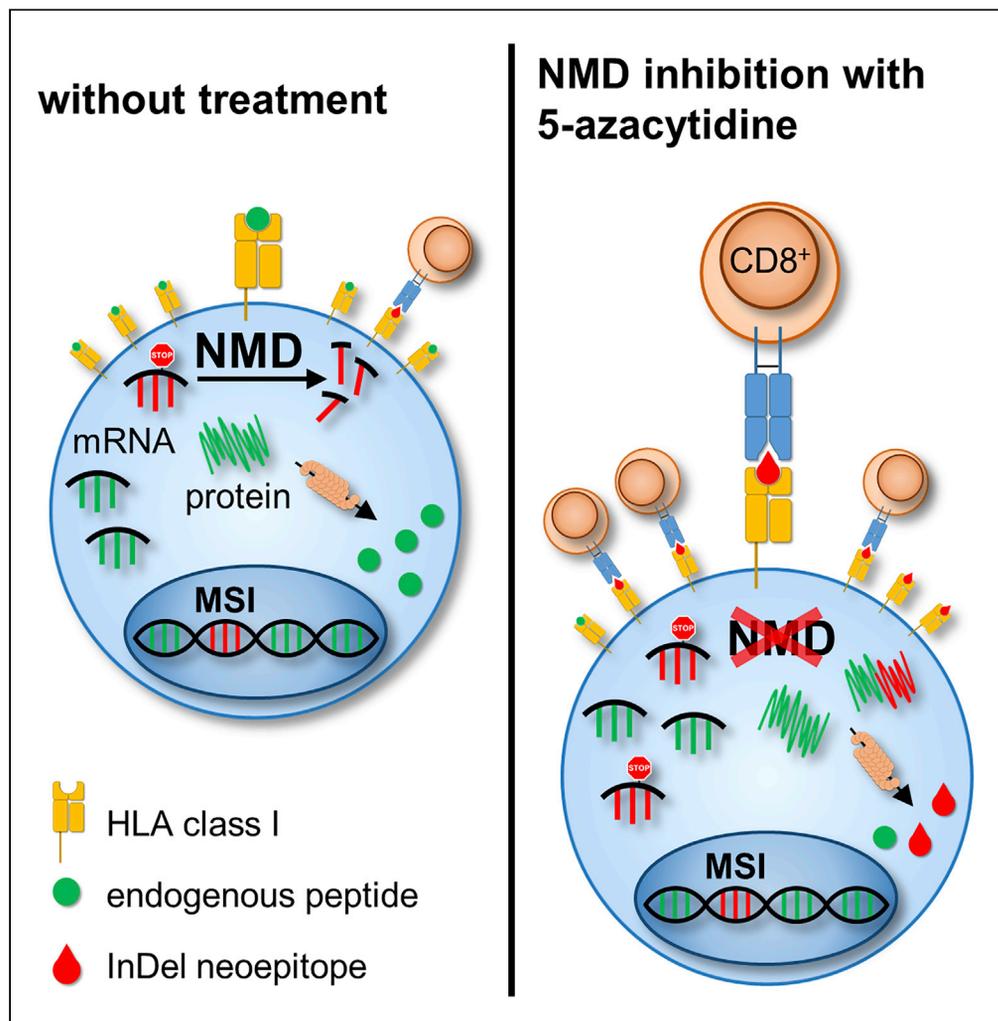# iScience

Article

NMD inhibition by 5-azacytidine augments presentation of immunogenic frameshift-derived neoepitopes

Jonas P. Becker, Dominic Helm, Mandy Rettel, ..., Magnus von Knebel Doeberitz, Matthias W. Hentze, Andreas E. Kulozik

andreas.kulozik@med. uni-heidelberg.de (A.E.K.) hentze@embl.org (M.W.H.)

Highlights

Immunopeptidomic neoepitope discovery workflow employing public DNA sequencing data

Identification of immunogenic frameshift-derived InDel neoepitopes in an MSI model

NMD inhibition increases HLA class I-mediated presentation of InDel neoepitopes

CKAP2 frameshift mutation is highly recurrent in different types of MSI cancer

iScience

Article

NMD inhibition by 5-azacytidine augments presentation of immunogenic frameshift-derived neoepitopes

Jonas P. Becker,[1,2,3,4] Dominic Helm,[5] Mandy Rettel,[6] Frank Stein,[6] Alejandro Hernandez-Sanchez,[1,2,7,8]
Katharina Urban,[1,2,7,8] Johannes Gebert,[1,2,7,8] Matthias Kloor,[1,2,7,8] Gabriele Neu-Yilik,[1,2,3,4]
Magnus von Knebel Doeberitz,[1,2,7,8] Matthias W. Hentze,[1,2,*] and Andreas E. Kulozik[1,2,3,4,9,*]

SUMMARY

Frameshifted protein sequences elicit tumor-specific T cell-mediated immune responses in microsatellite-unstable (MSI) cancers if presented by HLA class I molecules. However, their expression and presentation are limited by nonsense-mediated RNA decay (NMD). We employed an unbiased immunopeptidomics workflow to analyze MSI HCT-116 cells and identified >10,000 HLA class I-presented peptides including five frameshift-derived InDel neoepitopes. Notably, pharmacological NMD inhibition with 5-azacytidine stabilizes frameshift-bearing transcripts and increases the HLA class I-mediated presentation of InDel neoepitopes. The frameshift mutation underlying one of the identified InDel neoepitopes is highly recurrent in MSI colorectal cancer cell lines and primary patient samples, and immunization with the corresponding neoepitope induces strong CD8[+] T cell responses in an HLA-A*02:01 transgenic mouse model. Our data show directly that pharmacological NMD inhibition augments HLA class I-mediated presentation of immunogenic frameshift-derived InDel neoepitopes thus highlighting the clinical potential of NMD inhibition in anti-cancer immunotherapy strategies.

INTRODUCTION

Microsatellite-unstable colorectal cancers (MSI CRCs) account for approximately 15% of all CRCs, and hence about 275,000 cases per year (Bray et al., 2018). These cancers are characterized by the accumulation of somatic mutations, mainly small insertion/deletion (InDel) mutations in repetitive DNA stretches termed microsatellites (Cancer Genome Atlas, 2012). Importantly, two-thirds of all InDel mutations cause a shift of the reading frame, hence encoding tumor-specific protein sequences (Kloor and von Knebel Doeberitz, 2016). Human leukocyte antigen (HLA) class I-presented peptides (HLAp) derived from such frameshifted (FS) protein sequences (termed InDel neoepitopes) allow patrolling CD8[+] T cells to identify and target tumor cells presenting such InDel neoepitopes (Schumacher and Schreiber, 2015). In contrast to neoepitopes derived from a single amino acid (aa) change (termed SNP neoepitopes), FS protein sequences can potentially harbor several InDel neoepitopes with binding capacity to different HLA allotypes. Moreover, InDel neoepitopes have been suggested to possess higher immunogenicity caused by their fundamental difference to endogenous self-antigens originating from wild-type (WT) proteins (Turajlic et al., 2017). Several studies support the presence of InDel neoepitope-specific cytotoxic T cells in both healthy individuals and patients with MSI CRC (Leoni et al., 2020; Roudko et al., 2020; Schwitalle et al., 2008).

Recently, we and others showed the induction of neoepitope-directed immune responses after vaccination with shared, *in silico*-predicted neoepitopes in mice (Leoni et al., 2020) and a clinical trial with mismatch repair-deficient patients (Kloor et al., 2020). Moreover, it was shown that T cells, re-activated by immune checkpoint inhibition, target tumor-specific neoepitopes thus further emphasizing the crucial role of HLA class I-presented neoepitopes in immunotherapeutic strategies (Gubin et al., 2014). However, these strategies rely on the expression and presentation of such neoepitopes in sufficient quantities, and the expression of most InDel neoepitopes is limited by nonsense-mediated RNA decay (NMD). NMD is a translation-dependent quality control pathway that recognizes and degrades mRNAs with premature

[1]Molecular Medicine Partnership Unit (MMPU), Heidelberg University, 69120 Heidelberg, Germany

[2]European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany

[3]Department of Pediatric Oncology, Hematology and Immunology, Heidelberg University, 69120 Heidelberg, Germany

[4]Hopp Children's Cancer Center, National Center for Tumor Diseases (NCT), 69120 Heidelberg, Germany

[5]Genomics and Proteomics Core Facility, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

[6]European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany

[7]Department of Applied Tumor Biology, Institute of Pathology, Heidelberg University, 69120 Heidelberg, Germany

[8]Collaboration Unit Applied Tumor Biology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

[9]Lead contact

*Correspondence:
andreas.kulozik@med.uni-heidelberg.de (A.E.K.),
hentze@embl.org (M.W.H.)
https://doi.org/10.1016/j.isci.2021.102389

termination codons introduced by nonsense or frameshift mutations (Lavysh and Neu-Yilik, 2020). In MSI CRCs, the central NMD factors UPF1, UPF2, SMG1, SMG6, and SMG7 are expressed substantially more strongly compared with microsatellite-stable CRCs thus restricting the production of InDel neoepitopes and consequently their immune recognition (Bokhari et al., 2018; El-Bchiri et al., 2005, 2008). Our laboratory identified the approved drug 5-azacytidine (5AZA) as a potent NMD inhibitor that limits NMD without interfering with protein synthesis at therapeutic concentrations, thus distinguishing 5AZA from other known NMD inhibitors (Bhuvanagiri et al., 2014). Therefore, we hypothesized that therapeutic NMD inhibition by 5AZA could increase the production and presentation of InDel neoepitopes and thus tumor recognition by the host's immune system.

Leveraging the full potential of InDel neoepitopes in generating effective T cell responses for novel and possibly personalized immunotherapy strategies requires the reliable identification of these peptides. Neoepitopes are commonly identified by evaluating *in silico* candidates using T cell screening technologies. However, such analyses are strongly limited by high rates of false-positive results (Andreatta and Nielsen, 2016; Bassani-Sternberg et al., 2015; Cui et al., 2020; Freudenmann et al., 2018). Recent breakthroughs in the sensitivity and reproducibility of mass spectrometry (MS) help to overcome these disadvantages and enable the unbiased exploration of the global immunopeptidome presented by the HLA system. New methodological MS approaches such as dual-fragmentation by electron-transfer/higher-energy collision dissociation (EThcD) both expand the detectable immunopeptidome and increase the confidence in peptide identifications (Mommen et al., 2014). Furthermore, *de novo* peptide sequencing allows the identification of neoepitope sequences not included in standard proteomics databases (Schumacher et al., 2017). Here, we report the first unbiased, direct identification of previously unknown immunogenic InDel neoepitopes in MSI CRC by MS, and provide experimental evidence that NMD inhibition increases the HLA class I-mediated cell surface presentation of immunogenic InDel neoepitopes. Furthermore, the high frequency of the frameshift mutation underlying one such neoepitope in patients with MSI CRC highlights the potential of these findings for developing new immunotherapeutic strategies for MSI cancers.

RESULTS

Validation of the experimental system

The MSI CRC cell line HCT-116 was chosen as the model system for this study. HCT-116 cells express six different HLA class I alleles, including the common HLA-A*02:01 allele, allowing the presentation of a broad spectrum of peptides (Scholtalbers et al., 2015). NMD competence of HCT-116 cells was determined using a transiently transfected dual-luciferase reporter system (Boelz et al., 2006). HCT-116 cells exhibit a high NMD efficiency demonstrated by the substantially and highly significantly lower *Renilla* luciferase signal in cells transfected with the NS39 reporter (0.055 $\pm$ 0.038 normalized to WT reporter signal, p $\leq$ 0.0001; Figure S1). The NMD-restricting effect of 5AZA in HCT-116 cells was tested by assessment of transcript levels of known endogenous NMD targets by quantitative real-time PCR (qPCR). Treatment with 5 μM 5AZA for 24 h induced a significant stabilization (p $\leq$ 0.0001) of ATF3 (3.6-fold), SC35C (2.9-fold), and SC35D (2.7-fold) transcripts (Figure S1).

