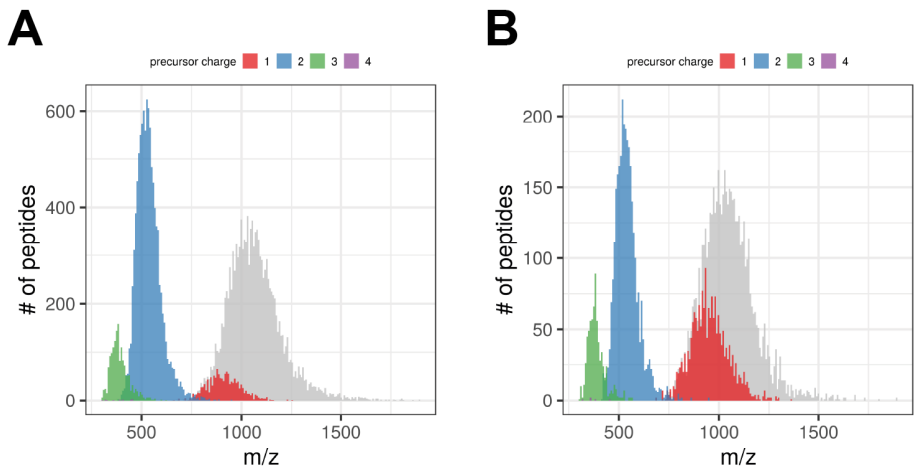


Figure 25: *m/z* distribution of peptides identified using single and dual fragmentation methods. *m/z* distribution of peptides with charge state +1 (red), +2 (blue), +3 (green), and +4 (purple) identified using single fragmentation (panel A). m/z distribution of peptides with charge state +1 (red), +2 (blue), +3 (green), and +4 (purple) identified using dual fragmentation with decision tree (panel B). Mass distribution is displayed in grey.

The second major drawback of dual fragmentation methods utilizing ETD is their slower speed compared to single fragmentation by collision-induced dissociation. This is exemplified in both, the number of peptide identifications as well as in the number of MS2 scans performed per MS1 scan (Table 22) and results in a lower coverage of the samples analyzed. Of note, in contrast to a previous report utilizing EThcD for immunopeptidome analysis (Mommen *et al.*, 2014), dual fragmentation did not significantly improve the identification rate (*i.e.* the number of peptide spectrum matches (PSMs) per MS2 scans) here.

Table 22: Summary of scan numbers and identification key figures for datasets recorded using single fragmentation and dual fragmentation methods. PSMs, peptide spectrum match.

fragmentation method	HCD		CID		EThcD		ETciD	
replicate	1	2	1	2	1	2	1	2
MS1 scans	1,721	1,721	1,689	1,689	1,672	1,670	1,672	1,669
MS2 scans	27,708	27,655	25,561	25,459	17,101	16,954	17,292	16,726
MS2/MS1	16.1	16.1	15.1	15.1	10.2	10.2	10	10.3
PSMs	10,647	10,618	10,221	10,467	7,686	7,669	6,653	6,949
% matched	38 %	38 %	40 %	41 %	45 %	45 %	38 %	42 %
peptides identified	7,414	7,371	7,244	7,327	3,521	3,501	3,071	3,199
peptides filtered	6,835	6,800	6,783	6,860	2,947	2,905	2,484	2,542
peptides combined	7,756		7,809		3,293		2,926	

Further analysis of the peptides identified using either HCD or CID revealed that peptides also identified by one of the dual fragmentation methods had a substantially higher intensity than those only identified

by single fragmentation (Figure 26) further showing that the reduced speed of dual fragmentation methods results in lower coverage, especially of low abundant peptides. Since mutation-derived neoepitopes are expected to be of low abundance, further optimization of the dual fragmentation methodology with a focus on HCD/EThcD fragmentation was pursued. HCD/EThcD was chosen based on the assumption that the higher dissociation energies employed for HCD result in more complex fragmentation patterns compared to CID and thus more information-rich spectra (Jedrychowski *et al.*, 2011; Ichou *et al.*, 2014) which will be beneficial for peptide identification. Moreover, it has been previously shown that HCD fragmentation outperforms CID fragmentation in Orbitrap instruments (Frese *et al.*, 2011; Tu *et al.*, 2016).

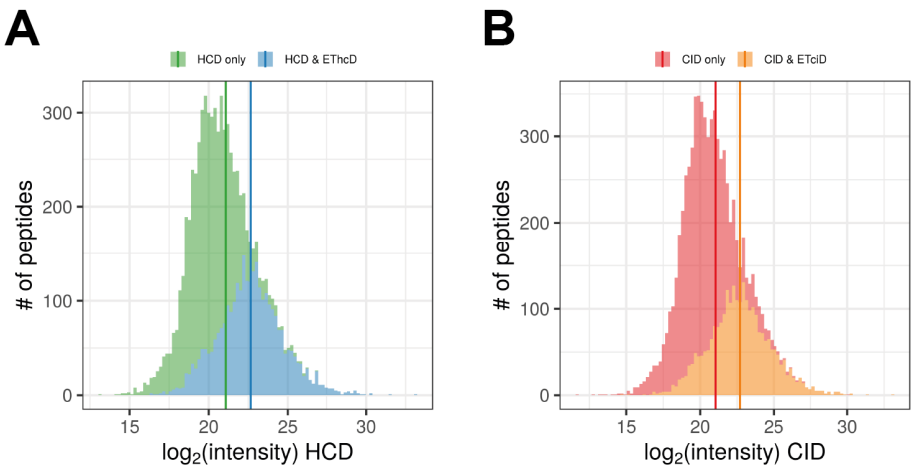


Figure 26: Low-intensity peptides are not identified by dual fragmentation methods. Intensity distributions of peptides identified by either HCD and EThcD (panel A), or CID and ETciD (panel B). Colors represent datasets in which peptides were identified: HCD only (green), HCD and EThcD (blue), CID only (red), CID and ETciD (orange). Vertical lines indicate the corresponding mean intensity.

3.4.2.2 Parameter optimization of EThcD fragmentation method

Although dual fragmentation methods such as EThcD and ETciD were found to provide a less comprehensive view of the analyzed immunopeptidome samples, further optimizations of the EThcD methodology were pursued since dual fragmentation provides more information-rich spectra thus providing benefit for unambiguous peptide identification. To this end, three different parameters have been optimized for the dual fragmentation method EThcD: gradient length, cycle time, and precursor selection using the peptide pool previously generated as sample material.

First, different gradient lengths for liquid chromatography were evaluated with the rationale that longer gradients lead to a proportional increase of acquired MS2 spectra and thus increases the chance of sampling unique peptides (Hsieh *et al.*, 2013). While the MS2/MS1 ratio stayed constant at about 10 MS2 spectra per MS1 spectrum, the expected increase in MS2 scans was observed with 16,279, 21,803, 27,403, 33,031, and 38,565 MS2 spectra recorded for gradients of 90, 120, 150, 180, and 210 minutes, respectively. In total 5,708 distinct peptides were found with the different gradient lengths of which 2,472 peptides were found in all five datasets. The number of total peptide identifications increased with the gradient length from 3,139 unique peptides found with the 90-minute gradient to 4,688 unique peptides found with the 210-minute gradient. The number of peptides found in both samples measured using HCD and EThcD increased with an increased gradient length (2,879 peptides identified

in HCD and EThcD with a 90-minute gradient versus 3,807 peptides identified in HCD and EThcD with a 210-minute gradient). Also, the mean intensity of peptides identified in both HCD and EThcD decreases with an increasing gradient length (Figure 27), showing the improved detection of lower abundant peptides by EThcD with a prolonged gradient. However, more than half of the peptides and especially those with low abundance identified by HCD were not identified by EThcD with any of the prolonged gradients tested. Further optimizations of the EThcD methodology were performed using the 180-minute gradient. It provided a substantial improvement regarding the number of identified peptides in general, those also identified in samples measured using HCD as well as low abundant peptides. However, in contrast to the number of recorded MS2 spectra, the number of identified peptides does not increase linearly with the gradient length. Moreover, further prolongation of gradients may lead to a flattening of chromatographic peaks hampering peptide quantification as well as detection of very low abundant peptides (Bantscheff *et al.*, 2007).

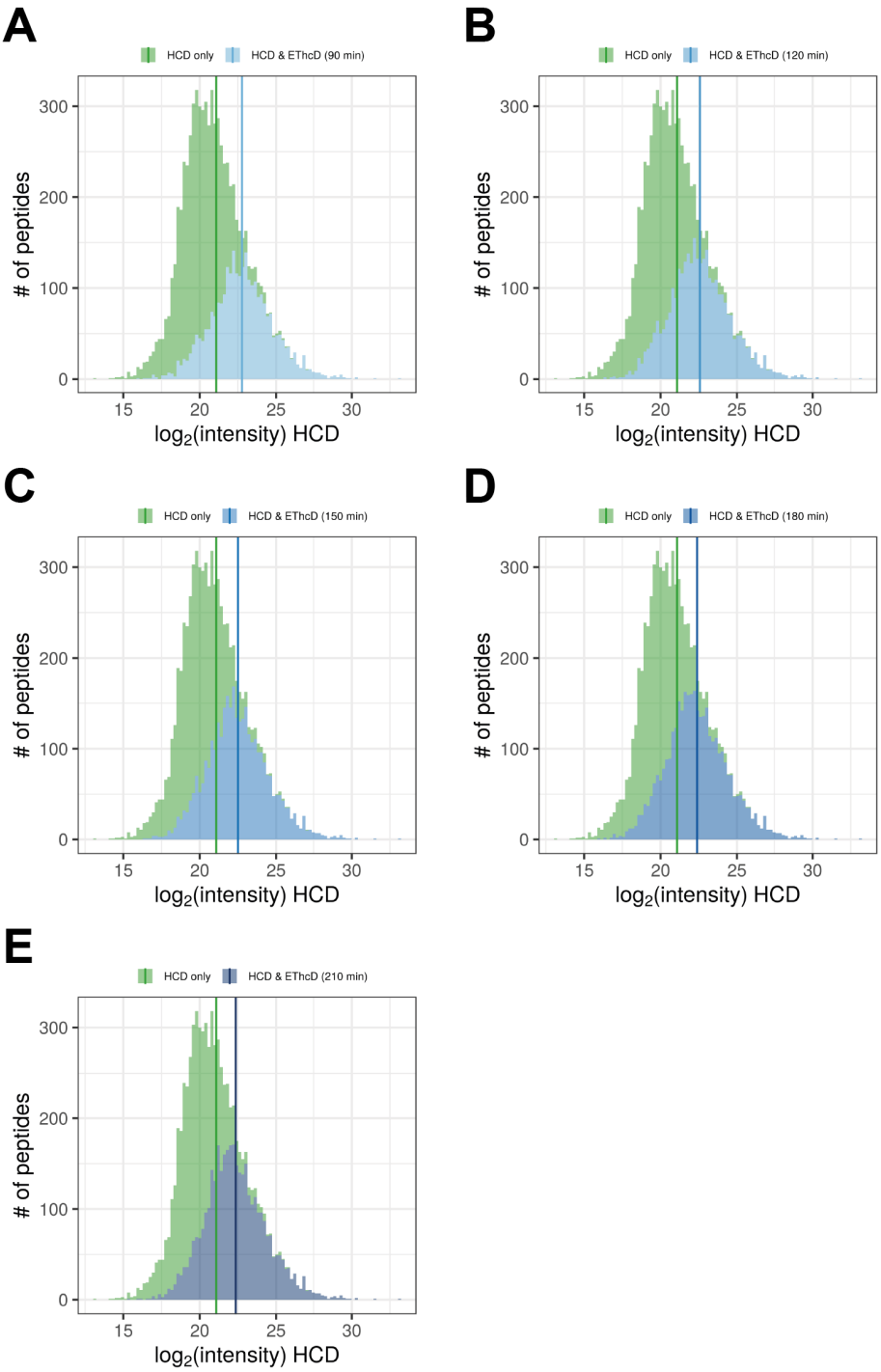


Figure 27: Effect of different gradient lengths on the identification of low abundant peptides. Intensity distributions of peptides identified by single fragmentation using HCD and dual fragmentation using EThcD and different gradient lengths ranging from 90 min (panel A) to 210 min (panel E). Colors represent datasets in which peptides were identified: HCD only (green), HCD and EThcD (blue), shades of blue indicate gradient lengths. Vertical lines indicate the corresponding mean intensity.

Next, different cycle times were compared. The cycle time is defined as the time used for the generation of MS2 scans of the top(n)-abundant precursor ions between two MS1 scans. For example, in the previously described HCD dataset, a cycle time of three seconds was used and 1,721 MS1 scans were generated during a 90-minute gradient. During the three seconds cycle time, on average 16.1 MS2 spectra were generated, *i.e.* fragment ion information of the 16 most abundant precursors was recorded.

The *m/z* values of already fragmented and measured precursors are subsequently dynamically excluded from further MS2 analysis for 20 seconds in order to allow fragmentation of less abundant precursors. Here, a cycle time of three seconds was compared to a cycle time of one second with the rationale that shorter cycle time will lead to the dynamic exclusion of highly abundant precursors thus increasing the chances for identification of less abundant peptide precursors. Shortening the cycle time to one second led to a decrease in peptide identifications compared to samples measured with three seconds cycle time (3,677 unique peptides versus 4,232 unique peptides). Moreover, the mean intensity of peptides identified with one second was slightly higher than for those identified using three seconds cycling time (22.62005 versus 22.43894; Figure 28). Again, the majority of low abundant peptides identified using HCD were neither identified in EThcD-fragmented samples with three seconds cycle time nor one second cycle time. Concluding from these data, shortening the cycle time does not provide a benefit for the identification of low abundant peptides and was therefore discarded. Further optimizations were performed using a 180-minute gradient and three seconds of cycle time.

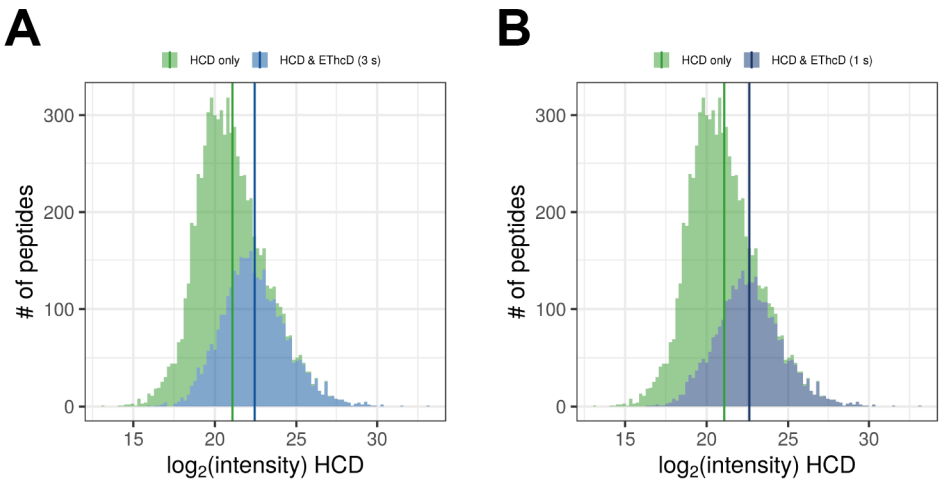


Figure 28: Effect of different cycle times on the identification of low abundant peptides. Intensity distributions of peptides identified by single fragmentation using HCD and dual fragmentation using EThcD and 3 s cycle time (panel A) and 1 s cycle time (panel B). Colors represent datasets in which peptides were identified: HCD only (green), HCD and EThcD (blue), shades of blue indicate cycle time. Vertical lines indicate the corresponding mean intensity.

Finally, different modes of precursor selection for EThcD were compared. Normally, precursors are selected according to their intensity from highest to lowest, using the so-called top(n) selection. Here, precursor selection starting from the least abundant (but detectable) precursor was tested with the rationale to generate data complementary to those produced using the standard top(n) method. In total 2,563 distinct peptides were identified in samples measured using this so-called “lowEThcD” approach of which 1,575 peptides were also identified using HCD. Notably, the mean intensity of peptides identified by lowEThcD was lower than any mean intensity obtained with the previously tested EThcD and top(n) selection methods (also termed “highEThcD”; Figure 29A and B). However, the majority of low abundant peptides was still only identified using HCD fragmentation. To verify the suitability of the lowEThcD approach for quantification, the intensities of peptides measured in both either HCD and highEThcD or HCD and lowEThcD were compared (Figure 29C and D). Peptide intensities measured

by HCD and highEThcD showed a strong correlation (R = 0.93) while peptide intensities measured by HCD and lowEThcD only showed a moderate correlation (R = 0.67).

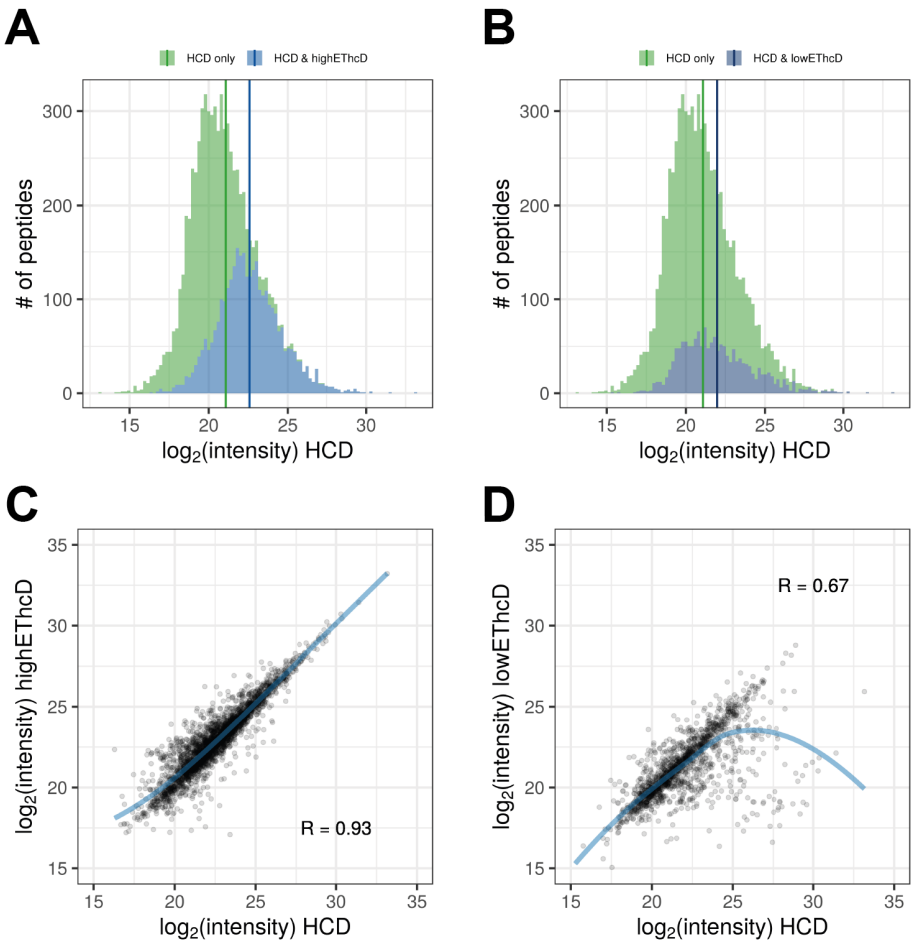


Figure 29: Effect of precursor selection method on the identification of low abundant peptides. Intensity distributions of peptides identified by single fragmentation using HCD and dual fragmentation using EThcD and standard precursor selection (highEThcD; panel A) and "low first" precursor selection (lowEThcD; panel B). Colors represent datasets in which peptides were identified: HCD only (green), HCD and EThcD (blue), shades of blue indicate gradient lengths. Vertical lines indicate the corresponding mean intensity. Intensity correlation of peptides measured using HCD and highEThcD (panel C) and of peptides measured using HCD and lowEThcD (panel D).

3.4.2.3 Summary of MS parameter optimization

The single fragmentation methods applying either CID or HCD provided the most comprehensive datasets during the parameter testing (Figure 30). For further evaluation of the effect of NMD inhibition by 5AZA on the immunopeptidome, HCD was chosen because dissociation by higher energy creates more information-rich spectra which may be beneficial for the unambiguous identification of mutation-derived neoepitopes. Furthermore, samples were analyzed by two additional methods utilizing EThcD. First, a 180-minute gradient in combination with EThcD and standard precursor selection was chosen. Using these settings, a large proportion of higher abundant peptides found by HCD is covered and EThcD provides more information-rich spectra allowing unambiguous peptide identification. Second, a 180-min gradient was chosen in combination with EThcD and reverse precursor selection, as this allowed at least a part of the low abundant peptides to be covered by a method with EThcD fragmentation. Moreover, the lowEThcD method was able to identify a subset of 781 low abundant

peptides that were not identified by HCD (Figure 30). However, based on the only moderate intensity correlation of peptides identified using HCD and lowEThcD, the lowEThcD is less suited for quantification purposes.

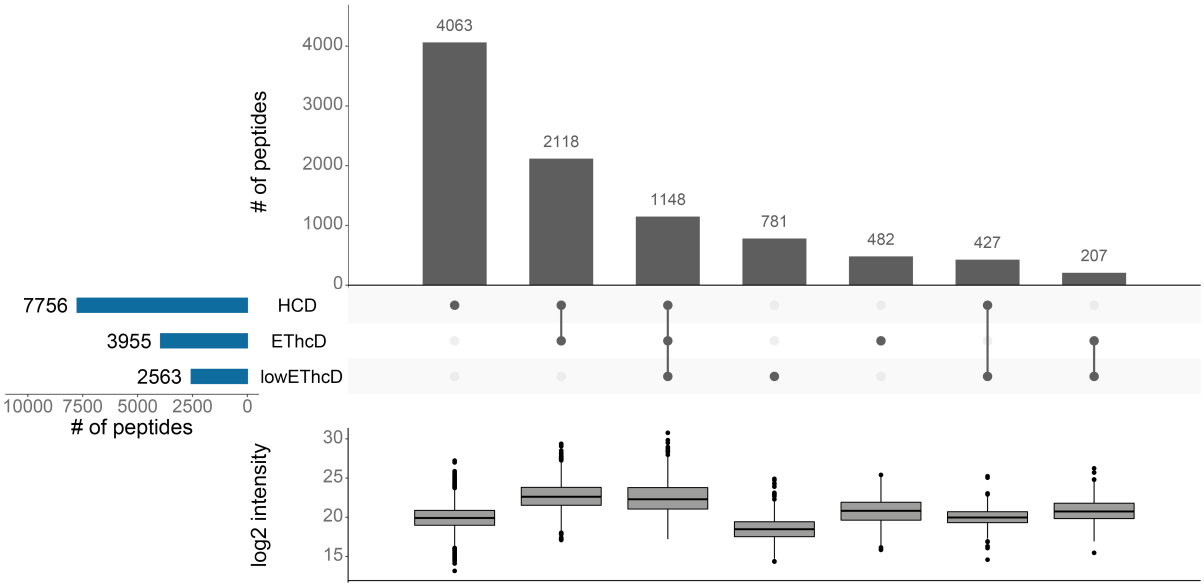


Figure 30: Overview of peptides identified during parameter optimization. Number of peptides identified in the HCD, EThcD, and lowEThcD datasets (vertical bar chart) and overlap between the three datasets (horizontal bar chart). Boxplot summary representing the intensity distribution of peptide subsets. Bars represent 25th to 75th percentiles, middle line represents median.

3.5 Investigation of the effect of 5AZA on the immunopeptidome of MSI CRC model cell lines

Using the previously established MS parameters for HCD, EThcD, and low EThcD, an immunopeptidomics pipeline (Figure 31) was set up to investigate the effect of NMD inhibition by the small molecule 5AZA on the immunopeptidome, and in particular InDel mutation-derived neoepitopes in the two MSI CRC model cell lines HCT-116 and RKO. Briefly, HLA class I-presented peptides were isolated using the high-throughput immunoprecipitation protocol and each sample of isolated peptides was analyzed by LC-MS/MS applying either HCD, EThcD, or lowEThcD fragmentation. The generated raw data were searched against the human proteome (endogenous peptides) as well as against the neoepitope databases and identified peptides were validated *in silico*. Neoepitope candidates were further validated both at the peptidomic and genomic level. Moreover, a targeted proteomics approach was used to enable the reliable quantification of InDel mutation-derived neoepitopes. Selected neoepitopes were further tested for their potential to elicit immune responses *in vivo*. The individual steps and their results are described in detail in the following sections.

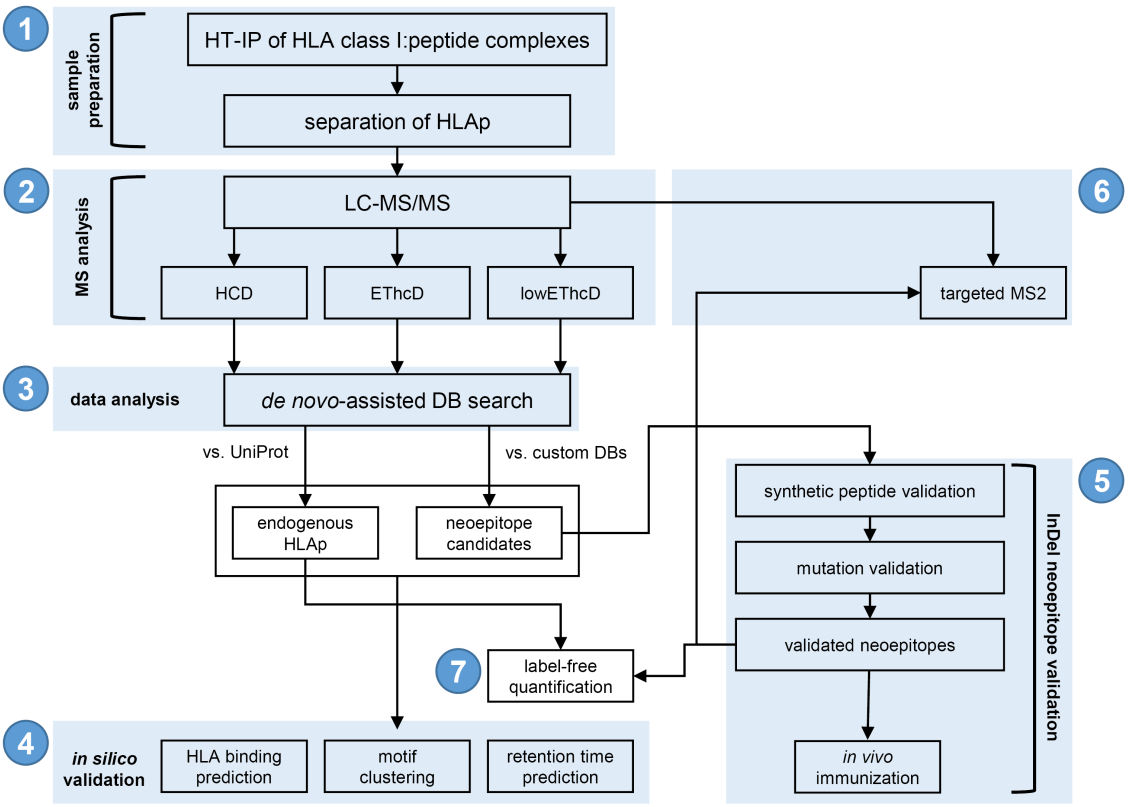


Figure 31: Immunopeptidomics workflow for the identification, validation, and quantification of InDel neoepitopes. After high-throughput immunoprecipitation of HLA class I:peptide complexes, HLAp are separated (1) and subjected to LC-MS/MS using different fragmentation and precursor selection methodologies (HCD, EThcD, lowEThcD; 2). Data analysis of raw data is performed using de novo-assisted database search against the UniProt database to identify endogenous HLAp and against custom databases containing neoepitope sequences (3). All identified peptides are validated bioinformatically (binding prediction, sequence clustering, retention time prediction; 4). Neoepitope candidates are further validated by comparison to synthetic peptide spectra and validation of underlying mutations. In vivo processing and presentation are tested by immunization of an “HLA-humanized” mouse model (5). Neoepitopes are measured again using a targeted MS2 approach (6). Label-free quantification is performed for endogenous HLAp and validated neoepitopes (7). HT-IP, high-throughput immunoprecipitation; HLAp, HLA class I-presented peptides; MS, mass spectrometry; LC-MS/MS, liquid chromatography-tandem mass spectrometry; DB, database.

3.5.1 Quality control of samples analyzed using the immunopeptidomics pipeline

Both NMD inhibition by 5AZA and success of immunoprecipitation were verified prior to mass spectrometric analysis. Treatment of cells with 5 µM 5AZA for 24 h induced a significant stabilization ($p \leq 0.0001$) of *ATF3* (3.67-fold), *SC35C* (4.90-fold), and *SC35D* (3.26-fold) mRNAs in HCT-116 cells. In RKO cells, mRNA abundance of *ATF3* (2.76-fold), *SC35C* (5.02-fold), and *SC35D* (5.51-fold) were also significantly increased after 5AZA treatment ($p \leq 0.0001$). In both cell lines, the abundance of the NMD-insensitive *SC35A* mRNA was only slightly and not significantly increased (Figure 32). These data show the successful and specific NMD inhibition by treatment with 5 µM 5AZA in samples used for mass spectrometry-based immunopeptidome analysis for both HCT-116 and RKO cells.

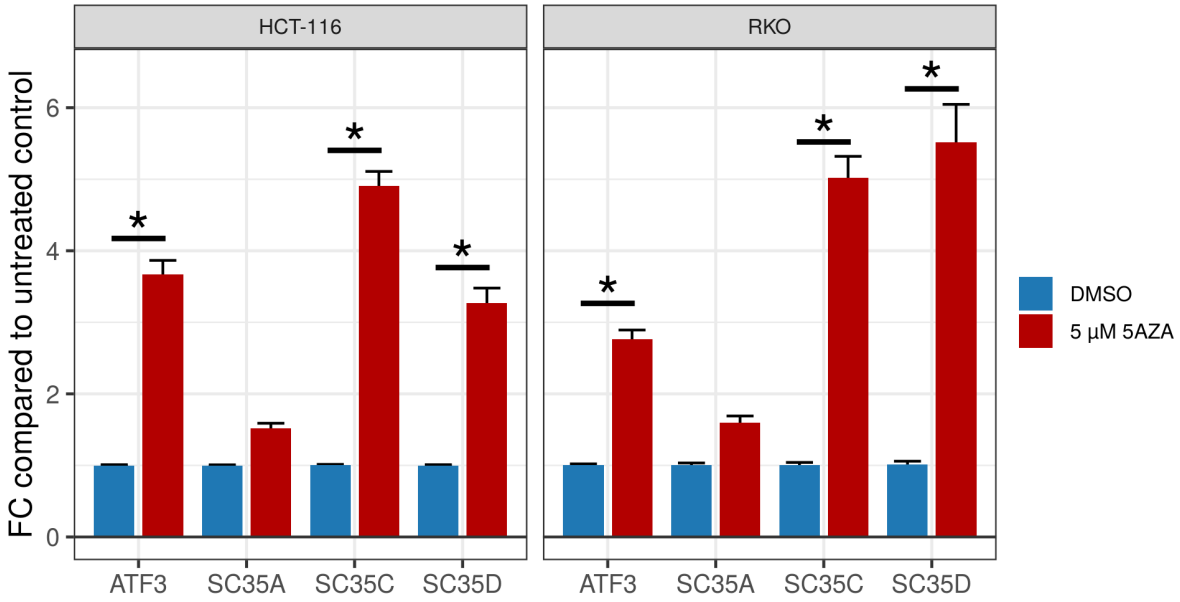


Figure 32: Validation of NMD inhibition by 5AZA in samples used for immunopeptidomics. Treatment with 5 µM 5AZA for 24 h increases mRNA abundance of endogenous NMD targets *ATF3*, *SC35C* and *SC35D* as determined by qPCR. Bars represent mean ± SD of three experiments, * $p \leq 0.0001$ (two-sided, unpaired t-test).

Next, the isolation of HLA molecules from HCT-116 and RKO cells was assessed by western blot analysis. The success of HLA immunoprecipitation is shown by enrichment of HLA class I heavy chains and B2M molecules in the elution fractions (lanes 9 and 10; Figure 33A and B). Although the signal intensity of these bands is relatively weak compared to the HLA and B2M signal in the input (lanes 1 and 2), a substantial HLA class I molecule enrichment in the eluate is reflected by the strongly reduced signal in the flow-through (lanes 3 and 4). In contrast to the contaminating tubulin, which is removed during the first washing step (lanes 5 and 6), the HLA signal was not found in the flow-through of this washing step. As outlined above, the weaker HLA class I signal intensity in eluates of the high-throughput immunoprecipitation is likely caused by the elution procedure which includes the separation of HLA class I molecules and bound peptides at low pH. Of note, treatment with 5AZA does not affect HLA class I expression in the two model systems used as the intensity of the corresponding bands of treated and untreated samples is similar (lanes 1 and 2).

In conclusion, the samples passed both quality control experiments and were subsequently analyzed by mass spectrometry.

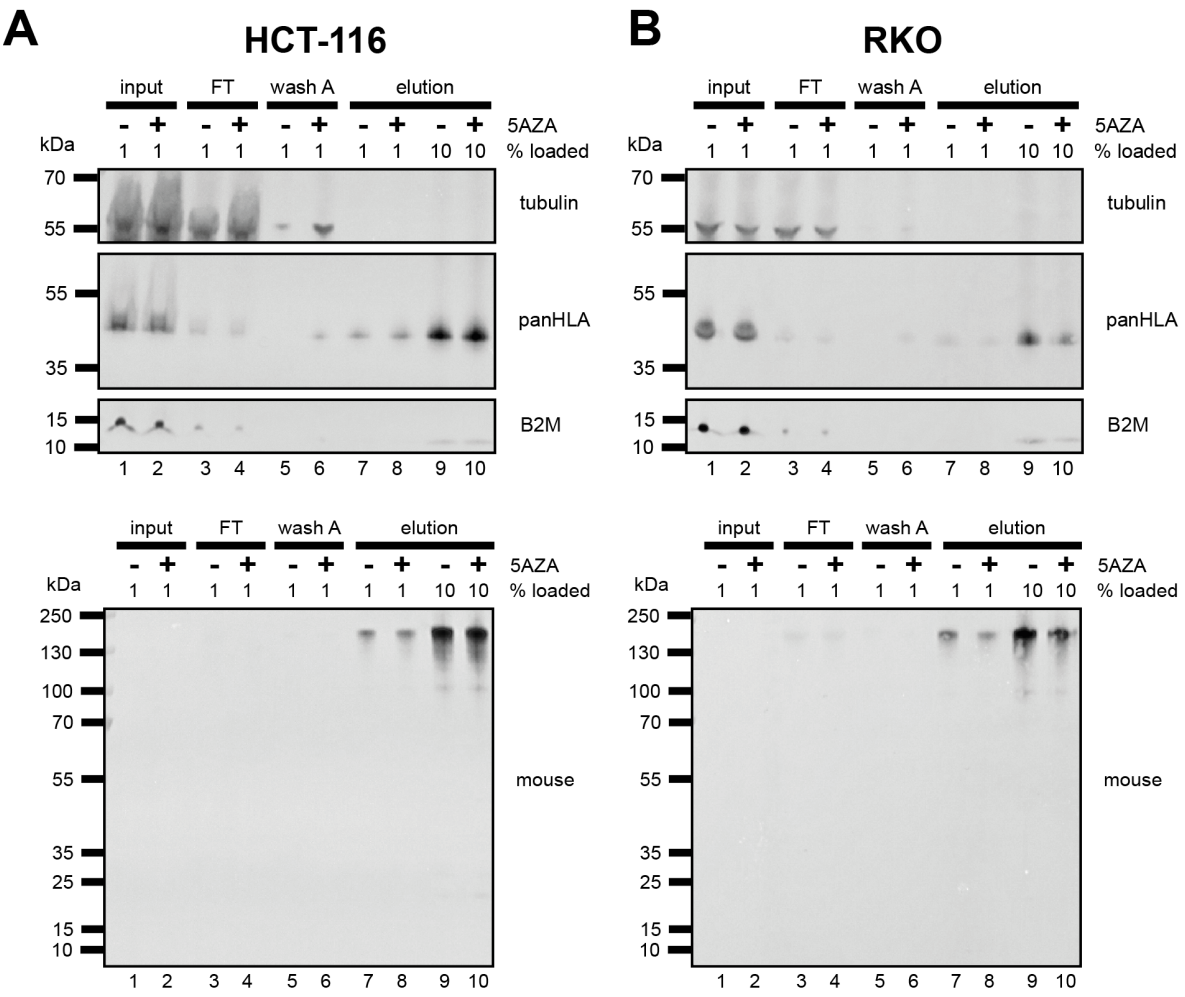


Figure 33: Selective purification of HLA class I molecules. Quality control by western blot analysis of HLA class I purification from HCT-116 cells (panel A) and RKO cells (panel B) by high-throughput immunoprecipitation. Samples were further analyzed using the immunopeptidomics pipeline. Tubulin was used as the loading control. B2M, β-2-microglobulin; FT, flowthrough.