MS-based immunopeptidomics offers the only unbiased method to directly identify (neo-)epitopes that are actually presented via HLA class I molecules on cancer cells. We extended a recently published, high-throughput workflow for the identification of HLA-presented peptides (Chong et al., 2018) to enable the identification of InDel neoepitopes and investigate the effect of NMD inhibition at the level of the immunopeptidome (Figure 1). Briefly, after immunoprecipitation (IP) with a pan-HLA antibody (Figure S2) and subsequent separation of the bound peptides from HLA class I molecules, HLAp were analyzed by liquid chromatography-tandem mss spectrometry (LC-MS/MS) applying different fragmentation methods. Higher-energy collisional dissociation (HCD) is the standard fragmentation mode for acquiring high-resolution data at a fast speed and therefore provides in-depth coverage of the immunopeptidome. EThcD, a combination of HCD and electron transfer dissociation, generates more complex fragmentation spectra leading to higher peptide sequence coverage to ensure high-confidence identification. Furthermore, we combined fragmentation by EThcD with precursor selection targeting low-abundance precursors (low-EThcD) first to compensate for the lower coverage caused by the slower acquisition frequency of EThcD fragmentation. Finally, the obtained MS raw data were subjected to a *de novo* sequencing-assisted database search, which improves both sensitivity and accuracy of peptide identifications and enables the identification of neoepitopes.
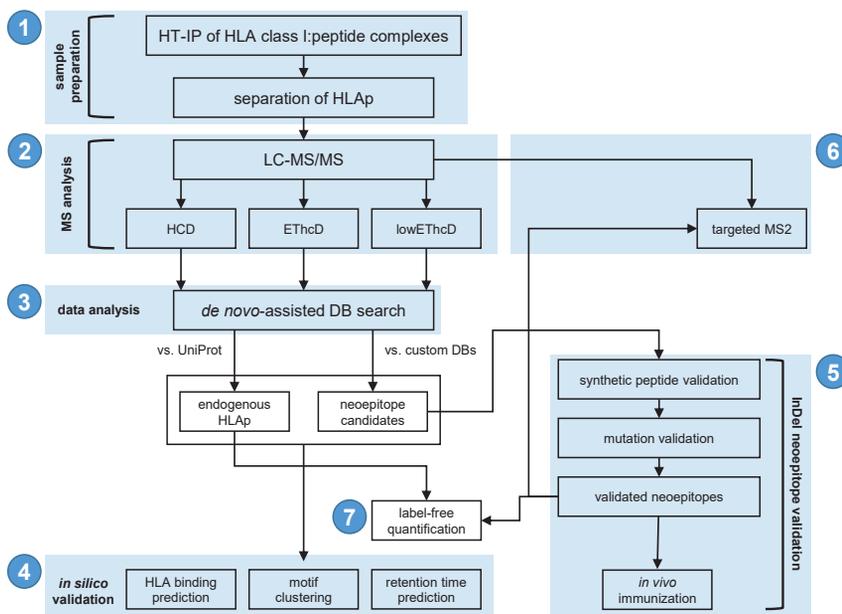
Figure 1. Immunopeptidomics workflow for the identification, validation, and quantification of InDel neoepitopes

After high-throughput IP of HLA class I:peptide complexes, HLAp are separated (1; see also Figure S2) 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; see also Figure S3). All identified peptides are validated bioinformatically (binding prediction, sequence clustering, retention time prediction; 4). Neoepitope candidates are further validated by comparison with 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.

Direct, mass spectrometry-based identification of HLA class I-presented peptides

The above-described workflow was used to analyze HLAp isolated from HCT-116 cells treated either with 5 μM 5AZA or the solvent DMSO control for 24 h. We identified a total of 10,030 unique HLAp at a stringent false discovery rate (FDR) of 1% (Figure 2A). Of these, 3,098 HLAp (31% of the total dataset) were identified both in datasets recorded using HCD and dual-fragmentation (EThcD/lowEThcD), whereas 4,907 HLAp (49%) were only identified using HCD fragmentation; 2,025 HLAp (20%) were only identified in the dual-fragmentation datasets. MS1 signals of identified and quantified peptides show a wide range of intensity spanning several orders of magnitude ($\log_2$ intensity: 12.75–32.55, mean $\log_2$ intensity: 20.19). Peptides identified by all three fragmentation methods show a slightly higher average intensity/abundance when compared with peptides identified by HCD fragmentation alone ($\log_2$ intensity: 21.58 versus 19.77). Of note, we identified a subset of 851 low-abundance peptides (mean $\log_2$ intensity: 18.52) using the low-EThcD method, which preferentially targets less abundant peptide precursors (Figure 2A). In summary, these findings illustrate the benefit of applying both different fragmentation methods and precursor selection strategies to increase the number of peptide identifications providing a more comprehensive view of the immunopeptidome. *In silico* quality controls were performed to ensure the quality of the obtained dataset. First, we calculated the sequence-specific hydrophobicity index (HI), which is an orthogonal parameter to validate correct peptide identifications and correlates with experimentally observed retention times (Krokhin, 2006). HI of identified peptides showed a tight correlation with observed retention times for all three fragmentation methods used (Pearson's correlation coefficient HCD: 0.96, EThcD: 0.96, lowEThcD: 0.95; Figure 2B). Next, we analyzed the HLA-associated properties of the identified peptides. The peptides showed an HLA class I-typical length distribution with mainly nonamers (Figure 2C). Investigation of the entire immunopeptidome using MS reduces the *a priori*-introduced bias of selectively surveying (neo-)

A

B

C

D

E
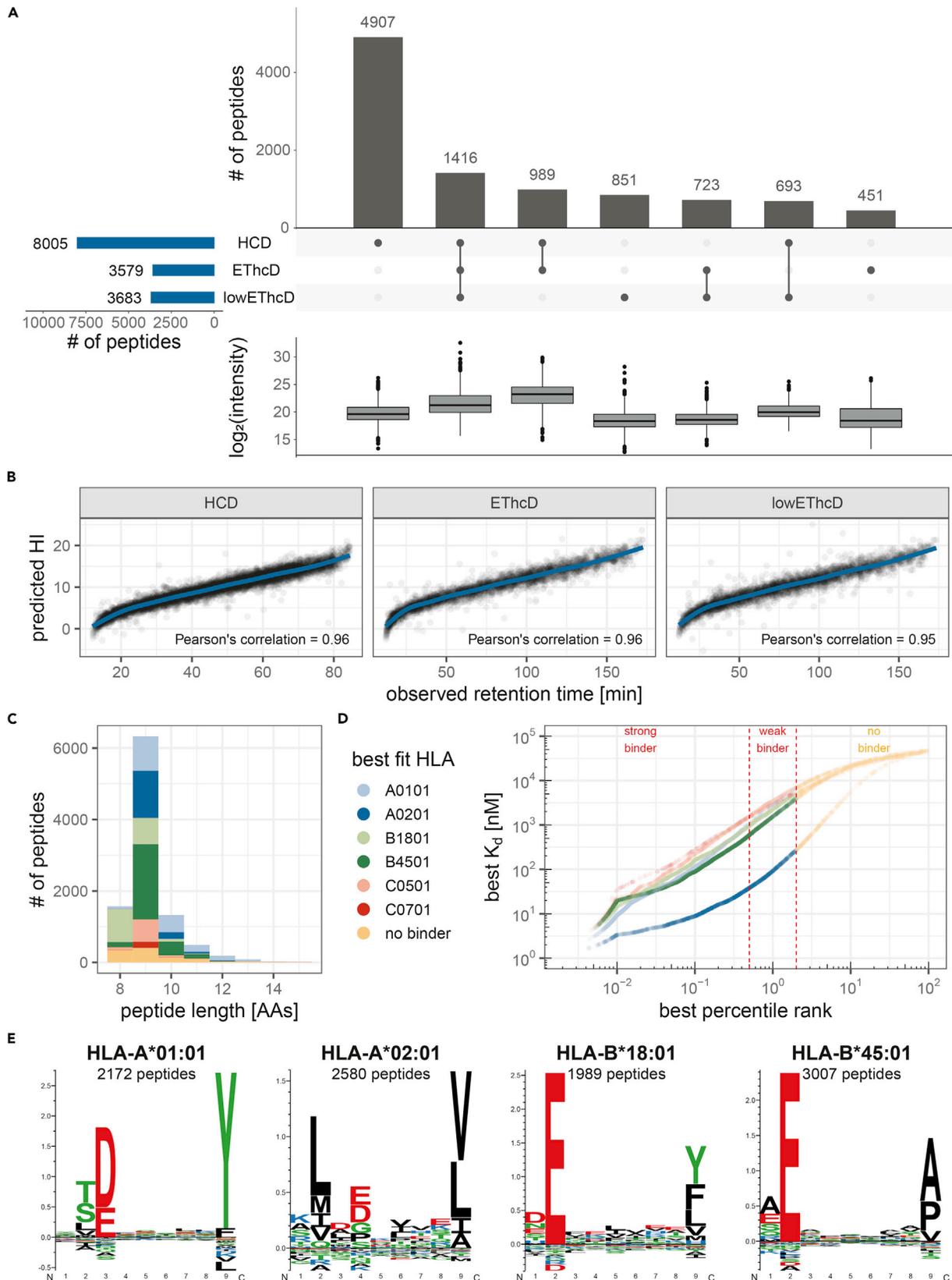
Figure 2. Quality control and characteristics of identified peptides

(A) Number of peptides identified using different MS fragmentation methods and overlapping sets. Boxplot summary representing intensity distribution for subsets of peptides. Bars of boxplot summary represent 25th to 75th percentiles, middle line represents median.

(B) Predicted hydrophobicity index (HI) against observed retention time of identified peptides for different MS fragmentation methods.

(C) 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.

(D) Binding prediction of all identified peptides. Threshold for strong binders is top 0.5% ranked, for weak binders top 2%.

(E) Sequence clustering of identified peptides to four distinct binding motifs matching HLA alleles expressed by HCT-116 cell line; 282 outliers (2.8%) were not clustered.

epitopes shortlisted by *in silico* binding predictions. However, binding prediction of identified HLAp *a posteriori* represents a suitable validation tool for immunopeptidomics datasets. Using NetMHCpan, we found that 90% of the identified peptides (8,988 peptides; 7,286 strong binder [SB], 1,702 weak binder [WB]) were predicted to bind at least one of the HLA class I molecules expressed on the HCT-116 cells (Figure 2D). Considering the similarity of HLAp, which is caused by their allele-specific binding-mediating anchor residues, we clustered peptide sequences into groups and identified four distinct motifs that correspond to the consensus binding motifs of HLA-A*01:01, HLA-A*02:01, HLA-B*18:01, and HLA-B*45:01 (Figure 2E). Although expressed on HCT-116 cells, consensus binding motifs for HLA-C*05:01 and HLA-C*07:01 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 (Bassani-Sternberg et al., 2015; Neisig et al., 1998). Finally, we analyzed the source proteins of the identified peptides. Of the 4,767 distinct source proteins, 2,524 (53%) were represented by only one, 1,054 (22%) by two, and 516 (11%) by three distinct HLAp at the cell surface. The remaining 673 source proteins (14%) were represented by four or more distinct HLAp. In general, source proteins were associated with a broad spectrum of cellular localization. In line with previous reports, the source proteins of the top 10% most abundant HLAp were significantly enriched in clusters for nuclear, cytoskeletal, and ribosomal proteins (Bassani-Sternberg et al., 2015; Walz et al., 2015). Taken together, these data validate our dataset as a representative view of the endogenous immunopeptidome of HCT-116 cells.

Identification and validation of HLA class I-presented InDel and SNP neoepitopes

After having created a high-quality, representative dataset of endogenous HLAp, we next sought to query this dataset for the existence of HLA class I-presented neoepitopes. To enable the identification of both InDel and SNP neoepitopes we constructed custom, cell line-specific databases based on publicly available sequencing data. InDel databases based on COSMIC and CCLE mutation data for the HCT-116 cell line contain 883 unique entries. The lengths of FS, tumor-specific protein sequences range from 13 to 393 aa (median: 31 aa), potentially generating 26,044 unique peptides with a length of nine aa, which is the preferred binding length of HLA class I molecules. Of these, 11% (2,743 peptides; 895 SB, 1,848 WB) were found to potentially bind at least one of the HLA alleles expressed by HCT-116 cells (Figure S3). The SNP database contains 1,260 previously reported *in silico*-predicted potential SNP neoepitopes (Scholtalbers et al., 2015).

In contrast to standard MS data analysis workflows, PEAKS reports spectra with high *de novo* sequencing scores, which were not matched to a UniProt database entry as "*de novo* only" spectra. These high-scoring "*de novo* only" spectra were subsequently searched against the custom SNP and InDel databases. After applying a stringent FDR of 1%, we identified nine InDel- and five SNP-neoepitope candidates (Figure 3A and Tables 1 and 2). Three of these five SNP neoepitopes have been reported previously (Bassani-Sternberg et al., 2015). To validate the identifications of the neoepitope candidates, we first used BLASTp with standard parameters for short input sequences to rule out that identified peptides (and leucine/isoleucine permutations of them) match known human amino acid sequences. We excluded one of the nine InDel neoepitope candidates matching the 14 aa of the WT protein part that was included during database generation (Figure 3A).

To confirm the identity/aa sequence of neoepitopes, we next compared their spectra with those obtained from synthetic peptide counterparts. The MS acquisition and data analysis workflow previously used for the identification of neoepitopes from cell line samples was applied to pools of synthetic peptides and confirmed the identities of five InDel neoepitopes and five SNP neoepitopes (Figures 3B and S4B). Intensities of matched fragment ions showed very high correlations (Pearson's correlation: 0.917–0.999) for correct identifications using both HCD and EThcD fragmentation methods, whereas this correlation was much
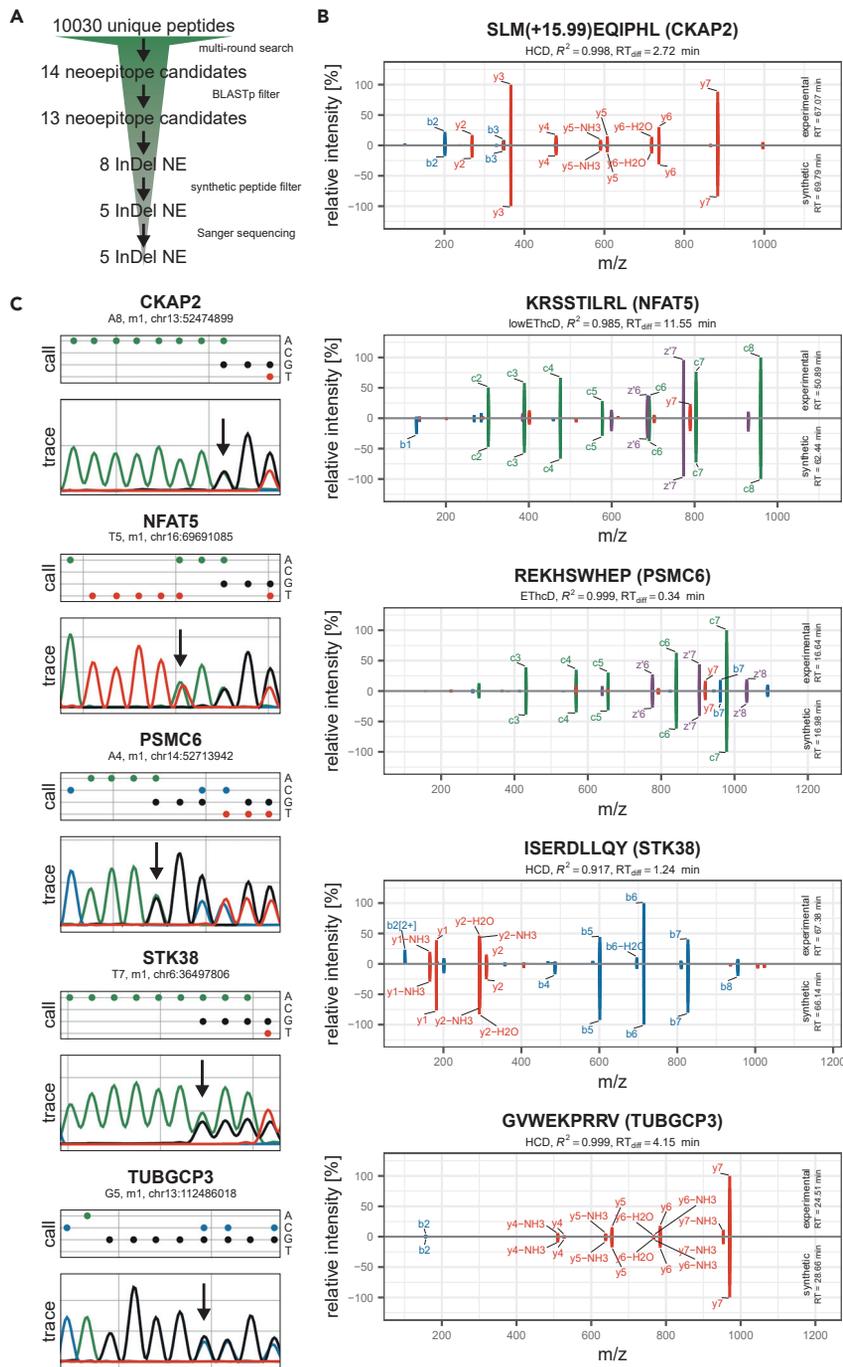
Figure 3. Validation of identified InDel neoepitopes

(A) Overview of validation procedure. Candidates were filtered using BLASTp to exclude peptides matching endogenous proteins. Spectra of candidates were compared with spectra recorded from synthetic peptides, and underlying frameshift mutations were confirmed by Sanger sequencing.

(B) Comparison of matched ions observed in candidate spectra (top) and synthetic peptide spectra (bottom). Top 10 most intense ions are labeled; retention time difference and correlation between experimental and synthetic peptide spectrum is reported. See also Table S2.

(C) Base calls and Sanger traces of underlying frameshift mutations. Positions of InDel mutations are indicated by an arrow. m1, minus one base pair deletion. See also Figure S5 and Tables S1 and S4.

Table 1. Overview of identified InDel neoepitopes.

Gene	Peptide	Location of mutation	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 (1,250), HLA-C*07:01 (1,665)
NFAT5	KRSSTILRL	chr16: 69691085	T5, m1	14	HLA-C*07:01 (203)
PSMC6	REKHSWHEP	chr14: 52713943	A4, m1	28	HLA-B*45:01 (1,218)
STK38	ISERDLLQY	chr6: 36497806	T7, m1	70	HLA-A*01:01 (33)
TUBGCP3	GVWEKPRRV	chr13: 112486018	G5, m1	108	no binder

HLA binding prediction was performed with NetMHCpan 4.0. Underlined aa of NFAT5-derived InDel neoepitope originate from wild-type NFAT5 protein sequence. m1, minus one base pair deletion. See also Figure 3 and Table S4.

lower for the four false-positive identifications (Pearson's correlation: 0.073–0.673). Furthermore, we observed substantial differences in retention times between experimental samples and synthetic peptide pools for the false-positive identifications. Binding prediction for identified InDel and SNP epitopes showed that all but one of the validated peptides are predicted to bind at least one of the HLA alleles expressed on HCT-116 cells. Furthermore, we confirmed the underlying mutations for all validated InDel and SNP neoepitopes in the genomic DNA using Sanger sequencing (Figures 3C and S4C). Taken together, these data show the validity of the identification of both InDel and SNP neoepitopes. Tables 1 and 2 provide an overview of all validated HLA class I-presented InDel and SNP neoepitopes.

CKAP2 frameshift mutation is recurrent in MSI CRC cell lines and patients

We next focused on the recurrence of frameshift mutations in the repeats of the five validated genes CKAP2, NFAT5, PSMC6, STK38, and TUBGCP3 by analyzing 24 MSI CRC cell lines (Table S1). In addition to HCT-116, the CKAP2 frameshift mutation was found in four other MSI CRC cell lines (KM12 [minus one base pair deletion (m1)], VaCo6 [m1], HROC24 [plus one base pair insertion (p1)], and LS411 [m1]). The TUBGC3 and STK38 frameshift mutations were identified in HCT-116 cells and in LoVo (m1) and HROC24 cells (m1), respectively. The NFAT5 and PSMC6 frameshift mutations were only found in HCT-116 cells. We next asked if the most recurrent CKAP2 mutation could also be identified in MSI CRC patient samples. To this end, we analyzed genomic tumor DNA obtained from 56 patients with MSI CRC and found m1 mutations in the described A8 repeat of the CKAP2 gene in 9 samples (16%). Finally, we evaluated the potential InDel neoepitopes arising from the confirmed frameshift mutations (Figure S5). Binding prediction for overlapping nonamers originating from FS protein sequences revealed multiple potential neoepitopes with promising binding affinities to common HLA supertypes (Doytchinova et al., 2004; Sette and Sidney, 1999; Sidney et al., 2008). Taken together, frameshift mutations were observed recurrently in cell lines derived from different tumors. CKAP2 frameshifts emerged to be the most interesting because these were found to be recurrent in cell lines and also in primary patient samples tested. Furthermore, the FS CKAP2 protein sequence harbors potential neoepitopes with binding potential to eight of twelve HLA supertypes tested.

NMD inhibition stabilizes frameshifted transcripts and augments HLA class I-mediated presentation of InDel neoepitopes

Apart from introducing frameshifts in the open reading frame and thus generating mRNAs encoding neoepitopes, InDel mutations typically trigger mRNA degradation by NMD (El-Bchiri et al., 2008). Therefore, the expression and consequently the presentation of most InDel neoepitopes in MSI CRC must be expected to be limited by NMD, reducing the usefulness of such neoepitopes for immunotherapy. We reasoned that NMD inhibition may stimulate the biosynthesis and the presentation of the InDel neoepitopes potentially increasing the cell's visibility to the immune system. We have previously identified the licensed drug 5AZA as a pharmacological inhibitor of NMD (Bhuvanagiri et al., 2014) and now tested its effect on the transcript and the peptidome level in HCT-116 cells. First, we confirmed the known effect of 5AZA on NMD efficiency by measuring the abundance of known endogenous NMD targets using qPCR. The NMD target mRNAs ATF3, ATF4, SC35C, SC35D, and UPP1 showed the expected upregulation between 1.8- and 4-fold following treatment with 5AZA (Figures 4A and S1). We then analyzed the abundance of the FS transcripts leading to the identified InDel neoepitopes. We found that 5AZA

Table 2. Overview of identified SNP neoepitopes.

Gene	Peptide	Location of mutation (GRCh37)	aa change	HLA binding prediction (K$_d$ in nM)
CHMP7	QTDQMVFN<u>T</u>Y[†]	chr8: 23116254	p.A324T	HLA-A*01:01 (16), HLA-B*18:01 (4,799)
NR1D1	YSDNSN<u>D</u>SF[†]	chr17: 38253572	p.G39D	HLA-A*01:01 (166), HLA-C*05:01 (23)
PCMT1	AAAPVVPQ<u>V</u>	chr6: 150117635	p.A226V	HLA-C*05:01 (2,282), HLA-C*07:01 (4,776)
RBBP7	EERVI<u>D</u>EEY[†]	chrX: 16887311	p.N61D	HLA-B*18:01 (107)
RGP1	RLDPGE<u>P</u>KSY	chr9: 35750729	p.S110P	HLA-A*01:01 (2,083), HLA-C*05:01 (6,887)

HLA binding prediction was performed with NetMHCpan 4.0. Mutated amino acids originating from SNPs are underlined. Peptides previously identified by mass spectrometry (Bassani-Sternberg et al., 2015) are marked with a dagger. See also Figure S4 and Table S5.

treatment highly significantly (p ≤ 0.0001) increased the abundance of *CKAP2* (2.0-fold), *PSMC6* (1.7-fold), *STK38* (1.4-fold), and *TUBGCP3* (1.9-fold) transcripts (Figure 4A). To further validate FS transcripts as *bona fide* NMD targets, we measured mRNA levels after small interfering RNA (siRNA)-mediated knockdown of the NMD core factor UPF1 (Lavysh and Neu-Yilik, 2020). We observed significant increases in mRNA abundance in four of the five targets (*CKAP2,* 1.4-fold; *PSMC6,* 1.5-fold; *NFAT5,* 1.4-fold; *TUBGCP3,* 1.6-fold; p ≤ 0.01; Figure 4A).