3.5.2 Exploring the immunopeptidome

3.5.2.1 Comprehensive identification of endogenous HLA class I-presented peptides

Samples prepared from $3x10^8$ cells either treated with 5 µM 5AZA or the solvent control DMSO for 24 h in biological triplicates were prepared for HCT-116 and RKO cells and analyzed using the three different MS methodologies. In total, 10,030 unique HLA class I-presented endogenous peptides were identified for HCT-116 using a stringent FDR of 1 % (Figure 34). Of these, about 80 % (8,005 peptides) were identified using HCD fragmentation while 3,579 (36 %) and 3,683 (37 %) unique peptides were identified using EThcD and lowEThcD fragmentation, respectively. 3,098 peptides (31 %) were identified in both datasets recorded using HCD or one of the dual fragmentation methods, while 4,907 peptides (49 %) were only identified in datasets recorded with HCD fragmentation. 2,025 peptides (20 %) were only identified in either the EThcD or lowEThcD dataset. The MS1 signals of identified and quantified peptides showed a wide range of intensity spanning several orders of magnitude (log_2 intensity: 12.75-32.55, mean log_2 intensity: 20.19). In comparison, peptides identified in all three datasets showed a slightly higher average intensity/abundance than peptides identified in the HCD dataset alone (log_2 intensity: 21.58 vs. 19.77). Of note, using the lowEThcD method a subset of 851 low abundant peptides

(mean log_2 intensity: 18.52) was identified which was not found in any of the other datasets. The 10,030 identified HLA class I-presented peptides originated from 4,767 unique source proteins. Of these, 2,524 (53 %) were represented by only one peptide, 1,054 (22 %) by two distinct peptides, and 516 (11 %) by three distinct peptides at the cell surface. The remaining 673 source proteins were represented by four or more distinct HLA class I-presented peptides. In general, source proteins of identified peptides were associated with a broad spectrum of cellular localization and the top 10 % most abundant peptides were significantly enriched in clusters for nuclear, cytoskeletal, and ribosomal proteins, which is in line with previous reports (Bassani-Sternberg *et al.*, 2015; Walz *et al.*, 2015).

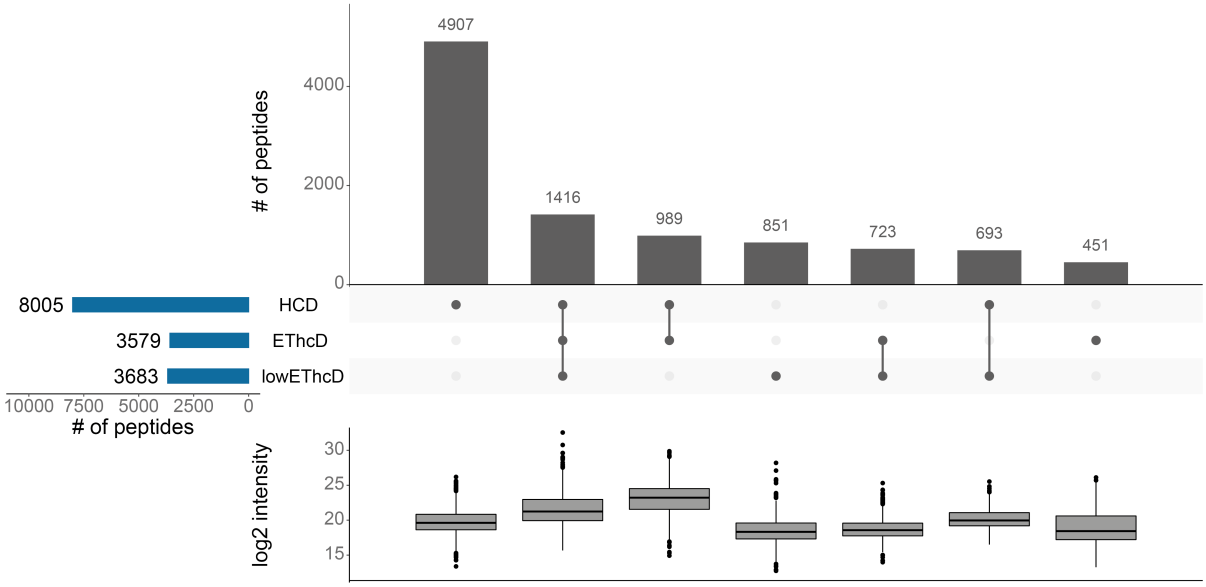


Figure 34: Overview of peptides identified from HCT-116 cells using the immunopeptidomics pipeline. Number of peptides identified in the HCD, EThcD, and lowEThcD datasets (vertical bar chart) and overlap between the three datasets (horizontal bar chart). Boxplot summary representing the intensity distribution of peptide subsets. Bars represent 25th to 75th percentiles, middle line represents median.

In silico quality controls were performed to ensure the quality of the obtained datasets and peptide identifications. First, sequence-specific hydrophobicity indices (HIs) were calculated as an orthogonal parameter for peptide identifications (Krokhin, 2006). The predicted HIs of identified peptides showed a tight correlation with the experimentally observed retention times for all three fragmentation methods used (Pearson's correlation coefficient HCD: 0.96, EThcD: 0.96, lowEThcD: 0.95) thus validating the peptide identifications (Figure 35A). The identified peptides showed an HLA class I-typical length distribution with mainly the preferred length of nine amino acids (Figure 35B). Typically, the identification of HLA class I-presented (neo)epitopes by methods other than MS involves shortlisting of candidate peptides using HLA binding prediction creating an *a priori* bias. However, binding prediction of HLA class I-presented peptides identified by MS *a posteriori* represents a suitable tool for further validation of immunopeptidomics datasets. 90 % of identified peptides (8,988 peptides; 7,286 strong binder (SB), 1,702 weak binder (WB)) were predicted to bind at least one of the six HLA class I molecules expressed on the HCT-116 cells (Figure 35C).

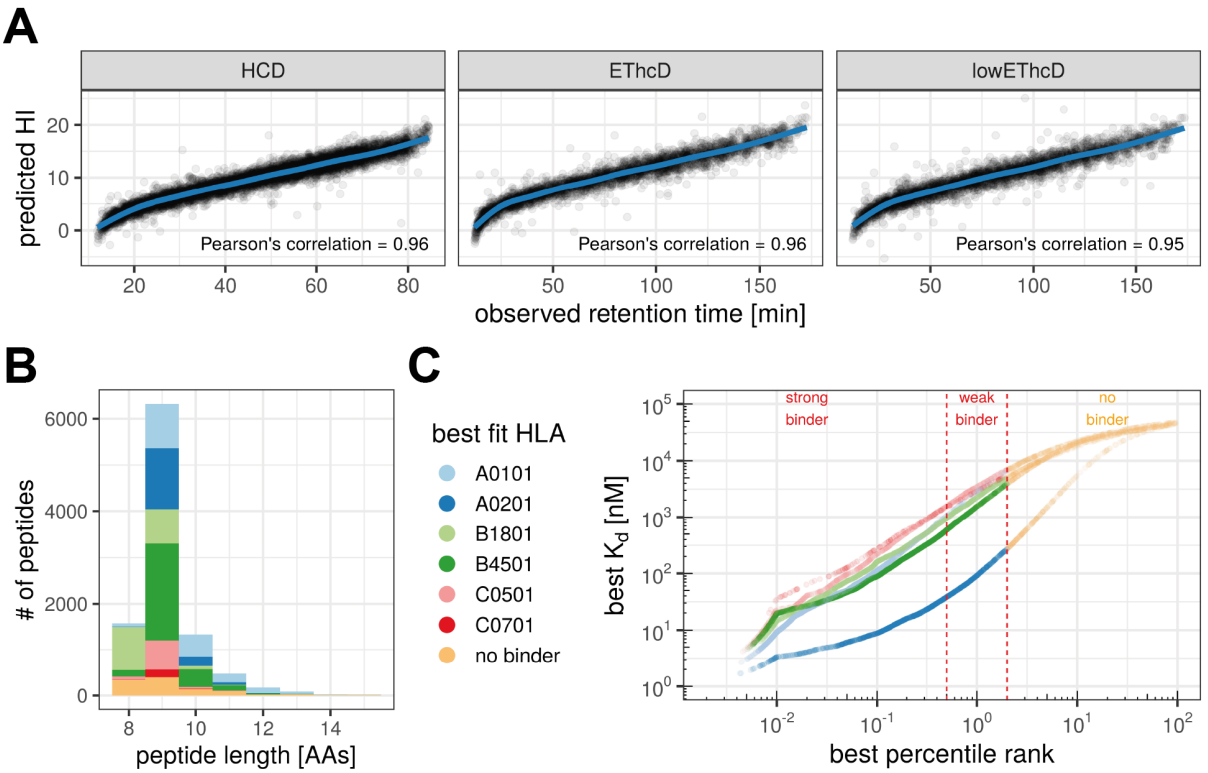


Figure 35: *In silico* quality controls of peptides identified from HCT-116 cells. Predicted hydrophobicity indices against observed retention times of identified peptides for the different MS methods used (panel A). Typical length distribution of HLA class I-presented peptides. Colors represent the fraction of peptides with predicted binding affinity to a particular HLA allele determined by NetMHCpan 4.0 (panel B). Binding prediction of all identified peptides predicted by NetMHCpan 4.0. Threshold for strong binders is top 0.5 % ranked, for weak binders top 2 % ranked (panel C).

Due to the allele-specific binding-mediating anchor residues of HLA molecules, the presented peptides display a high degree of sequence similarity. Given this similarity, identified peptide sequences were clustered into groups, and four distinct motifs corresponding to the consensus binding motifs of HLA-A*01:01, HLA-A*02:01, HLA-B*18:01, and HLA-B*45:01 were identified (Figure 36A). For all identified motifs, aa in the binding-mediating anchor positions (positions 2/3 and 9) were strongly enriched for the corresponding aa reported in the literature (Figure 36B). Of the 10,030 identified peptides, only 282 (2.8 %) were not assigned to one of the expressed HLA alleles. Although expressed on HCT-116 cells, consensus binding motifs for HLA-C*05:01 and HLA-C*07:01 (Figure 36B) could not be defined probably due to the low cell surface expression of the corresponding alleles and their motif redundancy to HLA-A and B alleles (Neisig *et al.*, 1998; Bassani-Sternberg *et al.*, 2015).

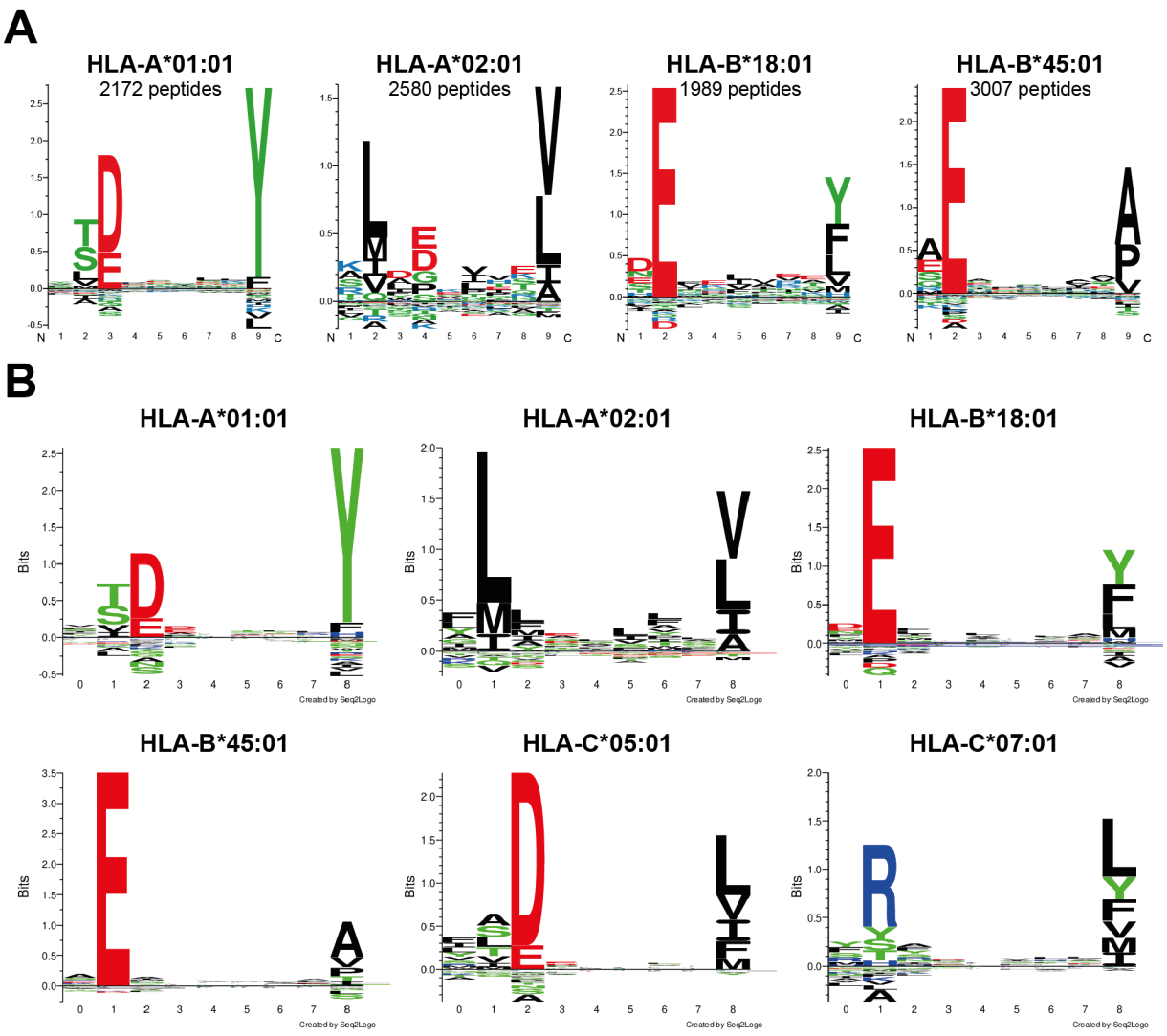


Figure 36: Sequences of identified peptides represent binding motifs of HLA alleles expressed by HCT-116 cells. Amino acid sequences of identified peptides clustered by GibbsCluster and visualized by Seq2Logo identify four distinct HLA binding motifs. 282 outliers (2.8 %) were not clustered (panel A). Literature binding motifs of HLA alleles expressed by HCT-116 cells (panel B).

For the RKO cell line, 9,294 unique HLA class I-presented endogenous peptides were identified using a stringent false discovery rate of 1 % (Figure 37). Again, fragmentation using HCD led to the highest number of peptide identifications (7,366 unique peptides; 79 % of the total dataset), while 3,426 (37 %) and 3,990 (43 %) unique peptides were identified in the EThcD and lowEThcD datasets, respectively. 2,620 peptides (28 %) were identified in both datasets recorded using HCD or one of the dual fragmentation methods. 4,746 peptides (51 %) were only identified using HCD, while 1,928 peptides (21 %) were only identified using one of the two dual fragmentation methods. Similar to the HCT-116 dataset, the MS1 signals of identified and quantified peptides in RKO showed a wide range of intensity spanning several orders of magnitude (log_2 intensity: 12.91-33.45, mean log2 intensity: 20.31). In comparison, peptides identified in all three datasets showed a slightly higher average intensity/abundance than peptides identified in the HCD dataset alone (log_2 intensity: 21.85 vs. 20.72). Also for the RKO cell line, a subset of 678 low abundant peptides (mean log_2 intensity: 18.52) was identified only by the lowEThcD method. The 9,294 identified HLA class I-presented peptides originate

from 4,506 distinct source proteins. 2,429 of the source proteins (54 %) were represented by only one peptide, while 906 (20 %) and 497 (11 %) source proteins were represented by two or three unique peptides, respectively. The remaining 620 (14 %) source proteins were represented by four or more distinct peptides. Again, source proteins exhibit a broad cellular localization with the source proteins of the 10 % most abundant peptides showing significant enrichment for nuclear, cytoskeletal, and ribosomal proteins.

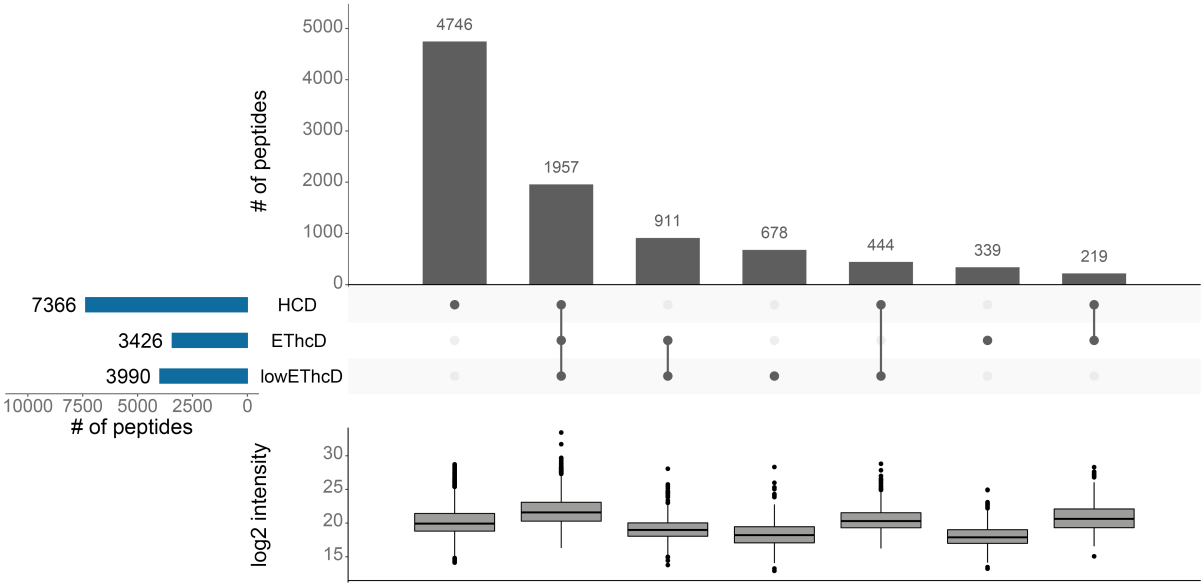


Figure 37: Overview of peptides identified from RKO cells using the immunopeptidomics pipeline. Number of peptides identified in the HCD, EThcD, and lowEThcD datasets (vertical bar chart) and overlap between the three datasets (horizontal bar chart). Boxplot summary representing the intensity distribution of peptide subsets. Bars represent 25th to 75th percentiles, middle line represents median.

In silico predicted HIs showed a tight correlation with observed retention times for peptides identified for the RKO cell line with all three fragmentation methods utilized (Pearson's correlation coefficient HCD: 0.95, EThcD: 0.94, lowEThcD: 0.94; Figure 38A) thus further validating the identifications. Moreover, the peptides showed an HLA class I-typical length distribution with mainly nonamers (Figure 38B) and 90 % of the identified peptides (8,336 peptides; 6,295 strong binder (SB), 2,041 weak binder (WB)) were predicted to bind to at least one of the four HLA alleles expressed at the cell surface of RKO cells (Figure 38C).

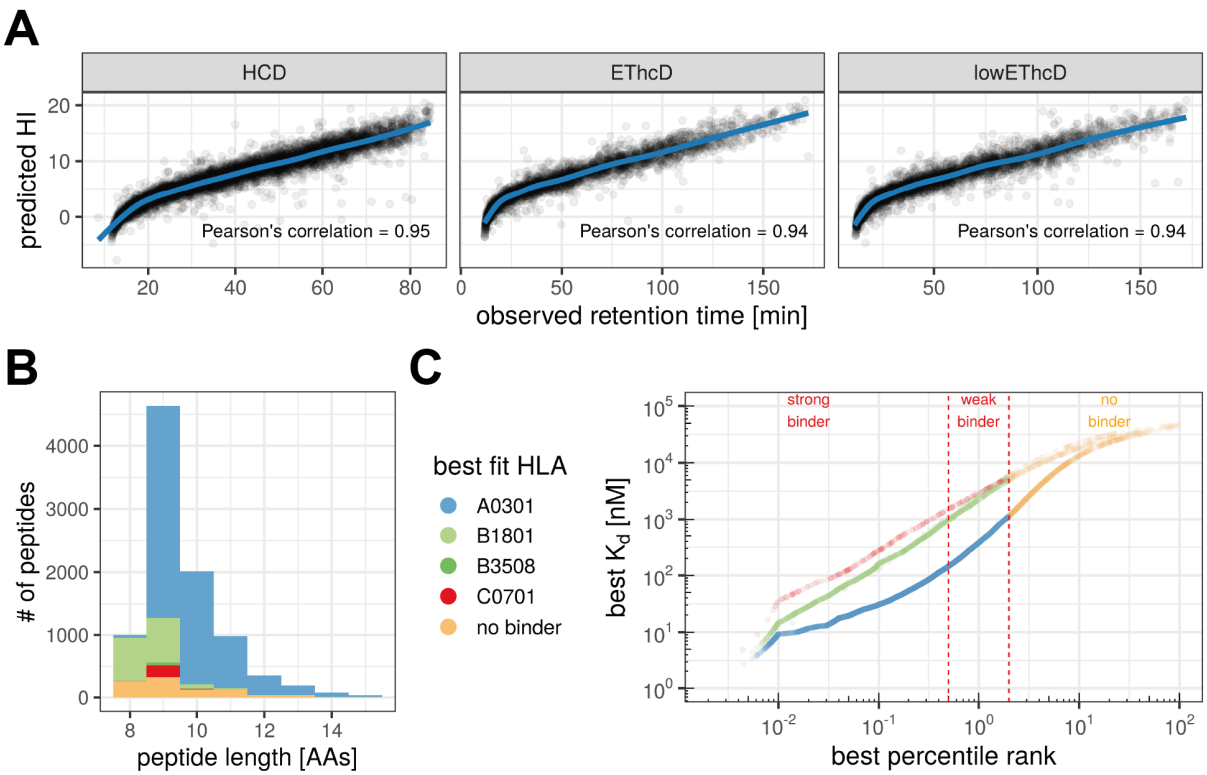


Figure 38: In *silico* quality controls of peptides identified from RKO cells. Predicted hydrophobicity indices against observed retention times of identified peptides for the different MS methods used (panel A). Typical length distribution of HLA class I-presented peptides. Colors represent the fraction of peptides with predicted binding affinity to a particular HLA allele determined by NetMHCpan 4.0 (panel B). Binding prediction of all identified peptides predicted by NetMHCpan 4.0. Threshold for strong binders is top 0.5 % ranked, for weak binders top 2 % ranked (panel C).

Clustering of peptide sequences revealed two distinct binding motifs corresponding to the consensus binding of HLA-A*03:01 and HLA-B*18:01, both of which are expressed by RKO cells (Figure 39). For the two identified motifs, aa in the binding-mediating anchor positions (positions 2/3 and 9) were strongly enriched for the corresponding aa reported in the literature (Figure 39B). Binding motifs of the other two HLA alleles, HLA-B*35:08 and HLA-C*07:01, could not be identified probably because of their low cell surface expression and high similarity to other HLA binding motifs, respectively (Neisig *et al.*, 1998; Bassani-Sternberg *et al.*, 2015). Of note, most of the peptides identified for RKO are predicted to bind to HLA-A*03:01 and this cluster contains the most peptides which is in line with other reports showing HLA-A*03:01 to be the most strongly expressed HLA allele in RKO cells (Scholtalbers *et al.*, 2015).

Taken together, the data presented here validate the two datasets as a representative view of the immunopeptidome of HCT-116 and RKO cell, respectively.

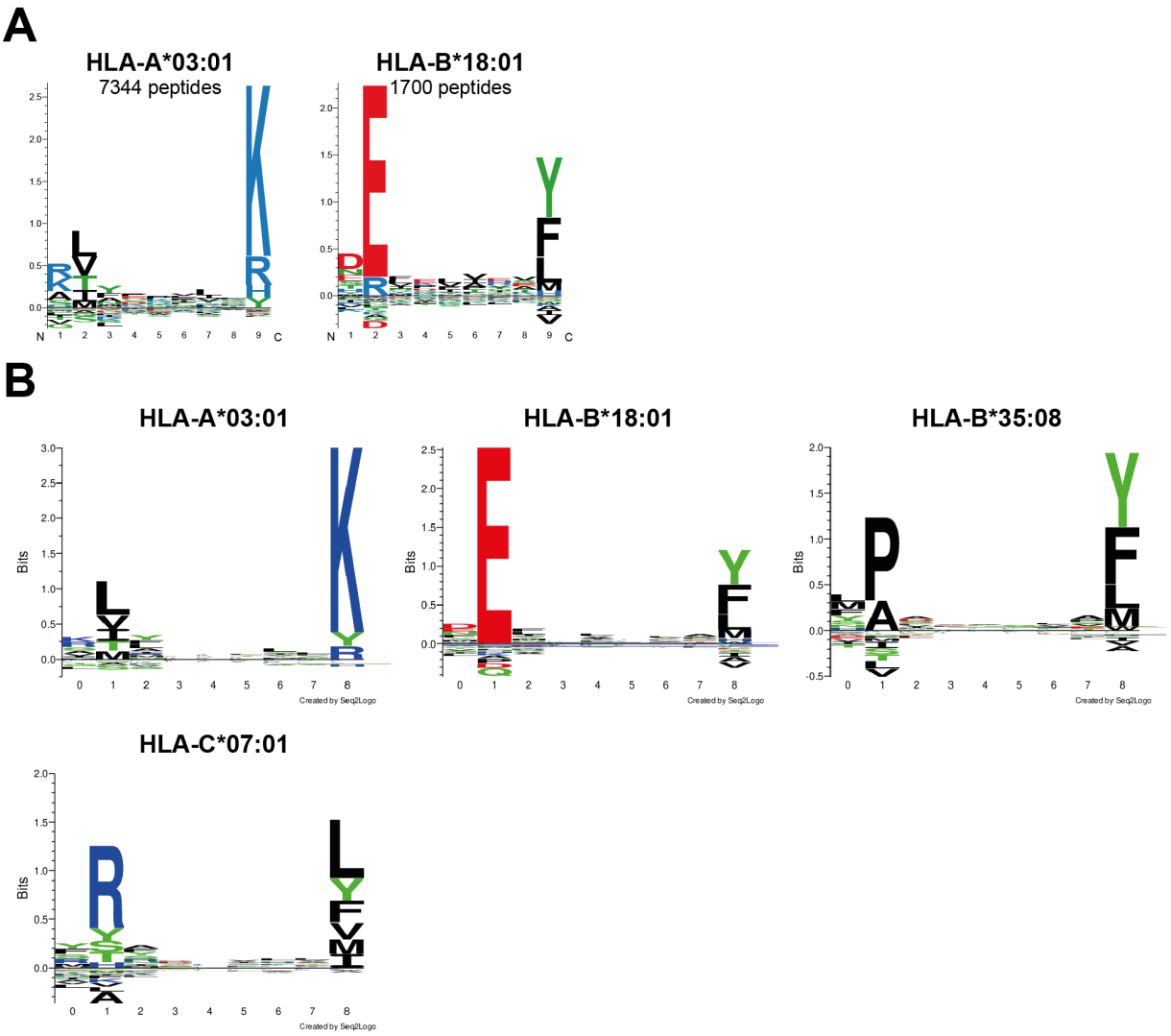


Figure 39: Sequences of identified peptides represent binding motifs of HLA alleles expressed by RKO cells. Amino acid sequences of identified peptides clustered by GibbsCluster and visualized by Seq2Logo identify four distinct HLA binding motifs. 251 outliers (2.7 %) were not clustered (panel A). Literature binding motifs of HLA alleles expressed by RKO cells (panel B).

3.5.2.2 Identification of HLA class I-presented mutation derived neoepitopes

3.5.2.2.1 MS-based identification and validation of mutation-derived neoepitopes

After successful identification of endogenous HLA class I-presented peptides and validation of the datasets as high-quality and comprehensive views of the immunopeptidome, the datasets were queried for the existence of mutation-derived neoepitopes. For this purpose, the *de novo* sequencing capability of the PEAKS software was utilized (Ma *et al.*, 2003; Zhang *et al.*, 2012). During *de novo* sequencing-assisted database search, PEAKS first performs *de novo* sequencing for all spectra of the raw file to create *de novo* sequence tags. These *de novo* sequence tags are then matched against the protein databases of interest to create first a protein shortlist to reduce computing time. Subsequently, the proteins in this shortlist are digested *in silico*, and generated peptide sequences are compared to the measured spectra to find peptide spectrum matches (PSMs). The PSMs are rated by a scoring function and the highest-scoring candidate sequences are assigned to the peptide shortlist. Entries of the peptide shortlist are further evaluated to find the highest-scoring peptide for each spectrum. Finally, results are

validated using a decoy fusion approach which provides the minimum peptide score meeting the user-specified FDR. Peptides with a score higher than this threshold and a database match are reported. Additionally, PEAKS reports so-called "*de novo* only" spectra which produced high-scoring *de novo* sequencing tags which, however, could not be matched to an entry in the provided database. The "*de novo* only" spectra can be extracted and used as input for a multi-round search with modified parameters such as a different protein database. Here, this approach was used to first search immunopeptidomics datasets against the human proteome to identify endogenous HLA class I-presented peptides. Spectra matching those peptides were then removed, thus limiting the search space, and only spectra with high-scoring *de novo* sequencing tags were searched against the databases containing mutation-derived neoepitopes (described in Section 2.6).

The "*de novo* only" spectra of both, the HCT-116 and RKO datasets were searched against the cell line-specific CCLE and COSMIC databases containing InDel mutation-derived protein sequences, against the cell line-specific TCLP databases containing SNP mutation-derived neoepitopes, as well as against the ReFrame database with protein sequences based on recurrent InDel mutations in MSI CRC and the MNR database containing the protein sequences resulting from one and two base pair InDel mutations in coding microsatellites with a length of six or more nucleotides. Here, 30 and 79 neoepitope candidates were identified at an FDR of 1 % in the HCT-116 and RKO cell lines, respectively. For HCT-116, one neoepitope candidate matched to the CCLE database, three candidates were found in both CCLE and COSMIC databases, four candidates were found in the COSMIC database, one candidate peptide was found in the COSMIC database and the MNR database, and 16 neoepitope candidates matched to the MNR database. The remaining five candidates were SNP neoepitopes matching to the TCLP database. No neoepitope candidate was identified in the HCT-116 cell line using the ReFrame database. For RKO, four neoepitope candidates were found using the CCLE database, one candidate was identified using the ReFrame database, and 69 neoepitope candidates were matching the MNR database. Again, five SNP neoepitope candidates were identified using the TCLP database. No neoepitope candidates were found in the RKO cell line using the COSMIC database.

As the first step of neoepitope validation, BLASTp was used to exclude that the identified peptides match known human aa sequences. During this step, also leucine/isoleucine permutations of the neoepitope candidates were searched against the human proteome because standard LC-MS/MS is not able to distinguish these aa because of their identical molecular mass. Based on this analysis, seven candidates each were excluded from further evaluation in the HCT-116 and RKO datasets, respectively. Of note, in all cases, the identified peptide sequences matched to non-mutated parts of the corresponding source proteins. This can be explained by the fact that during database generation the 14 normal aa upstream of the frameshifted sequence were added in order to include potential neoepitopes located at the transition between normal and frameshifted sequence. Next, synthetic peptide counterparts were obtained for the remaining 95 neoepitope candidates and subjected to the same MS acquisition and data analysis workflow previously used for the identification of neoepitopes from cell line samples.

For HCT-116, spectra of synthetic peptides were obtained for 22 of the 23 candidates and then further compared to the spectra of neoepitope candidates to confirm their identity and exact amino acid sequence (Figure 40 and Figure 41). To this end, the correlation of matched fragment ion intensities was computed using an in-house developed *R* script. Five of the candidates showed poor

correlation (0.073-0.801) between the spectra obtained from cell lines and synthetic peptides and were discarded as likely false-positive identifications. The remaining 17 neoepitope candidates for the HCT-116 cell line showed a Pearson's correlation larger than 0.9 and were further considered.

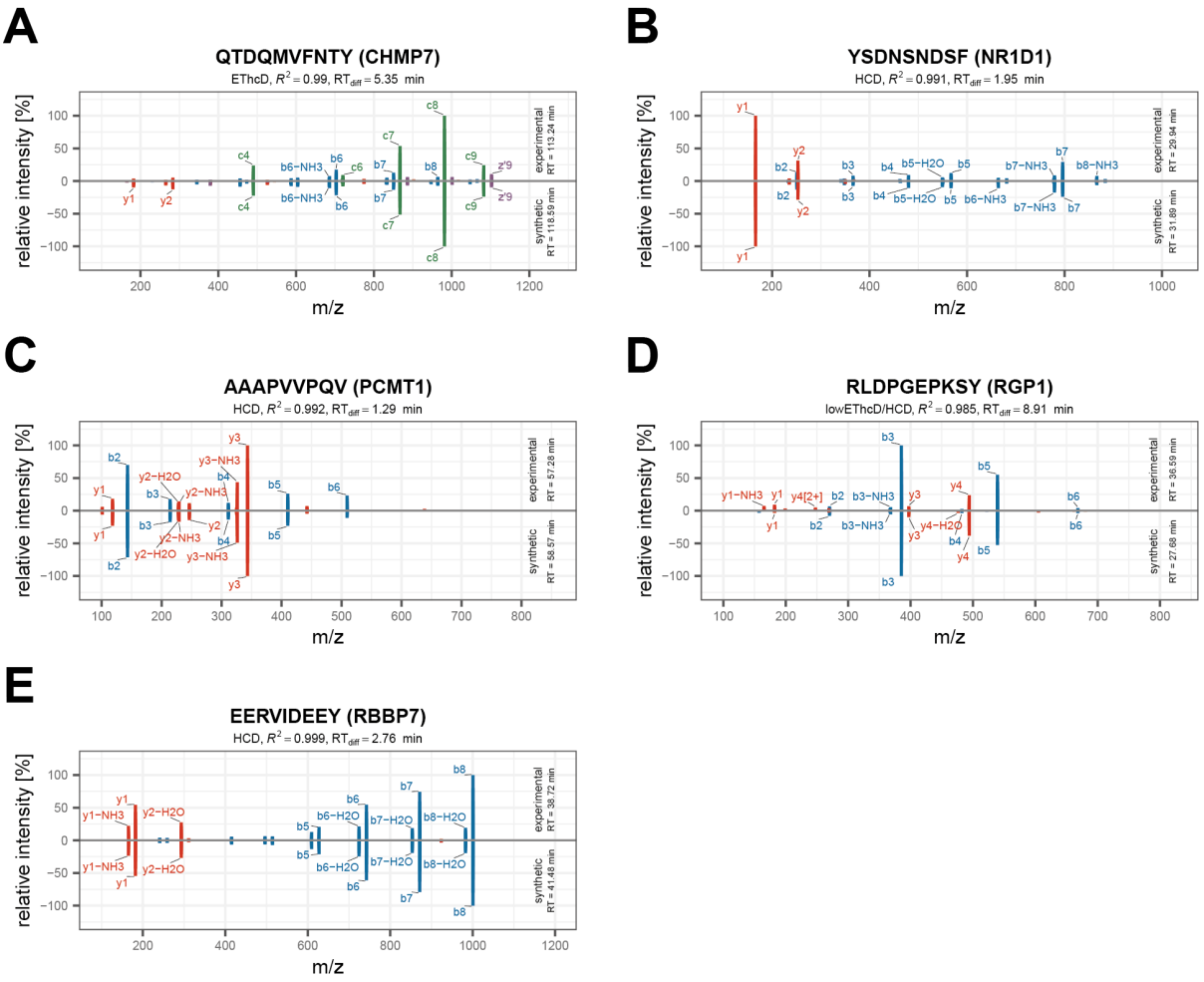


Figure 40: Synthetic peptide validation of SNP neoepitope candidates identified in HCT-116 cells. In each panel, matched ions observed in candidate spectra (top) are compared to matched ions observed in synthetic peptide spectrum (bottom). Top 10 most intense ions are labeled, the correlation between experimental and synthetic peptide spectrum and retention time difference are reported.

As the second criterium for correct identification, retention times between the synthetic peptide and the peptide identification obtained from the cell line sample were compared. Two of the InDel neoepitope candidates with high spectrum agreement showed a retention time difference larger than 10 minutes, and one of them was excluded from further analysis. The peptide KRSSTILRL had a retention time difference of 11.55 minutes (Figure 41B) for the two best matching spectra. However, it was further considered because the retention time of the synthetic peptide showed a very broad range overlapping the retention time observed for the peptide identification in the cell line sample and the predicted HI and observed retention time in the cell line sample for KRSSTILRL are close to the calculated regression line in Figure 35A (left panel) suggesting a correct identification. Moreover, the peptide KRSSTILRL is of special interest because it was only identified using the MS method utilizing lowEThcD fragmentation. The added benefit of dual fragmentation for unambiguous peptide identification is exemplarily shown

with the peptide REKHSWHEP (Figure 41E). Although it was identified in both, samples measured using HCD and EThcD fragmentation, the sequence coverage is much more complete in the EThcD spectrum (left panel) compared to the HCD spectrum (right panel). While the HCD spectrum contained only 13 matching fragment ions not covering the whole peptide sequence, the EThcD spectrum contains 23 matching fragment ions resulting in a more unambiguous identification. This is also translated into a higher identification score for the EThcD spectrum compared to the HCD spectrum (37.81 versus 13.85). Of the remaining 16 candidates, five were classified as SNP neoepitopes with only one amino acid changed (Figure 40) and eleven candidates were classified as InDel mutation-derived neoepitopes (Figure 41). Three of the identified SNP neoepitopes originating from substitutions in *CHMP7*, *NR1D1*, and *RBBP7* have been previously reported (Bassani-Sternberg *et al.*, 2015).