We next tested if the increase of transcript levels by NMD inhibition is translated into an increased presentation of InDel neoepitopes at the cell surface and performed label-free quantification of HLA-presented peptides isolated from HCT-116 cells treated with either 5AZA or DMSO. Given the difference in the number of quantifiable peptides, the quantification workflow was performed separately for each dataset recorded. Moreover, lowEThcD data were solely used for identification purposes and not for quantification as the correlation of raw intensity values between samples measured using lowEThcD and HCD was lower than between samples measured using EThcD and HCD (Figures S6A and S6B). We included only peptides that were identified and quantified in at least two out of three biological replicates per condition in the quantification dataset to minimize the need for imputing values. For the validated InDel neoepitopes, we complemented these untargeted data by integrating intensity values measured in a targeted MS2 analysis. Peptides were measured from the same samples and showed excellent reproducibility of intensities between untargeted and targeted MS2 data (Pearson's correlation coefficient: 0.93; Figure S6C). After filtering, normalization, and imputation of missing values, the final quantification dataset consisted of 5,072 distinct, quantifiable peptides (HCD: 4,634 peptides, EThcD: 1,376 peptides; overview of data processing in Figure S6D). Using the *limma* package, we identified a total of 838 differentially presented peptides upon NMD inhibition; 434 peptides showed an increased presentation, whereas 404 were less abundant (Figure 4C). The composition of the immunopeptidome is known to be influenced by the composition of the proteome (Bassani-Sternberg et al., 2015; Caron et al., 2011). We thus validated the effect of NMD inhibition on the immunopeptidome by confirming that the Gene Ontology (GO) term categories of the source proteins of differentially presented HLAp mirror those known to be affected by NMD inhibition at the proteome level (Sieber et al., 2016). GO term enrichment analysis for source genes of upregulated hit peptides following NMD inhibition revealed these to be involved in protein folding, endoplasmic reticulum stress response, unfolded protein response, proteasome-mediated APC-dependent catabolic process (i.e., breakdown of proteins by peptide bond hydrolysis), as well as antigen processing and presentation (Figure 4B). The effect of 5AZA treatment on the presentation of peptides originating from known NMD targets was confirmed by the significant upregulation of several *ATF3, ATF4,* and *UPP1* peptides (Figure 4D). We further validated the presentation of peptides originating from endogenous NMD targets by comparing source proteins of identified peptides with previously reported and ENSEMBL annotated NMD targets (Aliouat et al., 2019; Colombo et al., 2017; El-Bchiri et al., 2008; Tani et al., 2012). This analysis revealed several peptides originating from *ASAH1, CAV1, DDIT3, DDX5, HYOU1, JUN, PLIN3,* and *SLIRP* to be upregulated in the immunopeptidome of HCT-116 cells following NMD inhibition by 5AZA. We next studied endogenous peptides originating from the non-FS 5′ sequences of FS-bearing transcripts. This
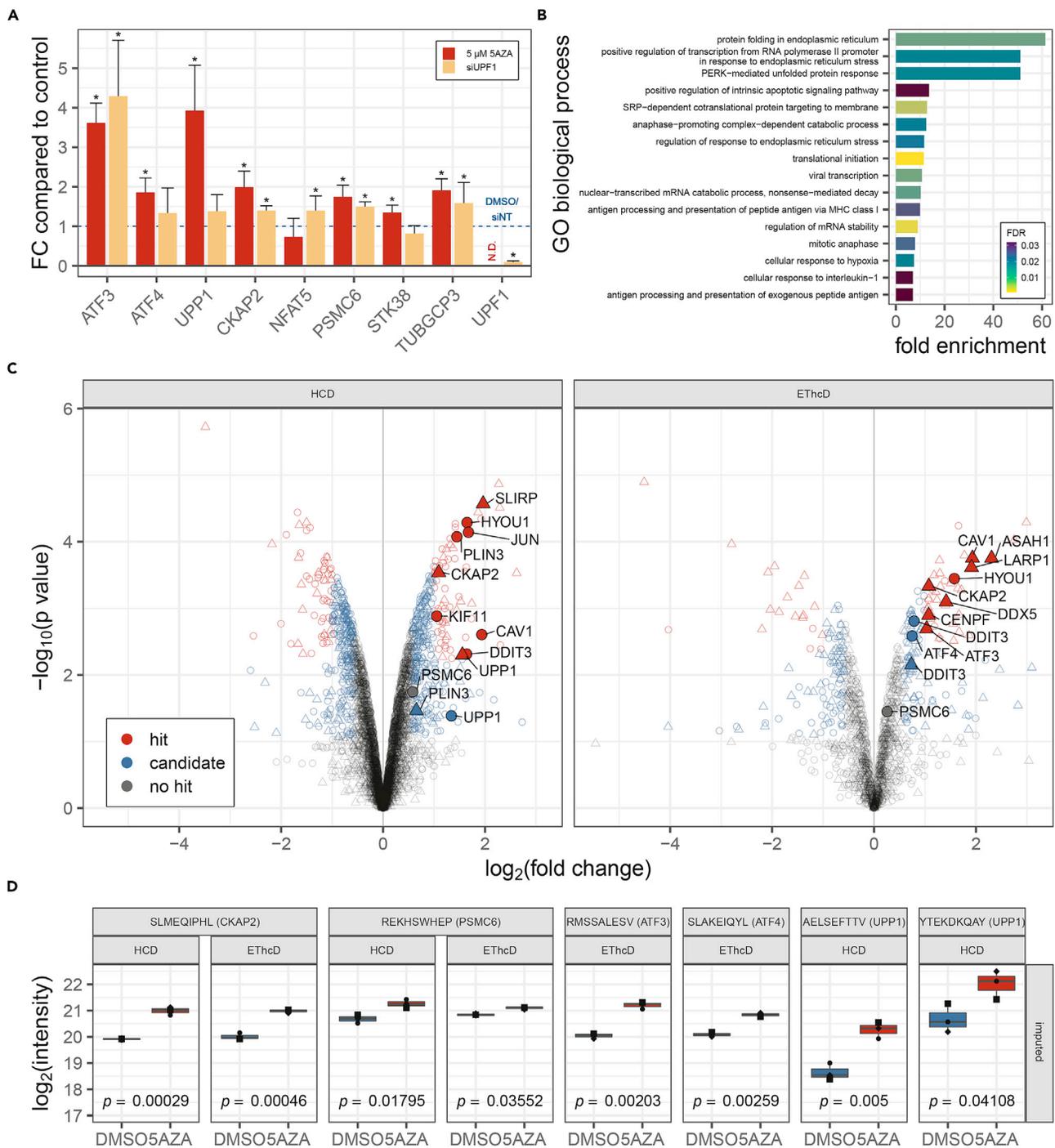
Figure 4. Treatment with 5AZA stabilizes NMD-targeted transcripts and augments HLA-mediated presentation of peptides originating from the encoded proteins

(A) qPCR analysis of endogenous NMD targets (*ATF3*, *ATF4*, *UPP1*) and InDel-mutated transcripts (*CKAP2*, *NFAT5*, *PSMC6*, *STK38*, *TUBGCP3*) after treatment with 5 μM 5AZA for 24 h (red) or siRNA-mediated KD of *UPF1* (orange). *UPF1* mRNA levels were determined as control for siRNA-mediated knockdown (N.D. = not determined). Each bar represents the mean ± SD of 3 experiments, *p ≤ 0.0001 (two-sided, unpaired t test). See also Figure S1.

(B) GO term enrichment for source genes of significantly upregulated hit peptides after 5AZA treatment for 24 h.

(C) Volcano plot summarizing limma analysis of label-free quantification of the immunopeptidome isolated from 5AZA-treated versus DMSO-treated HCT-116 cells. Upregulated InDel neoepitopes (*CKAP2*, *PSMC6*) and peptides originating from putative endogenous NMD targets are labeled with

Figure 4. Continued

the corresponding gene name. Color represents hit annotation; shape indicates if values were imputed (circle = no, triangle = yes). See also Figure S6 and Table S3.

(D) Representative plots showing changes in intensity for InDel neoepitopes SLMEQIPHL (*CKAP2*), REKHSWHEP (*PSMC6*), and selected peptides originating from known NMD targets *ATF3*, *ATF4*, and *UPP1* after treatment with 5AZA for 24 h. Bars represent 25th to 75th percentiles, middle line represents median, and points represent individual measurements of biological replicates. See also Table S3.

analysis revealed peptides originating from *CENPF*, *KIF11*, and *LARP1* to be upregulated following NMD inhibition by 5AZA (Figure 4C). Finally, we analyzed the effect of 5AZA treatment on the presentation of the validated InDel neoepitopes. The HLA class I-mediated presentation of the *CKAP2*-derived InDel neoepitope was significantly upregulated 2.1- to 2.13-fold (p ≤ 0.0005) and that of the *PSMC6*-derived InDel neoepitope was significantly (p ≤ 0.05) albeit less strongly (1.2- to 1.5-fold) upregulated both in the HCD and the EThcD datasets (Figure 4D). Taken together, these findings show that modulation of NMD efficiency in MSI CRC cells by the pharmacological NMD inhibitor 5AZA stabilizes FS-bearing and NMD-targeted transcripts and results in the increased cell surface presentation of HLAp, including InDel neoepitopes, derived thereof.
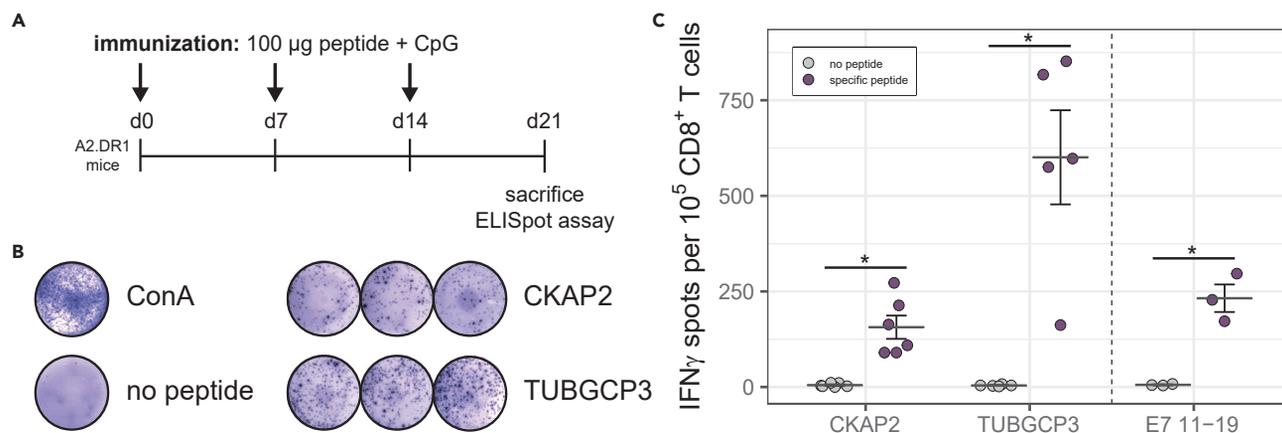
In vivo immunization with InDel neoepitopes induces CD8[+] T cell responses

We next analyzed the potential of InDel neoepitopes to induce specific T cell responses by performing *in vivo* immunizations in a humanized HLA-A*02:01-transgenic mouse model (Pajot et al., 2004). Binding predictions indicated that the *CKAP2*-derived InDel neoepitope is a strong HLA-A*02:01 binder (percentile rank: 0.0071, predicted IC50: 2.3736 nM), whereas InDel neoepitopes derived from *NFAT5*, *PSMC6*, and *STK38* were predicted to bind other alleles than HLA-A*02:01 (Table 1). Interestingly, the *TUBGCP3*-derived InDel neoepitope was not predicted to bind any of the HLA alleles expressed by HCT-116 cells by NetMHCpan but showed the strongest affinity to HLA-A*02:01 (percentile rank: 2.3267, predicted IC50: 6,217 nM). We have therefore included this InDel neoepitope for further testing.

As a first step, we immunized three mice with a mixture of peptides consisting of two potential HLA-A*02:01 binders (*CKAP2*- and *TUBGCP3*-derived InDel neoepitopes) and two "non-binders" (*NFAT5*-and *PSMC6*-derived InDel neoepitopes) as negative controls. Whole splenocytes were analyzed by *ex vivo* IFNγ enzyme-linked immune absorbent spot (ELISpot) assays. One 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 S7). These results were validated by immunizations of mice with a single peptide according to the immunization scheme shown in Figure 5A. The analysis of isolated splenic CD8[+] T cells following immunizations with either *CKAP2*- or *TUBGCP3*-derived InDel neoepitopes or the positive control HPV16 peptide E7 (aa 11–19) resulted in IFNγ-specific and highly significant responses in the ELISpot assay (Figures 5B and 5C). These data demonstrate that InDel neoepitopes are loaded on nascent HLA molecules and presented *in vivo* via HLA-A*02:01 molecules. Importantly, immunization can induce a specific CD8[+] T cell-mediated immune response.

DISCUSSION

The identification of tumor-specific neoepitopes represents a crucial step in the development of therapeutic cancer vaccines, and a high load of neoepitopes has been associated with effective immunotherapy (Schumacher and Schreiber, 2015). Although *in silico* predictions have been employed previously to identify cancer-specific neoepitopes, only a negligible fraction of candidates implicated by this approach are presented by HLA class I molecules and do not, therefore, elicit anti-tumor immune responses (Schmidt et al., 2017). MS of the immunopeptidome provides an unbiased view of actually presented peptides. Previous studies of neoepitopes using MS provided the proof of concept but focused on SNP-derived neoepitopes (Bassani-Sternberg et al., 2015), which induce less robust immune responses than InDel-derived neoepitopes thus limiting their therapeutic potential (Turajlic et al., 2017). Other studies have analyzed only selected InDel neoepitopes originating from specific, recurrent InDel mutations (Narayan et al., 2019; Roudko et al., 2020; van der Lee et al., 2019), or failed to detect frameshift-derived mutant sequences at the proteome level (Halvey et al., 2014). Here, we report the first unbiased, MS-based identification of immunogenic InDel neoepitopes using publicly available sequencing data. Moreover, we demonstrate that pharmacological inhibition of NMD using an approved drug stabilizes the corresponding mRNAs

Figure 5. *In vivo* immunization of A2.DR1 mice with InDel neoepitopes induces CD8+ T cell responses; see also **Figure S7**

(A) Immunization scheme. See also Table S2.

(B) Representative ELISpot assay results for isolated CD8+ T cells stimulated with ConA (assay positive control), no peptide control, and InDel neoepitopes SLMEQIPHL (*CKAP2*) and GVWEKPRRV (*TUBGCP3*).

(C) Quantitative analysis of ELISpot assays. Bars represent mean ± SEM of N = 6 (*CKAP2*), N = 5 (*TUBGCP3*), or N = 3 (E7 11–19, control peptides) experiments, *p ≤ 0.005 (two-sided, unpaired t test).

bearing FS open reading frames consequently increasing HLA class I-mediated presentation of InDel neo-epitopes at the cell surface.

We combined a previously established high-throughput procedure for IP of HLA:peptide complexes (Chong et al., 2018) with different fragmentation and precursor selection methods for LC-MS/MS to obtain a representative view of the immunopeptidome of MSI CRC cells. Samples analyzed using HCD fragmentation yielded the majority of HLAp identifications due to the higher MS2/MS1 rate and deeper sampling of the immunopeptidome compared with dual-fragmentation measurements, whereas lowEThcD fragmentation identified a subset of low-abundance HLAp, which were found neither by EThcD with standard precursor selection nor by HCD. Therefore, lowEThcD (or any other methodology targeting low-abundance precursors first) is suited to further expand the detectable HLA class I immunopeptidome.