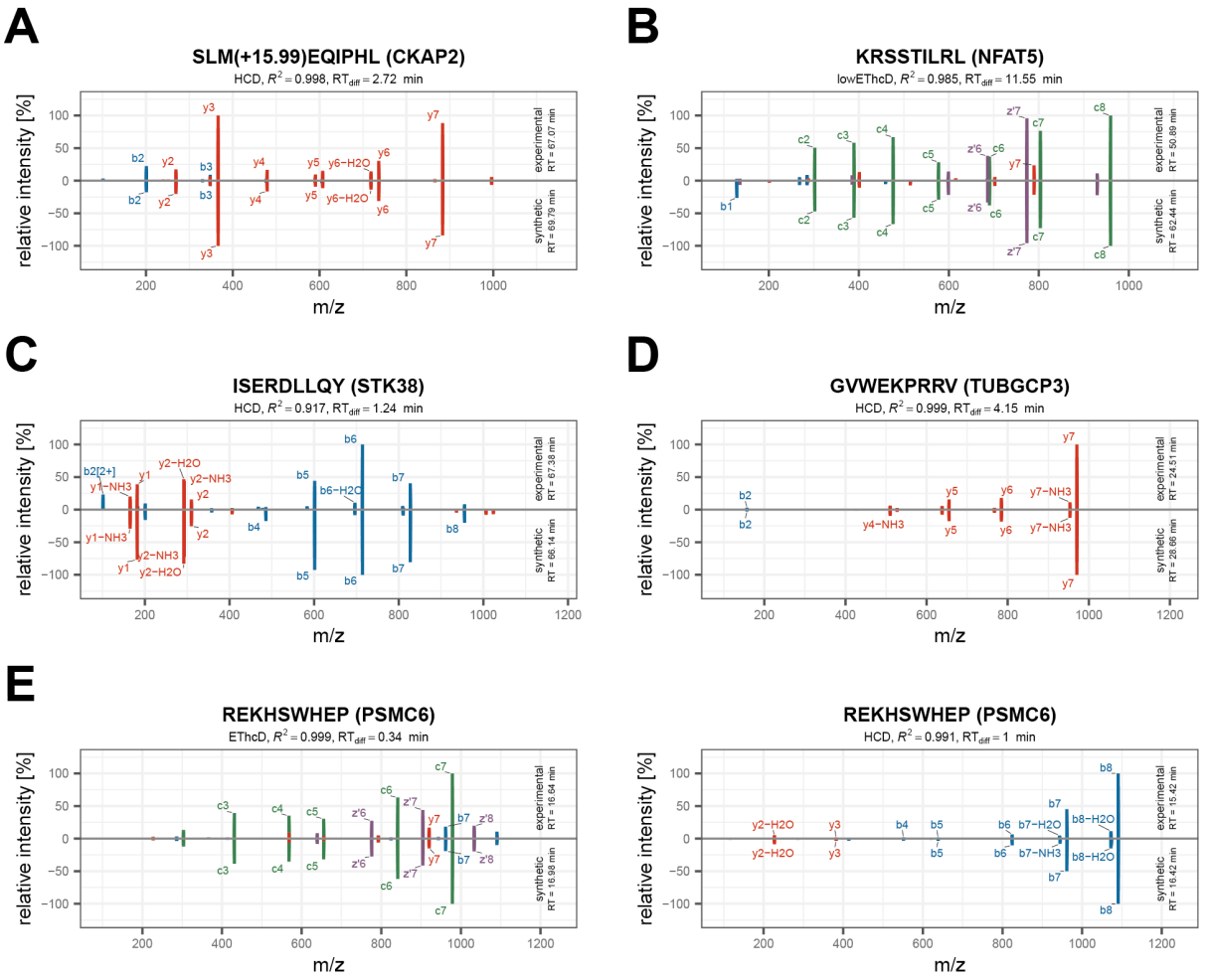


Figure 41: Synthetic peptide validation of representative InDel neoepitope candidates identified in HCT-116 cells. In each panel, matched ions observed in candidate spectra (top) are compared to matched ions observed in synthetic peptide spectrum (bottom). Top 10 most intense ions are labeled, the correlation between experimental and synthetic peptide spectrum and retention time difference are reported. For the *PSMC6*-derived peptide REKHSWHEP (panel E), both EThcD and HCD spectra comparisons are shown.

For RKO, spectra of synthetic counterparts were recorded for 70 of the 72 candidate peptides. Ten neoepitope candidates showed a correlation of matched fragment ion intensities smaller than 0.9 between synthetic peptide spectra and spectra obtained from cell line samples and were excluded from

further analysis. Eight additional candidates were excluded based on their differences in measured retention times in synthetic peptide samples and cell line samples. Of the remaining 52 neoepitope candidates, five were classified as SNP neoepitopes with only one amino acid changed (Figure 42) and 47 candidates were classified as InDel mutation-derived neoepitopes (Figure 43). Of note, all InDel neoepitope candidates identified for RKO and validated to this point matched either the ReFrame or MNR database, but not one of the cell line-specific databases generated based on publicly available sequencing data.

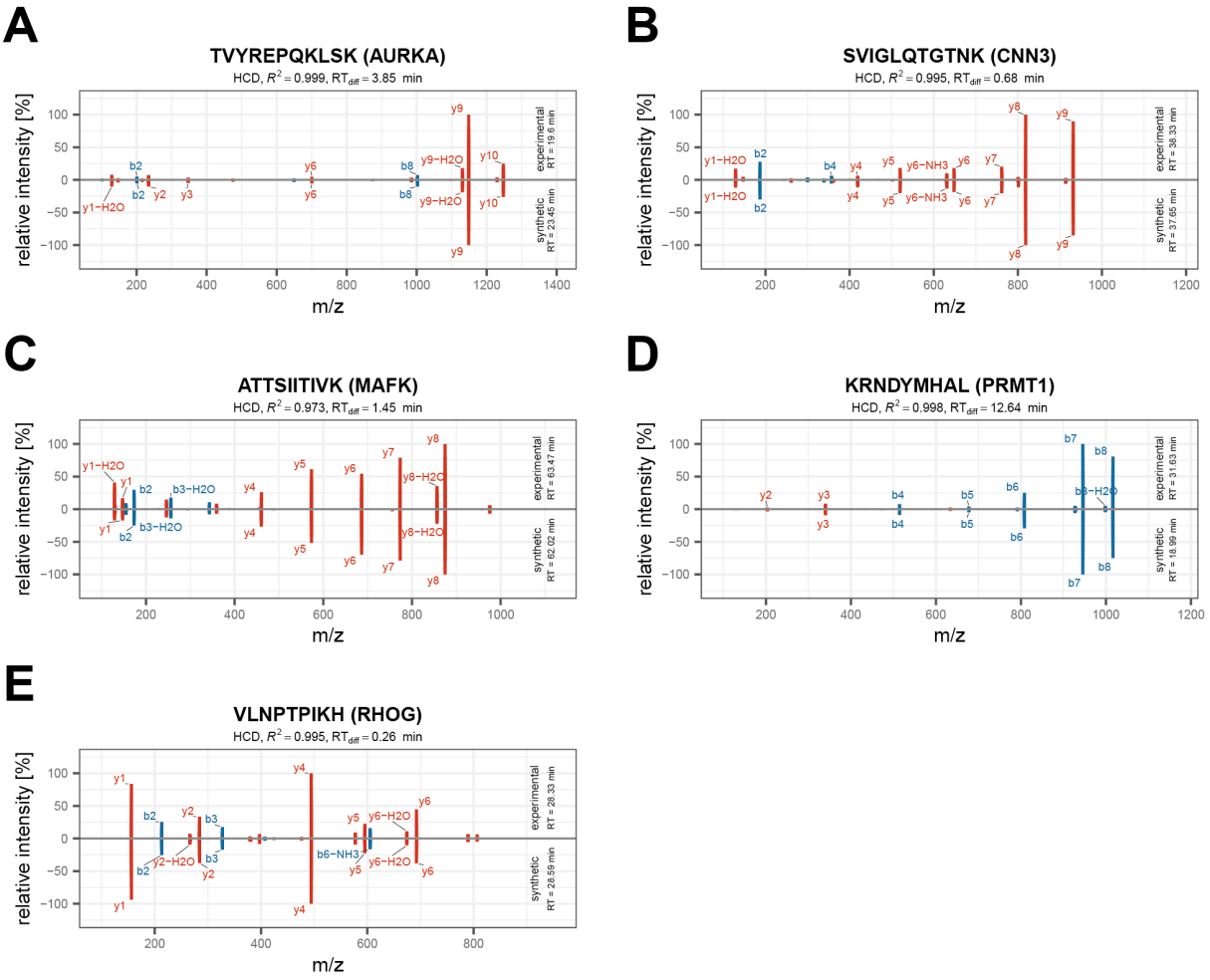


Figure 42: Synthetic peptide validation of SNP neoepitope candidates identified in RKO cells. In each panel, matched ions observed in candidate spectra (top) are compared to matched ions observed in synthetic peptide spectrum (bottom). Top 10 most intense ions are labeled, the correlation between experimental and synthetic peptide spectrum and retention time difference are reported.

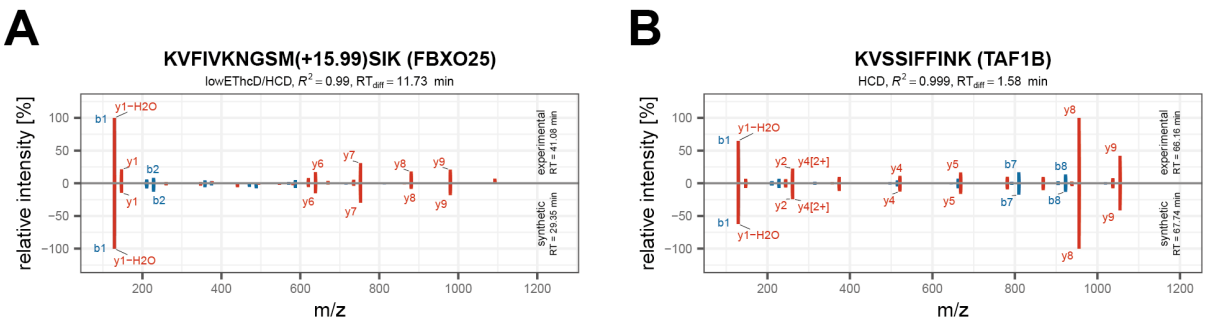


Figure 43: Synthetic peptide validation of representative InDel neoepitope candidates identified in RKO cells. In each panel, matched ions observed in candidate spectra (top) are compared to matched ions observed in synthetic peptide spectrum (bottom). Top 10 most intense ions are labeled, the correlation between experimental and synthetic peptide spectrum and retention time difference are reported.

3.5.2.2.2 Genomic validation of mutation-derived neoepitopes

For the mutation-derived neoepitopes that showed good intensity correlation with their synthetic counterpart, validation was performed at the genomic level using Sanger sequencing. To this end, primers were designed spanning the putative site of mutation based on the information included in the previously constructed neoepitope databases. For InDel mutation-derived neoepitopes, it was also sought to include as much of the upstream and downstream coding sequence as possible and, where feasible, the region encoding the neoepitope. Of the 16 neoepitope candidates identified in the HCT-116 cell line, the underlying mutations of the five SNP neoepitopes, as well as five out of the eleven InDel mutation-derived neoepitopes, were confirmed (Figure 44 and Figure 45). SNP neoepitope-generating mutations were all found to be heterozygous, except for the p.N61D mutation in RBBP7 (Figure 44).

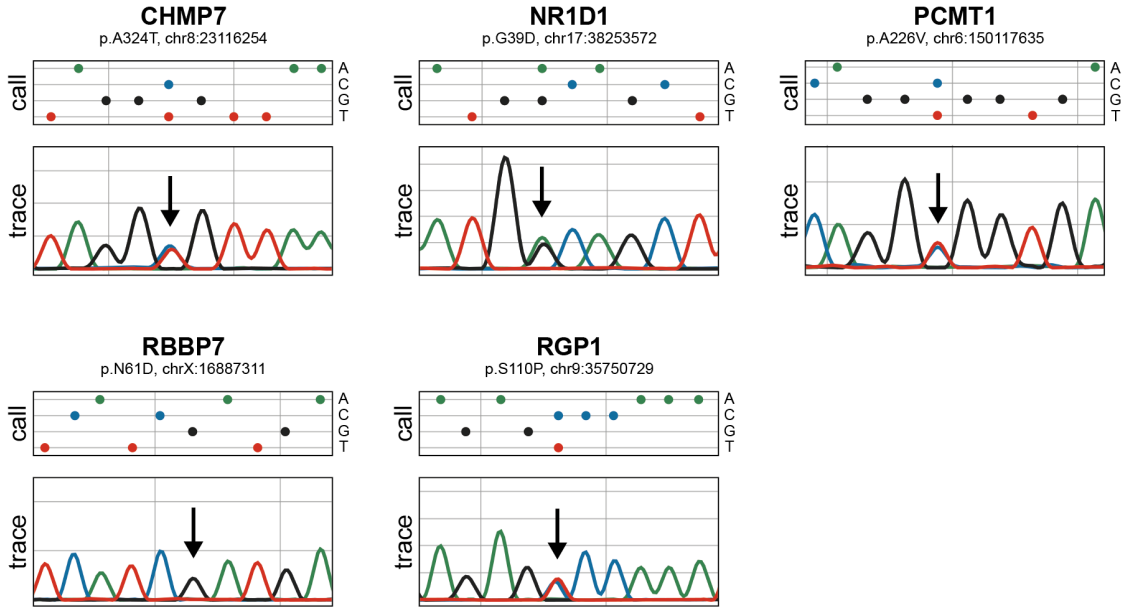


Figure 44: Genomic validation of SNP neoepitope candidates identified in HCT-116 cells. Base calls and Sanger traces of underlying SNPs as visualized by the Indigo webtool. Positions of SNP mutations are indicated by an arrow. *RBBP7* mutation is homozygous.

Underlying frameshift mutations of the InDel neoepitopes were confirmed for *CKAP2*, *NFAT5*, *PSMC6*, *STK38*, and *TUBGCP3* (Figure 45). All confirmed mutations were previously annotated either in the CCLE or the COSMIC database. Except for *CKAP2*, which has a one-base deletion in an A8 repeat, none of the other mutations was included in the MNR databases due to their repeat length. Moreover, none of the genes was listed in the ReFrame database of recurrent mutations in MSI cancers. For the seven candidates which were identified using the MNR database (*DYNC2LI1*, *EWSR1*, *MSL2*, *NMT2*, *RPS2*, *SIAH2*) the putative underlying mutations could not be confirmed by Sanger sequencing.

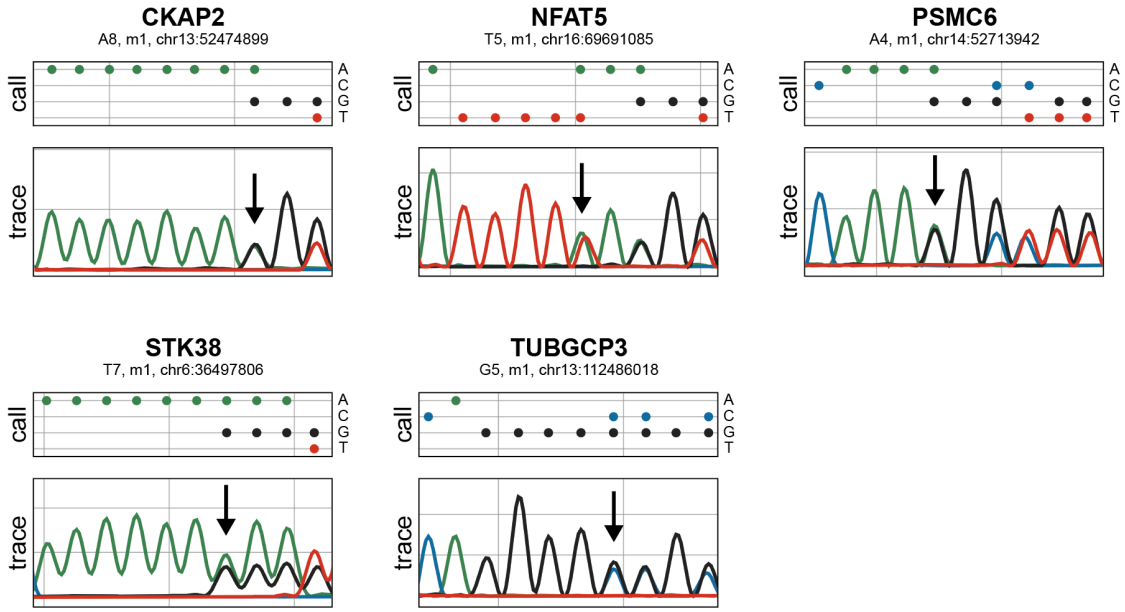


Figure 45: Genomic validation of representative InDel neoepitope candidates identified in HCT-116 cells. Base calls and Sanger traces of underlying frameshift mutations as visualized by the Indigo webtool. Positions of InDel mutations are indicated by an arrow. m1, one-base pair deletion.

For RKO, mutations underlying the five SNP neoepitopes (*AURKA*, *CNN3*, *MAFK*, *PRMT1*, *RHOG*) were confirmed by Sanger sequencing and found to be heterozygous (Figure 46).

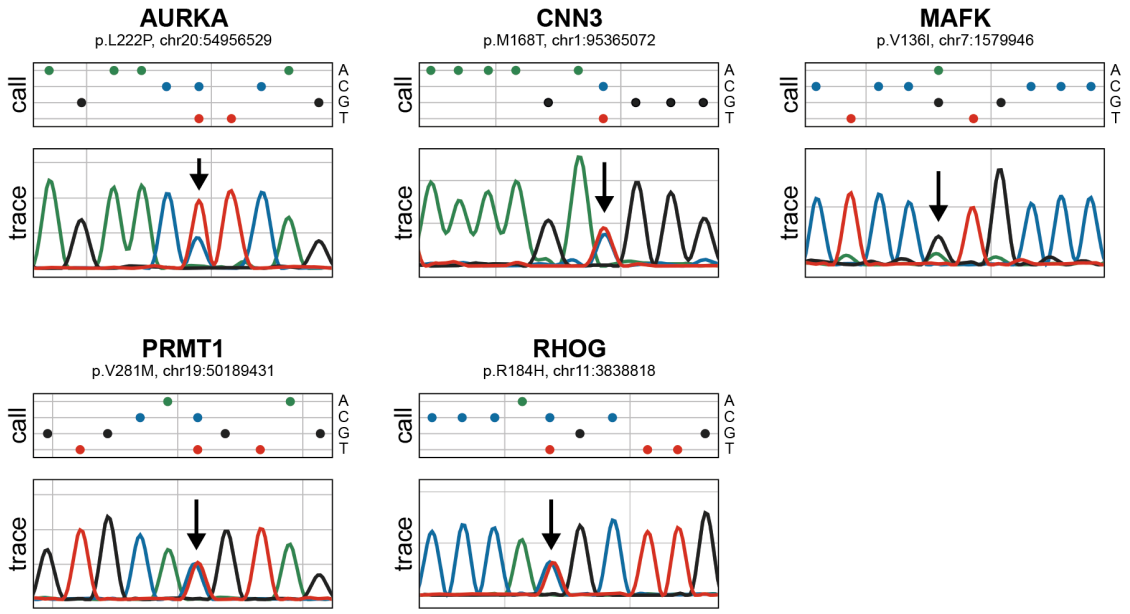


Figure 46: Genomic validation of SNP neoepitope candidates identified in RKO cells. Base calls and Sanger traces of underlying SNPs as visualized by the Indigo webtool. Positions of SNP mutations are indicated by an arrow.

For the 47 remaining InDel neoepitope candidates after synthetic peptide validation in the RKO cell line, underlying mutations were confirmed in only two genes, *FBXO25* and *TAF1B* (Figure 47). Here, the mutation calling for the A11 repeat of *TAF1B* remained ambiguous and could be either a one-base or two-base deletion. The identified peptide (KVSSIFFINK) originates from a one-base deletion while the mutation is reported to be a two-base deletion in the RKO cell line (Ballhausen *et al.*, 2020). Both the chromatogram trace and the base call indicate that the RKO cell line mainly exhibits a two-base deletion and only a small fraction exhibits a one-base deletion (Figure 47, right panel).

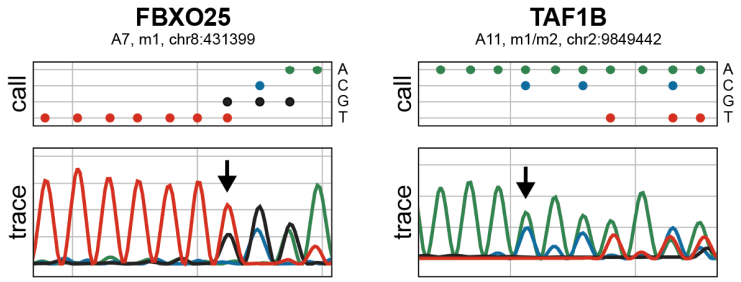


Figure 47: Genomic validation of representative InDel neoepitope candidates identified in HCT-116 cells. Base calls and Sanger traces of underlying frameshift mutations as visualized by the Indigo webtool. Positions of InDel mutations are indicated by an arrow. m1, one-base pair deletion; m2, two-base pair deletion.

3.5.2.2.3 Summary of identified and validated neoepitopes

In total, 17 HLA class I-presented neoepitopes were validated for HCT-116 and RKO cells. For HCT-116, five SNP and five InDel neoepitopes were validated, while five SNP and two InDel neoepitopes were confirmed for the RKO cell line. Table 23 - Table 26 provide an overview of the identified SNP and InDel neoepitopes in both HCT-116 and RKO cells.

Table 23: Overview of validated SNP neoepitopes in HCT-116 cells. HLA binding prediction was performed using NetMHCpan 4.0. Mutated aa originating from SNP are underlined. Peptides previously identified by mass spectrometry (Bassani-Sternberg et al., 2015) are marked with a dagger.

gene	peptide	location of mutation (GRCh37)	aa change	HLA binding prediction (K_d in nM)
CHMP7	QTDQMVFN**T**Y†	chr8: 23116254	p.A324T	HLA-A*01:01 (16), HLA-B*18:01 (4799)
NR1D1	YSDNSN**D**SF†	chr17: 38253572	p.G39D	HLA-A*01:01 (166), HLA-C*05:01 (23)
PCMT1	AAAPVVPQ**V**	chr6: 150117635	p.A226V	HLA-C*05:01 (2282), HLA-C*07:01 (4776)
RBBP7	EERVI**D**EEY†	chrX: 16887311	p.N61D	HLA-B*18:01 (107)
RGP1	RLDPGE**P**KSY	chr9: 35750729	p.S110P	HLA-A*01:01 (2083), HLA-C*05:01 (6887)

Table 24: Overview of validated SNP neoepitopes in RKO cells. HLA binding prediction was performed using NetMHCpan 4.0. Mutated aa originating from SNP are underlined.

gene	peptide	location of mutation (GRCh37)	aa change	HLA binding prediction (K_d in nM)
AURKA	TVYRE**P**QKLSK	chr20: 54956529	p.L222P	HLA-A*03:01 (140)
CNN3	SVIGLQ**T**GTNK	chr1: 95365072	p.M168T	HLA-A*03:01 (552)
MAFK	ATTS**I**ITIVK	chr7: 1579946	p.V136I	HLA-A*03:01 (102)
PRMT1	KRNDY**M**HAL	chr19: 50189431	p.V281M	HLA-C*07:01 (54)
RHOG	VLNPTPIK**H**	chr11: 3838818	p.R184H	HLA-A*03:01 (703)

In the following, the key characteristics of the identified InDel neoepitopes and their reported functions of the corresponding source proteins are summarized and, if reasonable, possible mechanisms by which functional inactivation of the source proteins might contribute to the tumor's fitness by increasing its survival and growth rate are considered.

The InDel neoepitope SLMEQIPHL originates from a one-base deletion in an A8 repeat of the Cytoskeleton Associated Protein 2 (*CKAP2*) gene of HCT-116 leading to a frameshifted stretch of 14 aa. CKAP2 is involved in maintaining genome stability (Case *et al.*, 2013) and CKAP2 overexpression in

p53-competent HCT-116 cells has been shown to enhance DNA damage-induced p53 function resulting in cell cycle arrest and apoptosis (Tsuchihara *et al.*, 2005).

The InDel neoepitope KRSSTILRL identified in HCT-116 originates from a one-base deletion in a T5 repeat of the Nuclear Factor of Activated T Cells 5 (*NFAT5*) gene leading to a frameshifted stretch of 14 aa. While the N-terminal part ("KRSSTI") of the identified InDel neoepitope originates from the NFAT5 wt aa sequence, the C-terminal aa "LRL" result from the frameshift mutation and enable the binding to the HLA-C*07:01 allele with leucine at the anchor position 9. Transcription factor activity of NFAT5 impacts various signaling pathways such as mTOR and Notch via downregulation of the negative regulator REDD1 (Zhou *et al.*, 2014) as well as on Wnt signaling by repressing β-catenin acetylation (Wang *et al.*, 2013). Impaired NFAT5 activity might, therefore, add to the tumor's fitness by activation of these pathways leading to uncontrolled cell growth and downregulation of apoptosis (Zou *et al.*, 2020).

The InDel neoepitope REKHSWHEP originates from a one-base deletion in an A4 repeat of the 26S proteasome regulatory subunit 10B (*PSMC6*) gene of HCT-116 leading to a frameshifted stretch of 28 aa. PSMC6 is part of the 19S proteasome regulatory subunit involved in the unfolding of ubiquitinated substrates and their translocation into the catalytic part of the proteasome (Boland *et al.*, 2016; Yedidi *et al.*, 2017). siRNA-mediated knockdown of PSMC6 *in vitro* in HCT-116 cells has been associated with reduced proteasomal activity and with reduced tumor growth and increased survival *in vivo* (Boland *et al.*, 2016). In contrast, low PSMC6 expression correlated with decreased disease-free and overall survival in patients with Stage II colon cancer (McCawley *et al.*, 2012). Reduced PSMC6 expression and consequently reduced proteasome activity, might hinder the processing and presentation of HLA class I-presented peptides and thereby add to the tumor's fitness.

The InDel neoepitope ISERDLLQY identified in HCT-116 originates from a one-base deletion in a T7 repeat of the Serine/Threonine Kinase 38 (*STK38*) gene leading to a frameshifted stretch of 70 aa. STK38 was shown to negatively regulate the activation of MEKK1/2 (Enomoto *et al.*, 2008) and, together with STK38L, is essential for Fas receptor-induced apoptosis (Vichalkovski *et al.*, 2008). Inactivation of this gene is therefore predicted to interfere with apoptosis thus increasing cell viability.

The InDel neoepitope GVWEKPRRV originates from a one-base deletion in a G5 repeat of the Tubulin Gamma Complex Associated Protein 3 (*TUBGCP3*) gene of HCT-116 leading to a frameshifted stretch of 108 aa. TUBGCP3 is part of the γ-tubulin complex and necessary for nucleation of microtubules at the centrosome (Dráberová *et al.*, 2015).

Table 25: Overview of validated InDel neoepitopes in HCT-116 cells. HLA binding prediction was performed using NetMHCpan 4.0. Underlined aa of the NFAT5-derived InDel neoepitope originate from the wild-type protein sequence. m1, one-base deletion.

gene	peptide	location of mutation (GRCh37)	repeat & mutation type	number of frameshift aa	HLA binding prediction (K_d in nM)
CKAP2	SLMEQIPHL	chr13: 52474899	A8, m1	14	HLA-A*02:01 (2), HLA-C*05:01 (1250), HLA-C*07:01 (1665)
NFAT5	**KRSSTI**LRL	chr16: 69691085	T5, m1	14	HLA-C*07:01 (203)
PSMC6	REKHSWHEP	chr14: 52713943	A4, m1	28	HLA-B*45:01 (1218)
STK38	ISERDLLQY	chr6: 36497806	T7, m1	70	HLA-A*01:01 (33)
TUBGCP3	GVWEKPRRV	chr13: 112486018	G5, m1	108	no binder

The InDel neoepitope KVFIVKNGSMSIK identified in RKO originates from a one-base deletion in an A7 repeat of the F-box Protein 25 (*FBXO25*) gene leading to a frameshifted stretch of 45 aa. F-box proteins are part of the E3 ubiquitin ligase complex SCF where they function as substrate receptors and thus involved in the ubiquitin-proteasome degradation pathway (Heo *et al.*, 2016; Randle and Laman, 2016). FBXO25 has been found to act as a tumor suppressor by mediating the ubiquitination and initiating the degradation of the proto-oncogene regulator ELK-1 and the anti-apoptotic HAX-1 (Teixeira *et al.*, 2013; Baumann *et al.*, 2014). Moreover, heterozygous deletion of *FBXO25* in mantle cell lymphoma cells was shown to promote their survival (Baumann *et al.*, 2014). Similarly, heterozygous inactivation of FBXO25 function by InDel mutations in microsatellite-instable cancers is conceivable and might add to the tumors' fitness.

The InDel neoepitope KVSSIFFINK originates from a one-base deletion in an A11 repeat of the TATA box-binding protein-associated factor RNA polymerase I subunit B (*TAF1B*) gene of RKO leading to a frameshifted stretch of 28 aa. TAF1B is involved in rRNA synthesis as part of the RNA polymerase I transcription initiation complex (Knutson and Hahn, 2011). Although the *TAF1B* gene has been suggested as a target gene in MSI CRCs almost 20 years ago and shows high mutation frequencies in MSI cancers (Kim *et al.*, 2002; Kloor and von Knebel Doeberitz, 2016), its role in tumorigenesis remains elusive.

Table 26: Overview of validated InDel neoepitopes in RKO cells. HLA binding prediction was performed using NetMHCpan 4.0. m1, one-base deletion.

gene	peptide	location of mutation (GRCh37)	repeat & mutation type	number of frameshift aa	HLA binding prediction (K_d in nM)
FBXO25	KVFIVKNGSMSIK	chr8: 431399	A7, m1	45	HLA-A*03:01 (40)
TAF1B	KVSSIFFINK	chr2: 9849442	A11, m1	28	HLA-A*03:01 (17)

3.5.3 Quantifying the immunopeptidome

After having obtained a comprehensive view of the HLA class I-presented peptides including both endogenous peptides and mutation-derived neoepitopes, the effect of 5AZA treatment on the immunopeptidome was evaluated using label-free quantification. However, because of the heuristic nature of precursor selection applied in data-dependent MS acquisition strategies, a major obstacle was identified in the overall number of quantified peptides in the recorded dataset of different biological replicates and with different MS fragmentation and precursor selection methods. “Quantifiable peptides” were defined as peptides for which intensity values were obtained in at least two out of three biological replicates per condition and MS methodology. Exemplified for the HCT-116 dataset, 5,204 endogenous peptides (52 %) were quantifiable as opposed to the 10,030 unique peptides identified in the whole dataset (Figure 48). Only 485 peptides (4.8 % of the whole dataset, 9.3 % of the quantifiable peptides) were covered by all three MS methodologies used. Given that a large proportion of peptides is only quantifiable using the HCD dataset (3,582 peptides), it was decided to perform label-free quantification separately for the different MS methodologies. For HCT-116, none of the identified InDel neoepitopes was quantifiable according to the previously defined strict inclusion criteria of "quantifiable peptides”. Intensity values for the *CKAP2*-derived peptide SLMEQIPHL and the *TUBGCP*-derived peptide GVWEKPRRV were obtained in four biological replicates measured with HCD (1X DMSO, 3X 5AZA). For the *PSMC6*-derived peptide REKHSWHEP intensity values were obtained in one sample of cells treated with 5AZA and measured with HCD as well as in two 5AZA-treated samples measured with the EThcD method. The *STK38*-derived peptide ISERDLLQY was quantified only in one HCD-measured 5AZA sample, while for the *NFAT5*-derived peptide KRSSTILRL, intensity values were only obtained for two biological replicates of 5AZA-treated cells measured with the lowEThcD method. In contrast, for the two identified and validated InDel neoepitopes in RKO, intensity values were obtained either in all samples measured using HCD (*TAF1B*-derived peptide KVSSIFFINK) or in all samples measured with EThcD (*FBXO25*-derived peptide KVFIVKNGSMSIK).

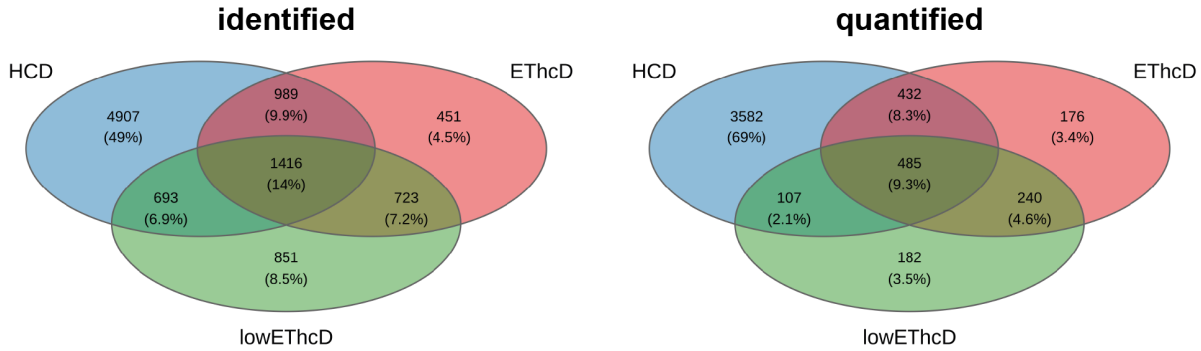


Figure 48: Representative overview of identified and quantified peptides for the HCT-116 cell line. Overlap of peptides identified (left panel) and quantified (right panel) using either HCD (blue), EThcD (red), or lowEThcD (green).

Moreover, as previously observed during MS parameter optimization, data obtained using the lowEThcD methodology showed only moderate intensity correlation with samples recorded using HCD fragmentation and standard precursor selection (Figure 49). While the lowEThcD dataset added valuable identification data, *e.g.* for the *NFAT5*-derived neoepitope KRSSTILRL, quantification data for

only 182 additional peptides not covered by the EThcD or HCD dataset is added. For this reason, the lowEThcD was solely used for identification but not quantification purposes.

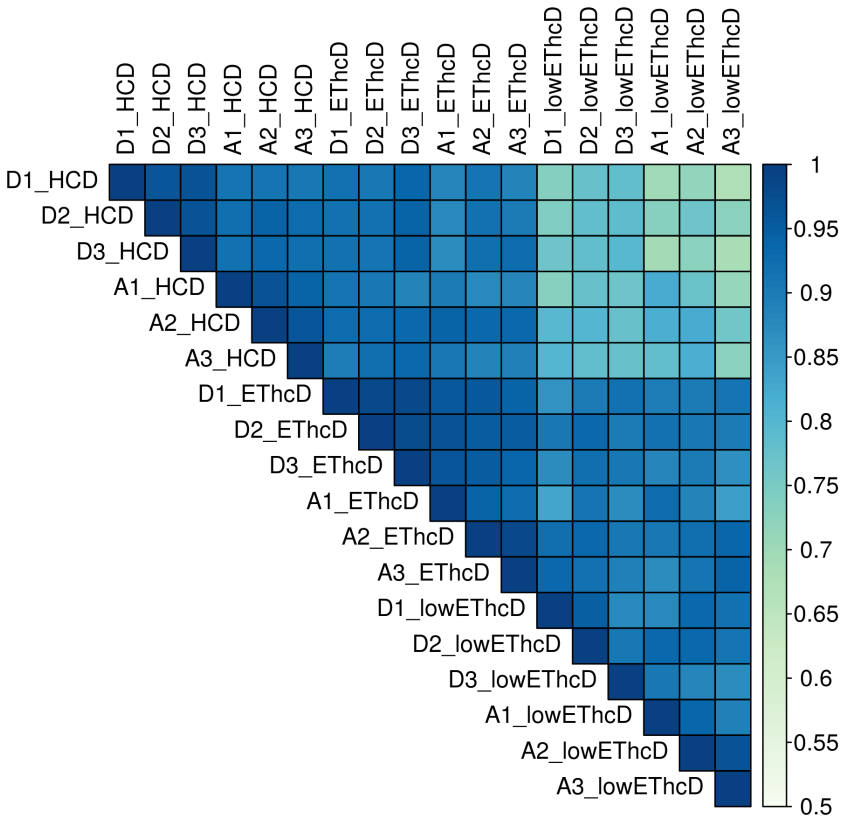


Figure 49: Representative correlation matrix for HCT-116 samples. Matrix showing correlation of raw intensity values between all datasets recorded for HCT-116. Samples are labeled with condition (D = DMSO, A = 5AZA), replicate number (1-3) and MS method (HCD, EThcD, lowEThcD). Color represents degree of correlation (light green = low, dark blue = high).

3.5.3.1 A targeted MS2 method for InDel neoepitope quantification

As previously discussed, label-free quantification of InDel neoepitopes in the HCT-116 cell line was not feasible using the existing data because of missing values. To address this issue, a targeted MS2 method was established to obtain quantification data for the five identified and validated InDel neoepitopes using the mass and retention time information obtained from data-dependent acquisition of both cell line samples and synthetic peptide samples (Figure 50A). In addition to the five InDel neoepitopes of interest, five “standard” peptides that were unchanged after 5AZA treatment in the initial dataset were included as controls and the previously reported HLA-A*02:01-restricted InDel neoepitope RLSSCVPVA derived from a recurrent one-base deletion in an A10 repeat of the *TGFBR2* gene was included (Sæterdal *et al.*, 2001; Woerner *et al.*, 2001).

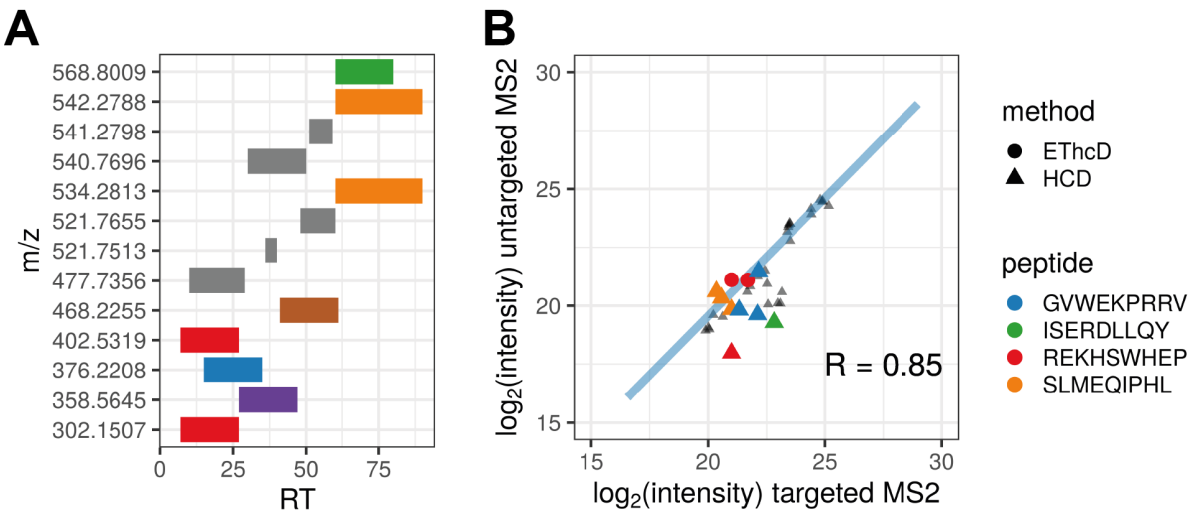
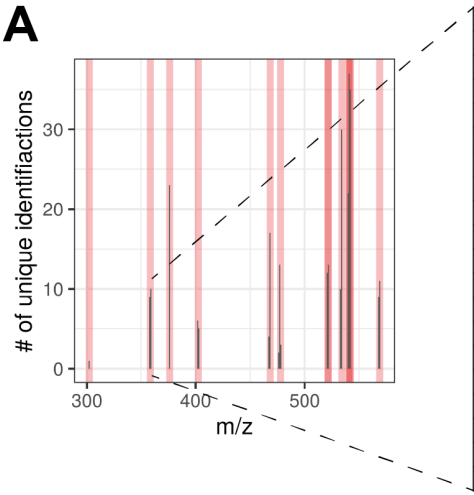


Figure 50: Targeted MS2 approach for InDel neoepitope quantification in HCT-116 cells. Graphical overview of targeted MS2 approach showing targeted m/z values and retention time windows. InDel neoepitopes are colored, grey represents standard peptides (panel A). Intensities of targeted InDel neoepitopes and standard peptides identified using a targeted MS approach versus intensities of targeted InDel neoepitopes and standard peptides identified using an untargeted MS2 approach (panel B).

The same samples previously measured were analyzed again using the targeted MS2 approach with HCD and EThcD fragmentation. Here, intensity values were obtained from all samples for the InDel neoepitope derived from *PSMC6* (REKHSWHEP), *STK38* (ISERDLLQY), and *NFAT5* (KRSSTILRL). For the *TUBGCP3*-derived InDel neoepitope GVWEKPRRV and the *CKAP2*-derived InDel neoepitope SLMEQIPHL intensity values were obtained from all samples but one and four, respectively. The intensities measured with the targeted MS2 approach correlated well with those previously obtained with the untargeted approach (Pearson's correlation coefficient: 0.85; Figure 50B). Further analysis of the targeted MS2 dataset using the previously established identification pipeline revealed that not only the peptides of interest were identified, but also 267 other peptides (Figure 51A). Of these, 177 were previously identified in the untargeted MS2 dataset and their intensities showed a high correlation between the targeted and untargeted MS2 approaches (Pearson's correlation coefficient: 0.93; Figure 51B). Considering this high degree of correlation, it was decided to integrate the intensity values obtained using the targeted MS2 approach into the dataset previously obtained using a data-dependent acquisition strategy.



peptide	length	mass	m/z	score	retention time	gene
NELERLRAA	9	1070.5833	357.868	23.9	34.53	NUMA1
ALQKRLDEV	9	1070.6084	357.877	28.85	27.87	RRBP1
EEQKIVVKV	9	1070.6335	357.8848	34.76	41.47	RL34
EEAQHFIKA	9	1071.5349	358.1853	54.28	29.01	AP4AT
MEQFVKKY	8	1071.5422	358.1879	43.92	33.38	TES
RADNHFLLS	9	1071.5461	358.1892	29.4	46.36	PKD1L1
KLDPHNHVL	9	1071.5825	358.2016	27.57	27.95	STIP1
RLVESKVEL	9	1071.6288	358.2173	43.58	45.71	SAR1A
SRSVIIREL	9	1071.64	358.2206	43.34	47.03	UBR5
YTDKIDRY	8	1072.5189	358.5139	42.75	28.13	TSN4
KEHLELYAA	9	1072.5553	358.5255	41.16	43.94	ABCAA
TEIEHKVAF	9	1072.5553	358.5258	50.61	37.99	SYNE2
TEVGHVIRY	9	1072.5665	358.5292	42.68	39.19	FAT1
TELDKIRQA	9	1072.5876	358.5359	46.36	30.73	DYHC1
KLSEKIDQL	9	1072.6128	358.5448	33.08	42.91	TRAF7
RRNPGVVIY	9	1072.6141	358.545	36.01	36.1	RM43
KLMKKLLEA	9	1072.6678	358.5645	26.63	32.17	SNAA
KRSSTILRL	9	1072.6716	358.5645	78.32	39.91	NFAT5
KLYGKPIRV	9	1072.6757	358.5659	47.8	29.35	SF3B4

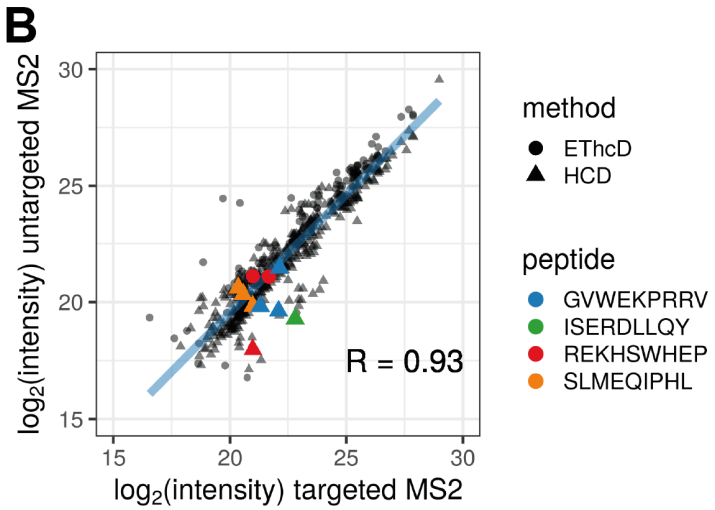


Figure 51: Targeted MS2 approach identifies additional, untargeted peptides. Representative overview of identified peptides for targeted m/z value of 358.5645 (panel A). Intensities of all peptides identified using a targeted MS approach versus intensities of all peptides identified using an untargeted MS2 approach (panel B).

3.5.3.2 NMD inhibition by 5AZA augments HLA class I-mediated presentation of InDel neoepitopes and peptides originating from NMD-targeted transcripts

Label-free quantification of HLA class I-presented peptides was performed for both the RKO and HCT-116 datasets. The RKO dataset contained 5,466 unique peptides, of which 3,193 were quantifiable only in the HCD dataset and 575 unique peptides were only quantifiable in the EThcD dataset. 849 distinct peptides were quantifiable in both datasets (Figure 52A). For the HCT-116 dataset, intensities of InDel neoepitopes obtained with the targeted MS2 approach were integrated into the data-dependent dataset as mentioned before. With this, the final dataset used for label-free quantification of the HCT-116 immunopeptidome contained a total of 5,072 unique peptides. Of these, 3,696 peptides were quantifiable only in the HCD dataset and 429 peptides only in the EThcD dataset. 947 peptides were quantifiable in both datasets (Figure 52B). Both datasets were subjected to a modified, *R*-based quantification workflow (Perez-Perri *et al.*, 2018; Backlund *et al.*, 2020) including filtering, removal of batch effects, variance stabilizing normalization, imputation of missing values, and finally differential expression analysis (data processing overview in Figure 52C and D).

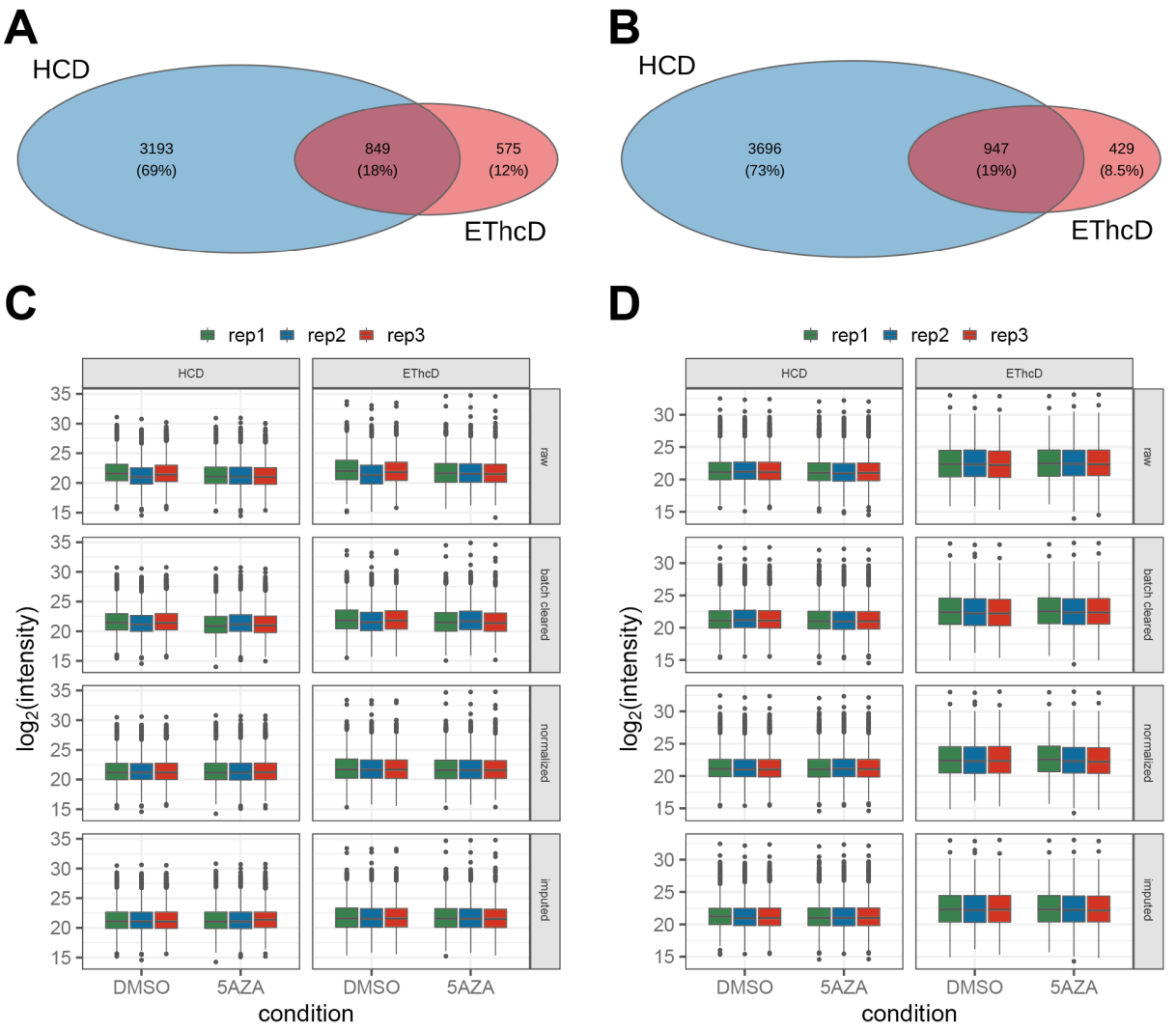


Figure 52: Overview of data processing for label-free quantification of HLA class I-presented peptides. Overlap of quantification datasets for the HCT-116 cell line (panel A) and RKO cell line (panel B) recorded using either HCD (blue) or EThcD (red). Graphical overview of data processing for HCT-116 data (panel C) and RKO data (panel D) from top to bottom: raw data, batch-cleared data, normalized data, normalized data including imputed data for missing values.

For RKO, the differential expression analysis revealed 1,581 peptides showing differential HLA class I-mediated presentation upon NMD inhibition with 5AZA. Of these, 774 peptides showed an increased presentation while 807 peptides were less abundant at the cell surface (Figure 53A). While the *FBXO25*-derived InDel neoepitope showed a slight, but non-significant downregulation after 5AZA treatment, the *TAF1B*-derived InDel neoepitope showed significant, albeit only moderate upregulation (1.4-fold; $p \leq 0.05$) after NMD inhibition by 5AZA (Figure 53B). The effect of 5AZA treatment on the presentation of peptides originating from well-established, previously reported, and ENSEMBL-annotated NMD targets (El-Bchiri *et al.*, 2008; Tani *et al.*, 2012; Colombo *et al.*, 2017; Aliouat *et al.*, 2019) was further analyzed. Here, peptides derived from *ANXA5* (9.9-fold), *DTWD1* (11.9-12.2-fold), *HYOU1* (1.7-fold), and *UPP1* (3.3-fold) were significantly more abundant ($p \leq 0.005$) at the cell surface of RKO cells treated with 5AZA compared to the untreated control (Figure 53C).

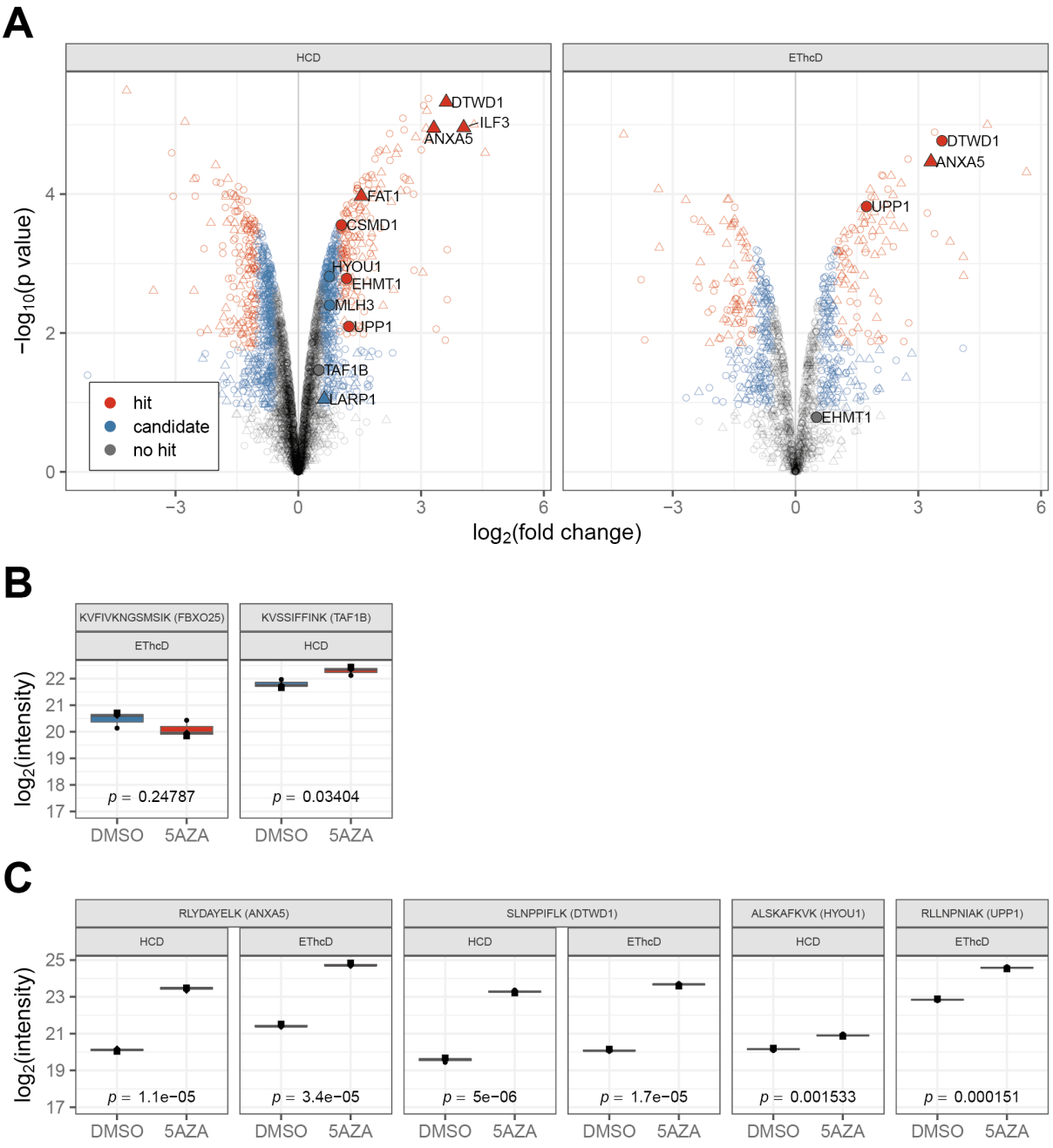


Figure 53: Treatment with 5AZA augments HLA class I-mediated presentation of peptides originating from NMD targeted transcripts in RKO cells. Volcano plot summarizing the *limma* analysis of label-free quantification of the immunopeptidome isolated from 5AZA-treated versus DMSO-treated RKO cells. Color represents hit annotation (red = hit, blue = candidate), shape indicates if values were imputed (circle = no, triangle = yes; panel A). Representative plots showing changes in intensity for identified InDel neoepitopes (panel B) and peptides originating from well-established and putative NMD targets (panel C) after treatment with 5AZA for 24 h. Bars represent 25th to 75th percentiles, middle line represents median, points represent individual measurements of biological replicates.

Next, endogenous peptides originating from the non-frameshifted 5' sequences of frameshift-bearing transcripts were investigated. Here, peptides originating from *CSMD1* (2.1-fold), *EHMT1* (1.4-2.3-fold), *FAT1* (2.9-fold), *LARP1* (1.55-fold), and *MLH3* (1.1-1.7-fold) were found to be upregulated (Figure 54A). Interestingly, the two endogenous peptides originating from the wt part of *TAF1B* were not affected by

NMD inhibition with 5AZA. As expected, the HLA class I-mediated presentation of SNP neoepitopes derived from *AURKA*, *PRMT1*, and *RHOG* was not affected by 5AZA treatment (Figure 54B).

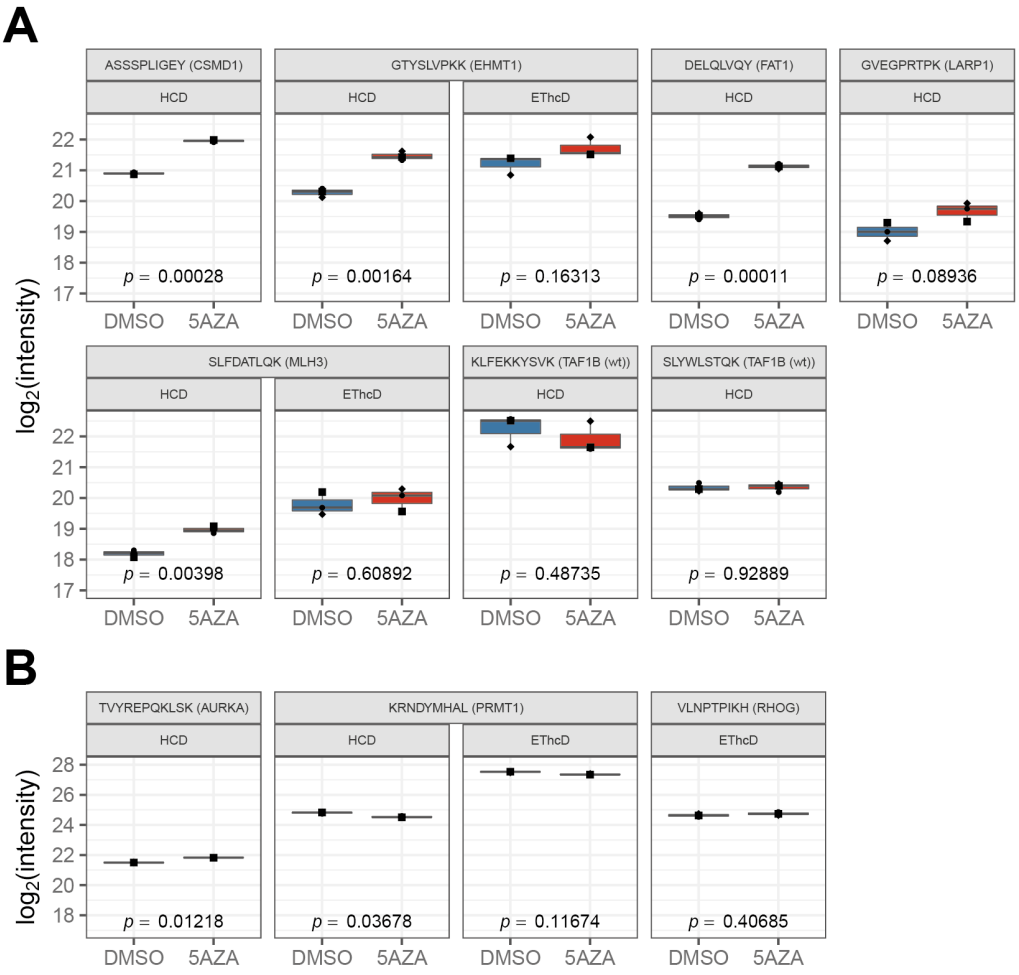


Figure 54: Treatment with 5AZA augments HLA class I-mediated presentation of peptides originating from wild-type parts of frameshift-mutated transcripts in RKO cells. Representative plots showing changes in intensity for peptides originating from wild-type parts of frameshift-mutated transcripts (panel A) and SNP neoepitopes (panel B) after treatment with 5AZA for 24 h. Bars represent 25th to 75th percentiles, middle line represents median, points represent individual measurements of biological replicates.

Label-free quantification of the immunopeptidome isolated from HCT-116 cells treated either with DMSO or 5AZA revealed 838 differentially presented peptides. 434 peptides showed an increased cell surface presentation, while 404 peptides were less abundant after treatment with 5AZA (Figure 55A). While the presentation of the InDel neoepitopes derived from *NFAT5*, *STK38*, and *TUBGCP3* remained unchanged or was slightly decreased after treatment with 5AZA, the *CKAP2*-derived and *PSMC6*-derived InDel neoepitopes showed an increased cell surface presentation. The HLA class I-mediated presentation of the *CKAP2*-derived InDel neoepitope SLMEQIPHL was significantly and substantially upregulated 2.1-2.13-fold ($p \leq 0.005$) and that of the PSMC6-derived InDel neoepitope significantly ($p \leq 0.05$) albeit less strongly (1.2-1.5-fold) in both the HCD and EThcD datasets (Figure 55B). Further analysis of source proteins of upregulated peptides revealed that NMD inhibition significantly, and in many cases strongly upregulates ($p \leq 0.05$) the HLA class I-mediated presentation of peptides

originating from known NMD targets such as *ASAH1* (4.9-fold), *ATF3* (2.0-fold), *ATF4* (1.2-1.7-fold), *CAV1* (3.8-fold), *DDIT3* (two peptides; 1.7-3.1-fold and 1.7-fold, respectively), *DDX5* (2.7-fold), *HYOU1* (3.0-3.1-fold), *JUN* (3.2-fold), *PLIN3* (2.7-fold), *SLIRP* (3.9-fold), and *UPP1* (two peptides; 2.9-fold and 2.5-fold, respectively; Figure 55C and Figure 56A). Analysis of peptides originating from wild-type sequences of frameshift-mutated transcripts revealed significant upregulation ($p \leq 0.0005$) of peptides originating from *CENPF* (2.1-fold), *KIF11* (2.1-fold), and *LARP1* (3.8-fold; Figure 56B). HLA class I-mediated presentation of SNP neoepitopes derived from *CHMP7*, *PCMT1*, and *RBBP7* was not significantly altered following 5AZA treatment (Figure 56C)

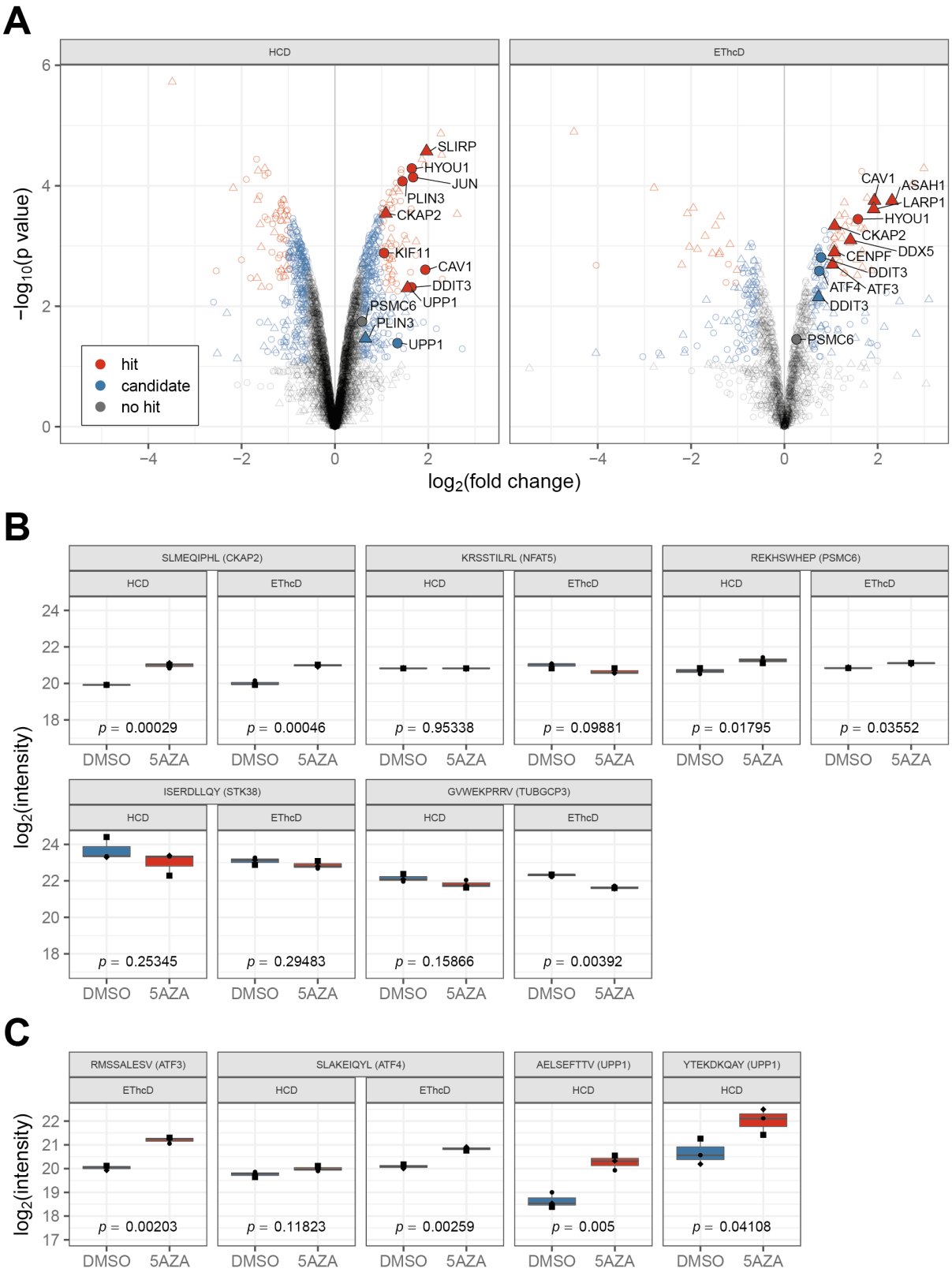


Figure 55: Treatment with 5AZA augments HLA class I-mediated presentation of peptides originating from NMD-targeted transcripts in HCT-116 cells. Volcano plot summarizing the *limma* analysis of label-free quantification of the immunopeptidome isolated from 5AZA-treated versus DMSO-treated HCT-116 cells. Color represents hit annotation (red = hit, blue = candidate), shape indicates if values were imputed (circle = no, triangle = yes; panel A). Representative plots showing changes in intensity for identified InDel neoepitopes (panel B) and peptides originating from well-established NMD targets (panel C) after treatment with 5AZA for 24 h. Bars represent 25th to 75th percentiles, middle line represents median, points represent individual measurements of biological replicates.

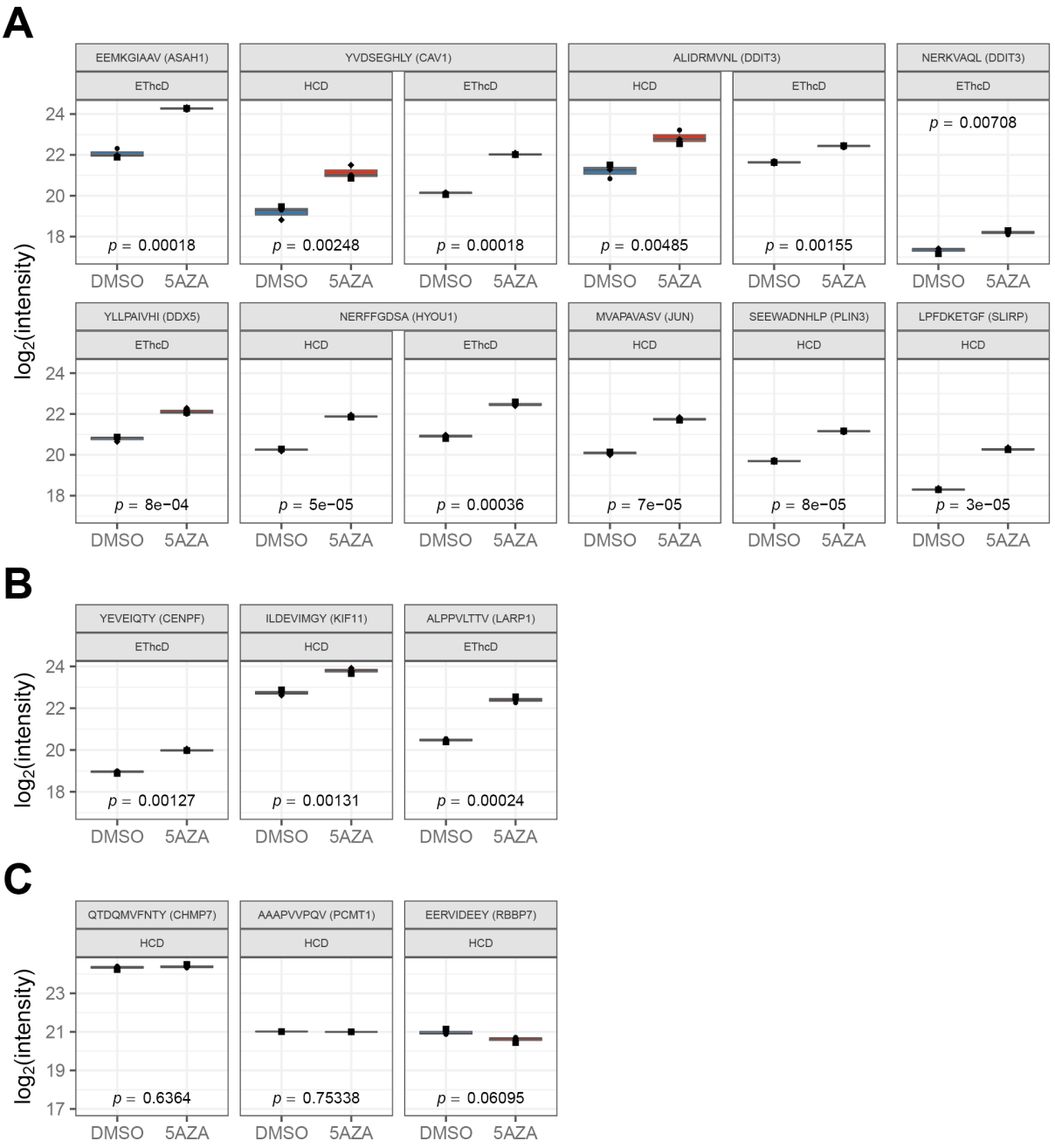


Figure 56: Treatment with 5AZA augments HLA class I-mediated presentation of peptides originating from NMD-targeted transcripts and wild-type parts of frameshift-mutated transcripts in HCT-116 cells. Representative plots showing changes in intensity for peptides originating putative NMD targets (panel A), wild-type parts of frameshift-mutated transcripts (panel B) and SNP neoepitopes (panel C) after treatment with 5AZA for 24 h. Bars represent 25th to 75th percentiles, middle line represents median, points represent individual measurements of biological replicates.

It is known that the composition of the proteome is reflected by the composition of the immunopeptidome (Caron *et al.*, 2011; Bassani-Sternberg *et al.*, 2015). The effect of NMD inhibition on the immunopeptidome was thus further investigated by GO term enrichment analysis of source proteins of upregulated hit peptides. This analysis showed that GO term categories of differentially presented peptides mirror those known to be affected by NMD inhibition at the proteome level (Sieber *et al.*, 2016) including protein folding, ER stress response, unfolded protein response, and proteasome-mediated

APC-dependent catabolic processes (*i.e.* breakdown of proteins by peptide bond hydrolysis; Figure 57). Moreover, peptides originating from proteins involved in antigen processing and presenting were upregulated suggesting that increased production of C-terminally truncated proteins caused by NMD inhibition also affects peptide presentation by the HLA class I system. Taken together, these data show that pharmacological NMD inhibition by 5AZA results in an increased cell surface presentation of peptides originating from NMD-targeted transcripts, including InDel neoepitopes.

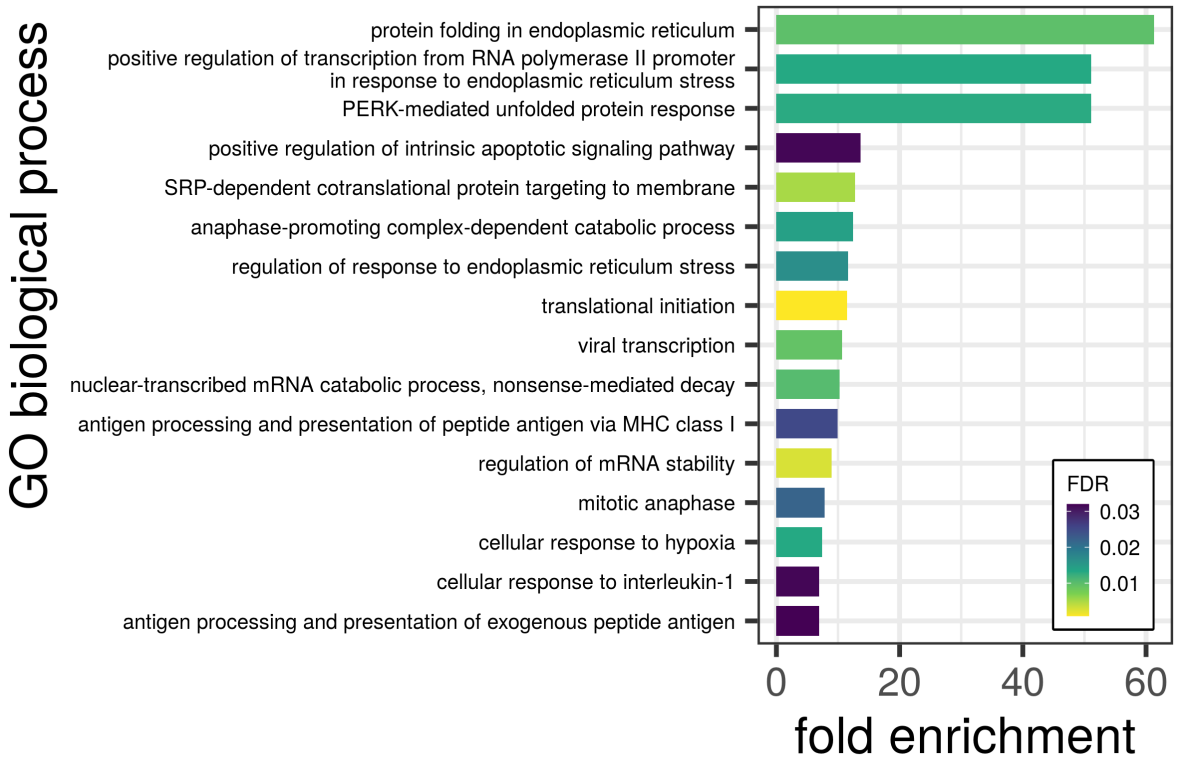


Figure 57: The immunopeptidome mirrors the proteome. GO term enrichment analysis of source proteins of significantly upregulated HLA class I-presented peptides after NMD inhibition by 5AZA. Colors represent FDR.