Next, we performed a multi-round database search, matching high-scoring "*de novo* only" spectra, which did not match any known human protein sequences, against custom InDel and SNP neoepitope databases and identified 14 neoepitope candidates. Eleven of these candidates were identified using standard HCD fragmentation, whereas three could only be identified using EThcD or lowEThcD fragmentation, emphasizing the added value of combining different fragmentation methods. As a first important finding, our work thus demonstrates that the combination of different fragmentation and precursor selection method-ologies can increase the number of identified HLA class I-presented (neo-)epitopes. Five of the eight iden-tified InDel neoepitope candidates were successfully validated on both the genomic and peptidomic level demonstrating the higher success rate of an unbiased, MS-based approach for the identification of neoe-pitopes compared with *in silico* predictions and indirect immunological readouts.

As a second major finding, NMD inhibition by a clinically achievable 5AZA concentration (Stresemann and Lyko, 2008) significantly augments HLA class I-mediated presentation of peptides originating from NMD-sensitive transcripts, including InDel neoepitopes, thus increasing the likelihood of immune recognition. Notably, we found the m1 mutation resulting in the *CKAP2*-derived neoepitope to be highly recurrent in different MSI CRC cell lines and patient samples. In addition to the MSI CRC cell lines analyzed in this study, the *CKAP2* frameshift mutation leading to the identified InDel neoepitope was found to be present in 13 other samples of the CCLE panel (Barretina et al., 2012; Ghandi et al., 2019). These include 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). All these cancer cell lines except for SNU-1040 and Kasumi-2 exhibit microsatellite instability (Barretina et al., 2012; Ghandi et al., 2019) further supporting the previously reported importance of this frameshift mutation in MSI cancers (Leoni et al., 2020;

Spaanderman et al., 2020). While the frameshift-bearing *CKAP2* transcript has previously been classified as likely being NMD-resistant based on the localization of the m1 mutation (Thermann et al., 1998), we directly demonstrate NMD sensitivity of this transcript by pharmacological NMD inhibition with 5AZA and by RNA interference of the key NMD factor UPF1. In agreement with previous reports (Lindeboom et al., 2016; Neu-Yilik et al., 2011) these findings indicate that sequence features alone are not sufficient to predict NMD sensitivity and require experimental validation. Notably, in addition to the FS NMD targets many HLAp up-regulated by 5AZA treatment originated from source proteins involved in stress response mechanisms. These findings confirm our previously reported data showing that many stress-related transcripts are controlled by NMD and support the hypothesis that NMD inhibition augments the expression of physiological C-terminally truncated proteins (Sieber et al., 2016).

Finally, we show directly in a humanized HLA-A*02:01-transgenic mouse model that the identified HLA-A*02:01-restricted InDel neoepitopes derived from *CKAP2* and *TUBGCP3* can induce strong CD8[+] T cell responses. These results are in agreement with a recent study reporting the isolation and reactivity of human T cells directed against the *CKAP2*-derived InDel neoepitope (Leoni et al., 2020). Frameshift-derived InDel neoepitopes have previously been shown to elicit T cell-mediated cytotoxicity (Schwitalle et al., 2008), and the InDel neoepitopes identified here might be employed for the development of personalized vaccines or engineered T cell therapies.

Although we have chosen MSI CRC as a proof-of-concept model disease, microsatellite instability is increasingly recognized in several other malignancies, which might benefit from increased neoepitope presentation after NMD inhibition (Bonneville et al., 2017). Although immune checkpoint blockade was shown to be effective in cancers with high mutational burden (Marcus et al., 2019; Yarchoan et al., 2017), we envision that 5AZA treatment expands the repertoire of T cells directed against frameshift-derived InDel neoepitopes and may thus sensitize such tumors for immunotherapeutic interventions. In contrast to SNP neoepitopes, FS transcripts encode multiple InDel neoepitopes. Therefore, NMD inhibition is expected to increase HLA class I-mediated presentation of InDel neoepitopes independently of both, a patient's HLA allotype, and tumor mutational landscape. Based on the data in this study, it is not possible to propose an explicit threshold for inducing a CD8[+] T cell cytotoxic response directed toward the identified InDel neoepitopes. Although early reports suggest, that even a single HLA class I-presented neoepitope can trigger an immune response (Foote and Eisen, 2000), other factors such as TCR affinity, HLA:peptide complex stability, as well as neoepitope foreignness are important for the activation of T cells (McGranahan and Swanton, 2019; Wells et al., 2020). Moreover, recent reports have shown that the fraction of cells presenting the epitope also has an important role in effective T cell-mediated tumor rejection (Gejman et al., 2018; McGranahan et al., 2016). In our approach, treatment with 5AZA augments the HLA class I-mediated presentation of InDel neoepitopes thus potentially increasing the likelihood of effective T cell activation.

In summary, we demonstrate the feasibility of high-throughput immunopeptidomics for the identification of clinically frequent immunogenic InDel neoepitopes in MSI cancers and show that pharmacological NMD inhibition augments HLA class I-mediated presentation of such neoepitopes. The data reported here thus suggest that immunotherapeutic strategies for MSI cancers may benefit from a combination with NMD inhibition by turning cancer cells into easily identifiable targets for tumor-specific T cells.

Limitations of the study

It is important to note that the number of neoepitopes identified by our approach may be limited. Variant calling of InDel mutations remains challenging even with advanced biocomputational algorithms (Ballhausen et al., 2020), and the usage of public mutation databases may limit the identification of InDel neoepitopes (Freudenmann et al., 2018). Indeed, the length of mutated microsatellites resulting in identified InDel neoepitopes was short (4–8 bp) and we did not document InDel neoepitopes derived from well-established InDel mutations occurring in longer repetitive sequences of the HCT-116 cell line as these mutations are not documented in either the COSMIC or CCLE database. To address this limitation we also searched our MS raw data against a database of frequent InDel mutations in MSI CRC (Ballhausen et al., 2020), but this approach did not yield additional InDel neoepitopes. Potentially, this limitation may be overcome in future studies by employing long-read sequencing technologies (Nakano et al., 2017).

Moreover, we did not identify the known HLA-A*02:01-restricted *TGFBR2*-derived InDel neoepitope RLSSCVPVA (Saeterdal et al., 2001; Schwitalle et al., 2008) in our targeted MS approach. Based on a recently

reported analysis of 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) which is most likely lost during the IP procedure. However, future immunopeptidomics studies might follow a similar approach and integrate validated sequences or spectral libraries of InDel neoepitopes identified in this study to further improve InDel neoepitope identification.

Finally, our quantification approach only allows relative quantification between the two conditions tested, but no absolute quantification of InDel neoepitopes. Future studies employing targeted MS analysis such as PRM or SRM (Blatnik et al., 2018; Tan et al., 2011) may utilize isotopically labeled counterparts of the InDel neoepitopes identified in this study to further investigate their absolute amounts in MSI CRC cell line and patient samples.

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Andreas E. Kulozik (andreas.kulozik@med.uni-heidelberg.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The MS data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol et al., 2019) with the dataset identifier PRIDE: PXD021755. All scripts are available from the lead contact upon request.

METHODS

All methods can be found in the accompanying transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102389.

AUTHOR CONTRIBUTION

J.P.B., M.W.H., and A.E.K. designed the project. J.P.B. performed most of the experiments, performed the data analysis, generated the figures, and drafted the manuscript. D.H. contributed to the experimental design and establishment of MS methodologies. D.H. and M.R. performed MS runs. F.S. contributed to the quantitative MS data analysis. K.U. and A.H.-S. performed *in vivo* immunizations and ELISpot assays. J.G., M.K., G.N.-Y., and M.v.K.D. contributed to the conceptualization of the study and interpretation of the results. M.W.H. and A.E.K. coordinated the study and finalized the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

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. 17, 227–239, https://doi.org/10.1080/15476286.2019.1674595.

Andreatta, M., and Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 32, 511–517, https://doi.org/10.1093/bioinformatics/btv639.

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, https://doi.org/10.1038/s41467-020-18514-5.

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, https://doi.org/10.1038/nature11003.

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, https://doi.org/10.1074/mcp.M114.042812.

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, https://doi.org/10.15252/emmm.201404461.

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, https://doi.org/10.1002/pmic.201700390.

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, https://doi.org/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, https://doi.org/10.1038/s41389-018-0079-x.

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 Precis. Oncol. 1, 1–15, https://doi.org/10.1200/po.17.00073.

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, https://doi.org/10.3322/caac.21492.

Cancer Genome Atlas, N. (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature 487, 330–337, https://doi.org/10.1038/nature11252.

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, https://doi.org/10.1038/msb.2011.68.

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, https://doi.org/10.1074/mcp.TIR117.000383.

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, https://doi.org/10.1261/rna.059055.116.

Cui, J.J., Wang, L.Y., Tan, Z.R., Zhou, H.H., Zhan, X., and Yin, J.Y. (2020). Mass spectrometry-based personalized drug therapy. Mass Spectrom. Rev. 39, 523–552, https://doi.org/10.1002/mas.21620.

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

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, https://doi.org/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, https://doi.org/10.1371/journal.pone.0002583.

Foote, J., and Eisen, H.N. (2000). Breaking the affinity ceiling for antibodies and T cell receptors. Proc. Natl. Acad. Sci. 97, 10679–10681, https://doi.org/10.1073/pnas.97.20.10679.

Freudenmann, L.K., Marcu, A., and Stevanovic, S. (2018). Mapping the tumour human leukocyte antigen (HLA) ligandome by mass spectrometry.

Immunology 154, 331–345, https://doi.org/10.1111/imm.12936.

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, e41090, https://doi.org/10.7554/eLife.41090.

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, https://doi.org/10.1038/s41586-019-1186-3.

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, https://doi.org/10.1038/nature13988.

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 Res. 74, 387–397, https://doi.org/10.1158/0008-5472.CAN-13-2488.

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, https://doi.org/10.1080/2162402X.2017.1302631.

Kloor, M., Reuschenbach, M., Pauligk, C., Karbach, J., Rafiyan, M.-R., Al-Batran, S.-E., Tariverdian, M., Jäger, E., and von Knebel Doeberitz, M. (2020). A frameshift peptide neoantigen-based vaccine for mismatch repair-deficient cancers: a phase I/IIa clinical trial. Clin. Cancer Res. 26, 4503–4510, https://doi.org/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, https://doi.org/10.1016/j.trecan.2016.02.004.

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, https://doi.org/10.1021/ac060777w.

Lavysh, D., and Neu-Yilik, G. (2020). UPF1-Mediated RNA decay-danse macabre in a cloud. Biomolecules 10, 999, https://doi.org/10.3390/biom10070999.

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 Res. 80, 3972–3982, https://doi.org/10.1158/0008-5472.CAN-20-1072.

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, https://doi.org/10.1038/ng.3664.

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, https://doi.org/10.1158/1078-0432.Ccr-18-4070.

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 *351*, 1463–1469, https://doi.org/10.1126/science.aaf1490.

McGranahan, N., and Swanton, C. (2019). Neoantigen quality, not quantity. Sci. Transl. Med. *11*, eaax7918, https://doi.org/10.1126/scitranslmed.aax7918.

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. *111*, 4507–4512, https://doi.org/10.1073/pnas.1321458111.

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, https://doi.org/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, https://doi.org/10.1371/journal.pone.0219547.

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, https://doi.org/10.1261/rna.2401811.

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, https://doi.org/10.1002/eji.200425463.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support

for quantification data. Nucleic Acids Res. *47*. D442–d450. https://doi.org/10.1093/nar/gky1106.

Roudko, V., Bozkus, C.C., Orfanelli, T., McClain, C.B., Carr, C., O'Donnell, T., Chakraborty, L., Samstein, R., Huang, K.-l., Blank, S.V., et al. (2020). Shared immunogenic poly-epitope frameshift mutations in microsatellite unstable tumors. Cell *183*, 1634–1649.e17, https://doi.org/10.1016/j.cell.2020.11.004.

Saeterdal, I., Bjorheim, J., Lislerud, K., Gjertsen, M.K., Bukholm, I.K., Olsen, O.C., Nesland, J.M., Eriksen, J.A., Moller, M., Lindblom, A., and Gaudernack, G. (2001). Frameshift-mutation-derived peptides as tumor-specific antigens in inherited and spontaneous colorectal cancer. Proc. Natl. Acad. Sci. U S A *98*, 13255–13260, https://doi.org/10.1073/pnas.231326898.

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, https://doi.org/10.1074/jbc.M117.789511.

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, https://doi.org/10.1186/s13073-015-0240-5.

Schumacher, F.R., Delamarre, L., Jhunjhunwala, S., Modrusan, Z., Phung, Q.T., Elias, J.E., and Lill, J.R. (2017). Building proteomic tool boxes to monitor MHC class I and class II peptides. Proteomics *17*, 1600061, https://doi.org/10.1002/pmic.201600061.