3.5.4 NMD inhibition by 5AZA stabilizes frameshift-bearing transcripts coding for InDel neoepitopes

The NMD-inhibitory effect of 5AZA was previously shown by the stabilization of endogenous NMD target mRNAs of *ATF3*, *SC35C*, and *SC35D* (Figure 7 and Figure 32). Subsequently, the effect of 5AZA treatment on frameshift-bearing transcripts encoding InDel neoepitopes as well as endogenous NMD-targeted transcripts coding for upregulated HLA class I-presented peptides was investigated in HCT-116 cells. NMD inhibition by 5AZA significantly stabilized ($p \leq 0.01$) the transcript levels of the endogenous NMD targets *ATF3* (3.6-fold), *ATF4* (1.8-fold), and *UPP1* (3.9-fold; Figure 58A). 5AZA treatment significantly increased ($p \leq 0.0001$) the abundance of frameshift-bearing *CKAP2* (2.0-fold), *PSMC6* (1.7-fold), *STK38* (1.4-fold), and *TUBGCP3* (1.9-fold) transcripts (Figure 58A). To further validate the frameshift-bearing transcripts as *bona fide* NMD targets, the effect of siRNA-mediated knockdown of the NMD key player UPF1 (Lavysh and Neu-Yilik, 2020) on their abundance was determined. siRNA-mediated knockdown of UPF1 was highly significant both on the mRNA as well as on the protein level (Figure 58A and B) and significantly stabilized ($p \leq 0.01$) the frameshift-bearing

transcripts of *CKAP2* (1.4-fold), *PSMC6* (1.5-fold), *NFAT5* (1.4-fold), and *TUBGCP3* (1.6-fold; Figure 58A). Taken together, these data show that NMD inhibition either by treatment with 5AZA or siRNA-mediated knockdown of UPF1 stabilizes frameshift-bearing transcripts encoding InDel neoepitopes.

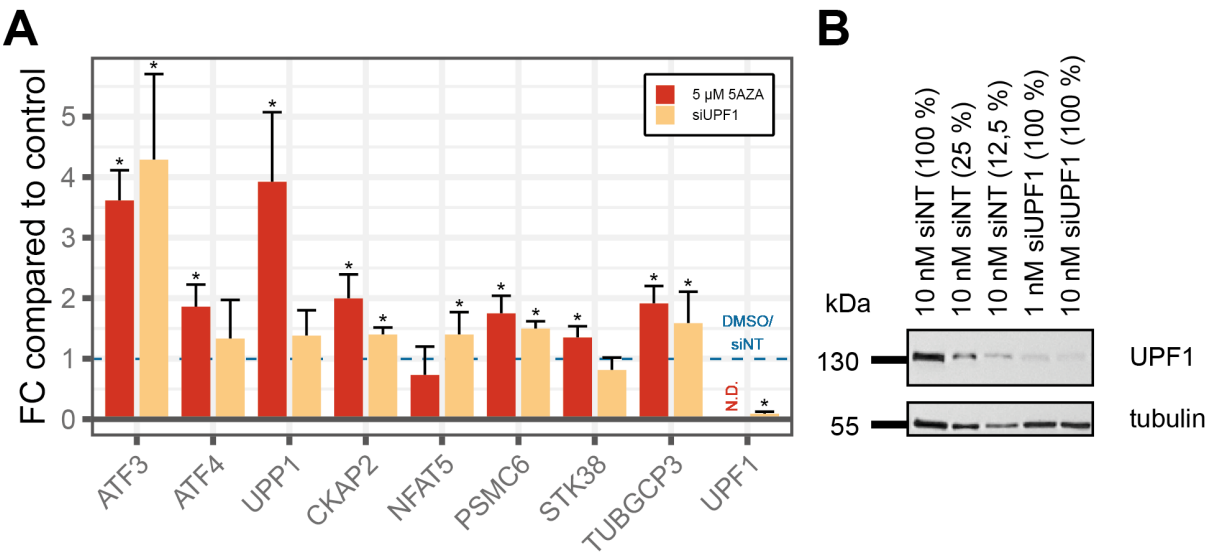


Figure 58: Validation of frameshift-mutated transcripts as bona fide NMD targets. qPCR analysis of endogenous NMD targets and frameshift-mutated transcripts after treatment with 5 µM 5AZA (red) or siRNA-mediated knockdown of *UPF1* (orange). UPF1 mRNA levels were determined as control for siRNA-mediated knockdown of *UPF1*. Bars represent mean ± SD of three experiments, *p ≤ 0.0001 (two-sided, unpaired t-test; panel A). Validation of siRNA-mediated knockdown of *UPF1* on the protein level. Tubulin served as loading control (panel B).

3.6 Evaluating the immunological and clinical potential of InDel neoepitopes

3.6.1 *In vivo* immunization with InDel neoepitopes induce CD8⁺ T cell responses

The potential of identified InDel neoepitopes to induce specific T cell responses was analyzed by *in vivo* immunizations in an HLA-A*02:01-transgenic mouse model (Pajot *et al.*, 2004). The HLA-A2.1/HLA-DR1-transgenic H-2 class I/class II-knockout mouse model used does not express any murine MHC molecules. Instead, these mice express human HLA-A*02:01 and HLA-DR1 molecules and are thus ideal to investigate HLA-restricted CD8⁺ and CD4⁺ T cell responses. Binding predictions indicated that the *CKAP2*-derived InDel neoepitope is a strong HLA-A*02:01 binder (percentile rank: 0.0071, predicted IC_{50}: 2.3736 nM), while the InDel neoepitopes derived from *NFAT5*, *PSMC6*, and *STK38* were predicted to bind other alleles than HLA-A*02:01 (Table 25). Interestingly, the *TUBGCP3*-derived InDel neoepitope was not predicted to bind any of the HLA alleles expressed by HCT-116 cells using the standard NetMHCpan thresholds but showed the strongest affinity to HLA-A*02:01 (percentile rank: 2.3267, predicted IC_{50}: 6217 nM) and was therefore included for further evaluation. First, mice were immunized three times with 7 days interval using a pool of five peptides consisting of two potential HLA-A*02:01 binders (*CKAP2*- and *TUBGCP3*-derived InDel neoepitopes), two “non-binders” as negative controls (*NFAT5*- and *PSMC6*-derived InDel neoepitopes), and a known HPV epitope (E7 11-19) as the positive control. Seven days after the last injection, mice were sacrificed and whole splenocytes were analyzed by *ex vivo* IFNγ ELISpot assays (Figure 59A). Here, one out of three mice generated a peptide-specific T cell response against the *CKAP2*-derived InDel neoepitope, and two out of three mice generated a peptide-specific T cell response against the *TUBGCP3*-derived InDel neoepitope. As expected, immunization did not induce peptide-specific T cells for predicted “non-binder” InDel neoepitopes derived from *NFAT5* and *PSMC6* (Figure 59B and C).

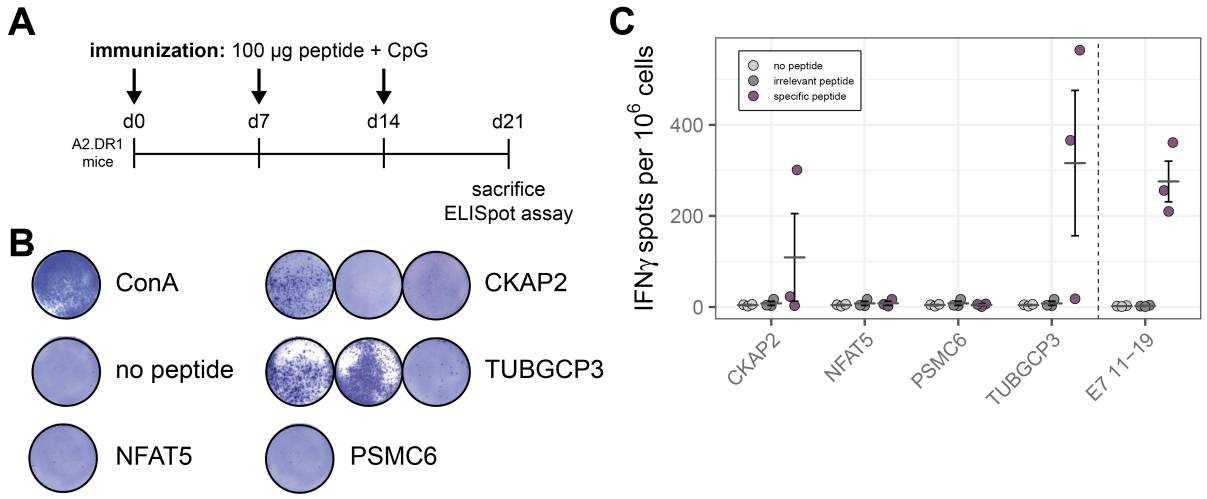


Figure 59: *In vivo* immunization of A2.DR1 mice with pooled InDel neoepitopes induces an immune response. Mice were immunized with peptide pools according to the immunization scheme (panel A). Representative ELISpot assay results for whole splenocytes stimulated with ConA (assay positive control), no peptide control, and InDel neoepitopes (panel B). Quantitative analysis of ELISpot assays. Bars represent mean ± SEM of three experiments (panel C).

The immune responses observed in whole splenocytes were further characterized by immunization with a single peptide and *ex vivo* analysis of CD8⁺ T cells by IFNγ ELISpot assays (Figure 60A). Here, immunizations with either the *CKAP2*- or *TUBGCP3*-derived InDel neoepitopes or the viral positive

control peptide E7 11-19 resulted in IFNγ-specific and highly significant T cell responses (Figure 60B and C). Taken together, these data demonstrate that the identified InDel neoepitopes derived from *CKAP2* and *TUBGCP3* are processed and presented *in vivo* via HLA-A*02:01 molecules and that immunization with these peptides can induce CD8⁺ T cell-mediated immune responses.

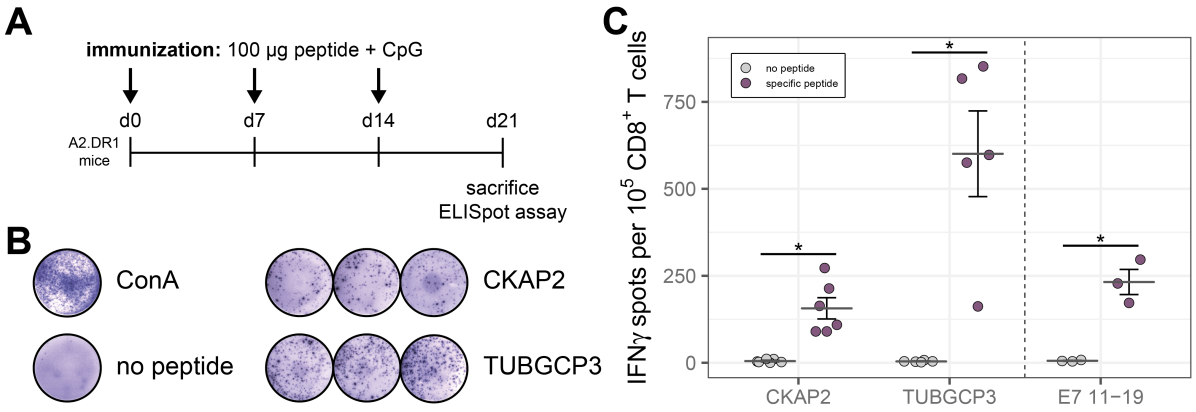


Figure 60: *In vivo* immunization of A2.DR1 mice with single InDel neoepitopes induces a CD8⁺ T cell response. Mice were immunized with single peptides according to the immunization scheme (panel A). Representative ELISpot assay results for isolated CD8⁺ T cells stimulated with ConA (assay positive control), no peptide control, and InDel neoepitopes (panel B). Quantitative analysis of ELISpot assays. Bars represent mean ± SEM of six (CKAP2), five (TUBGCP3), and three experiments (E7 11-19), respectively. $p \leq 0.005$ (two-sided, unpaired t-test; panel C).

3.6.2 *CKAP2* frameshift mutation is recurrent in MSI cancer cell lines and patients

As a first step to further evaluate the clinical potential of the identified InDel neoepitopes, the recurrence of frameshift mutations in the repeats of *CKAP2*, *NFAT5*, *PSMC6*, *STK38*, *TAF1B*, and *TUBGCP3* was analyzed. To this end, 24 MSI CRC cell lines were analyzed using gel capillary electrophoresis and the ReFrame tool (Ballhausen *et al.*, 2020; Table 27). In addition to HCT-116, the *CKAP2* frameshift mutation was found in four other MSI CRC cell lines, namely KM12 (one-base deletion), VaCo6 (one-base deletion), HROC24 (one-base insertion), and LS411 (one-base deletion; Figure 61). The *TUBGC3* and *STK38* frameshift mutations were identified in HCT-116 cells and in LoVo (one-base deletion) and HROC24 cells (one-base deletion), respectively. The *NFAT5* and *PSMC6* frameshift mutations were only found in HCT-116 cells. The recurrence of the frameshift mutation leading to the identified *TAF1B*-derived InDel neoepitope is well established (Kim *et al.*, 2002; Kloor *et al.*, 2020a) and its high mutation frequency was confirmed in the MSI CRC cell line panel tested (Table 27 and Figure 62).

Table 27: InDel mutation analysis by gel capillary electrophoresis in MSI CRC cell lines.

cell line	*CKAP2*	*NFAT5*	*PSMC6*	*STK38*	*TAF1B*	*TUBGCP3*
HCT-116	**m1**	**m1**	**m1**	**m1**	**m2/m1**	**m1**
RKO	wt	wt	wt	wt	**m2/m1**	wt
Coga1	wt	wt	wt	wt	**m1**	wt
Colo60H	wt	wt	wt	wt	**m2/m1**	wt
DLD1	wt	wt	wt	wt	wt	wt
HCT15	wt	wt	wt	wt	wt	wt
HDC135	wt	wt	wt	wt	**m1**	wt
HDC143	wt	wt	wt	wt	**m1**	wt
HROC24	**p1**	wt	wt	**m1**	**m1**	wt
K073A	wt	wt	wt	wt	**m1**	wt
KM12	**m1**	wt	wt	wt	**m1**	wt
LIM1215	wt	wt	wt	wt	**m1**	wt
LIM2405	wt	wt	wt	wt	**m1**	wt
LIM2412	wt	wt	wt	wt	**m1**	wt
LIM2537	wt	wt	wt	wt	wt	wt
LIM2551	wt	wt	wt	wt	**m1**	wt
LoVo	wt	wt	wt	wt	**m1**	**m1**
LS174T	wt	wt	wt	wt	**m2/m1**	wt
LS411	**m1**	wt	wt	wt	**m1**	wt
TC7	wt	wt	wt	wt	**m2/m1**	wt
TC71	wt	wt	wt	wt	**m1**	wt
VaCo457	wt	wt	wt	wt	**m1**	wt
VaCo5	wt	wt	wt	wt	**m1**	wt
VaCo6	**m1**	wt	wt	wt	**m1/m2**	wt
total	**5/24 (20.83 %)**	**1/24 (4.17 %)**	**1/24 (4.17 %)**	**2/24 (8.3 %)**	**21/24 (87.5 %)**	**2/24 (8.3 %)**

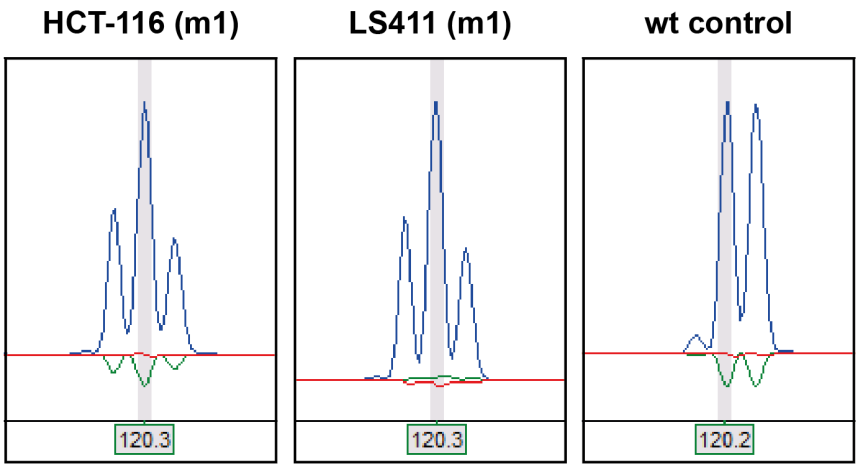


Figure 61: Exemplary results of gel capillary electrophoresis for *CKAP2*. The size of the main wild-type product is 120 bp (highlighted in grey for wt control).

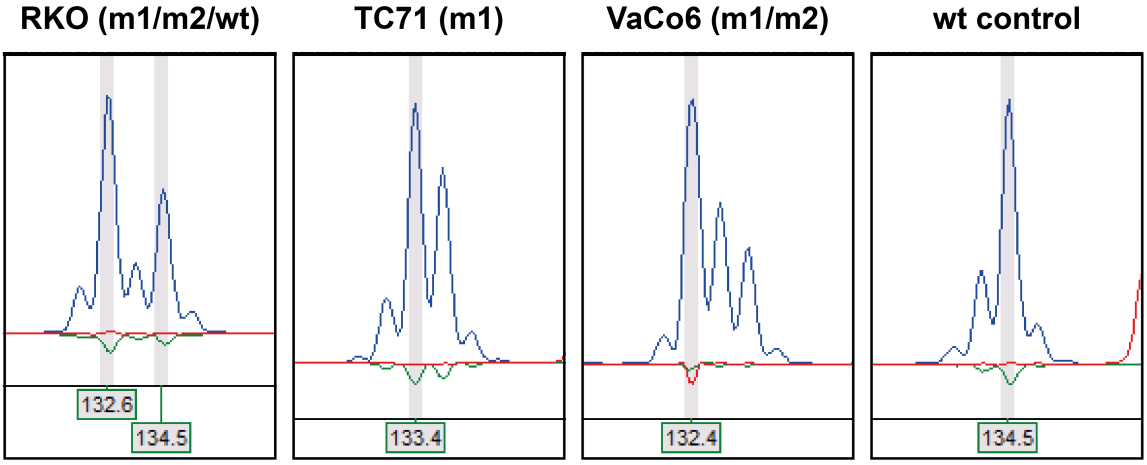


Figure 62: Exemplary results of gel capillary electrophoresis for *TAF1B*. The size of the main wild-type product is 134 bp (highlighted in grey for wt control).

Considering the unexpectedly high mutation recurrency in the A8 repeat of *CKAP2* in the MSI CRC cell line panel tested, it was investigated if the corresponding mutation could also be identified in other cell lines as well as in MSI CRC patient samples. Further analysis of the CCLE database (Barretina *et al.*, 2012; Ghandi *et al.*, 2019) revealed 13 other cell lines carrying the described mutation in *CKAP2*. These included cancer cell lines obtained from various sites such as the endometrium (HEC-151, HEC-59), large intestine (GP5d, SNU-1040, SNU-C2A, SNU-C2B), ovary (TOV-21G), stomach (23132/87, TGBC11TKB), and hematopoietic and lymphoid tissues (Jurkat, Kasumi-2, MN-60, WSU-NHL). Of note, all of these cancer cell lines except for SNU-1040 and Kasumi-2 exhibit microsatellite instability (Barretina et al., 2012; Ghandi et al., 2019). Next, genomic tumor DNA samples from 54 patients were analyzed as previously described and the *CKAP2* mutation was found in nine samples (16 %). Finally, the potential InDel neoepitopes arising from confirmed frameshift mutations were evaluated by performing binding predictions for all overlapping nonamers originating from frameshifted protein sequences. The confirmed InDel mutations generate frameshifted protein sequences with a length between 14 (*CKAP2*, *NFAT5*) and 108 aa (*TUBGCP3*). All frameshifted aa sequences harbor multiple promising InDel neoepitopes with promising binding affinities to common HLA supertypes (Sette and

Sidney, 1999; Doytchinova *et al.*, 2004; Sidney *et al.*, 2008; Figure 63 and Figure 64). Taken together, frameshift mutations were observed recurrently in cell lines and MSI CRC patient samples. Besides the well-established *TAF1B* mutation, *CKAP2* frameshift mutations emerged to be the most interesting because of their high degree of recurrence in both MSI CRC cell lines and patient samples. Furthermore, although the identified InDel mutation only leads to a frameshifted protein sequence of 14 aa, it harbors potential neoepitopes with promising binding affinities for eight of the twelve HLA supertypes tested.

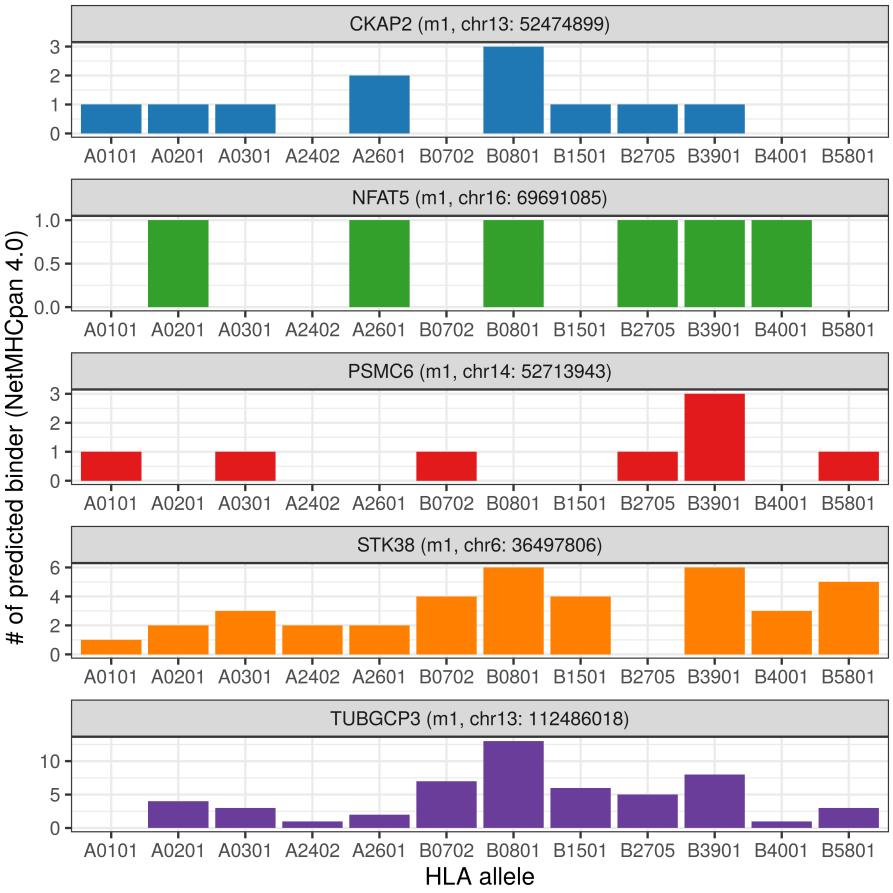


Figure 63: Frameshift parts of source proteins of InDel neoepitopes identified in HCT-116 cells generate numerous nonamers predicted to bind HLA supertype alleles. Binding prediction for overlapping nonamers originating from frameshift part of mutated source proteins was performed using NetMHCpan 4.0. Threshold for binders was top 2 % ranked.

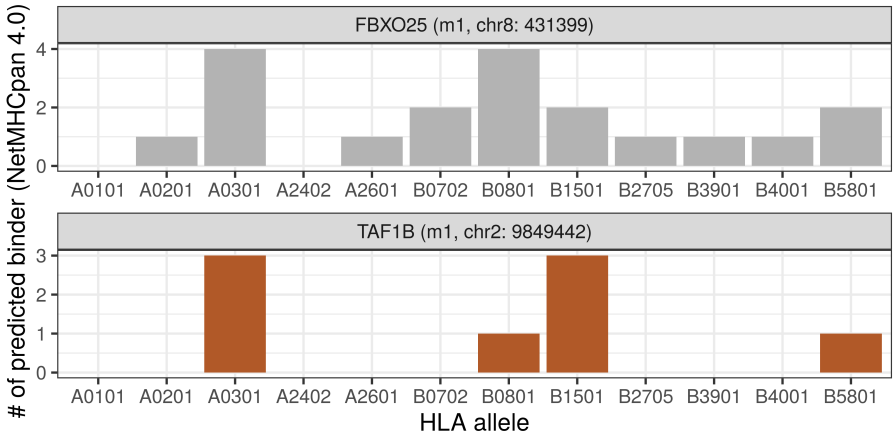


Figure 64: Frameshift parts of source proteins of InDel neoepitopes identified in RKO cells generate numerous nonamers predicted to bind HLA supertype alleles. Binding prediction for overlapping nonamers originating from frameshift part of mutated source proteins was performed using NetMHCpan 4.0. Threshold for binders was top 2 % ranked.

4 Discussion

4.1 Establishment of an immunopeptidomics pipeline for neoepitope discovery

HLA class I-mediated presentation of tumor-specific, mutation-derived neoepitopes represent a crucial factor for the success of immunotherapeutic approaches such as immune checkpoint blockade or cancer vaccination (Schumacher and Schreiber, 2015). Until recently, the investigation of such neoepitopes mostly relied on *in silico* predictions or *in vitro* binding assays. However, only a negligible fraction of these candidates is actually presented by HLA class I molecules at the surface of tumor cells and thus the number of predicted neoepitopes eliciting anti-tumor immune responses *in vivo* is limited (Schmidt *et al.*, 2017). With the advent of immunotherapy and the notion that T cells re-activated by immune checkpoint inhibition target tumor-specific neoepitopes, the unambiguous identification of these neoepitopes gained new interest. Here, in particular, mass spectrometry-based investigation of the immunopeptidome holds promise for the direct identification of actually processed and presented, tumor-specific neoepitopes. Previous studies on neoepitopes using mass spectrometry mainly focused on mutation-derived neoepitopes due to SNPs (Bassani-Sternberg *et al.*, 2016). However, SNP-derived neoepitopes are thought to induce less robust immune responses due to their higher “similarity to self” compared to InDel mutation-derived neoepitopes (Turajlic *et al.*, 2017). Mass spectrometry-based studies on InDel neoepitopes conducted to date focused either on selected neoepitopes originating from specific, recurrent InDel mutations such as mutated nucleophosmin in acute myeloid leukemia (Narayan *et al.*, 2019; van der Lee *et al.*, 2019) or failed to detect frameshift-derived protein sequences at the proteome level (Halvey *et al.*, 2014). Within the scope of this thesis, a mass spectrometry-based immunopeptidomics workflow tailored for InDel neoepitope discovery was established and utilized for the unbiased identification of HLA class I-presented InDel neoepitopes in two MSI CRC model cell lines.

The sample preparation has been described as the Achilles heel of immunopeptidomics analysis with reported peptide losses of up to 99.5 % (Hassan *et al.*, 2014; Kuznetsov *et al.*, 2020; Stopfer *et al.*, 2020). As the basic principle of the protocol did not change much since it was introduced in 1992 (Hunt *et al.*, 1992), previously established protocols for antibody purification and crosslinking (Bassani-Sternberg, 2018), as well as high-throughput immunoprecipitation of HLA class I:peptide complexes and subsequent separation of HLA class I molecules and HLA class I-presented peptides (Chong *et al.*, 2018), were utilized. Although other lysis conditions, matrix materials, antibody:matrix ratios, matrix:sample ratios, as well as washing and elution buffers were tested during the initial optimization of the batch immunoprecipitation procedure, the parameters described by Chong and colleagues for the high-throughput immunoprecipitation yielded the best results both in terms of HLA molecule extraction from lysates as determined by western blot analysis and peptide identifications by mass spectrometry. Nevertheless, the number and nature of both endogenous HLA class I-presented peptides and neoepitopes identified during this study might be biased by the choice of the sample preparation protocol (Nicastri *et al.*, 2020). An example of potential peptide loss in the presented work is the non-detection of the well-known HLA-A*02:01-restricted *TGFBR2*-derived InDel neoepitope RLSSCVPVA (Sæterdal *et al.*, 2001; Schwitalle *et al.*, 2008) using the targeted MS2 approach. Based on a recent report on T cell responses against this InDel neoepitope, this negative result is likely explained by the low stability of the corresponding HLA class I:peptide complex (Inderberg *et al.*, 2017) leading to peptide loss during the immunoprecipitation procedure. Another point which should be

emphasized is the amount of antibody and input material used which could be further optimized to prevent clogging and increased pressures of LC columns during MS analysis. Pre-elution of non-crosslinked antibody was tested, although this did only negligibly reduce its co-elution during HLA class I:peptide complex elution. Moreover, pre-elution under acidic conditions might lead to denaturation of the chemically crosslinked antibody and thus reduce HLA class I:peptide complex yield. Large amounts of input material, as used in this study, will also lead to increased co-elution of B2M molecules during the separation of HLA class I molecules and HLA class I-presented peptides using C_{18} cartridges. Here, lower concentrations of organic solvent for peptide elution were tested, however, did not reduce clogging and increased pressure of LC columns during MS analysis. Another possibility, which was, however, not evaluated within the scope of this thesis, could be the use of C_8 instead of C_{18} cartridges. C_8 cartridges exhibit a lower hydrophobicity and thus lower retention of small molecules such as peptides and could be used for better peptide separation from large HLA molecules. Moreover, smaller pore sizes of the silica cartridges or additional HPLC separation might further decrease the amount of co-eluted proteins (Purcell *et al.*, 2019; Nicastri *et al.*, 2020).

Different fragmentation and precursor selection methods were used for the LC-MS/MS analysis of immunopeptidomes. Single fragmentation by either HCD or CID as well as dual fragmentation by either EThcD or ETciD were compared based on previous reports suggesting that dual fragmentation leads to an increased identification rate (Mommen *et al.*, 2014). In the present work, standard fragmentation with either HCD or CID and top(n) precursor selection yielded the majority of HLA class I-presented peptide identifications. In contrast to previous reports (Mommen *et al.*, 2014), dual fragmentation by either EThcD or ETciD did not increase the identification rate and was comparable to the identification rates of samples measured with single fragmentation by either HCD or CID. Importantly, single fragmentation methods exhibit a much higher MS2/MS1 rate and thus allow a deeper sampling of the immunopeptidome and especially low abundant peptides which were missed by dual fragmentation methods. Different MS parameters such as prolonged gradient lengths or different cycle times were tested to increase the number of peptide identifications using dual fragmentation by EThcD, however, single fragmentation by HCD still outperformed all methods tested in terms of peptide identifications. Moreover, precursor selection targeting low abundant precursors first was tested for EThcD fragmentation to increase the immunopeptidome coverage, especially for low abundant peptides. Using this lowEThcD method led to the identification of a low abundant peptide subset which was neither identified by EThcD using top(n) precursor selection nor by HCD, suggesting that the described lowEThcD method (or any other methodology targeting low abundant precursors first) is suited to further expand the detectable immunopeptidome.

In general, each peptide is presented 100-10,000 times per cell (Bleek and Nathenson, 1990; Hunt *et al.*, 1992), and cells express between 5,000 and 3,000,000 HLA molecules at their cell surface (Schuster *et al.*, 2017; Lanoix *et al.*, 2018). It is thus estimated that each cell can present 10,000-30,000 unique peptides (Kuznetsov *et al.*, 2020). Here, approximately 10,000 unique peptides were identified for both the HCT-116 and RKO cell lines which is well in line with this estimation as well as with previously reported studies on the immunopeptidome. Given the heuristic nature of abundance-dependent precursor selection and fragmentation of data-dependent mass spectrometry used in this study, more extensive data acquisition on samples derived from HCT-116 and RKO cells might identify additional

peptides and neoepitopes. Moreover, data-independent acquisition has been introduced for the analysis of immunopeptidomes (Caron *et al.*, 2011; Ritz *et al.*, 2017; Pak *et al.*, 2021). Here, all peptides within a defined *m/z* window are subjected to fragmentation in an unbiased manner generating highly convoluted MS2 spectra. This process is iterated across the full *m/z* range to generate “digital maps” of the analyzed samples. To identify peptides from such spectra, they are compared to a spectral library containing fragmentation and retention time information of the peptides of interest (Guo *et al.*, 2015; Ludwig *et al.*, 2018). Until recently, the generation of a high-quality spectral library required *a priori* information on the peptides of interest from data-dependent mass spectrometry analysis. However, library-free approaches have been introduced recently, theoretically allowing the identification of neoepitopes without *a priori* knowledge (Tsou *et al.*, 2015; Tran *et al.*, 2019; Demichev *et al.*, 2020). Of note, in this study, 90 additional peptides were identified in the targeted MS2 dataset that have not been identified in any of the previously generated datasets using data-dependent acquisition with different fragmentation and precursor selection methods, providing evidence that data-independent acquisition of immunopeptidomes might further increase the number of HLA class I-presented peptides and neoepitopes identified.

In order to identify such HLA class I-presented neoepitopes by mass spectrometry, their sequence must be known. To this end, publicly available sequencing data from the CCLE and COSMIC data repositories were utilized to construct cell line-specific neoepitope databases. As SNP neoepitope databases have been previously published (Scholtalbers *et al.*, 2015), the main focus here was on the construction of InDel neoepitope databases. The established bioinformatics pipeline uses Ensembl transcript identifier in conjunction with mutation information in standard mutation nomenclature (Ogino *et al.*, 2007) at the level of coding DNA (*e.g.* c.123delG denotes deletion of G at position 123) and is thus easily adaptable for the analysis of other cell lines for which sequencing information is publicly available. Moreover, it is anticipated that the established pipeline for database generation can be used in future proteogenomic studies combining immunopeptidome and sequencing data of patient samples for the identification of patient-specific neoepitopes. Although the usage of public sequencing data represents a suitable substitute for sequencing when analyzing well-established cancer cell lines (Freudenmann *et al.*, 2018), it is important to note that variant calling especially for InDel mutations remains challenging even with advanced biocomputational algorithms (O'Rawe *et al.*, 2013; Narzisi and Schatz, 2015). Indeed, several well-established InDel mutations occurring in longer repetitive sequences are not documented in the CCLE or COSMIC database, including *e.g.* a one-base deletion in *AIM2* of HCT-116 cells (Woerner *et al.*, 2007), a one-base deletion in *CASP5* of HCT-116 cells (Woerner *et al.*, 2001), and a one-base deletion in *TGFBR2* of HCT-116 cells (reported in COSMIC but not in CCLE) and in RKO cells (Markowitz *et al.*, 1995). To address this limitation, two additional InDel neoepitope databases were constructed: the ReFrame database based on previously published data on recurrent frameshift mutations in MSI cancers (Ballhausen *et al.*, 2020) and the MNR database containing protein sequences derived from *in silico* mutation of all coding mononucleotide repeats larger than six nucleotides.

A *de novo* sequencing-assisted, multi-round database search was performed for the identification of both endogenous HLA class I-presented peptides and neoepitopes and yielded 109 candidate peptides. In total, 17 neoepitopes were identified and validated by synthetic peptide comparison and Sanger sequencing. Compared to the average number of only one identifiable neoepitope per $1.8x10^3$ non-

synonymous mutations or per $1.1x10^4$ identified peptides proposed by Freudenmann and colleagues (Freudenmann *et al.*, 2018), the relatively large number of neoepitopes identified in this study reflects both the high mutational burden of MSI CRC cancers as well as the increased number of potential neoepitopes due to frameshift mutations. Ten of these neoepitopes are derived from single nucleotide polymorphisms while the remaining seven are derived from InDel mutations. Interestingly, searching the mass spectrometry raw data against the ReFrame and MNR databases yielded only a few new InDel neoepitopes identifications which could be validated by both synthetic peptide comparison and Sanger sequencing. For the ReFrame database, this could be explained by the general observation that recurrent mutations are biased towards producing peptides that are poorly presented by the HLA alleles expressed by the corresponding tumor cell (Marty *et al.*, 2017). For InDel mutations this can be further exploited by cancer cells as functional inactivation of a gene usually can be achieved by either one- or two-base InDel mutations leading to completely different frameshifted protein sequences. Here, mutations are positively selected which either lead to transcripts targeted by NMD (Spaanderman *et al.*, 2020) or which produce only short neoantigen sequences. For the latter, a prominent example is the *TAF1B* gene that was found to exhibit a significantly elevated number of two-base deletions (Ballhausen *et al.*, 2020). While a one-base deletion would lead to a frameshifted sequence of 28 aa, a two-base deletion in the described mononucleotide repeat leads to a frameshifted sequence of only two aa. In general, the length of mutated microsatellites resulting in identified InDel neoepitopes was, with exception of the A11 repeat in *TAF1B*, rather short (4-8 bp) further pointing in the direction that InDel neoepitopes may arise from passenger rather than from recurrent driver mutations. Future studies might incorporate data obtained with long-read sequencing technologies in order to further improve InDel mutation calling in repetitive sequences (Nakano *et al.*, 2017) and construct more comprehensive, sample-specific InDel neoepitope databases. Searching the mass spectrometry raw data against the MNR database yielded a total of 85 InDel neoepitope candidates, however only the *FBXO25*-derived InDel neoepitope could be validated. Interestingly, only eleven of these candidates were identified as false identifications after synthetic peptide validation which can be explained by an increased rate of false positives due to the large size of the MNR database (Holmes and Huber, 2019). For the remaining candidates, genomic validation by Sanger sequencing was unsuccessful. Although it was aimed to sequence the largest possible fragments, it can not be excluded that other mutations leading to the found InDel neoepitopes were missed. Another explanation for the occurrence of these peptides without an apparent genomic template could be proteasome-catalyzed peptide splicing, a process in which a protein is first broken into peptides that are re-ligated again (Vigneron *et al.*, 2004). The existence and contribution of spliced peptides to the immunopeptidome have been a matter of debate. It is not known if they represent cancer-specific epitopes (Liepe *et al.*, 2016; Faridi *et al.*, 2018; Mylonas *et al.*, 2018; Liepe *et al.*, 2019; Rolfs *et al.*, 2019) and their further consideration was not within the scope of this thesis. Of note, considering only SNP neoepitope candidates and InDel neoepitope candidates matching database entries based on sequencing data from CCLE and COSMIC, all SNP neoepitopes and five out of ten InDel neoepitopes could be validated both on the peptidomic and genomic level demonstrating the high success rate of an unbiased, mass spectrometry-based approach for neoepitope discovery. While 14 of the identified neoepitopes were identified using standard fragmentation by HCD and top(n) precursor selection, the *NFAT5*- and *FBXO25*-derived InDel neoepitopes as well as the *RGP1*-derived SNP neoepitope could only be identified using either the EThcD or lowEThcD methodology, further

emphasizing the added value of combining different fragmentation and precursor selection methods. Moreover, spectra recorded using dual fragmentation provided more sequence coverage thus allowing less ambiguous peptide identification.

4.2 Pharmacological NMD inhibition augments the HLA class I-mediated presentation of immunogenic InDel neoepitopes

Immunotherapeutic strategies exploiting InDel neoepitopes rely on their expression and HLA class I-mediated presentation at the surface of tumor cells. In most cases, frameshift mutations generating InDel neoepitopes also lead to PTCs which trigger the RNA quality control pathway NMD. The mRNAs targeted by NMD are rapidly degraded thus preventing the expression of InDel neoepitopes on their HLA class I-mediated cell surface presentation and it has previously been shown that MSI CRC cancers express increased amounts of the NMD factors UPF1, UPF2, SMG1, SMG6, and SMG7 as compared to microsatellite-stable colorectal cancers thus effectively restricting the production of InDel neoepitopes and consequently their immune recognition (El-Bchiri *et al.*, 2005; El-Bchiri *et al.*, 2008; Bokhari *et al.*, 2018). Here, it was hypothesized that NMD inhibition by the previously identified and clinically approved drug 5AZA could stabilize frameshift-bearing transcripts encoding InDel neoepitopes and increase the expression as well as HLA class I-mediated cell surface presentation of such InDel neoepitopes.

Label-free quantification was performed to evaluate the effect of NMD inhibition on the immunopeptidome of HCT-116 and RKO cells and identified more than 800 and 1500 differentially presented peptides, respectively. In general, source proteins of upregulated peptides are involved in the ER stress response, protein folding, the unfolded protein response, and proteasome-mediated APC-dependent catabolic processes as well as HLA class I antigen processing and presentation. The obtained results mirror the previously reported effects of NMD inhibition at the proteome level (Sieber *et al.*, 2016) and are well in line with the hypothesized effect of NMD inhibition on the immunopeptidome. Here, NMD inhibition is expected to stabilize faulty transcripts leading to enhanced expression of C-terminally truncated and potentially mutant protein sequences. These C-terminally truncated proteins cannot be folded correctly and are rapidly degraded by the proteasome and thus added to the pool of peptides available for HLA class I-mediated cell surface presentation. Importantly, it was shown for the first time that the HLA-class I-mediated cell surface presentation of InDel mutation-derived neoepitopes is upregulated after pharmacological NMD inhibition. Additionally, several other peptides encoded by endogenous NMD targets as well as wild-type parts of frameshifted transcripts were upregulated after treatment with 5AZA providing further evidence that pharmacological NMD inhibition affects the HLA class I-mediated presentation of peptides originating from transcripts targeted by NMD. Of note, InDel neoepitopes were also identified and quantified in samples not treated with 5AZA. This can be explained by the fact that NMD does not lead to complete degradation of faulty mRNAs and small amounts of C-terminally truncated proteins might be synthesized. Studies on defective ribosomal products (DRiPs) further postulate that these non-functional and rapidly degraded products are a major source of HLA class I-presented peptides (Apcher *et al.*, 2011; Yewdell, 2011).

The abundance of the frameshift-bearing transcripts was further assessed in response to NMD inhibition by either 5AZA treatment or siRNA-mediated knockdown of UPF1 to validate them as *bona fide* NMD

targets. *CKAP2*, *STK38,* and *TUBGCP3* transcripts were significantly stabilized after both pharmacological and siRNA-mediated NMD inhibition. Of note, the *CKAP2* transcript has previously been classified as likely being resistant to NMD based on the localization of the mutation leading to the InDel neoepitope (Thermann *et al.*, 1998; Spaanderman *et al.*, 2020). In agreement with other reports (Neu-Yilik *et al.*, 2011; Lindeboom *et al.*, 2016) this contradictory result indicates that sequence features alone are not sufficient to predict transcript degradation by NMD and that further experimental validation is required for the identification of *bona fide* NMD targets. The discrepancy in stabilization between pharmacological and siRNA-mediated NMD inhibition for *NFAT5* and *STK38* transcripts can likely be explained by other effects of 5AZA treatment or UPF1 knockdown such as an increased stress response due to NMD inhibition or DNA hypomethylation by 5AZA (Stresemann and Lyko, 2008).

The clinical potential of the identified InDel neoepitopes was assessed by evaluating their potential to induce T cell responses *in vivo*. The HLA*02:01-restricted InDel neoepitopes derived from *TUBGCP3* and *CKAP2* induced strong immune $CD8^+$ T cell responses in an HLA-humanized mouse model immunized with the corresponding peptides. Interestingly, applying the standard thresholds of NetMHCpan 4.0, the *TUBGCP3*-derived peptide was not predicted to be an HLA-A*02:01 binder which is in line with the observation that commonly used thresholds for binding predictions convey only a limited sensitivity of approximately 40 % (Bonsack *et al.*, 2019). However, applying a more conservative, experimentally validated threshold as suggested by Bonsack and colleagues for binding predictions of nonamers to HLA-A*02:01 (Bonsack *et al.*, 2019), the identified *TUBGCP3*-derived InDel neoepitope would be classified as a binder. Nevertheless, this observation further shows that an unbiased, mass spectrometry-based approach is superior to neoepitope discovery approaches relying on *in silico* predictions as the *TUBGCP3*-derived InDel neoepitope would have been missed by such approaches. Of note, the used *in vivo* immunization approach only allows a statement as to whether the InDel neoepitopes are recognized by T cells and induce their proliferation. However, it remains to be determined if these T cells are also capable of lysing tumor cells presenting such InDel neoepitopes. Future studies may utilize advanced targeted mass spectrometry approaches such as SRM or PRM (Tan *et al.*, 2011; Blatnik *et al.*, 2018) and isotopically labeled counterparts to investigate absolute InDel neoepitope amounts and determine thresholds needed for T cell-mediated tumor cell lysis. While early reports suggested that even a single HLA class I-presented neoepitope can trigger an effective immune response (Foote and Eisen, 2000), other factors such as TCR affinity, HLA class I:peptide complex stability, and neoepitope foreignness are important for successful T cell activation (McGranahan and Swanton, 2019; Wells *et al.*, 2020). Moreover, it has been shown that the fraction of cells presenting a neoepitope is crucial for T cell-mediated tumor rejection (McGranahan *et al.*, 2016; Gejman *et al.*, 2018).

Recurrence analysis revealed that the frameshift mutation described for *CKAP2* was found to be highly recurrent in both an MSI cancer cell line panel and MSI CRC patient samples. Also, the frameshift mutation in *CKAP2* was found in 13 other samples in the CCLE obtained from different sites, 11 of which showed MSI. Taken together, these findings further support the previously reported importance of the *CKAP2* frameshift mutation in MSI cancer (Leoni *et al.*, 2020; Spaanderman *et al.*, 2020) and the frequent recurrence of the mutation together with the HLA-A02:01-restriction of the identified *CKAP2*-derived InDel neoepitope predestines it for further clinical evaluation *e.g.* in cancer vaccination approaches.

4.3 Conclusions and perspectives

In this study, an immunopeptidomics workflow for neoepitope discovery combining high-throughput sample preparation and advanced mass spectrometry was established. This workflow was then used to identify HLA class I-presented InDel neoepitopes in an unbiased, mass spectrometry-based manner for the first time. As an important technical conclusion, it was shown that the combination of different fragmentation and precursor selection methods for immunopeptidomics expands the detectable immunopeptidome and increases the number of identified neoepitopes. Moreover, it was shown that the usage of cell line-specific, sequencing-based neoepitope databases without *a priori* shortlisting by HLA binding prediction yields high-quality neoepitope candidates of which most could be validated on both the peptidomic and genomic level. The established workflow lays the foundation for future immunopeptidomics analysis of both, samples from established cell lines and patients. Recently, vaccination with shared, *in silico* predicted InDel neoepitopes has been shown to induce neoepitope-directed immune responses in a clinical trial with MSI CRC patients (Kloor *et al.*, 2020a) and the InDel neoepitopes identified within the scope of this thesis might be further evaluated in personalized vaccines or engineered T cell therapies. In particular, the HLA-A*02:01-restricted *CKAP2*-derived InDel neoepitope originating from a recurrent mutation in cancers with microsatellite instability, a genetic predisposition increasingly recognized in various malignancies (Bonneville *et al.*, 2017), is a promising candidate. Notably, a recently reported “off-the-shelf” cancer vaccine incorporated the frameshifted protein sequence derived from the described *CKAP2* mutation and showed *in vitro* processing and presentation of the encoded InDel neoepitope by antigen-presenting cells as well as activation of human $CD8^+$ T cells (Leoni *et al.*, 2020) using cell-based assays complementary to the utilized immunopeptidomics approach in this study. The genetic vaccine is currently under clinical investigation in combination with a PD-1 checkpoint inhibtor (NCT04041310).

As the second major finding, it was shown that pharmacological NMD inhibition with clinically achievable 5AZA concentrations (Stresemann and Lyko, 2008) significantly augments the HLA class I-mediated presentation of InDel neoepitopes. Although immune checkpoint blockade targeting neoepitopes has been shown to be highly efficient in cancers with high mutational burden (Yarchoan *et al.*, 2017a; Marcus *et al.*, 2019; Osipov *et al.*, 2020), it is envisioned that 5AZA treatment acts synergistically with these approaches by expanding the T cell repertoire directed against frameshift-derived InDel neoepitopes and sensitizing such tumors for other immunotherapeutic interventions (Figure 65). In this study, 5AZA treatment selectively increased the HLA class I-mediated presentation of InDel neoepitopes encoded by NMD-targeted transcripts. In contrast to transcripts with SNPs, these transcripts typically encode multiple InDel neoepitopes with binding potential to different HLA alleles. Therefore, NMD inhibition is expected to increase HLA class I-mediated InDel neoepitope presentation independently of a patients’ HLA allotype and the tumor mutational landscape.

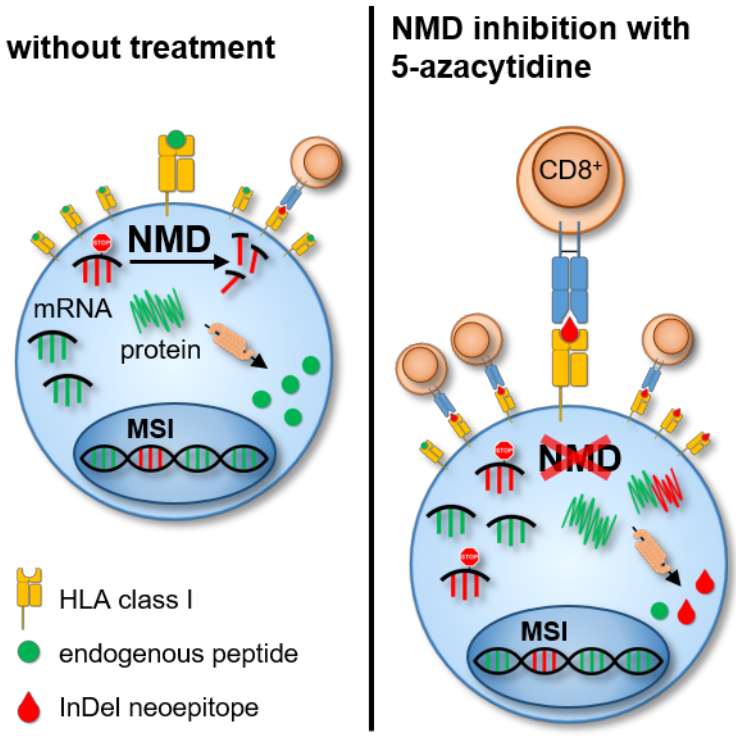


Figure 65: Pharmacological NMD inhibition turns cancer cells into easily identifiable T cell targets. Under normal conditions, the expression of InDel mutation-derived neoepitopes and their cell surface presentation by HLA class I molecules is limited by NMD (left side). Pharmacological NMD inhibition with 5AZA stabilizes FS-bearing transcripts and increases the expression of frameshifted protein sequences. Consequently, the HLA class I-mediated presentation of InDel neoepitopes and thus recognition by $CD8^+$ T cells is augmented (right side).

For the further development of NMD inhibition as a treatment option for cancers with MSI the following questions have to be addressed: Are the identified InDel neoepitopes recognized by $CD8^+$ T cells and does this recognition lead to tumor cell destruction? Is the observed tumor cell destruction enhanced in combination with NMD inhibition by 5AZA? To address these questions, future experiments should include in vitro killing assays (Schwitalle *et al.*, 2008). Here, $CD8^+$ T cells recognizing the HLA-A:02*01-restricted *CKAP2*-derived InDel neoepitopes have to be isolated from peripheral blood as previously described (Leoni *et al.*, 2020). Next, the isolated $CD8^+$ T cells are co-incubated with HCT-116 cells pretreated with either DMSO or 5AZA, and cytotoxic activity is subsequently measured by ^{51}Cr release assay or flow cytometry (Maletzki *et al.*, 2013; Poh *et al.*, 2020). Transfer of syngeneic cells transduced with minigenes coding for the identified InDel neoepitopes into the previously used HLA-humanized mouse model as described by Schumacher and colleagues (Schumacher *et al.*, 2014) could be used to further assess the preventive effect of InDel neoepitope vaccination. Additionally, the InDel neoepitopes could be encoded in an NMD context to evaluate the benefit of NMD inhibition by 5AZA *in vivo*. Alternatively, the effect of NMD inhibition *in vivo* could be tested using DNA MMR-deficient mouse models (Kucherlapati *et al.*, 2010; Lee *et al.*, 2016). However, in such models it has to be considered that coding microsatellites are only partially conserved between murine and human genes (Woerner *et al.*, 2015). As a first experimental step, the immunopeptidomics workflow established in this thesis could be utilized to identify murine InDel neoepitopes, before the effect of NMD inhibition on tumorigenesis is assessed.

In summary, the presented work demonstrates the feasibility of high-throughput immunopeptidomics for the identification of immunogenic InDel neoepitopes and provides an expandable platform for future studies of the immunopeptidome. Furthermore, it was shown that pharmacological NMD inhibition by 5AZA augments the HLA class I-mediated presentation of such InDel neoepitopes suggesting that future immunotherapeutic strategies for MSI cancers may benefit from a combination with NMD inhibition by turning cancer cells into easily identifiable targets for InDel neoepitope-specific $CD8^+$ T cells.

5 References

Adams, E.J., and Luoma, A.M. (2013). The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. Annu Rev Immunol *31*, 529-561. 10.1146/annurev-immunol-032712-095912.

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature *422*, 198-207. 10.1038/nature01511.

Aguado, B., Bahram, S., Beck, S., Campbell, R.D., Forbes, S.A., Geraghty, D., Guillaudeux, T., Hood, L., Horton, R., Inoko, H., et al. (1999). Complete sequence and gene map of a human major histocompatibility complex. Nature *401*, 921-923. 10.1038/44853.

Aguilar, M.-I., and Hearn, M.T.W. (1996). [1] High-resolution reversed-phase high-performance liquid chromatography of peptides and proteins. In Methods in enzymology, (Academic Press), pp. 3-26. 10.1016/S0076-6879(96)70003-4.

Aliouat, A., Hatin, I., Bertin, P., Francois, P., Stierle, V., Namy, O., Salhi, S., and Jean-Jean, O. (2019). Divergent effects of translation termination factor eRF3A and nonsense-mediated mRNA decay factor UPF1 on the expression of uORF carrying mRNAs and ribosome protein genes. RNA Biol, 1-13. 10.1080/15476286.2019.1674595.

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature *432*, 112-118. 10.1038/nature03060.

Andreatta, M., Alvarez, B., and Nielsen, M. (2017a). GibbsCluster: unsupervised clustering and alignment of peptide sequences. Nucleic Acids Res *(Epub ahead of print)*. 10.1093/nar/gkx248.

Andreatta, M., Alvarez, B., and Nielsen, M. (2017b). GibbsCluster: unsupervised clustering and alignment of peptide sequences. Nucleic Acids Res *45*, W458-w463. 10.1093/nar/gkx248.

Annibaldis, G., Dreos, R., Domanski, M., Carl, S., and Mühlemann, O. (2019). Ribosome recycling factor ABCE1 depletion inhibits nonsense-mediated mRNA decay by promoting stop codon readthrough. bioRxiv. 10.1101/870097.

Apcher, S., Daskalogianni, C., Lejeune, F., Manoury, B., Imhoos, G., Heslop, L., and Fahraeus, R. (2011). Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. Proc Natl Acad Sci U S A *108*, 11572-11577. 10.1073/pnas.1104104108.

Arana, M.E., and Kunkel, T.A. (2010). Mutator phenotypes due to DNA replication infidelity. Semin Cancer Biol *20*, 304-311. 10.1016/j.semcancer.2010.10.003.

Arya, M., Shergill, I.S., Williamson, M., Gommersall, L., Arya, N., and Patel, H.R. (2005). Basic principles of real-time quantitative PCR. Expert review of molecular diagnostics *5*, 209-219. 10.1586/14737159.5.2.209.

Auld, D.S., Lovell, S., Thorne, N., Lea, W.A., Maloney, D.J., Shen, M., Rai, G., Battaile, K.P., Thomas, C.J., Simeonov, A., et al. (2010). Molecular basis for the high-affinity binding and stabilization of firefly luciferase by PTC124. Proc Natl Acad Sci U S A *107*, 4878-4883. 10.1073/pnas.0909141107.

Auld, D.S., Thorne, N., Maguire, W.F., and Inglese, J. (2009). Mechanism of PTC124 activity in cell-based luciferase assays of nonsense codon suppression. Proc Natl Acad Sci U S A *106*, 3585-3590. 10.1073/pnas.0813345106.

Backlund, M., Stein, F., Rettel, M., Schwarzl, T., Perez-Perri, J.I., Brosig, A., Zhou, Y., Neu-Yilik, G., Hentze, M.W., and Kulozik, A.E. (2020). Plasticity of nuclear and cytoplasmic stress responses of RNA-binding proteins. Nucleic Acids Res. 10.1093/nar/gkaa256.

Ballhausen, A., Przybilla, M.J., Jendrusch, M., Haupt, S., Pfaffendorf, E., Seidler, F., Witt, J., Hernandez Sanchez, A., Urban, K., Draxlbauer, M., et al. (2020). The shared frameshift mutation landscape of

microsatellite-unstable cancers suggests immunoediting during tumor evolution. Nat Commun *11*, 4740. 10.1038/s41467-020-18514-5.

Bantscheff, M., Lemeer, S., Savitski, M.M., and Kuster, B. (2012). Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem *404*, 939-965. 10.1007/s00216-012-6203-4.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem *389*, 1017-1031. 10.1007/s00216-007-1486-6.

Bardhan, K., Anagnostou, T., and Boussiotis, V.A. (2016). The PD1:PD-L1/2 Pathway from Discovery to Clinical Implementation. Front Immunol *7*, 550. 10.3389/fimmu.2016.00550.

Barnstable, C.J., Bodmer, W.F., Brown, G., Galfre, G., Milstein, C., Williams, A.F., and Ziegler, A. (1978). Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. Cell *14*, 9-20.

Barnum, S.R. (2017). Complement: A primer for the coming therapeutic revolution. Pharmacol Ther *172*, 63-72. 10.1016/j.pharmthera.2016.11.014.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603-607. 10.1038/nature11003.

Bassani-Sternberg, M. (2018). Mass Spectrometry Based Immunopeptidomics for the Discovery of Cancer Neoantigens. Methods Mol Biol *1719*, 209-221. 10.1007/978-1-4939-7537-2_14.

Bassani-Sternberg, M., Braunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., Weber, J., Slotta-Huspenina, J., Specht, K., et al. (2016). Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. Nat Commun *7*, 13404. 10.1038/ncomms13404.

Bassani-Sternberg, M., and Coukos, G. (2016). Mass spectrometry-based antigen discovery for cancer immunotherapy. Curr Opin Immunol *41*, 9-17. 10.1016/j.coi.2016.04.005.

Bassani-Sternberg, M., Digklia, A., Huber, F., Wagner, D., Sempoux, C., Stevenson, B.J., Thierry, A.-C., Michaux, J., Pak, H., Racle, J., et al. (2019). A Phase Ib Study of the Combination of Personalized Autologous Dendritic Cell Vaccine, Aspirin, and Standard of Care Adjuvant Chemotherapy Followed by Nivolumab for Resected Pancreatic Adenocarcinoma—A Proof of Antigen Discovery Feasibility in Three Patients. Front Immunol *10*. 10.3389/fimmu.2019.01832.

Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L.J., and Mann, M. (2015). Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol Cell Proteomics *14*, 658-673. 10.1074/mcp.M114.042812.

Baumann, U., Fernández-Sáiz, V., Rudelius, M., Lemeer, S., Rad, R., Knorn, A.M., Slawska, J., Engel, K., Jeremias, I., Li, Z., et al. (2014). Disruption of the PRKCD-FBXO25-HAX-1 axis attenuates the apoptotic response and drives lymphomagenesis. Nat Med *20*, 1401-1409. 10.1038/nm.3740.

Becker, J.P., Helm, D., Rettel, M., Stein, F., Hernandez-Sanchez, A., Urban, K., Gebert, J., Kloor, M., Neu-Yilik, G., von Knebel Doeberitz, M., et al. (2021). NMD inhibition by 5-azacytidine augments presentation of immunogenic frameshift-derived neoepitopes. iScience. 10.1016/j.isci.2021.102389.

Bhuvanagiri, M., Lewis, J., Putzker, K., Becker, J.P., Leicht, S., Krijgsveld, J., Batra, R., Turnwald, B., Jovanovic, B., Hauer, C., et al. (2014). 5-azacytidine inhibits nonsense-mediated decay in a MYC-dependent fashion. EMBO Mol Med *6*, 1593-1609. 10.15252/emmm.201404461.

Bhuvanagiri, M., Schlitter, A.M., Hentze, M.W., and Kulozik, A.E. (2010). NMD: RNA biology meets human genetic medicine. Biochem J *430*, 365-377. 10.1042/BJ20100699.

Bichmann, L., Nelde, A., Ghosh, M., Heumos, L., Mohr, C., Peltzer, A., Kuchenbecker, L., Sachsenberg, T., Walz, J.S., Stevanovic, S., et al. (2019). MHCquant: Automated and Reproducible Data Analysis for Immunopeptidomics. J Proteome Res *18*, 3876-3884. 10.1021/acs.jproteome.9b00313.

Biemann, K. (1988). Contributions of mass spectrometry to peptide and protein structure. Biomedical & Environmental Mass Spectrometry *16*, 99-111. 10.1002/bms.1200160119.

Bilich, T., Nelde, A., Bauer, J., Walz, S., Roerden, M., Salih, H.R., Weisel, K., Besemer, B., Marcu, A., Lubke, M., et al. (2020). Mass spectrometry-based identification of a B-cell maturation antigen-derived T-cell epitope for antigen-specific immunotherapy of multiple myeloma. Blood Cancer J *10*, 24. 10.1038/s41408-020-0288-3.

Blatnik, R., Mohan, N., Bonsack, M., Falkenby, L.G., Hoppe, S., Josef, K., Steinbach, A., Becker, S., Nadler, W.M., Rucevic, M., et al. (2018). A Targeted LC-MS Strategy for Low-Abundant HLA Class-I-Presented Peptide Detection Identifies Novel Human Papillomavirus T-Cell Epitopes. Proteomics *18*, e1700390. 10.1002/pmic.201700390.

Bleek, G.M.V., and Nathenson, S.G. (1990). Isolation of an endogenously processed immunodominant viral peptide from the class I H–2Kb molecule. Nature *348*, 213-216. 10.1038/348213a0.

Boehm, T. (2011). Design principles of adaptive immune systems. Nat Rev Immunol *11*, 307-317. 10.1038/nri2944.

Boehm, V., Haberman, N., Ottens, F., Ule, J., and Gehring, Niels H. (2014). 3′ UTR Length and Messenger Ribonucleoprotein Composition Determine Endocleavage Efficiencies at Termination Codons. Cell Reports *9*, 555-568. 10.1016/j.celrep.2014.09.012.

Boehm, V., Kueckelmann, S., Gerbracht, J.V., Britto-Borges, T., Altmüller, J., Dieterich, C., and Gehring, N.H. (2020). Nonsense-mediated mRNA decay relies on "two-factor authentication" by SMG5-SMG7. bioRxiv. 10.1101/2020.07.07.191437.

Boelz, S., Neu-Yilik, G., Gehring, N.H., Hentze, M.W., and Kulozik, A.E. (2006). A chemiluminescence-based reporter system to monitor nonsense-mediated mRNA decay. Biochem Biophys Res Commun *349*, 186-191. 10.1016/j.bbrc.2006.08.017.

Bokhari, A., Jonchere, V., Lagrange, A., Bertrand, R., Svrcek, M., Marisa, L., Buhard, O., Greene, M., Demidova, A., Jia, J., et al. (2018). Targeting nonsense-mediated mRNA decay in colorectal cancers with microsatellite instability. Oncogenesis *7*, 70. 10.1038/s41389-018-0079-x.

Boland, C.R., and Goel, A. (2010). Microsatellite instability in colorectal cancer. Gastroenterology *138*, 2073-2087 e2073. 10.1053/j.gastro.2009.12.064.

Boland, C.R., Thibodeau, S.N., Hamilton, S.R., Sidransky, D., Eshleman, J.R., Burt, R.W., Meltzer, S.J., Rodriguez-Bigas, M.A., Fodde, R., Ranzani, G.N., and Srivastava, S. (1998). A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition: Development of International Criteria for the Determination of Microsatellite Instability in Colorectal Cancer. Cancer Research *58*, 5248-5257.

Boland, K., Flanagan, L., McCawley, N., Pabari, R., Kay, E.W., McNamara, D.A., Murray, F., Byrne, A.T., Ramtoola, Z., Concannon, C.G., and Prehn, J.H. (2016). Targeting the 19S proteasomal subunit, Rpt4, for the treatment of colon cancer. Eur J Pharmacol *780*, 53-64. 10.1016/j.ejphar.2016.03.031.

Bonilla, F.A., and Oettgen, H.C. (2010). Adaptive immunity. J Allergy Clin Immunol *125*, S33-40. 10.1016/j.jaci.2009.09.017.

Bonneville, R., Krook, M.A., Kautto, E.A., Miya, J., Wing, M.R., Chen, H.-Z., Reeser, J.W., Yu, L., and Roychowdhury, S. (2017). Landscape of Microsatellite Instability Across 39 Cancer Types. JCO precision oncology *1*, 1-15. 10.1200/po.17.00073.

Bonsack, M., Hoppe, S., Winter, J., Tichy, D., Zeller, C., Kupper, M.D., Schitter, E.C., Blatnik, R., and Riemer, A.B. (2019). Performance Evaluation of MHC Class-I Binding Prediction Tools Based on an

Experimentally Validated MHC-Peptide Binding Data Set. Cancer immunology research *7*, 719-736. 10.1158/2326-6066.CIR-18-0584.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry *72*, 248-254. 10.1016/0003-2697(76)90527-3.

Brattain, M.G., Fine, W.D., Khaled, F.M., Thompson, J., and Brattain, D.E. (1981). Heterogeneity of malignant cells from a human colonic carcinoma. Cancer Research *41*, 1751-1756.

Brattain, M.G., Levine, A.E., Chakrabarty, S., Yeoman, L.C., Willson, J.K., and Long, B. (1984). Heterogeneity of human colon carcinoma. Cancer metastasis reviews *3*, 177-191. 10.1007/bf00048384.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin *68*, 394-424. 10.3322/caac.21492.

Brentjens, R.J., Davila, M.L., Riviere, I., Park, J., Wang, X., Cowell, L.G., Bartido, S., Stefanski, J., Taylor, C., Olszewska, M., et al. (2013). CD19-targeted T cells rapidly induce molecular remissions in adults with chemotherapy-refractory acute lymphoblastic leukemia. Science translational medicine *5*, 177ra138. 10.1126/scitranslmed.3005930.

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

Brogna, S., and Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol *16*, 107-113. 10.1038/nsmb.1550.

Browning, M., Petronzelli, F., Bicknell, D., Krausa, P., Rowan, A., Tonks, S., Murray, N., Bodmer, J., and Bodmer, W. (1996). Mechanisms of loss of HLA class I expression on colorectal tumor cells. Tissue Antigens *47*, 364-371.

Brubaker, S.W., Bonham, K.S., Zanoni, I., and Kagan, J.C. (2015). Innate immune pattern recognition: a cell biological perspective. Annu Rev Immunol *33*, 257-290. 10.1146/annurev-immunol-032414-112240.

Brunet, J.-F., Denizot, F., Luciani, M.-F., Roux-Dosseto, M., Suzan, M., Mattei, M.-G., and Golstein, P. (1987). A new member of the immunoglobulin superfamily—CTLA-4. Nature *328*, 267-270. 10.1038/328267a0.

Burnet, F.M. (1970). The concept of immunological surveillance. Progress in experimental tumor research *13*, 1-27. 10.1159/000386035.

Burnet, M. (1957). Cancer: a biological approach. III. Viruses associated with neoplastic conditions. IV. Practical applications. British medical journal *1*, 841-847. 10.1136/bmj.1.5023.841.

Burnet, M. (1964). Immunological Factors In The Process Of Carcinogenesis. British medical bulletin *20*, 154-158. 10.1093/oxfordjournals.bmb.a070310.

Busch, W. (1868). Aus der Sitzung der medicinischen Section vom 13 November 1867. Berl Klin Wochenschr *5*, 137.

Butler, J.M., Buel, E., Crivellente, F., and McCord, B.R. (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis *25*, 1397-1412. 10.1002/elps.200305822.

Caron, E., Vincent, K., Fortier, M.H., Laverdure, J.P., Bramoulle, A., Hardy, M.P., Voisin, G., Roux, P.P., Lemieux, S., Thibault, P., and Perreault, C. (2011). The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. Mol Syst Biol *7*, 533. 10.1038/msb.2011.68.

Carter, M.S., Doskow, J., Morris, P., Li, S., Nhim, R.P., Sandstedt, S., and Wilkinson, M.F. (1995). A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is

reversed by protein synthesis inhibitors in vitro. J Biol Chem *270*, 28995-29003. 10.1074/jbc.270.48.28995.

Case, C.M., Sackett, D.L., Wangsa, D., Karpova, T., McNally, J.G., Ried, T., and Camps, J. (2013). CKAP2 ensures chromosomal stability by maintaining the integrity of microtubule nucleation sites. PLoS One *8*, e64575. 10.1371/journal.pone.0064575.

Chakrabarti, S., Bonneau, F., Schüssler, S., Eppinger, E., and Conti, E. (2014). Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5-SMG7 and SMG6. Nucleic Acids Res *42*, 9447-9460. 10.1093/nar/gku578.

Chamieh, H., Ballut, L., Bonneau, F., and Le Hir, H. (2008). NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nat Struct Mol Biol *15*, 85-93. 10.1038/nsmb1330.

Charif, D., and Lobry, J.R. (2007). SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In Structural Approaches to Sequence Evolution: Molecules, Networks, Populations, U. Bastolla, M. Porto, H.E. Roman, and M. Vendruscolo, eds. (Springer Berlin Heidelberg), pp. 207-232. 10.1007/978-3-540-35306-5_10.

Chen, C., Hou, J., Tanner, J.J., and Cheng, J. (2020a). Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis. Int J Mol Sci *21*. 10.3390/ijms21082873.

Chen, L., and Flies, D.B. (2013). Molecular mechanisms of T cell co-stimulation and co-inhibition. Nat Rev Immunol *13*, 227-242. 10.1038/nri3405.

Chen, L., Shi, H., Koftori, D., Sekine, T., Nicastri, A., Ternette, N., and Bowness, P. (2020b). Identification of an unconventional sub-peptidome bound to the Behçet's disease - associated HLA–B*51:01 that is regulated by endoplasmic reticulum aminopeptidase 1 (ERAP1). Mol Cell Proteomics, mcp.RA119.001617. 10.1074/mcp.RA119.001617.

Chen, R., Fulton, K.M., Twine, S.M., and Li, J. (2019). Identification of Mhc Peptides Using Mass Spectrometry for Neoantigen Discovery and Cancer Vaccine Development. Mass Spectrom Rev. 10.1002/mas.21616.

Cheruiyot, A., Li, S., Nickless, A., Roth, R., Fitzpatrick, J.A.J., and You, Z. (2018). Compound C inhibits nonsense-mediated RNA decay independently of AMPK. PLoS One *13*, e0204978. 10.1371/journal.pone.0204978.

Chi, A., Huttenhower, C., Geer, L.Y., Coon, J.J., Syka, J.E., Bai, D.L., Shabanowitz, J., Burke, D.J., Troyanskaya, O.G., and Hunt, D.F. (2007). Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proc Natl Acad Sci U S A *104*, 2193-2198. 10.1073/pnas.0607084104.

Chiu, S.Y., Serin, G., Ohara, O., and Maquat, L.E. (2003). Characterization of human Smg5/7a: a protein with similarities to Caenorhabditis elegans SMG5 and SMG7 that functions in the dephosphorylation of Upf1. RNA Biol *9*, 77-87. 10.1261/rna.2137903.

Chomczynski, P., and Sacchi, N. (2006). The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. Nature Protocols *1*, 581-585. 10.1038/nprot.2006.83.

Chong, C., Marino, F., Pak, H., Racle, J., Daniel, R.T., Müller, M., Gfeller, D., Coukos, G., and Bassani-Sternberg, M. (2018). High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferonγ-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome. Mol Cell Proteomics *17*, 533-548. 10.1074/mcp.TIR117.000383.

Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B.J., et al. (2019). Integrated Proteogenomic Deep Sequencing and Analytics Accurately Identify Non-Canonical Peptides in Tumor Immunopeptidomes. bioRxiv, 758680. 10.1101/758680.

Choucair, K., Duff, J.R., Cassidy, C.S., Albrethsen, M.T., Kelso, J.D., Lenhard, A., Staats, H., Patel, R., Brunicardi, F.C., Dworkin, L., and Nemunaitis, J. (2019). Natural killer cells: a review of biology, therapeutic potential and challenges in treatment of solid tumors. Future oncology (London, England) *15*, 3053-3069. 10.2217/fon-2019-0116.

Clay, T.M., Custer, M.C., Sachs, J., Hwu, P., Rosenberg, S.A., and Nishimura, M.I. (1999). Efficient transfer of a tumor antigen-reactive TCR to human peripheral blood lymphocytes confers anti-tumor reactivity. J Immunol *163*, 507-513.

Coley, W.B. (1891). II. Contribution to the Knowledge of Sarcoma. Annals of surgery *14*, 199-220. 10.1097/00000658-189112000-00015.

Colombo, M., Karousis, E.D., Bourquin, J., Bruggmann, R., and Mühlemann, O. (2017). Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. RNA *23*, 189-201. 10.1261/rna.059055.116.

Couzin-Frankel, J. (2013). Cancer Immunotherapy. Science (New York, N.Y.) *342*, 1432-1433. 10.1126/science.342.6165.1432.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature biotechnology *26*, 1367-1372. 10.1038/nbt.1511.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res *10*, 1794-1805. 10.1021/pr101065j.

Cramer, R. (2009). MALDI MS. Methods Mol Biol *564*, 85-103. 10.1007/978-1-60761-157-8_5.

Dang, Y., Low, W.K., Xu, J., Gehring, N.H., Dietz, H.C., Romo, D., and Liu, J.O. (2009). Inhibition of nonsense-mediated mRNA decay by the natural product pateamine A through eukaryotic initiation factor 4AIII. J Biol Chem *284*, 23613-23621. 10.1074/jbc.M109.009985.

Dekker, E., Tanis, P.J., Vleugels, J.L.A., Kasi, P.M., and Wallace, M.B. (2019). Colorectal cancer. The Lancet *394*, 1467-1480. 10.1016/s0140-6736(19)32319-0.

Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods *17*, 41-44. 10.1038/s41592-019-0638-x.

Desjardins, P., Hansen, J.B., and Allen, M. (2009). Microvolume Protein Concentration Determination using the NanoDrop 2000c Spectrophotometer. JoVE, e1610. doi:10.3791/1610.

Dobosz, P., and Dzieciatkowski, T. (2019). The Intriguing History of Cancer Immunotherapy. Front Immunol *10*, 2965. 10.3389/fimmu.2019.02965.

Doytchinova, I.A., Guan, P., and Flower, D.R. (2004). Identifiying human MHC supertypes using bioinformatic methods. J Immunol *172*, 4314-4323. 10.4049/jimmunol.172.7.4314.

Dráberová, E., D'Agostino, L., Caracciolo, V., Sládková, V., Sulimenko, T., Sulimenko, V., Sobol, M., Maounis, N.F., Tzelepis, E., Mahera, E., et al. (2015). Overexpression and Nucleolar Localization of γ-Tubulin Small Complex Proteins GCP2 and GCP3 in Glioblastoma. Journal of neuropathology and experimental neurology *74*, 723-742. 10.1097/nen.0000000000000212.

Drewinko, B., Romsdahl, M.M., Yang, L.Y., Ahearn, M.J., and Trujillo, J.M. (1976). Establishment of a human carcinoembryonic antigen-producing colon adenocarcinoma cell line. Cancer Research *36*, 467-475.

Dunn, G.P., Bruce, A.T., Ikeda, H., Old, L.J., and Schreiber, R.D. (2002). Cancer immunoediting: from immunosurveillance to tumor escape. Nature immunology *3*, 991-998. 10.1038/ni1102-991.