Schumacher, T.N., and Schreiber, R.D. (2015). Neoantigens in cancer immunotherapy. Science *348*, 69–74, https://doi.org/10.1126/science.aaa4971.

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, https://doi.org/10.1053/j.gastro.2008.01.015.

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, https://doi.org/10.1007/s002510050594.

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, https://doi.org/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, https://doi.org/10.1074/mcp.M115.054056.

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 *2*. https://www.biorxiv.org/content/10.1101/2020.03.26.010231v2, (Accessed 30 March 2020).

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

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, https://doi.org/10.1002/pmic.201000531.

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, https://doi.org/10.4161/rna.22360.

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. EMBO J. *17*, 3484–3494, https://doi.org/10.1093/emboj/17.12.3484.

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. Lancet Oncol. *18*, 1009–1021, https://doi.org/10.1016/s1470-2045(17)30516-8.

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, https://doi.org/10.1172/JCI97482.

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, https://doi.org/10.1182/blood-2015-04-640532.

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 *183*, 818–834.e13, https://doi.org/10.1016/j.cell.2020.09.015.

Yarchoan, M., Hopkins, A., and Jaffee, E.M. (2017). Tumor mutational burden and response rate to PD-1 inhibition. N. Engl. J. Med. *377*, 2500–2501, https://doi.org/10.1056/NEJMc1713444.

## Supplemental information

NMD inhibition by 5-azacytidine augments

presentation of immunogenic

frameshift-derived neoepitopes

Jonas P. Becker, Dominic Helm, Mandy Rettel, Frank Stein, Alejandro Hernandez-Sanchez, Katharina Urban, Johannes Gebert, Matthias Kloor, Gabriele Neu-Yilik, Magnus von Knebel Doeberitz, Matthias W. Hentze, and Andreas E. Kulozik

Supplemental Information

Supplemental figures and legends
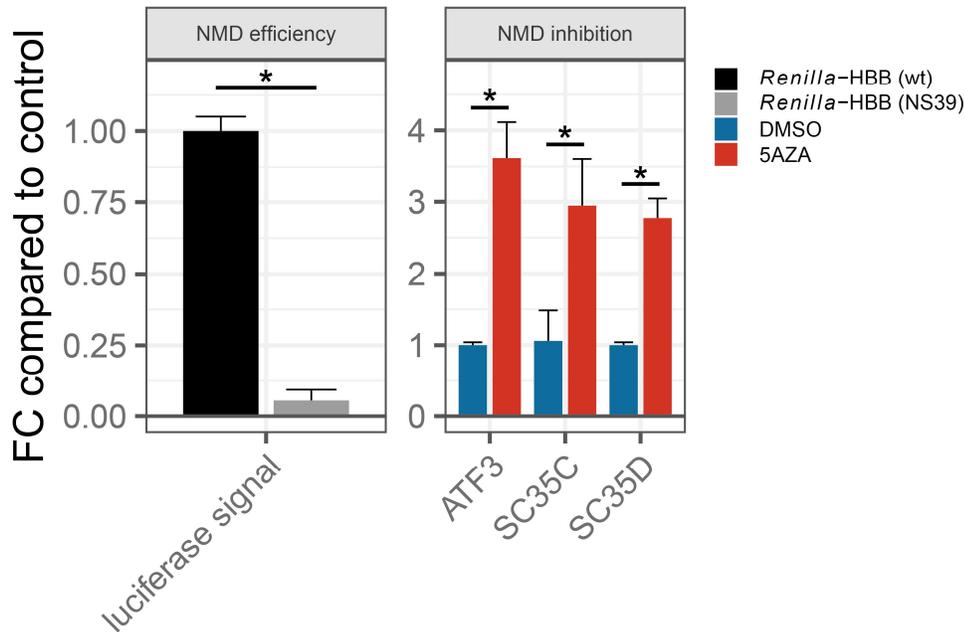
Supplemental figure 1: NMD is efficient in HCT-116 cells and can be inhibited by 5AZA; related to Fig 4.

NMD efficiency determined by significant downregulation of luciferase signal from *Renilla*-HBB (NS39) construct compared to *Renilla*-HBB (wt) construct (left panel). Each bar represents mean ± SD of 3 experiments, *$p \leq 0.0001$ (two-sided, unpaired t-test). Treatment with 5 µM 5AZA for 24 h increases mRNA abundance of endogenous NMD targets *ATF3*, *SC35C*, and *SC35C* as determined by qPCR (right panel). Each bar represents mean ± SD of 3 experiments, *$p \leq 0.0001$ (two-sided, unpaired t-test). See also Table S7.
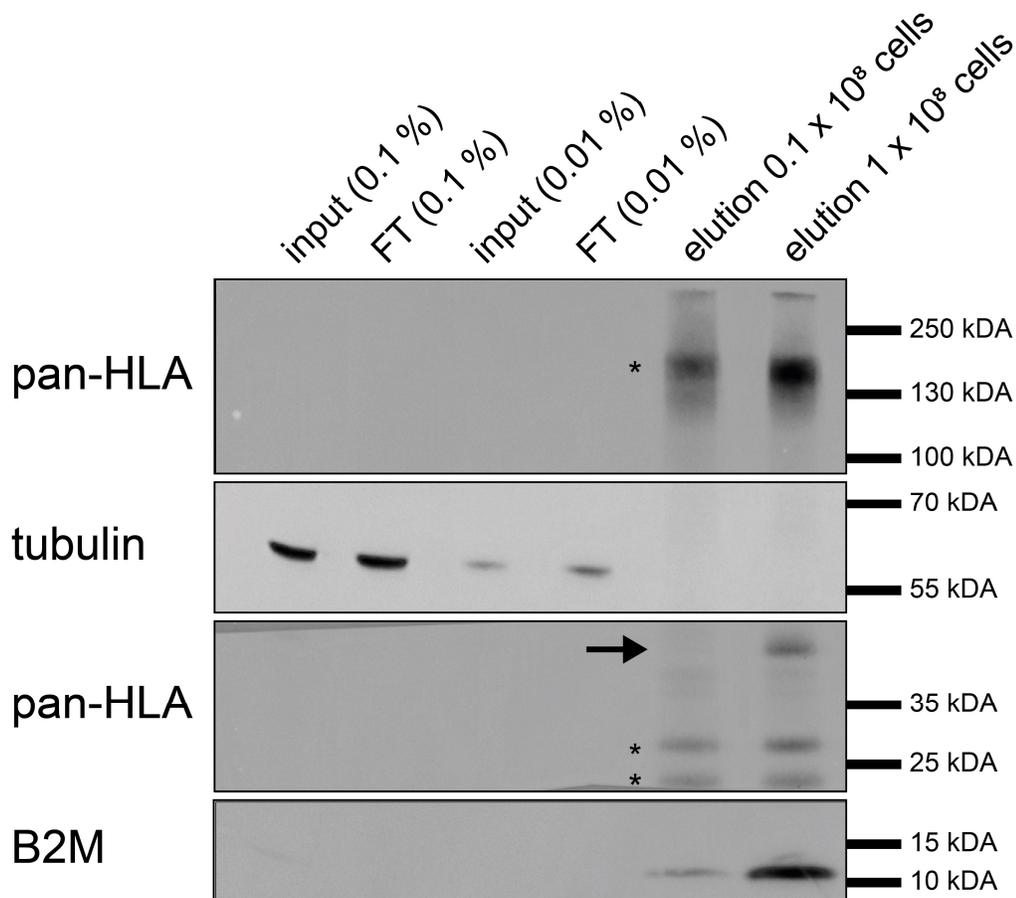
Supplemental figure 2: Selective purification of HLA molecules by high-throughput immunoprecipitation; related to Fig 1.

Western blot analysis of HLA molecules purified by immunoprecipitation with W6/32 antibody. Arrow indicates expected band of HLA molecules. Asterisks indicate background bands of eluted W6/32 antibody. B2M, beta-2-microglobulin.
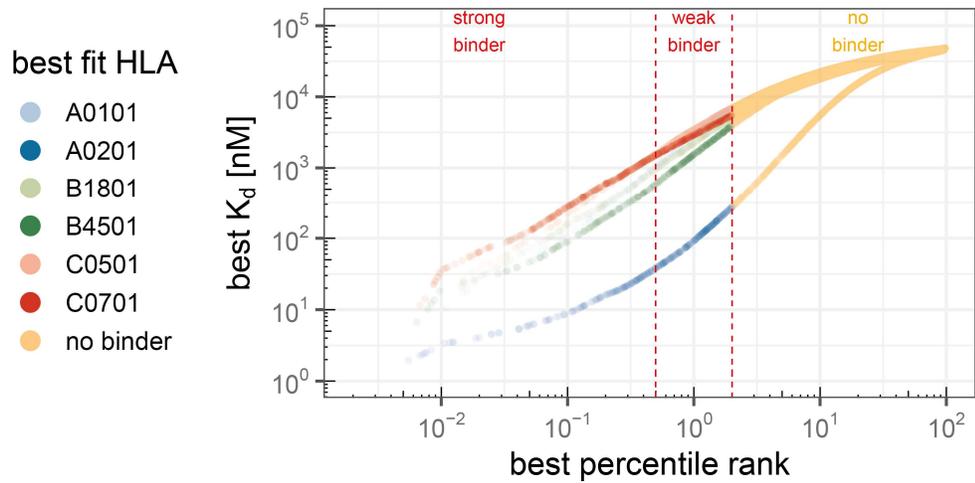
Supplemental figure 3: Frameshift inducing InDel mutations generate 2782 potential InDel neoepitopes; related to Fig 1.

Binding prediction of all potential nonamers resulting from InDel mutations annotated in COSMIC and CCLE databases to HLA alleles expressed by HCT-116 cells. Threshold for strong binders is top 0.5% ranked, for weak binders top 2%.
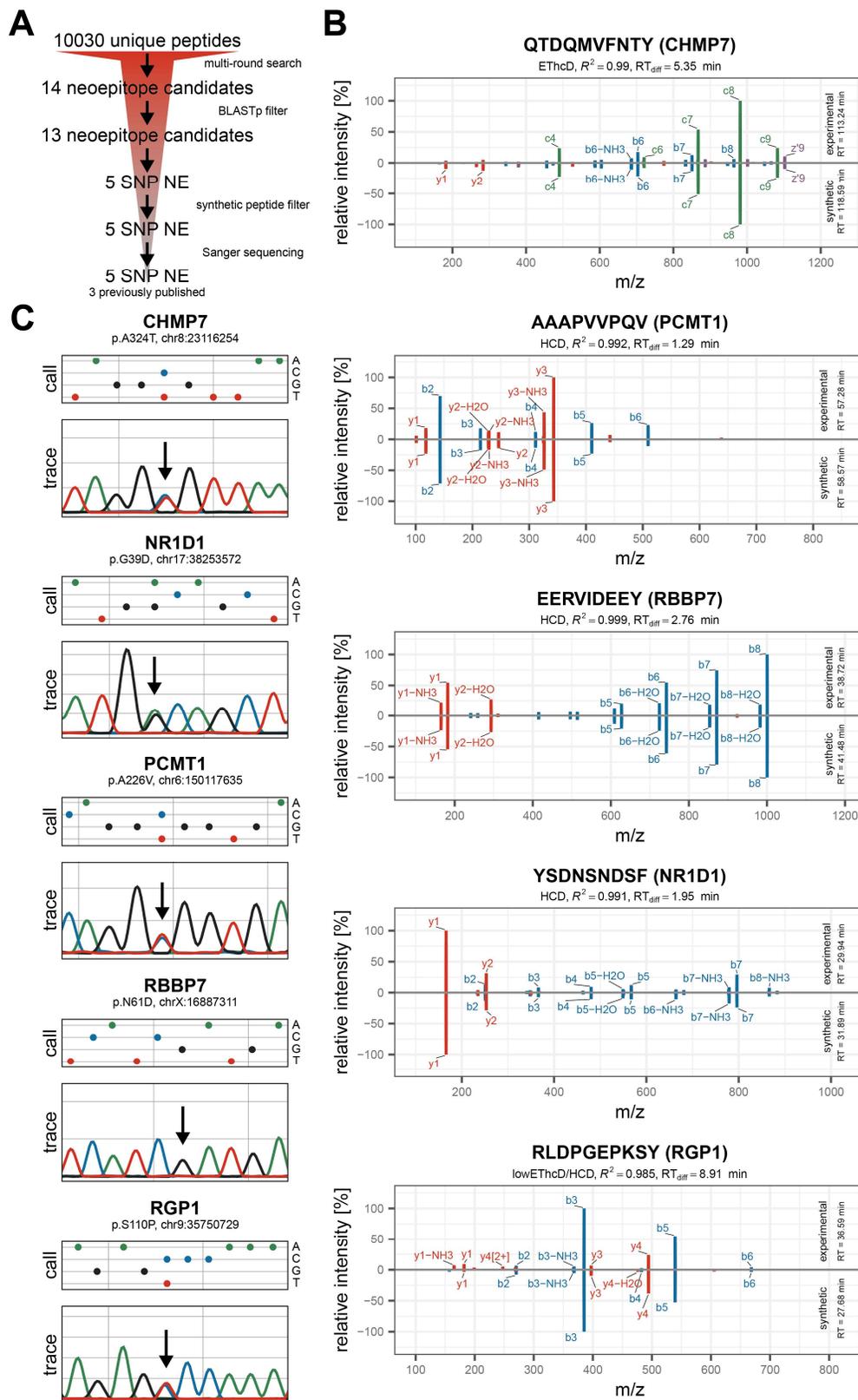
Supplemental figure 4: Validation of identified SNP neoepitopes; related to Table 2.