Dunn, G.P., Old, L.J., and Schreiber, R.D. (2004). The three Es of cancer immunoediting. Annu Rev Immunol *22*, 329-360. 10.1146/annurev.immunol.22.012703.104803.

Durand, S., Cougot, N., Mahuteau-Betzer, F., Nguyen, C.H., Grierson, D.S., Bertrand, E., Tazi, J., and Lejeune, F. (2007). Inhibition of nonsense-mediated mRNA decay (NMD) by a new chemical molecule reveals the dynamic of NMD factors in P-bodies. The Journal of cell biology *178*, 1145-1160. 10.1083/jcb.200611086.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics (Oxford, England) *21*, 3439-3440. 10.1093/bioinformatics/bti525.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature Protocols *4*, 1184-1191. 10.1038/nprot.2009.97.

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

El-Bchiri, J., Buhard, O., Penard-Lacronique, V., Thomas, G., Hamelin, R., and Duval, A. (2005). Differential nonsense mediated decay of mutated mRNAs in mismatch repair deficient colorectal cancers. Hum Mol Genet *14*, 2435-2442. 10.1093/hmg/ddi245.

El-Bchiri, J., Guilloux, A., Dartigues, P., Loire, E., Mercier, D., Buhard, O., Sobhani, I., de la Grange, P., Auboeuf, D., Praz, F., et al. (2008). Nonsense-mediated mRNA decay impacts MSI-driven carcinogenesis and anti-tumor immunity in colorectal cancers. PLoS One *3*, e2583. 10.1371/journal.pone.0002583.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods *4*, 207-214. 10.1038/nmeth1019.

Elias, J.E., and Gygi, S.P. (2010). Target-decoy search strategy for mass spectrometry-based proteomics. Methods Mol Biol *604*, 55-71. 10.1007/978-1-60761-444-9_5.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. Nat Rev Genet *5*, 435-445. 10.1038/nrg1348.

Eng, J.K., McCormack, A.L., and Yates, J.R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom *5*, 976-989. 10.1016/1044-0305(94)80016-2.

Enomoto, A., Kido, N., Ito, M., Morita, A., Matsumoto, Y., Takamatsu, N., Hosoi, Y., and Miyagawa, K. (2008). Negative regulation of MEKK1/2 signaling by serine-threonine kinase 38 (STK38). Oncogene *27*, 1930-1938. 10.1038/sj.onc.1210828.

Falk, K., Rötzschke, O., Stevanović, S., Jung, G., and Rammensee, H.-G. (1991). Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature *351*, 290-296. 10.1038/351290a0.

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., et al. (2018). A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. Science immunology *3*. 10.1126/sciimmunol.aar3947.

Fehleisen, F. (1882). Ueber die Züchtung der Erysipelkokken auf künstlichem Nährboden und ihre Übertragbarkeit auf den Menschen. Dtsch Med Wochenschr *8*, 553-554.

Fehres, C.M., Unger, W.W., Garcia-Vallejo, J.J., and van Kooyk, Y. (2014). Understanding the biology of antigen cross-presentation for the design of vaccines against cancer. Front Immunol *5*, 149. 10.3389/fimmu.2014.00149.

Fenn, J., Mann, M., Meng, C., Wong, S., and Whitehouse, C. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science (New York, N.Y.) *246*, 64-71. 10.1126/science.2675315.

Findeisen, P., Kloor, M., Merx, S., Sutter, C., Woerner, S.M., Dostmann, N., Benner, A., Dondog, B., Pawlita, M., Dippold, W., et al. (2005). T25 repeat in the 3' untranslated region of the CASP2 gene: a sensitive and specific marker for microsatellite instability in colorectal cancer. Cancer Research *65*, 8072-8078. 10.1158/0008-5472.Can-04-4146.

Finn, O.J. (2018). The dawn of vaccines for cancer prevention. Nat Rev Immunol *18*, 183-194. 10.1038/nri.2017.140.

Flajnik, M.F., and Kasahara, M. (2010). Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet *11*, 47-59. 10.1038/nrg2703.

Foote, J., and Eisen, H.N. (2000). Breaking the affinity ceiling for antibodies and T cell receptors. Proc Natl Acad Sci U S A *97*, 10679-10681. 10.1073/pnas.97.20.10679.

Freeman, G.J., Long, A.J., Iwai, Y., Bourque, K., Chernova, T., Nishimura, H., Fitz, L.J., Malenkovich, N., Okazaki, T., Byrne, M.C., et al. (2000). Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. J Exp Med *192*, 1027-1034. 10.1084/jem.192.7.1027.

Frese, C.K., Altelaar, A.F., Hennrich, M.L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A.J., and Mohammed, S. (2011). Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J Proteome Res *10*, 2377-2388. 10.1021/pr1011729.

Frese, C.K., Altelaar, A.F., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A.J., and Mohammed, S. (2012). Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. Anal Chem *84*, 9668-9673. 10.1021/ac3025366.

Freudenmann, L.K., Marcu, A., and Stevanovic, S. (2018). Mapping the tumour human leukocyte antigen (HLA) ligandome by mass spectrometry. Immunology *154*, 331-345. 10.1111/imm.12936.

Gaczynska, M., Rock, K.L., and Goldberg, A.L. (1993). γ-Interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes. Nature *365*, 264-267. 10.1038/365264a0.

Gatto, L., and Lilley, K.S. (2012). MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics (Oxford, England) *28*, 288-289. 10.1093/bioinformatics/btr645.

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. Molecular Cell *20*, 65-75. 10.1016/j.molcel.2005.08.012.

Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W., and Kulozik, A.E. (2003). Y14 and hUpf3b Form an NMD-Activating Complex. Molecular Cell *11*, 939-949. 10.1016/S1097-2765(03)00142-4.

Gejman, R.S., Chang, A.Y., Jones, H.F., DiKun, K., Hakimi, A.A., Schietinger, A., and Scheinberg, D.A. (2018). Rejection of immunogenic tumor clones is limited by clonal fraction. Elife *7*. 10.7554/eLife.41090.

Gfeller, D., Guillaume, P., Michaux, J., Pak, H.S., Daniel, R.T., Racle, J., Coukos, G., and Bassani-Sternberg, M. (2018). The Length Distribution and Multiple Specificity of Naturally Presented HLA-I Ligands. J Immunol *201*, 3705-3716. 10.4049/jimmunol.1800914.

Ghandi, M., Huang, F.W., Jane-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., 3rd, Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature *569*, 503-508. 10.1038/s41586-019-1186-3.

Giannakis, M., Mu, X.J., Shukla, S.A., Qian, Z.R., Cohen, O., Nishihara, R., Bahl, S., Cao, Y., Amin-Mansour, A., Yamauchi, M., et al. (2016). Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. Cell Reports *15*, 857-865. 10.1016/j.celrep.2016.03.075.

Gonzalez-Hilarion, S., Beghyn, T., Jia, J., Debreuck, N., Berte, G., Mamchaoui, K., Mouly, V., Gruenert, D.C., Deprez, B., and Lejeune, F. (2012). Rescue of nonsense mutations by amlexanox in human cells. Orphanet journal of rare diseases *7*, 58. 10.1186/1750-1172-7-58.

Grasso, C.S., Giannakis, M., Wells, D.K., Hamada, T., Mu, X.J., Quist, M., Nowak, J.A., Nishihara, R., Qian, Z.R., Inamura, K., et al. (2018). Genetic Mechanisms of Immune Evasion in Colorectal Cancer. Cancer Discov *8*, 730-749. 10.1158/2159-8290.CD-17-1327.

Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. Nature *481*, 306-313. 10.1038/nature10762.

Greenbaum, J., Sidney, J., Chung, J., Brander, C., Peters, B., and Sette, A. (2011). Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenetics *63*, 325-335. 10.1007/s00251-011-0513-0.

Grommé, M., Uytdehaag, F.G., Janssen, H., Calafat, J., van Binnendijk, R.S., Kenter, M.J., Tulp, A., Verwoerd, D., and Neefjes, J. (1999). Recycling MHC class I molecules and endosomal peptide loading. Proc Natl Acad Sci U S A *96*, 10326-10331. 10.1073/pnas.96.18.10326.

Gross, G., Waks, T., and Eshhar, Z. (1989). Expression of immunoglobulin-T-cell receptor chimeric molecules as functional receptors with antibody-type specificity. Proc Natl Acad Sci U S A *86*, 10024-10028. 10.1073/pnas.86.24.10024.

Gubin, M.M., Zhang, X., Schuster, H., Caron, E., Ward, J.P., Noguchi, T., Ivanova, Y., Hundal, J., Arthur, C.D., Krebber, W.J., et al. (2014). Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature *515*, 577-581. 10.1038/nature13988.

Guillaume, P., Picaud, S., Baumgaertner, P., Montandon, N., Schmidt, J., Speiser, D.E., Coukos, G., Bassani-Sternberg, M., Filippakopoulos, P., and Gfeller, D. (2018). The C-terminal extension landscape of naturally presented HLA-I ligands. Proc Natl Acad Sci U S A *115*, 5083-5088. 10.1073/pnas.1717277115.

Guo, T., Kouvonen, P., Koh, C.C., Gillet, L.C., Wolski, W.E., Rost, H.L., Rosenberger, G., Collins, B.C., Blum, L.C., Gillessen, S., et al. (2015). Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. Nat Med *21*, 407-413. 10.1038/nm.3807.

Haag, A.M. (2016). Mass Analyzers and Mass Spectrometers. Adv Exp Med Biol *919*, 157-169. 10.1007/978-3-319-41448-5_7.

Hagemann, S., Heil, O., Lyko, F., and Brueckner, B. (2011). Azacytidine and decitabine induce gene-specific and non-random DNA demethylation in human cancer cell lines. PLoS One *6*, e17388. 10.1371/journal.pone.0017388.

Halvey, P.J., Wang, X., Wang, J., Bhat, A.A., Dhawan, P., Li, M., Zhang, B., Liebler, D.C., and Slebos, R.J. (2014). Proteogenomic analysis reveals unanticipated adaptations of colorectal tumor cells to deficiencies in DNA mismatch repair. Cancer Research *74*, 387-397. 10.1158/0008-5472.CAN-13-2488.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70. 10.1016/s0092-8674(00)81683-9.

Hanahan, D., and Weinberg, Robert A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646-674. 10.1016/j.cell.2011.02.013.

Hargadon, K.M., Johnson, C.E., and Williams, C.J. (2018). Immune checkpoint blockade therapy for cancer: An overview of FDA-approved immune checkpoint inhibitors. Int Immunopharmacol *62*, 29-39. 10.1016/j.intimp.2018.06.001.

Hassan, C., Kester, M.G., Oudgenoeg, G., de Ru, A.H., Janssen, G.M., Drijfhout, J.W., Spaapen, R.M., Jimenez, C.R., Heemskerk, M.H., Falkenburg, J.H., and van Veelen, P.A. (2014). Accurate quantitation of MHC-bound peptides by application of isotopically labeled peptide MHC complexes. J Proteomics *109*, 240-244. 10.1016/j.jprot.2014.07.009.

Hato, T., and Dagher, P.C. (2015). How the Innate Immune System Senses Trouble and Causes Trouble. Clin J Am Soc Nephrol *10*, 1459-1469. 10.2215/CJN.04680514.

Hecht, E.S., Scigelova, M., Eliuk, S., and Makarov, A. (2019). Fundamentals and Advances of Orbitrap Mass Spectrometry. In Encyclopedia of Analytical Chemistry, pp. 1-40. 10.1002/9780470027318.a9309.pub2.

Heo, J., Eki, R., and Abbas, T. (2016). Deregulation of F-box proteins and its consequence on cancer development, progression and metastasis. Semin Cancer Biol *36*, 33-51. 10.1016/j.semcancer.2015.09.015.

Hilf, N., Kuttruff-Coqui, S., Frenzel, K., Bukur, V., Stevanovic, S., Gouttefangeas, C., Platten, M., Tabatabai, G., Dutoit, V., van der Burg, S.H., et al. (2019). Actively personalized vaccination trial for newly diagnosed glioblastoma. Nature *565*, 240-245. 10.1038/s41586-018-0810-y.

Ho, N.I., Huis In 't Veld, L.G.M., Raaijmakers, T.K., and Adema, G.J. (2018). Adjuvants Enhancing Cross-Presentation by Dendritic Cells: The Key to More Effective Vaccines? Front Immunol *9*, 2874. 10.3389/fimmu.2018.02874.

Hoff, F.W., Hu, C.W., Qutub, A.A., Qiu, Y., Graver, E., Hoang, G., Chauhan, M., de Bont, E., and Kornblau, S.M. (2018). Mycoplasma contamination of leukemic cell lines alters protein expression determined by reverse phase protein arrays. Cytotechnology *70*, 1529-1535. 10.1007/s10616-018-0244-2.

Hogg, J.R., and Goff, S.P. (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. Cell *143*, 379-389. 10.1016/j.cell.2010.10.005.

Holbrook, J.A., Neu-Yilik, G., Hentze, M.W., and Kulozik, A.E. (2004). Nonsense-mediated decay approaches the clinic. Nat Genet *36*, 801-808. 10.1038/ng1403.

Holmes, S., and Huber, W. (2019). Modern Statistics for Modern Biology (Cambridge University Press).

Howard, M., Frizzell, R.A., and Bedwell, D.M. (1996). Aminoglycoside antibiotics restore CFTR function by overcoming premature stop mutations. Nat Med *2*, 467-469. 10.1038/nm0496-467.

Hsieh, E.J., Bereman, M.S., Durand, S., Valaskovic, G.A., and MacCoss, M.J. (2013). Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples. J Am Soc Mass Spectrom *24*, 148-153. 10.1007/s13361-012-0508-6.

Hu, Z., Ott, P.A., and Wu, C.J. (2018). Towards personalized, tumour-specific, therapeutic vaccines for cancer. Nat Rev Immunol *18*, 168-182. 10.1038/nri.2017.131.

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics (Oxford, England) *18*, S96-S104. 10.1093/bioinformatics/18.suppl_1.S96.

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

Ichou, F., Schwarzenberg, A., Lesage, D., Alves, S., Junot, C., Machuron-Mandard, X., and Tabet, J.C. (2014). Comparison of the activation time effects and the internal energy distributions for the CID, PQD and HCD excitation modes. Journal of mass spectrometry : JMS *49*, 498-508. 10.1002/jms.3365.

Imai, K., and Yamamoto, H. (2008). Carcinogenesis and microsatellite instability: the interrelationship between genetics and epigenetics. Carcinogenesis *29*, 673-680. 10.1093/carcin/bgm228.

Inderberg, E.M., Walchli, S., Myhre, M.R., Trachsel, S., Almasbak, H., Kvalheim, G., and Gaudernack, G. (2017). T cell therapy targeting a public neoantigen in microsatellite instable colon cancer reduces in vivo tumor growth. Oncoimmunology *6*, e1302631. 10.1080/2162402X.2017.1302631.

Ishida, Y., Agata, Y., Shibahara, K., and Honjo, T. (1992). Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. The EMBO Journal *11*, 3887-3895.

Isken, O., Kim, Y.K., Hosoda, N., Mayeur, G.L., Hershey, J.W., and Maquat, L.E. (2008). Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. Cell *133*, 314-327. 10.1016/j.cell.2008.02.030.

Ivanov, P.V., Gehring, N.H., Kunz, J.B., Hentze, M.W., and Kulozik, A.E. (2008). Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. The EMBO Journal *27*, 736-747. 10.1038/emboj.2008.17.

Iwai, Y., Ishida, M., Tanaka, Y., Okazaki, T., Honjo, T., and Minato, N. (2002). Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. Proc Natl Acad Sci U S A *99*, 12293-12297. 10.1073/pnas.192461099.

Jedrychowski, M.P., Huttlin, E.L., Haas, W., Sowa, M.E., Rad, R., and Gygi, S.P. (2011). Evaluation of HCD- and CID-type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics*. Mol Cell Proteomics *10*, M111.009910. 10.1074/mcp.M111.009910.

Jiricny, J. (2006). The multifaceted mismatch-repair system. Nature reviews. Molecular cell biology *7*, 335-346. 10.1038/nrm1907.

Jonchere, V., Marisa, L., Greene, M., Virouleau, A., Buhard, O., Bertrand, R., Svrcek, M., Cervera, P., Goloudina, A., Guillerm, E., et al. (2018). Identification of Positively and Negatively Selected Driver Gene Mutations Associated With Colorectal Cancer With Microsatellite Instability. Cell Mol Gastroenterol Hepatol *6*, 277-300. 10.1016/j.jcmgh.2018.06.002.

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. J Immunol *199*, 3360-3368. 10.4049/jimmunol.1700893.

Kalos, M., Levine, B.L., Porter, D.L., Katz, S., Grupp, S.A., Bagg, A., and June, C.H. (2011). T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. Science translational medicine *3*, 95ra73. 10.1126/scitranslmed.3002842.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem *60*, 2299-2301. 10.1021/ac00171a028.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev *20*, 355-367. 10.1101/gad.1389006.

Keating, G.M. (2012). Azacitidine: a review of its use in the management of myelodysplastic syndromes/acute myeloid leukaemia. Drugs *72*, 1111-1136. 10.2165/11209430-000000000-00000.

Kebarle, P., and Verkerk, U.H. (2009). Electrospray: from ions in solution to ions in the gas phase, what we know now. Mass Spectrom Rev *28*, 898-917. 10.1002/mas.20247.

Keeling, K.M., Wang, D., Conard, S.E., and Bedwell, D.M. (2012). Suppression of premature termination codons as a therapeutic approach. Critical reviews in biochemistry and molecular biology *47*, 444-463. 10.3109/10409238.2012.694846.

Keir, M.E., Butte, M.J., Freeman, G.J., and Sharpe, A.H. (2008). PD-1 and its ligands in tolerance and immunity. Annu Rev Immunol *26*, 677-704. 10.1146/annurev.immunol.26.021607.090331.

Khong, H.T., and Restifo, N.P. (2002). Natural selection of tumor variants in the generation of “tumor escape” phenotypes. Nature immunology *3*, 999-1005. 10.1038/ni1102-999.

Kim, M.S., and Pandey, A. (2012). Electron transfer dissociation mass spectrometry in proteomics. Proteomics *12*, 530-542. 10.1002/pmic.201100517.

Kim, N.-G., Rhee, H., Li, L.S., Kim, H., Lee, J.-S., Kim, J.-H., Kim, N.K., and Kim, H. (2002). Identification of MARCKS, FLJ11383 and TAF1B as putative novel target genes in colorectal carcinomas with microsatellite instability. Oncogene *21*, 5081-5087. 10.1038/sj.onc.1205703.

Kim, T.M., Laird, P.W., and Park, P.J. (2013a). The landscape of microsatellite instability in colorectal and endometrial cancer genomes. Cell *155*, 858-868. 10.1016/j.cell.2013.10.015.

Kim, W.K., Park, M., Park, M., Kim, Y.J., Shin, N., Kim, H.K., You, K.T., and Kim, H. (2013b). Identification and selective degradation of neopeptide-containing truncated mutant proteins in the tumors with high microsatellite instability. Clin Cancer Res *19*, 3369-3382. 10.1158/1078-0432.CCR-13-0684.

Kincaid, E.Z., Murata, S., Tanaka, K., and Rock, K.L. (2016). Specialized proteasome subunits have an essential role in the thymic selection of CD8(+) T cells. Nature immunology *17*, 938-945. 10.1038/ni.3480.

Klein, G. (1966). Tumor Antigens. Annual Review of Microbiology *20*, 223-252. 10.1146/annurev.mi.20.100166.001255.

Kloor, M., Michel, S., Buckowitz, B., Ruschoff, J., Buttner, R., Holinski-Feder, E., Dippold, W., Wagner, R., Tariverdian, M., Benner, A., et al. (2007). Beta2-microglobulin mutations in microsatellite unstable colorectal tumors. Int J Cancer *121*, 454-458. 10.1002/ijc.22691.

Kloor, M., Reuschenbach, M., Pauligk, C., Karbach, J., Rafiyan, M.-R., Al-Batran, S.-E., Tariverdian, M., Jäger, E., and von Knebel Doeberitz, M. (2020a). A Frameshift Peptide Neoantigen-Based Vaccine for Mismatch Repair-Deficient Cancers: A Phase I/IIa Clinical Trial. Clin Cancer Res *26*, 4503-4510. 10.1158/1078-0432.Ccr-19-3517.

Kloor, M., Reuschenbach, M., Pauligk, C., Karbach, J., Rafiyan, M.R., Al-Batran, S.E., Tariverdian, M., Jaeger, E., and von Knebel Doeberitz, M. (2020b). A Phase I/IIa trial of a frameshift peptide neoantigen-based vaccine for mismatch repair-deficient cancers. Clin Cancer Res. 10.1158/1078-0432.CCR-19-3517.

Kloor, M., and von Knebel Doeberitz, M. (2016). The Immune Biology of Microsatellite-Unstable Cancer. Trends Cancer *2*, 121-133. 10.1016/j.trecan.2016.02.004.

Knutson, B.A., and Hahn, S. (2011). Yeast Rrn7 and human TAF1B are TFIIB-related RNA polymerase I general transcription factors. Science (New York, N.Y.) *333*, 1637-1640. 10.1126/science.1207699.

Koch, U., and Radtke, F. (2011). Mechanisms of T cell development and transformation. Annu Rev Cell Dev Biol *27*, 539-562. 10.1146/annurev-cellbio-092910-154008.

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., Lanier, B.J., et al. (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*, 4099-4102. 10.1182/blood-2010-04-281931.

Konermann, L., Ahadi, E., Rodriguez, A.D., and Vahidi, S. (2013). Unraveling the mechanism of electrospray ionization. Anal Chem *85*, 2-9. 10.1021/ac302789c.

Kovacsovics-Bankowski, M., and Rock, K. (1995). A phagosome-to-cytosol pathway for exogenous antigens presented on MHC class I molecules. Science (New York, N.Y.) *267*, 243-246. 10.1126/science.7809629.

Krokhin, O.V. (2006). Sequence-Specific Retention Calculator. Algorithm for Peptide Retention Prediction in Ion-Pair RP-HPLC: Application to 300- and 100-Å Pore Size C18 Sorbents. Anal Chem *78*, 7785-7795. 10.1021/ac060777w.

Krummel, M.F., and Allison, J.P. (1995). CD28 and CTLA-4 have opposing effects on the response of T cells to stimulation. J Exp Med *182*, 459-465. 10.1084/jem.182.2.459.

Kucherlapati, M.H., Lee, K., Nguyen, A.A., Clark, A.B., Hou, H., Jr., Rosulek, A., Li, H., Yang, K., Fan, K., Lipkin, M., et al. (2010). An Msh2 conditional knockout mouse for studying intestinal cancer and testing anticancer agents. Gastroenterology *138*, 993-1002 e1001. 10.1053/j.gastro.2009.11.009.

Kumar, G. (2018). Principle and Method of Silver Staining of Proteins Separated by Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis. Methods Mol Biol *1853*, 231-236. 10.1007/978-1-4939-8745-0_26.

Kurosaki, T., and Maquat, L.E. (2013). Rules that govern UPF1 binding to mRNA 3' UTRs. Proc Natl Acad Sci U S A *110*, 3357-3362. 10.1073/pnas.1219908110.

Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nature reviews. Molecular cell biology. 10.1038/s41580-019-0126-2.

Kuznetsov, A., Voronina, A., Govorun, V., and Arapidi, G. (2020). Critical Review of Existing MHC I Immunopeptidome Isolation Methods. Molecules *25*. 10.3390/molecules25225409.

Lanoix, J., Durette, C., Courcelles, M., Cossette, E., Comtois-Marotte, S., Hardy, M.P., Cote, C., Perreault, C., and Thibault, P. (2018). Comparison of the MHC I Immunopeptidome Repertoire of B-Cell Lymphoblasts Using Two Isolation Methods. Proteomics *18*, e1700251. 10.1002/pmic.201700251.

Lavysh, D., and Neu-Yilik, G. (2020). UPF1-Mediated RNA Decay-Danse Macabre in a Cloud. Biomolecules *10*, 999. 10.3390/biom10070999.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. The EMBO Journal *20*, 4987-4997. 10.1093/emboj/20.17.4987.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. The EMBO Journal *19*, 6860-6869. 10.1093/emboj/19.24.6860.

Leach, D.R., Krummel, M.F., and Allison, J.P. (1996). Enhancement of Antitumor Immunity by CTLA-4 Blockade. Science (New York, N.Y.) *271*, 1734-1736. 10.1126/science.271.5256.1734.

Lee, K., Tosti, E., and Edelmann, W. (2016). Mouse models of DNA mismatch repair in cancer research. DNA Repair *38*, 140-146. 10.1016/j.dnarep.2015.11.015.

Leeds, P., Wood, J.M., Lee, B.S., and Culbertson, M.R. (1992). Gene products that promote mRNA turnover in Saccharomyces cerevisiae. Molecular and Cellular Biology *12*, 2165-2177. 10.1128/mcb.12.5.2165.

Leoni, G., D'Alise, A.M., Cotugno, G., Langone, F., Garzia, I., De Lucia, M., Fichera, I., Vitale, R., Bignone, V., Tucci, F.G., et al. (2020). A Genetic Vaccine Encoding Shared Cancer Neoantigens to Treat Tumors with Microsatellite Instability. Cancer Research *80*, 3972-3982. 10.1158/0008-5472.CAN-20-1072.

Levinson, G., and Gutman, G.A. (1987). High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12. Nucleic Acids Res *15*, 5323-5338. 10.1093/nar/15.13.5323.

Liepe, J., Marino, F., Sidney, J., Jeko, A., Bunting, D.E., Sette, A., Kloetzel, P.M., Stumpf, M.P., Heck, A.J., and Mishto, M. (2016). A large fraction of HLA class I ligands are proteasome-generated spliced peptides. Science (New York, N.Y.) *354*, 354-358. 10.1126/science.aaf4384.

Liepe, J., Sidney, J., Lorenz, F.K.M., Sette, A., and Mishto, M. (2019). Mapping the MHC Class I-Spliced Immunopeptidome of Cancer Cells. Cancer immunology research *7*, 62-76. 10.1158/2326-6066.CIR-18-0424.

Lindeboom, R.G., Supek, F., and Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. Nat Genet *48*, 1112-1118. 10.1038/ng.3664.

Lindeboom, R.G.H., Vermeulen, M., Lehner, B., and Supek, F. (2019). The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. Nat Genet. 10.1038/s41588-019-0517-5.

Litchfield, K., Reading, J.L., Lim, E.L., Xu, H., Liu, P., Al-Bakir, M., Wong, Y.N.S., Rowan, A., Funt, S.A., Merghoub, T., et al. (2020). Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. Nat Commun *11*, 3800. 10.1038/s41467-020-17526-5.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif.) *25*, 402-408. 10.1006/meth.2001.1262.

Llosa, N.J., Cruise, M., Tam, A., Wicks, E.C., Hechenbleikner, E.M., Taube, J.M., Blosser, R.L., Fan, H., Wang, H., Luber, B.S., et al. (2015). The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. Cancer Discov *5*, 43-51. 10.1158/2159-8290.CD-14-0863.

Loh, B., Jonas, S., and Izaurralde, E. (2013). The SMG5–SMG7 heterodimer directly recruits the CCR4–NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. Genes & Development *27*, 2125-2138. 10.1101/gad.226951.113.

Lowry, O.H., Rosebrough, N.J., Farr, A.L., and Randall, R.J. (1951). Protein measurement with the Folin phenol reagent. J Biol Chem *193*, 265-275.

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., et al. (2014). Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions. Clin Cancer Res *20*, 3401-3410. 10.1158/1078-0432.Ccr-14-0433.

Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., and Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol *14*, e8126. 10.15252/msb.20178126.

Lund, O., Nielsen, M., Kesmir, C., Petersen, A.G., Lundegaard, C., Worning, P., Sylvester-Hvid, C., Lamberth, K., Roder, G., Justesen, S., et al. (2004). Definition of supertypes for HLA molecules using clustering of specificity matrices. Immunogenetics *55*, 797-810. 10.1007/s00251-004-0647-4.

Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., and Lajoie, G. (2003). PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom *17*, 2337-2342. 10.1002/rcm.1196.

Maher, J., Brentjens, R.J., Gunset, G., Rivière, I., and Sadelain, M. (2002). Human T-lymphocyte cytotoxicity and proliferation directed by a single chimeric TCRzeta /CD28 receptor. Nature biotechnology *20*, 70-75. 10.1038/nbt0102-70.

Mak, T.W., Saunders, M.E., and Jett, B.D. (2014a). Chapter 5 - B Cell Development, Activation and Effector Functions. In Primer to the Immune Response (Second Edition), T.W. Mak, M.E. Saunders, and B.D. Jett, eds. (Academic Cell), pp. 111-142.

10.1016/B978-0-12-385245-8.00005-4.

Mak, T.W., Saunders, M.E., and Jett, B.D. (2014b). Chapter 6 - The Major Histocompatibility Complex. In Primer to the Immune Response (Second Edition), T.W. Mak, M.E. Saunders, and B.D. Jett, eds. (Academic Cell), pp. 143-159. 10.1016/B978-0-12-385245-8.00006-6.

Mak, T.W., Saunders, M.E., and Jett, B.D. (2014c). Chapter 7 - Antigen Processing and Presentation. In Primer to the Immune Response (Second Edition), T.W. Mak, M.E. Saunders, and B.D. Jett, eds. (Academic Cell), pp. 161-179. 10.1016/B978-0-12-385245-8.00007-8.

Mak, T.W., Saunders, M.E., and Jett, B.D. (2014d). Chapter 8 - The T Cell Receptor: Proteins and Genes. In Primer to the Immune Response (Second Edition), T.W. Mak, M.E. Saunders, and B.D. Jett, eds. (Academic Cell), pp. 181-196. 10.1016/B978-0-12-385245-8.00008-X.

Mak, T.W., Saunders, M.E., and Jett, B.D. (2014e). Chapter 9 - T Cell Development, Activation and Effector Functions. In Primer to the Immune Response (Second Edition), T.W. Mak, M.E. Saunders, and B.D. Jett, eds. (Academic Cell), pp. 197-226. 10.1016/B978-0-12-385245-8.00009-1.

Makarov, A. (2000). Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal Chem *72*, 1156-1162. 10.1021/ac991131p.

Maletzki, C., Schmidt, F., Dirks, W.G., Schmitt, M., and Linnebacher, M. (2013). Frameshift-derived neoantigens constitute immunotherapeutic targets for patients with microsatellite-instable haematological malignancies: frameshift peptides for treating MSI+ blood cancers. Eur J Cancer *49*, 2587-2595. 10.1016/j.ejca.2013.02.035.

Mann, M. (2019). The ever expanding scope of electrospray mass spectrometry-a 30 year journey. Nat Commun *10*, 3744. 10.1038/s41467-019-11747-z.

Marcus, L., Lemery, S.J., Keegan, P., and Pazdur, R. (2019). FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors. Clin Cancer Res *25*, 3753-3758. 10.1158/1078-0432.Ccr-18-4070.

Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R., Zborowska, E., Kinzler, K., Vogelstein, B., and et, a. (1995). Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science (New York, N.Y.) *268*, 1336-1338. 10.1126/science.7761852.

Martin, L., Grigoryan, A., Wang, D., Wang, J., Breda, L., Rivella, S., Cardozo, T., and Gardner, L.B. (2014). Identification and characterization of small molecules that inhibit nonsense-mediated RNA decay and suppress nonsense p53 mutations. Cancer Research *74*, 3104-3113. 10.1158/0008-5472.CAN-13-2235.

Martins, C.P.B., Bromirski, M., Prieto Conaway, M.C., and Makarov, A.A. (2016). Orbitrap Mass Spectrometry. In Applications of Time-of-Flight and Orbitrap Mass Spectrometry in Environmental, Food, Doping, and Forensic Analysis, pp. 3-18. 10.1016/bs.coac.2016.01.001.

Marty Pyke, R., Thompson, W.K., Salem, R.M., Font-Burgada, J., Zanetti, M., and Carter, H. (2018). Evolutionary Pressure against MHC Class II Binding Cancer Mutations. Cell *175*, 416-428 e413. 10.1016/j.cell.2018.08.048.

Marty, R., Kaabinejadian, S., Rossell, D., Slifker, M.J., van de Haar, J., Engin, H.B., de Prisco, N., Ideker, T., Hildebrand, W.H., Font-Burgada, J., and Carter, H. (2017). MHC-I Genotype Restricts the Oncogenic Mutational Landscape. Cell *171*, 1272-1283 e1215. 10.1016/j.cell.2017.09.050.

Maude, S.L., Laetsch, T.W., Buechner, J., Rives, S., Boyer, M., Bittencourt, H., Bader, P., Verneris, M.R., Stefanski, H.E., Myers, G.D., et al. (2018). Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. New England Journal of Medicine *378*, 439-448. 10.1056/NEJMoa1709866.

McCawley, N., Conlon, S., Hector, S., Cummins, R.J., Dicker, P., Johnston, P.G., Kay, E.W., McNamara, D.A., Prehn, J.H., and Concannon, C.G. (2012). Analyzing proteasomal subunit expression reveals Rpt4 as a prognostic marker in stage II colorectal cancer. Int J Cancer *131*, E494-500. 10.1002/ijc.26468.

McElroy, S.P., Nomura, T., Torrie, L.S., Warbrick, E., Gartner, U., Wood, G., and McLean, W.H.I. (2013). A Lack of Premature Termination Codon Read-Through Efficacy of PTC124 (Ataluren) in a Diverse Array of Reporter Assays. PLoS Biol *11*, e1001593. 10.1371/journal.pbio.1001593.

McGranahan, N., Furness, A.J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S.K., Jamal-Hanjani, M., Wilson, G.A., Birkbak, N.J., Hiley, C.T., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science (New York, N.Y.) *351*, 1463-1469. 10.1126/science.aaf1490.

McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birkbak, N.J., Veeriah, S., Van Loo, P., Herrero, J., et al. (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. Cell *171*, 1259-1271 e1211. 10.1016/j.cell.2017.10.001.

McGranahan, N., and Swanton, C. (2019). Neoantigen quality, not quantity. Science Transl Med *11*, eaax7918. 10.1126/scitranslmed.aax7918.

Merle, N.S., Church, S.E., Fremeaux-Bacchi, V., and Roumenina, L.T. (2015a). Complement System Part I - Molecular Mechanisms of Activation and Regulation. Front Immunol *6*, 262. 10.3389/fimmu.2015.00262.

Merle, N.S., Noe, R., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V., and Roumenina, L.T. (2015b). Complement System Part II: Role in Immunity. Front Immunol *6*, 257. 10.3389/fimmu.2015.00257.

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., and Thomas, P.D. (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nature Protocols *14*, 703-721. 10.1038/s41596-019-0128-8.

Michalak, M., Katzenmaier, E.-M., Roeckel, N., Woerner, S.M., Fuchs, V., Warnken, U., Yuan, Y.P., Bork, P., Neu-Yilik, G., Kulozik, A., et al. (2020). (Phospho)proteomic Profiling of Microsatellite Unstable CRC Cells Reveals Alterations in Nuclear Signaling and Cholesterol Metabolism Caused by Frameshift Mutation of NMD Regulator UPF3A. International Journal of Molecular Sciences *21*, 5234. 10.3390/ijms21155234.

Miller, C.J., Kassem, H.S., Pepper, S.D., Hey, Y., Ward, T.H., and Margison, G.P. (2003). Mycoplasma infection significantly alters microarray gene expression profiles. BioTechniques *35*, 812-814. 10.2144/03354mt02.

Mittal, D., Gubin, M.M., Schreiber, R.D., and Smyth, M.J. (2014). New insights into cancer immunoediting and its three component phases--elimination, equilibrium and escape. Curr Opin Immunol *27*, 16-25. 10.1016/j.coi.2014.01.004.

Mlecnik, B., Bindea, G., Angell, H.K., Maby, P., Angelova, M., Tougeron, D., Church, S.E., Lafontaine, L., Fischer, M., Fredriksen, T., et al. (2016). Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. Immunity *44*, 698-711. 10.1016/j.immuni.2016.02.025.

Mogensen, T.H. (2009). Pathogen recognition and inflammatory signaling in innate immune defenses. Clin Microbiol Rev *22*, 240-273, Table of Contents. 10.1128/CMR.00046-08.

Mommen, G.P., Frese, C.K., Meiring, H.D., van Gaans-van den Brink, J., de Jong, A.P., van Els, C.A., and Heck, A.J. (2014). Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A *111*, 4507-4512. 10.1073/pnas.1321458111.

Morales, A., Eidinger, D., and Bruce, A.W. (1976). Intracavitary Bacillus Calmette-guerin in the Treatment of Superficial Bladder Tumors. Journal of Urology *116*, 180-182. doi:10.1016/S0022-5347(17)58737-6.

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., et al. (2006). Cancer regression in patients after transfer of genetically engineered lymphocytes. Science (New York, N.Y.) *314*, 126-129. 10.1126/science.1129003.

Mühlemann, O. (2008). Recognition of nonsense mRNA: towards a unified model. Biochem Soc Trans *36*, 497-501. 10.1042/BST0360497.

Müller, A.M.S., and Florek, M. (2014). 5-Azacytidine/5-Azacitidine. Recent Results Cancer Res *201*, 299-324. 10.1007/978-3-642-54490-3_19.

Murata, S., Sasaki, K., Kishimoto, T., Niwa, S., Hayashi, H., Takahama, Y., and Tanaka, K. (2007). Regulation of CD8+ T cell development by thymus-specific proteasomes. Science (New York, N.Y.) *316*, 1349-1353. 10.1126/science.1141915.

Murphy, K., and Weaver, C. (2017). Janeway's immunobiology.

Mylonas, R., Beer, I., Iseli, C., Chong, C., Pak, H.S., Gfeller, D., Coukos, G., Xenarios, I., Muller, M., and Bassani-Sternberg, M. (2018). Estimating the Contribution of Proteasomal Spliced Peptides to the HLA-I Ligandome. Mol Cell Proteomics *17*, 2347-2357. 10.1074/mcp.RA118.000877.

Nakano, K., Shiroma, A., Shimoji, M., Tamotsu, H., Ashimine, N., Ohki, S., Shinzato, M., Minami, M., Nakanishi, T., Teruya, K., et al. (2017). Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. Hum Cell *30*, 149-161. 10.1007/s13577-017-0168-8.

Narayan, R., Olsson, N., Wagar, L.E., Medeiros, B.C., Meyer, E., Czerwinski, D., Khodadoust, M.S., Zhang, L., Schultz, L., Davis, M.M., et al. (2019). Acute myeloid leukemia immunopeptidome reveals HLA presentation of mutated nucleophosmin. PLoS One *14*, e0219547. 10.1371/journal.pone.0219547.

Narzisi, G., and Schatz, M.C. (2015). The challenge of small-scale repeats for indel discovery. Front Bioeng Biotechnol *3*, 8. 10.3389/fbioe.2015.00008.

Neefjes, J., Jongsma, M.L., Paul, P., and Bakke, O. (2011). Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol *11*, 823-836. 10.1038/nri3084.

Neisig, A., Melief, C.J.M., and Neefjes, J. (1998). Reduced Cell Surface Expression of HLA-C Molecules Correlates with Restricted Peptide Binding and Stable TAP Interaction. J Immunol *160*, 171-179.

Neu-Yilik, G., Amthor, B., Gehring, N.H., Bahri, S., Paidassi, H., Hentze, M.W., and Kulozik, A.E. (2011). Mechanism of escape from nonsense-mediated mRNA decay of human beta-globin transcripts with nonsense mutations in the first exon. RNA *17*, 843-854. 10.1261/rna.2401811.

Neu-Yilik, G., Raimondeau, E., Eliseev, B., Yeramala, L., Amthor, B., Deniaud, A., Huard, K., Kerschgens, K., Hentze, M.W., Schaffitzel, C., and Kulozik, A.E. (2017). Dual function of UPF3B in early and late translation termination. The EMBO Journal *36*, 2968-2986. 10.15252/embj.201797079.

Newton, K., and Dixit, V.M. (2012). Signaling in innate immunity and inflammation. Cold Spring Harb Perspect Biol *4*. 10.1101/cshperspect.a006049.

Nicastri, A., Liao, H., Muller, J., Purcell, A.W., and Ternette, N. (2020). The Choice of HLA-Associated Peptide Enrichment and Purification Strategy Affects Peptide Yields and Creates a Bias in Detected Sequence Repertoire. Proteomics, e1900401. 10.1002/pmic.201900401.

Niessen, W.M., and Falck, D. (2015). Introduction to mass spectrometry, a tutorial. Analyzing Biomolecular Interactions by Mass Spectrometry, 1-54.

Noor, Z., Ahn, S.B., Baker, M.S., Ranganathan, S., and Mohamedali, A. (2020). Mass spectrometry-based protein identification in proteomics-a review. Brief Bioinform. 10.1093/bib/bbz163.

O'Donnell, T.J., Rubinsteyn, A., Bonsack, M., Riemer, A.B., Laserson, U., and Hammerbacher, J. (2018). MHCflurry: Open-Source Class I MHC Binding Affinity Prediction. Cell Syst *7*, 129-132 e124. 10.1016/j.cels.2018.05.014.

O'Rawe, J., Jiang, T., Sun, G., Wu, Y., Wang, W., Hu, J., Bodily, P., Tian, L., Hakonarson, H., Johnson, W.E., et al. (2013). Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. Genome Med *5*, 28. 10.1186/gm432.

Ogino, S., Gulley, M.L., den Dunnen, J.T., and Wilson, R.B. (2007). Standard mutation nomenclature in molecular diagnostics: practical and educational challenges. The Journal of molecular diagnostics : JMD *9*, 1-6. 10.2353/jmoldx.2007.060081.

Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F., and Ohno, S. (2011). N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. Nucleic Acids Res *40*, 1251-1266. 10.1093/nar/gkr791.

Olsen, J.V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007). Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods *4*, 709-712. 10.1038/nmeth1060.

Olsen, J.V., Schwartz, J.C., Griep-Raming, J., Nielsen, M.L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., et al. (2009). A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. Mol Cell Proteomics *8*, 2759-2769. 10.1074/mcp.M900375-MCP200.

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. Mol Cell Proteomics *1*, 376-386. 10.1074/mcp.M200025-MCP200.

Osipov, A., Lim, S.J., Popovic, A., Azad, N.S., Laheru, D.A., Zheng, L., Jaffee, E.M., Wang, H., and Yarchoan, M. (2020). Tumor Mutational Burden, Toxicity, and Response of Immune Checkpoint Inhibitors Targeting PD(L)1, CTLA-4, and Combination: A Meta-regression Analysis. Clin Cancer Res *26*, 4842-4851. 10.1158/1078-0432.CCR-20-0458.

Ott, P.A., Hu, Z., Keskin, D.B., Shukla, S.A., Sun, J., Bozym, D.J., Zhang, W., Luoma, A., Giobbie-Hurder, A., Peter, L., et al. (2017). An immunogenic personal neoantigen vaccine for patients with melanoma. Nature *547*, 217-221. 10.1038/nature22991.

Ottens, F., and Gehring, N.H. (2016). Physiological and pathophysiological role of nonsense-mediated mRNA decay. Pflugers Arch *468*, 1013-1028. 10.1007/s00424-016-1826-5.

Pajot, A., Michel, M.L., Fazilleau, N., Pancre, V., Auriault, C., Ojcius, D.M., Lemonnier, F.A., and Lone, Y.C. (2004). A mouse model of human adaptive immune functions: HLA-A2.1-/HLA-DR1-transgenic H-2 class I-/class II-knockout mice. Eur J Immunol *34*, 3060-3069. 10.1002/eji.200425463.

Pak, H., Michaux, J., Huber, F., Chong, C., Stevenson, B.J., Müller, M., Coukos, G., and Bassani-Sternberg, M. (2021). Sensitive immunopeptidomics by leveraging available large-scale multi-HLA spectral libraries, data-independent acquisition and MS/MS prediction. Mol Cell Proteomics. 10.1016/j.mcpro.2021.100080.

Pardoll, D.M. (2012). The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer *12*, 252-264. 10.1038/nrc3239.

Parham, P., Barnstable, C.J., and Bodmer, W.F. (1979). Use of a monoclonal antibody (W6/32) in structural studies of HLA-A,B,C, antigens. J Immunol *123*, 342-349.

Pastor, F., Kolonias, D., Giangrande, P.H., and Gilboa, E. (2010). Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay. Nature *465*, 227-230. 10.1038/nature08999.

Pawlicka, K., Kalathiya, U., and Alfaro, J. (2020). Nonsense-Mediated mRNA Decay: Pathologies and the Potential for Novel Therapeutics. Cancers *12*. 10.3390/cancers12030765.

Perez-Perri, J.I., Rogell, B., Schwarzl, T., Stein, F., Zhou, Y., Rettel, M., Brosig, A., and Hentze, M.W. (2018). Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. Nat Commun *9*, 4408. 10.1038/s41467-018-06557-8.

Peterson, G.L. (1979). Review of the folin phenol protein quantitation method of Lowry, Rosebrough, Farr and Randall. Analytical Biochemistry *100*, 201-220. 10.1016/0003-2697(79)90222-7.

Platten, M., Bunse, L., Wick, A., Bunse, T., Le Cornet, L., Harting, I., Sahm, F., Sanghvi, K., Tan, C.L., Poschke, I., et al. (2021). A vaccine targeting mutant IDH1 in newly diagnosed glioma. Nature. 10.1038/s41586-021-03363-z.

Poh, C.M., Zheng, J., Channappanavar, R., Chang, Z.W., Nguyen, T.H.O., Rénia, L., Kedzierska, K., Perlman, S., and Poon, L.L.M. (2020). Multiplex Screening Assay for Identifying Cytotoxic CD8(+) T Cell Epitopes. Front Immunol *11*, 400. 10.3389/fimmu.2020.00400.

Popp, M.W., and Maquat, L.E. (2018). Nonsense-mediated mRNA Decay and Cancer. Curr Opin Genet Dev *48*, 44-50. 10.1016/j.gde.2017.10.007.

Pulak, R., and Anderson, P. (1993). mRNA surveillance by the Caenorhabditis elegans smg genes. Genes Dev *7*, 1885-1897. 10.1101/gad.7.10.1885.

Purcell, A.W., Ramarathinam, S.H., and Ternette, N. (2019). Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. Nature Protocols. 10.1038/s41596-019-0133-y.

Racle, J., Michaux, J., Rockinger, G.A., Arnaud, M., Bobisse, S., Chong, C., Guillaume, P., Coukos, G., Harari, A., Jandus, C., et al. (2019). Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. Nature biotechnology. 10.1038/s41587-019-0289-6.

Ramarathinam, S.H., Croft, N.P., Illing, P.T., Faridi, P., and Purcell, A.W. (2018). Employing proteomics in the study of antigen presentation: an update. Expert Rev Proteomics *15*, 637-645. 10.1080/14789450.2018.1509000.

Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., and Stevanovic, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics *50*, 213-219.

Randle, S.J., and Laman, H. (2016). F-box protein interactions with the hallmark pathways in cancer. Seminars in Cancer Biology *36*, 3-17. 10.1016/j.semcancer.2015.09.013.

Rausch, T., Fritz, M.H., Untergasser, A., and Benes, V. (2020). Tracy: basecalling, alignment, assembly and deconvolution of sanger chromatogram trace files. BMC Genomics *21*, 230. 10.1186/s12864-020-6635-8.

Reyes, G.X., Schmidt, T.T., Kolodner, R.D., and Hombauer, H. (2015). New insights into the mechanism of DNA mismatch repair. Chromosoma *124*, 443-462. 10.1007/s00412-015-0514-0.

Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. 10.1093/nar/gkaa379.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47. 10.1093/nar/gkv007.

Ritz, D., Kinzi, J., Neri, D., and Fugmann, T. (2017). Data-Independent Acquisition of HLA Class I Peptidomes on the Q Exactive Mass Spectrometer Platform. Proteomics *17*. 10.1002/pmic.201700177.

Robbins, P.F., Lu, Y.C., El-Gamil, M., Li, Y.F., Gross, C., Gartner, J., Lin, J.C., Teer, J.K., Cliften, P., Tycksen, E., et al. (2013). Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. Nat Med *19*, 747-752. 10.1038/nm.3161.

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., et al. (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 : official journal of the American Society of Clinical Oncology *29*, 917-924. 10.1200/jco.2010.32.2537.

Robinson, J., Barker, D.J., Georgiou, X., Cooper, M.A., Flicek, P., and Marsh, S.G.E. (2020). IPD-IMGT/HLA Database. Nucleic Acids Res *48*, D948-D955. 10.1093/nar/gkz950.

Roche, P.A., and Furuta, K. (2015). The ins and outs of MHC class II-mediated antigen processing and presentation. Nat Rev Immunol *15*, 203-216. 10.1038/nri3818.

Rock, K.L., Reits, E., and Neefjes, J. (2016). Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol *37*, 724-737. 10.1016/j.it.2016.08.010.

Roepstorff, P., and Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomedical mass spectrometry *11*, 601. 10.1002/bms.1200111109.

Rolfs, Z., Müller, M., Shortreed, M.R., Smith, L.M., and Bassani-Sternberg, M. (2019). Comment on “A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands”. Science Immunol *4*, eaaw1622. 10.1126/sciimmunol.aaw1622.

Rosenberg, S.A., Packard, B.S., Aebersold, P.M., Solomon, D., Topalian, S.L., Toy, S.T., Simon, P., Lotze, M.T., Yang, J.C., Seipp, C.A., and et al. (1988). Use of tumor-infiltrating lymphocytes and interleukin-2 in the immunotherapy of patients with metastatic melanoma. A preliminary report. N Engl J Med *319*, 1676-1680. 10.1056/nejm198812223192527.

Rosenberg, S.A., and Restifo, N.P. (2015). Adoptive cell transfer as personalized immunotherapy for human cancer. Science (New York, N.Y.) *348*, 62-68. 10.1126/science.aaa4967.

Rosenberg, S.A., Yang, J.C., and Restifo, N.P. (2004). Cancer immunotherapy: moving beyond current vaccines. Nat Med *10*, 909-915. 10.1038/nm1100.

Rosenthal, R., Cadieux, E.L., Salgado, R., Bakir, M.A., Moore, D.A., Hiley, C.T., Lund, T., Tanic, M., Reading, J.L., Joshi, K., et al. (2019). Neoantigen-directed immune escape in lung cancer evolution. Nature *567*, 479-485. 10.1038/s41586-019-1032-7.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., et al. (2004). Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. Mol Cell Proteomics *3*, 1154-1169. 10.1074/mcp.M400129-MCP200.

Roth-Walter, F., Riemer, A.B., Jensen-Jarolim, E., and Stockinger, H. (2014). Common Concepts of Immune Defense. In Comparative Medicine, pp. 219-266. 10.1007/978-3-7091-1559-6_13.

Rötzschke, O., Falk, K., Deres, K., Schild, H., Norda, M., Metzger, J., Jung, G., and Rammensee, H.-G. (1990). Isolation and analysis of naturally processed viral peptides as recognized by cytotoxic T cells. Nature *348*, 252-254. 10.1038/348252a0.

Roy, B., Friesen, W.J., Tomizawa, Y., Leszyk, J.D., Zhuo, J., Johnson, B., Dakka, J., Trotta, C.R., Xue, X., Mutyam, V., et al. (2016). Ataluren stimulates ribosomal selection of near-cognate tRNAs to promote nonsense suppression. Proc Natl Acad Sci U S A *113*, 12508-12513. 10.1073/pnas.1605336113.

Rozenbrand, J., and van Bennekom, W.P. (2011). Silica-based and organic monolithic capillary columns for LC: recent trends in proteomics. J Sep Sci *34*, 1934-1944. 10.1002/jssc.201100294.

Ruiz-Banobre, J., and Goel, A. (2019). DNA Mismatch Repair Deficiency and Immune Checkpoint Inhibitors in Gastrointestinal Cancers. Gastroenterology *156*, 890-903. 10.1053/j.gastro.2018.11.071.

Sæterdal, I., Gjertsen, M., Straten, P., Eriksen, J., and Gaudernack, G. (2001). A TGFβRII frameshift-mutation-derived CTL epitope recognised by HLA-A2-restricted CD8 + T cells. Cancer Immunology, Immunotherapy *50*, 469-476. 10.1007/s002620100222.

Saha, I., Mazzocco, G., and Plewczynski, D. (2013). Consensus classification of human leukocyte antigen class II proteins. Immunogenetics *65*, 97-105. 10.1007/s00251-012-0665-6.

Sahin, U., Derhovanessian, E., Miller, M., Kloke, B.P., Simon, P., Lower, M., Bukur, V., Tadmor, A.D., Luxemburger, U., Schrors, B., et al. (2017). Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature *547*, 222-226. 10.1038/nature23003.

Sahin, U., Oehm, P., Derhovanessian, E., Jabulowsky, R.A., Vormehr, M., Gold, M., Maurus, D., Schwarck-Kokarakis, D., Kuhn, A.N., Omokoko, T., et al. (2020). An RNA vaccine drives immunity in checkpoint-inhibitor-treated melanoma. Nature. 10.1038/s41586-020-2537-9.

Sandra, K., Moshir, M., D'Hondt, F., Verleysen, K., Kas, K., and Sandra, P. (2008). Highly efficient peptide separations in proteomics Part 1. Unidimensional high performance liquid chromatography. J Chromatogr B Analyt Technol Biomed Life Sci *866*, 48-63. 10.1016/j.jchromb.2007.10.034.

Schaefer, L. (2014). Complexity of danger: the diverse nature of damage-associated molecular patterns. J Biol Chem *289*, 35237-35245. 10.1074/jbc.R114.619304.

Schaefer, M., Hagemann, S., Hanna, K., and Lyko, F. (2009). Azacytidine inhibits RNA methylation at DNMT2 target sites in human cancer cell lines. Cancer Research *69*, 8127-8132. 10.1158/0008-5472.Can-09-0458.

Schmidt, J., Guillaume, P., Dojcinovic, D., Karbach, J., Coukos, G., and Luescher, I. (2017). In silico and cell-based analyses reveal strong divergence between prediction and observation of T-cell-recognized tumor antigen T-cell epitopes. J Biol Chem *292*, 11840-11849. 10.1074/jbc.M117.789511.

Schoenberger, S.P. (2018). Is It Possible to Develop Cancer Vaccines to Neoantigens, What Are the Major Challenges, and How Can These Be Overcome? Targeting the Right Antigens in the Right Patients. Cold Spring Harb Perspect Biol *10*. 10.1101/cshperspect.a028837.

Scholtalbers, J., Boegel, S., Bukur, T., Byl, M., Goerges, S., Sorn, P., Loewer, M., Sahin, U., and Castle, J.C. (2015). TCLP: an online cancer cell line catalogue integrating HLA type, predicted neo-epitopes, virus and gene expression. Genome Med *7*, 118. 10.1186/s13073-015-0240-5.

Schumacher, T., Bunse, L., Pusch, S., Sahm, F., Wiestler, B., Quandt, J., Menn, O., Osswald, M., Oezen, I., Ott, M., et al. (2014). A vaccine targeting mutant IDH1 induces antitumour immunity. Nature *512*, 324-327. 10.1038/nature13387.

Schumacher, T.N., and Schreiber, R.D. (2015). Neoantigens in cancer immunotherapy. Science (New York, N.Y.) *348*, 69-74. 10.1126/science.aaa4971.

Schuster, H., Peper, J.K., Bösmüller, H.-C., Röhle, K., Backert, L., Bilich, T., Ney, B., Löffler, M.W., Kowalewski, D.J., Trautwein, N., et al. (2017). The immunopeptidomic landscape of ovarian carcinomas. Proc Natl Acad Sci U S A *114*, E9942-E9951. 10.1073/pnas.1707658114.

Schwitalle, Y., Kloor, M., Eiermann, S., Linnebacher, M., Kienle, P., Knaebel, H.P., Tariverdian, M., Benner, A., and von Knebel Doeberitz, M. (2008). Immune response against frameshift-induced neopeptides in HNPCC patients and healthy HNPCC mutation carriers. Gastroenterology *134*, 988-997. 10.1053/j.gastro.2008.01.015.

Second, T.P., Blethrow, J.D., Schwartz, J.C., Merrihew, G.E., MacCoss, M.J., Swaney, D.L., Russell, J.D., Coon, J.J., and Zabrouskov, V. (2009). Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. Anal Chem *81*, 7757-7765. 10.1021/ac901278y.

Sen, G.C., and Ransohoff, R.M. (1993). Interferon-Induced Antiviral Actions and Their Regulation. In pp. 57-102. 10.1016/s0065-3527(08)60083-4.

Sette, A., and Sidney, J. (1999). Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics *50*, 201-212. 10.1007/s002510050594.

Sharma, P., and Allison, J.P. (2015). The future of immune checkpoint therapy. Science (New York, N.Y.) *348*, 56-61. 10.1126/science.aaa8172.

Shum, E.Y., Jones, S.H., Shao, A., Dumdie, J., Krause, M.D., Chan, W.K., Lou, C.H., Espinoza, J.L., Song, H.W., Phan, M.H., et al. (2016). The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay. Cell *165*, 382-395. 10.1016/j.cell.2016.02.046.

Sidney, J., Peters, B., Frahm, N., Brander, C., and Sette, A. (2008). HLA class I supertypes: a revised and updated classification. BMC Immunol *9*, 1. 10.1186/1471-2172-9-1.

Sieber, J., Hauer, C., Bhuvanagiri, M., Leicht, S., Krijgsveld, J., Neu-Yilik, G., Hentze, M.W., and Kulozik, A.E. (2016). Proteomic Analysis Reveals Branch-specific Regulation of the Unfolded Protein Response by Nonsense-mediated mRNA Decay. Mol Cell Proteomics *15*, 1584-1597. 10.1074/mcp.M115.054056.

Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. PLoS Biol *6*, e111. 10.1371/journal.pbio.0060111.

Sinitcyn, P., Rudolph, J.D., and Cox, J. (2018). Computational Methods for Understanding Mass Spectrometry–Based Shotgun Proteomics Data. Annual Review of Biomedical Data Science *1*, 207-234. 10.1146/annurev-biodatasci-080917-013516.

Spaanderman, I.T., Peters, F.S., Jongejan, A., Redeker, E.J., Punt, C.J.A., and Bins, A.D. (2020). Framing the potential of public frameshift peptides as immunotherapy targets in colon cancer. bioRxiv. Published online March 30, 2020. doi.org/10.1101/2020.03.26.010231.

Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G., Korf, I., Lapp, H., et al. (2002). The Bioperl toolkit: Perl modules for the life sciences. Genome research *12*, 1611-1618. 10.1101/gr.361602.

Stevanović, S., Draper, L.M., Langhan, M.M., Campbell, T.E., Kwong, M.L., Wunderlich, J.R., Dudley, M.E., Yang, J.C., Sherry, R.M., Kammula, U.S., et al. (2015). Complete regression of metastatic cervical cancer after treatment with human papillomavirus-targeted tumor-infiltrating T cells. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *33*, 1543-1550. 10.1200/jco.2014.58.9093.

Stopfer, L.E., Mesfin, J.M., Joughin, B.A., Lauffenburger, D.A., and White, F.M. (2020). Multiplexed relative and absolute quantitative immunopeptidomics reveals MHC I repertoire alterations induced by CDK4/6 inhibition. bioRxiv. 10.1101/2020.03.03.968750.

Stresemann, C., and Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. Int J Cancer *123*, 8-13. 10.1002/ijc.23607.

Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J., and Hunt, D.F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A *101*, 9528-9533. 10.1073/pnas.0402700101.

Tan, C.T., Croft, N.P., Dudek, N.L., Williamson, N.A., and Purcell, A.W. (2011). Direct quantitation of MHC-bound peptide epitopes by selected reaction monitoring. Proteomics *11*, 2336-2340. 10.1002/pmic.201000531.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., and Matsuo, T. (1988). Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry *2*, 151-153. 10.1002/rcm.1290020802.

Tange, O. (2018). GNU Parallel 2018. 10.5281/zenodo.1146014.

Tani, H., Imamachi, N., Salam, K.A., Mizutani, R., Ijiri, K., Irie, T., Yada, T., Suzuki, Y., and Akimitsu, N. (2012). Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. RNA Biol *9*, 1370-1379. 10.4161/rna.22360.

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2018). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res *47*, D941-D947. 10.1093/nar/gky1015.

Taylor, J.A., and Johnson, R.S. (1997). Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom *11*, 1067-1075. 10.1002/(sici)1097-0231(19970615)11:9<1067::Aid-rcm953>3.0.Co;2-l.

Teixeira, F.R., Manfiolli, A.O., Soares, C.S., Baqui, M.M., Koide, T., and Gomes, M.D. (2013). The F-box protein FBXO25 promotes the proteasome-dependent degradation of ELK-1 protein. J Biol Chem *288*, 28152-28162. 10.1074/jbc.M113.504308.

Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeier, C., Hentze, M.W., and Kulozik, A.E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. The EMBO Journal *17*, 3484-3494. 10.1093/emboj/17.12.3484.

Thomas, L. (1959). Discussion. In Cellular and humoral aspects of the hypersensitive states, H.S. Lawrence, ed. (Hoeber-Harper), pp. 529-532.

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem *75*, 1895-1904. 10.1021/ac0262560.

Thomsen, M.C., and Nielsen, M. (2012). Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. Nucleic Acids Res *40*, W281-287. 10.1093/nar/gks469.

Tougeron, D., Fauquembergue, E., Rouquette, A., Le Pessot, F., Sesboue, R., Laurent, M., Berthet, P., Mauillon, J., Di Fiore, F., Sabourin, J.C., et al. (2009). Tumor-infiltrating lymphocytes in colorectal cancers with microsatellite instability are correlated with the number and spectrum of frameshift mutations. Mod Pathol *22*, 1186-1195. 10.1038/modpathol.2009.80.

Tran, E., Robbins, P.F., and Rosenberg, S.A. (2017). 'Final common pathway' of human cancer immunotherapy: targeting random somatic mutations. Nature immunology *18*, 255-262. 10.1038/ni.3682.

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., et al. (2014). Cancer Immunotherapy Based on Mutation-Specific CD4+ T Cells in a Patient with Epithelial Cancer. Science (New York, N.Y.) *344*, 641-645. 10.1126/science.1251102.

Tran, N.H., Qiao, R., Xin, L., Chen, X., Liu, C., Zhang, X., Shan, B., Ghodsi, A., and Li, M. (2019). Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. Nat Methods *16*, 63-66. 10.1038/s41592-018-0260-3.

Treangen, T.J., and Salzberg, S.L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet *13*, 36-46. 10.1038/nrg3117.

Trzaska, C., Amand, S., Bailly, C., Leroy, C., Marchand, V., Duvernois-Berthet, E., Saliou, J.-M., Benhabiles, H., Werkmeister, E., Chassat, T., et al. (2020). 2,6-Diaminopurine as a highly potent corrector of UGA nonsense mutations. Nature Communications *11*, 1509. 10.1038/s41467-020-15140-z.

Tsou, C.-C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A.-C., and Nesvizhskii, A.I. (2015). DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat Methods *12*, 258-264. 10.1038/nmeth.3255.

Tsuchihara, K., Lapin, V., Bakal, C., Okada, H., Brown, L., Hirota-Tsuchihara, M., Zaugg, K., Ho, A., Itie-YouTen, A., Harris-Brandts, M., et al. (2005). Ckap2 Regulates Aneuploidy, Cell Cycling, and Cell Death in a p53-Dependent Manner. Cancer Research *65*, 6685-6691. 10.1158/0008-5472.Can-04-4223.

Tu, C., Li, J., Shen, S., Sheng, Q., Shyr, Y., and Qu, J. (2016). Performance Investigation of Proteomic Identification by HCD/CID Fragmentations in Combination with High/Low-Resolution Detectors on a Tribrid, High-Field Orbitrap Instrument. PLoS One *11*, e0160160. 10.1371/journal.pone.0160160.

Turajlic, S., Litchfield, K., Xu, H., Rosenthal, R., McGranahan, N., Reading, J.L., Wong, Y.N.S., Rowan, A., Kanu, N., Al Bakir, M., et al. (2017). Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. The Lancet Oncology *18*, 1009-1021. 10.1016/s1470-2045(17)30516-8.

Tynan, F.E., Borg, N.A., Miles, J.J., Beddoe, T., El-Hassen, D., Silins, S.L., van Zuylen, W.J., Purcell, A.W., Kjer-Nielsen, L., McCluskey, J., et al. (2005). High resolution structures of highly bulged viral epitopes bound to major histocompatibility complex class I. Implications for T-cell receptor engagement and T-cell immunodominance. J Biol Chem *280*, 23900-23909. 10.1074/jbc.M503060200.

Umar, A., Boland, C.R., Terdiman, J.P., Syngal, S., de la Chapelle, A., Rüschoff, J., Fishel, R., Lindor, N.M., Burgart, L.J., Hamelin, R., et al. (2004). Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst *96*, 261-268. 10.1093/jnci/djh034.

Usuki, F., Yamashita, A., Higuchi, I., Ohnishi, T., Shiraishi, T., Osame, M., and Ohno, S. (2004). Inhibition of nonsense-mediated mRNA decay rescues the phenotype in Ullrich's disease. Annals of Neurology *55*, 740-744. 10.1002/ana.20107.

Vaddepally, R.K., Kharel, P., Pandey, R., Garje, R., and Chandra, A.B. (2020). Review of Indications of FDA-Approved Immune Checkpoint Inhibitors per NCCN Guidelines with the Level of Evidence. Cancers *12*. 10.3390/cancers12030738.

van der Lee, D.I., Reijmers, R.M., Honders, M.W., Hagedoorn, R.S., de Jong, R.C., Kester, M.G., van der Steen, D.M., de Ru, A.H., Kweekel, C., Bijen, H.M., et al. (2019). Mutated nucleophosmin 1 as immunotherapy target in acute myeloid leukemia. J Clin Invest *129*, 774-785. 10.1172/JCI97482.

van Duikeren, S., Fransen, M.F., Redeker, A., Wieles, B., Platenburg, G., Krebber, W.J., Ossendorp, F., Melief, C.J., and Arens, R. (2012). Vaccine-induced effector-memory CD8+ T cell responses predict therapeutic efficacy against tumors. J Immunol *189*, 3397-3403. 10.4049/jimmunol.1201540.

van Rooij, N., van Buuren, M.M., Philips, D., Velds, A., Toebes, M., Heemskerk, B., van Dijk, L.J.A., Behjati, S., Hilkmann, H., El Atmioui, D., et al. (2013). Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *31*, e439-e442. 10.1200/JCO.2012.47.7521.

Vanderwalde, A., Spetzler, D., Xiao, N., Gatalica, Z., and Marshall, J. (2018). Microsatellite instability status determined by next-generation sequencing and compared with PD-L1 and tumor mutational burden in 11,348 patients. Cancer Med *7*, 746-756. 10.1002/cam4.1372.

Vesely, M.D., Kershaw, M.H., Schreiber, R.D., and Smyth, M.J. (2011). Natural innate and adaptive immunity to cancer. Annu Rev Immunol *29*, 235-271. 10.1146/annurev-immunol-031210-101324.

Vichalkovski, A., Gresko, E., Cornils, H., Hergovich, A., Schmitz, D., and Hemmings, B.A. (2008). NDR kinase is activated by RASSF1A/MST1 in response to Fas receptor stimulation and promotes apoptosis. Curr Biol *18*, 1889-1895. 10.1016/j.cub.2008.10.060.

Viegas, M.H., Gehring, N.H., Breit, S., Hentze, M.W., and Kulozik, A.E. (2007). The abundance of RNPS1, a protein component of the exon junction complex, can determine the variability in efficiency of the Nonsense Mediated Decay pathway. Nucleic Acids Res *35*, 4542-4551. 10.1093/nar/gkm461.

Vigneron, N., Stroobant, V., Chapiro, J., Ooms, A., Degiovanni, G., Morel, S., van der Bruggen, P., Boon, T., and Van den Eynde, B.J. (2004). An Antigenic Peptide Produced by Peptide Splicing in the Proteasome. Science (New York, N.Y.) *304*, 587-590. 10.1126/science.1095522.

Vyas, J.M., Van der Veen, A.G., and Ploegh, H.L. (2008). The known unknowns of antigen processing and presentation. Nat Rev Immunol *8*, 607-618. 10.1038/nri2368.

Walz, S., Stickel, J.S., Kowalewski, D.J., Schuster, H., Weisel, K., Backert, L., Kahn, S., Nelde, A., Stroh, T., Handel, M., et al. (2015). The antigenic landscape of multiple myeloma: mass spectrometry (re)defines targets for T-cell-based immunotherapy. Blood *126*, 1203-1213. 10.1182/blood-2015-04-640532.

Wang, Q., Zhou, Y., Rychahou, P., Liu, C., Weiss, H.L., and Evers, B.M. (2013). NFAT5 represses canonical Wnt signaling via inhibition of beta-catenin acetylation and participates in regulating intestinal cell differentiation. Cell Death Dis *4*, e671. 10.1038/cddis.2013.202.

Weeratna, R.D., Makinen, S.R., McCluskie, M.J., and Davis, H.L. (2005). TLR agonists as vaccine adjuvants: comparison of CpG ODN and Resiquimod (R-848). Vaccine *23*, 5263-5270. 10.1016/j.vaccine.2005.06.024.

Welch, E.M., Barton, E.R., Zhuo, J., Tomizawa, Y., Friesen, W.J., Trifillis, P., Paushkin, S., Patel, M., Trotta, C.R., Hwang, S., et al. (2007). PTC124 targets genetic disorders caused by nonsense mutations. Nature *447*, 87-91. 10.1038/nature05756.

Wells, D.K., van Buuren, M.M., Dang, K.K., Hubbard-Lucey, V.M., Sheehan, K.C.F., Campbell, K.M., Lamb, A., Ward, J.P., Sidney, J., Blazquez, A.B., et al. (2020). Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. Cell. 10.1016/j.cell.2020.09.015.

Wieczorek, M., Abualrous, E.T., Sticht, J., Alvaro-Benito, M., Stolzenberg, S., Noe, F., and Freund, C. (2017). Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. Front Immunol *8*, 292. 10.3389/fimmu.2017.00292.

Wiese, S., Reidegeld, K.A., Meyer, H.E., and Warscheid, B. (2007). Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. Proteomics *7*, 340-350. 10.1002/pmic.200600422.

Wilm, M. (2011). Principles of electrospray ionization. Mol Cell Proteomics *10*, M111 009407. 10.1074/mcp.M111.009407.

Woerner, S.M., Gebert, J., Yuan, Y.P., Sutter, C., Ridder, R., Bork, P., and von Knebel Doeberitz, M. (2001). Systematic identification of genes with coding microsatellites mutated in DNA mismatch repair-deficient cancer cells. Int J Cancer *93*, 12-19.

Woerner, S.M., Kloor, M., Schwitalle, Y., Youmans, H., Doeberitz, M., Gebert, J., and Dihlmann, S. (2007). The putative tumor suppressor AIM2 is frequently affected by different genetic alterations in microsatellite unstable colon cancers. Genes Chromosomes Cancer *46*, 1080-1089. 10.1002/gcc.20493.

Woerner, S.M., Kloor, M., von Knebel Doeberitz, M., and Gebert, J.F. (2006). Microsatellite instability in the development of DNA mismatch repair deficient tumors. Cancer Biomarkers *2*, 69-86. 10.3233/CBM-2006-21-208.

Woerner, S.M., Tosti, E., Yuan, Y.P., Kloor, M., Bork, P., Edelmann, W., and Gebert, J. (2015). Detection of coding microsatellite frameshift mutations in DNA mismatch repair-deficient mouse intestinal tumors. Mol Carcinog *54*, 1376-1386. 10.1002/mc.22213.

Woerner, S.M., Yuan, Y.P., Benner, A., Korff, S., von Knebel Doeberitz, M., and Bork, P. (2010). SelTarbase, a database of human mononucleotide-microsatellite mutations and their potential impact to tumorigenesis and immunology. Nucleic Acids Res *38*, D682-689. 10.1093/nar/gkp839.

Yamashita, A., Chang, T.-C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C.-Y.A., and Shyu, A.-B. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat Struct Mol Biol *12*, 1054-1063. 10.1038/nsmb1016.

Yamashita, A., Izumi, N., Kashima, I., Ohnishi, T., Saari, B., Katsuhata, Y., Muramatsu, R., Morita, T., Iwamatsu, A., Hachiya, T., et al. (2009). SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. Genes & Development *23*, 1091-1105. 10.1101/gad.1767209.

Yarchoan, M., Hopkins, A., and Jaffee, E.M. (2017a). Tumor Mutational Burden and Response Rate to PD-1 Inhibition. N Engl J Med *377*, 2500-2501. 10.1056/NEJMc1713444.

Yarchoan, M., Johnson, B.A., 3rd, Lutz, E.R., Laheru, D.A., and Jaffee, E.M. (2017b). Targeting neoantigens to augment antitumour immunity. Nat Rev Cancer *17*, 209-222. 10.1038/nrc.2016.154.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134. 10.1186/1471-2105-13-134.

Yedidi, R.S., Wendler, P., and Enenkel, C. (2017). AAA-ATPases in Protein Degradation. Front Mol Biosci *4*, 42. 10.3389/fmolb.2017.00042.

Yewdell, J.W. (2011). DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. Trends Immunol *32*, 548-558. 10.1016/j.it.2011.08.001.

You, K.T., Li, L.S., Kim, N.G., Kang, H.J., Koh, K.H., Chwae, Y.J., Kim, K.M., Kim, Y.K., Park, S.M., Jang, S.K., and Kim, H. (2007). Selective translational repression of truncated proteins from frameshift mutation-derived mRNAs in tumors. PLoS Biol *5*, e109. 10.1371/journal.pbio.0050109.

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A., and Ma, B. (2012). PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics *11*, M111 010587. 10.1074/mcp.M111.010587.

Zhang, Y., and Zhang, Z. (2020). The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. Cell Mol Immunol *17*, 807-821. 10.1038/s41423-020-0488-6.

Zhou, Y., Wang, Q., Weiss, H.L., and Evers, B.M. (2014). Nuclear factor of activated T-cells 5 increases intestinal goblet cell differentiation through an mTOR/Notch signaling pathway. Mol Biol Cell *25*, 2882-2890. 10.1091/mbc.E14-05-0998.

Zitvogel, L., Tesniere, A., and Kroemer, G. (2006). Cancer despite immunosurveillance: immunoselection and immunosubversion. Nat Rev Immunol *6*, 715-727. 10.1038/nri1936.

Zolg, D.P., Wilhelm, M., Schnatbaum, K., Zerweck, J., Knaute, T., Delanghe, B., Bailey, D.J., Gessulat, S., Ehrlich, H.-C., Weininger, M., et al. (2017). Building ProteomeTools based on a complete synthetic human proteome. Nat Methods *14*, 259. 10.1038/nmeth.4153.

Zou, Z., Tao, T., Li, H., and Zhu, X. (2020). mTOR signaling pathway and mTOR inhibitors in cancer: progress and challenges. Cell Biosci *10*, 31. 10.1186/s13578-020-00396-1.