(A) Overview of validation procedure. Candidates were filtered using BLASTp to exclude peptides matching endogenous proteins. Spectra of candidates were compared to spectra recorded from synthetic peptides and underlying SNPs were confirmed by Sanger sequencing.

(B) Comparison of matched ions observed in candidate spectra (top) and synthetic peptide spectra (bottom). Top 10 most intense ions are labeled, retention time difference and correlation between experimental and synthetic peptide spectrum is reported. *RGP1* peptide is singly charged and was therefore compared to HCD synthetic spectra. See also Table S2.

(C) Base calls and sanger traces of underlying SNPs. Positions of SNPs are indicated by an arrow. *RBBP7* SNP is homozygous. See also Table S5.
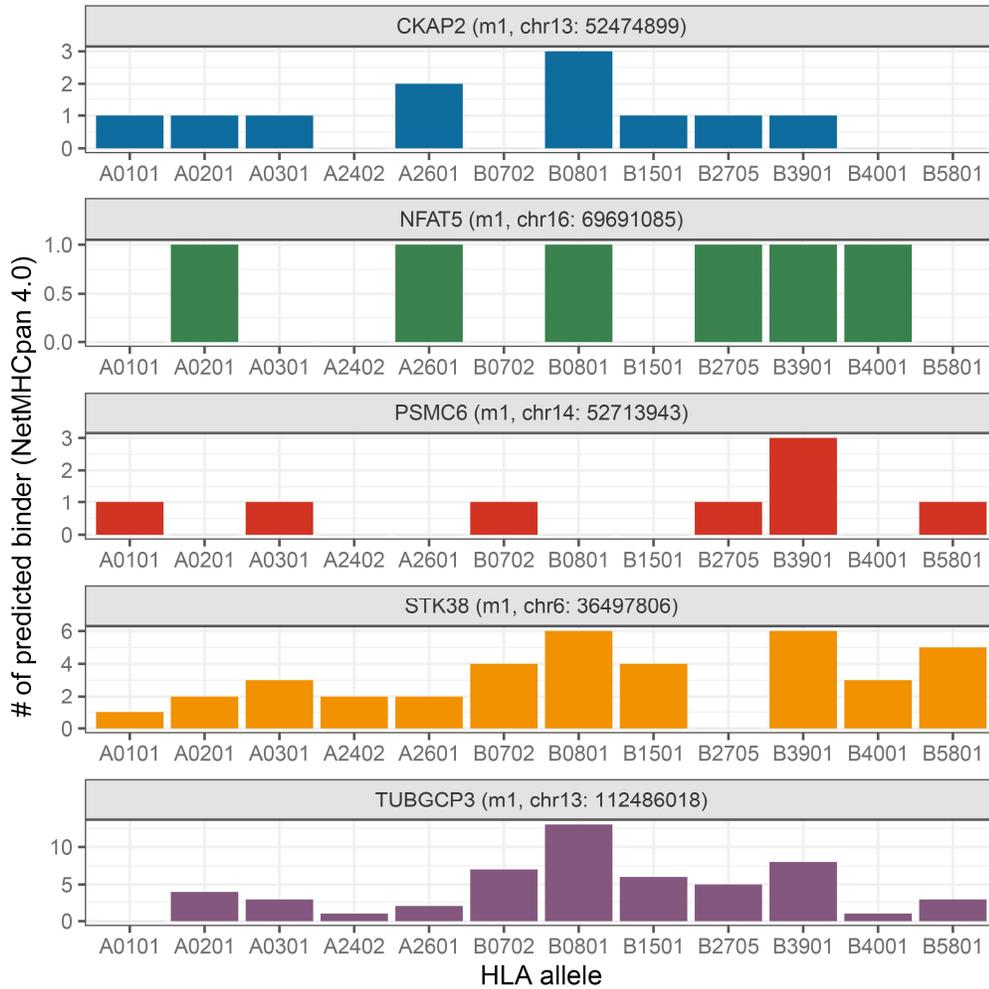
Supplemental figure 5: Frameshift parts of source proteins of identified InDel neoepitope generate numerous nonamers predicted to bind HLA supertype alleles; related to Fig 3.

Binding prediction for overlapping nonamers originating from frameshift part of mutated source proteins was performed using NetMHCpan 4.0.
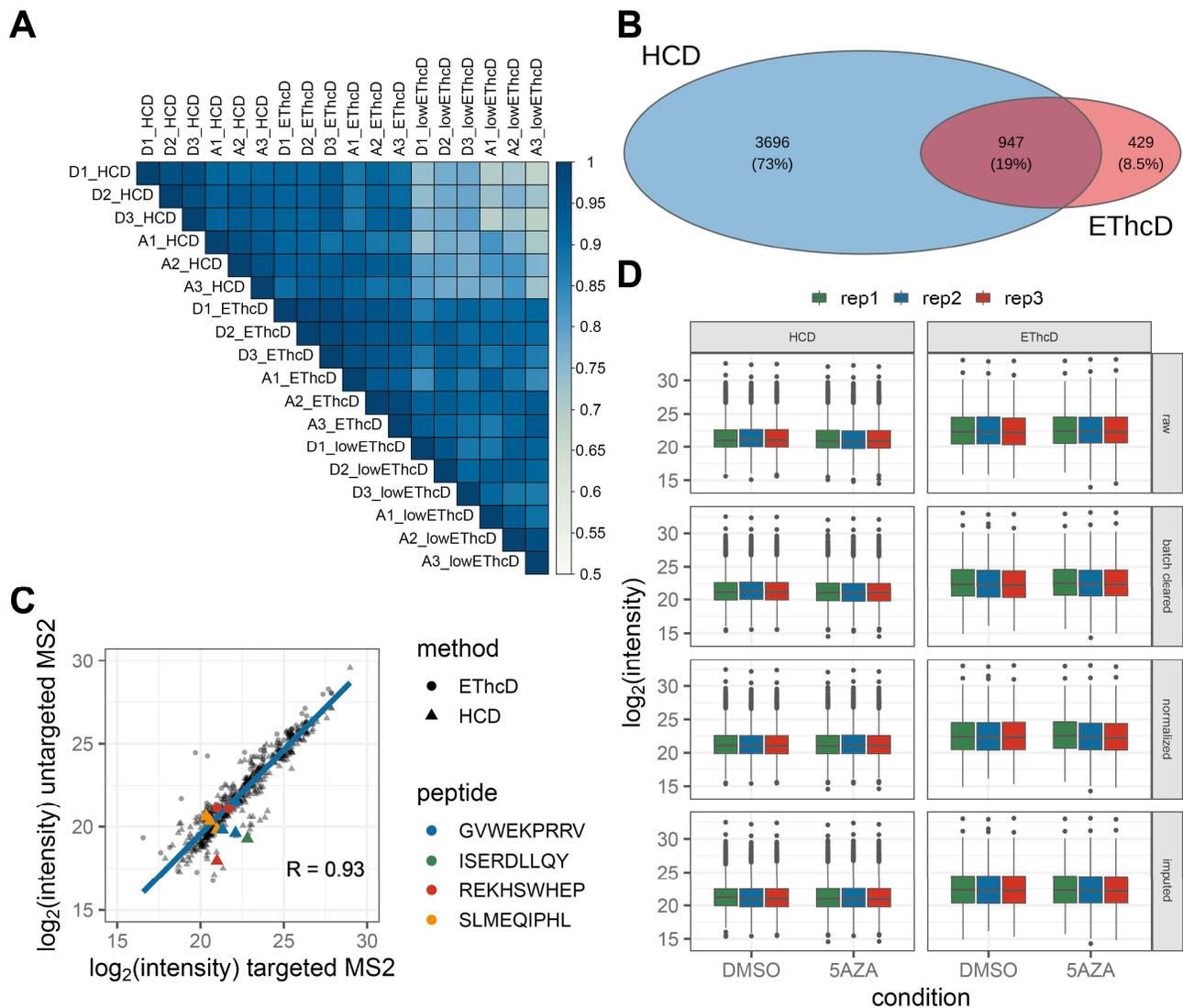
Supplemental figure 6: Overview of quantification datasets and dataset processing workflow; related to Fig 4.

(A) Correlation matrix showing correlation of raw intensity values between all recorded datasets. Samples are labeled with condition (D = DMSO, A = 5AZA), replicate number (1-3) and MS fragmentation method (HCD, EThcD, lowEThcD).

(B) Overlap between datasets of quantifiable peptides (measured in at least two out of three replicates per condition) for datasets recorded with HCD and EThcD fragmentation.

(C) Intensities of peptides identified using a targeted MS2 approach versus intensities of peptides identified using an untargeted MS2 approach (HCD/EThcD). Datapoint shape represents MS fragmentation method. Targeted InDel neoepitopes are colored. Peptide KRSSTILRL (*NFAT5*) was not measured with untargeted MS2 approach (HCD/EThcD). See also Table S3.

(D) Data processing overview. Top panel shows distribution of raw intensity values for each replicate, second shows intensities after batch clearing, third after normalization and bottom panel after imputation of missing values.
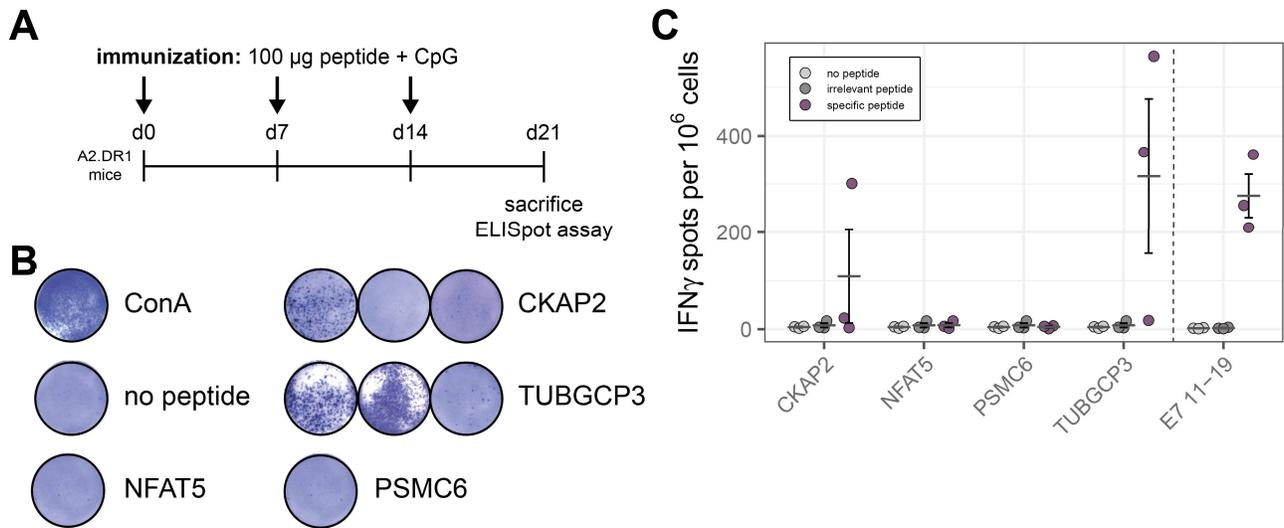
Supplemental figure 7: *In vivo* immunization of A2.DR1 mice with pooled InDel neoepitopes; related to Fig 5.

(A) Immunization scheme. See also Table S2.

(B) Representative ELISpot assay results for whole spleenocytes stimulated with ConA (assay positive control), no peptide control, and InDel neoepitopes.

(C) Quantitative analysis of ELISpot assays. Bars represent mean ± SEM of N = 3.

Supplemental tables

Supplemental table 1: InDel mutation analysis by gel capillary electrophoresis in MSI CRC cell lines; related to Fig 3.

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

Frequency of underlying frameshift mutations of identified InDel neoepitopes was tested in

24 MSI CRC cell lines. m1, minus one base pair deletion; p1, plus one base pair insertion.

See also Table S6.

Supplemental table 2: Synthetic peptides used for validation and *in vivo* immunization; related to Fig 3/5/S4/S7.

gene	sequence	type	used for
AXIN2	ac-GATAGTPAP	InDel neoepitope	synthetic peptide validation
CDC42EP5	PPRPAAAP	InDel neoepitope	synthetic peptide validation
CHMP7	QTDQMVFNTY	SNP neoepitope	synthetic peptide validation
CKAP2	SLMEQIPHL	InDel neoepitope	synthetic peptide validation, *in vivo* immunization
E7	YMLDLQPET	HPV16 viral epitope	in vivo immunization (positive control)
LMTK3	ac-VGGGFPPPPP	InDel neoepitope	synthetic peptide validation
NFAT5	KRSSTILRL	InDel neoepitope	synthetic peptide validation, *in vivo* immunization
NR1D1	YSDNSNDSF	SNP neoepitope	synthetic peptide validation
PCMT1	AAAPVVPQV	SNP neoepitope	synthetic peptide validation
PSMC6	REKHSWHEP	InDel neoepitope	synthetic peptide validation, *in vivo* immunization
RBBP7	EERVIDEEY	SNP neoepitope	synthetic peptide validation
RGP1	RLDPGEPKSY	SNP neoepitope	synthetic peptide validation
STK38	ISERDLLQY	InDel neoepitope	synthetic peptide validation
TUBGCP3	GVWEKPRRV	InDel neoepitope	synthetic peptide validation, *in vivo* immunization

ac, acetylation.

Supplemental table 3: Target m/z list of InDel neoepitopes for targeted MS2 method; related to Fig 4/S6.

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

Supplemental table 4: Primer used for validation of underlying frameshift mutations of identified InDel neoepitopes by Sanger sequencing; related to Fig 3 and Table 1.

target	forward primer	reverse primer
CKAP2	AAGTTTCTCACTTCGGTGAGCTT	TGCATTAGGGCGGCATACAA
NFAT5	TACCAGGTTTGTATCTCATGCTAC	TGGAAGTCACTATGTGGGCAAT
PSMC6	ATGGTTTTACCTAGCATGGAAGTCT	CAATACCAACCTGGTGGTCCAT
STK38	GGTCTAGGCTCTCACGGCTA	GCTTGAGATGTGCTGAAAGGC
TUBGCP3	GGGGAATACGTTTGTGGGTTG	CAGTGCAACGAACATCACCC

Supplemental table 5: Primer used for validation of underlying SNPs of identified SNP neoepitopes by Sanger sequencing; related to Fig S4 and Table 2.

target	forward primer	reverse primer
CHMP7	GGTTGCCTTTGCCTTTCCAG	TGGCCCCTTTCTGTACCTCT
RGP1	TTGCCGTGCTAGTCTTGTTCA	TGACTGACTGACCCCGAAAG
PCMT1	GTTTTTCTTTCAGTGGGGGATGG	TCCCAATTCTACTGTTGTCCTAGT
RBB7	AGAGCAGTTAGTTGTCCGTGT	GGAAGCCACTGAACGGTAAGA
NR1D1	GGTGGCGTCATCACCTACAT	GCCACTTGTAGACTCCCAGG

Supplemental table 6: Primers used for mutation analysis of gDNA isolated from MSI CRC cell lines and patients by gel capillary electrophoresis; related to Table S1.

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

Labeled primers are indicated.

Supplemental table 7: Primer used for qPCR analysis of endogenous NMD targets and frameshift-bearing transcripts; related to Fig 4/S1.

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

Transparent Methods

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 DKFZ have been approved by FELASA and accredited. For the experiments, mice were assigned to age-matched and sex-matched groups.

Human Tissues

Human tissues 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-583/2016). For all tissue samples, MSI status has been determined previously based on the National Cancer Institute/ICGHNPCC reference marker panel (Boland et al., 1998) and CAT25 as an additional mononucleotide marker (Findeisen et al., 2005). MSI is defined by instability in at least 30% of tested markers.

Cell culture

All human CRC cell lines have been described previously (Michalak et al., 2020; Woerner et al., 2001; Woerner et al., 2007). HCT-116 cells (ATCC® CCL-247™) cells were maintained in RPMI1640 medium supplemented with 10% FCS and 1% P/S. HLA-I types 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 (Scholtalbers et al., 2015) and confirmed by sequencing in the DKMS Life Science Lab GmbH (Dresden, Germany). HB95 cells (ATCC® HB-95™) were maintained in CELLine bioreactor flasks in RPMI1640 medium supplemented with 10% FCS and 1% P/S. All cell lines were tested negative for mycoplasma contamination.

NMD efficiency assay

NMD efficiency of HCT-116 cells was assessed in triplicates using a previously published dual-luciferase reporter system consisting of the human hemoglobin subunit beta (*HBB*) gene with or without an NMD-triggering nonsense mutation fused in-frame to the 3' end of the *Renilla* luciferase gene (Boelz et al., 2006). *Renilla-HBB* wild type (WT) or nonsense (NS39) reporter constructs were transiently transfected into HCT-116 cells using JetPRIME reagent (Polyplus transfection®). Cells were lysed with Passive Lysis Buffer (Promega) and dual-luciferase assay was performed using the Dual-Luciferase Reporter Assay System (Promega). *Renilla* luciferase signals were normalized to co-expressed firefly luciferase signals. The ratio between the normalized signals obtained from the *Renilla-HBB* NS39 reporter and the WT reporter were used to determine the NMD efficiency.

Antibody purification and coupling

Antibody purification and coupling were performed as described previously (Bassani-Sternberg, 2018). Briefly, columns were prepared by washing of empty Poly-Prep® Chromatography Columns (9 ml, Bio-Rad Laboratories) with 1 column volume (CV) of 1% SDS and 4 CV of ddH$_2$O. Next, 2 ml of Sepharose-Protein A conjugate 4B beads

(Invitrogen) were added and washed with 1 CV of 100 mM Tris-HCl (pH 8.0). W6/32 antibody was purified from HB95 cell culture supernatant clarified by centrifugation. 1 CV of cell culture supernatant was loaded to prepared Sepharose-Protein A columns and incubated for 10 min. Columns were washed with 1 CV of 100 mM Tris-HCl (pH 8.0) and 20 mM Tris-HCl (pH 8.0). 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) for neutralization. Antibody concentration of eluates was measured using NanoDrop 2000. For antibody coupling, 10 mg of purified W6/32 antibody was added to prepared Sepharose-Protein A columns and incubated for 30 min at room temperature (RT) with rotation. Next, beads were washed with 1 CV of 0.2 M sodium borate buffer (pH 9.0). For chemical crosslinking, beads were resuspended in 2 ml of 0.2 M sodium borate buffer (pH 9.0) containing 20 mM dimethyl pimelimidate (Thermo Fisher Scientific) and incubated for 30 min at RT with rotation. Next, columns were washed with 0.5 CV of 0.2 M ethanolamine (pH 8.0) and then incubated in 0.5 CV of 0.2 M ethanolamine (pH 8.0) for 120 min at RT with rotation. Prepared Sepharose-Protein A-W6/32 beads washed with 1 CV of PBS with 0.02% NaN_3 and stored in 2 ml of PBS with 0.02% NaN_3 until further usage.

Treatment conditions and immunoprecipitation sample preparation

For each of the 3 biological replicates 7.3×10^6 HCT-116 cells were seeded per 150 mm dish in RPMI1640 medium supplemented with 10% FCS and 1% penicillin/streptomycin. After 48 h, cells were treated either with a final concentration of 5 µM 5AZA (Sigma-Aldrich) or with DMSO (negative control; MACS Milteny Biotec) for 24 h. Cells were harvested by scraping in cold PBS and aliquots of 1×10^8 cells were stored as snap-frozen

dry pellets at 20 °C until usage in immunoprecipitation experiments. Furthermore, samples for treatment validation by quantitative real-time PCR were collected and stored as snap-frozen dry pellets until usage.

High-throughput purification buffers

Immunoprecipitation lysis buffer was described previously and contained 0.25% sodium deoxycholate (Sigma-Aldrich), 0.2 M iodoacetamide (Sigma-Aldrich), 1 mM EDTA (Carl Roth), 1 mM PMSF (Sigma-Aldrich), 1% octyl-β-D-glucopyranoside (Sigma-Aldrich) in PBS (Sigma-Aldrich). 1 cOmplete EDTA-free Protease Inhibitor Cocktail tablet (Roche) was added per 50 ml of IP lysis buffer. Wash buffer 1 contained 150 mM sodium chloride in 20 mM Tris-hydrochloride (pH 8.0). Wash buffer 2 contained 400 mM sodium chloride in 20 mM Tris-hydrochloride (pH 8.0). Wash buffer 3 contained 20 mM Tris hydrochloride (pH 8.0).

High-throughput purification of HLA class I-peptides

Immunoprecipitation of HLA class I:peptide complexes and separation of HLAp was performed as previously described (Chong et al., 2018) omitting pre-clear and HLA class II plates. Snap-frozen dry pellets were lysed immediately before immunoprecipitation in 1 ml of cold immunoprecipitation lysis buffer per 1×10^8 cells for 1 h on ice. After 30 min, thawed pellets were gently resuspended. Lysates were cleared by centrifugation with 21,130 x g at 4 °C for 30 min. If not indicated otherwise, steps were performed using a Positive Pressure-96 Processor (Waters) with 3-5 psi. Briefly, empty 96-well filter plates with 3 µm glass fiber and 10 µm polypropylene membranes (Agilent) were washed with 1

ml/well 100% acetonitrile (Fisher Scientific) and 1 ml/well 0.1% trifluoroacetic acid and then equilibrated with 2 ml/well of 0.1 M Tris-hydrochloride (pH 8.0). Next, Protein A-Sepharose-W6/32 beads were added to a final bed volume of 75 µl/well in 1 ml of 0.1 M Tris-hydrochloride (pH 8.0) and conditioned with 400 µl/well of IP lysis buffer. Supernatants from three cell pellets per sample ($3x10^8$ cells in total) were pooled, distributed equally in three wells ("technical IP replicates") and allowed to flow through by gravity at 4 °C. All subsequent steps were performed at room temperature. Plates were washed 8X with 1 ml/well of wash buffer 1, 8X with 1 ml/well of wash buffer 2, again 8X with 1 ml/well of wash buffer 1 and finally 4X with 1 ml/well of wash buffer 3. Sep-Pak tC18 plates with 100 mg sorbent/well (Waters) were conditioned with 1 ml/well of 80% acetonitrile (Fisher Scientific) in 0.1% trifluoroacetic acid (Fisher Scientific) and 2 ml/well of 0.1% trifluoroacetic acid and used to separate HLA class I peptides from HLA class I heavy chains and β2m molecules. For this purpose, the filter plate was stacked on top of the conditioned Sep-Pak tC18 plate and HLA class I complexes were eluted with 2X 500 µl/well of 1% trifluoroacetic acid at 1-1.5 psi. Next, the Sep-Pak tC18 plate was washed with 2X 1 ml/well of 0.1% trifluoroacetic acid. HLA class I peptides were eluted with 2X 400 µl/well of 28% acetonitrile in 0.1% trifluoroacetic acid in a 2 ml collection plate (Waters) at 1-1.5 psi, eluted peptides from technical IP replicates were pooled, dried by vacuum centrifugation and stored at -80 °C until analysis. HLA class I heavy chains and β2m molecules were eluted with 2X 300 µl/well of 80% acetonitrile in 0.1% trifluoroacetic acid in a 2 ml collection plate (Waters), dried by vacuum centrifugation and used for SDS-PAGE and western blot analysis.

LC-MS/MS analysis

For LC-MS/MS analysis pooled samples were resuspended in 30 µl of 0.1% formic acid and 4.5 µl were used per injection. Lyophilized synthetic peptides for validation of neoepitopes were purchased from JPT Peptide Technologies (Berlin, Germany) and diluted to a concentration of 100 fmol/ul and 3 µl were used per injection. The mass spectrometric analysis was conducted using an UltiMate™ 3000 RSLCnano system (Thermo Fisher Scientific) directly coupled to an Orbitrap Fusion Lumos (Thermo Fisher Scientific). Peptides were loaded onto the trapping cartridge (µ-Precolumn C18 PepMap 100, 5µm, 300 µm i.d. x 5 mm, 100 Å) for 3 min at 30 µL/min (0.05% TFA in water). Peptides were eluted and separated on an analytical column (nanoEase MZ HSS T3 column, 100 Å, 1.8 µm, 75 µm x 250 mm) with a constant flow of 0.3 µL/min using solvent A (0.1% formic acid in LC-MS grade water) and solvent B (0.1% formic acid in LC-MS grade acetonitrile). Total analysis time for the HCD method was 90 min 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 re-equilibration step to initial conditions. The LC system was coupled online to the mass spectrometer using a Nanospray-Flex ion source (Thermo Fisher Scientific) and a Pico-Tip Emitter 360 µm OD x 20 µm ID; 10 µm tip (New Objective). The MS 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 – 1650 *m/z* at a resolution of 120,000 (RF Lens 30%, AGC target 4e5 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. The quadrupole isolation window

was set to 1.2 m/z and peptides were fragmented via HCD (30%). Fragment mass spectra were recorded 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 between MS scans. Dynamic exclusion was set to 20 s. Additionally, samples were analyzed using two HCD/EThcD decision tree methods. Here, the instrument fragmented precursors with a charge state of +1 using the parameters of the HCD method. Charge states 2 – 7 were fragmented using ETD (Calibrated Charge-Dependent ETD Parameters) with supplemental activation enabled (HCD, 30%). For MS/MS spectra acquisition, either high abundant (EThcD) or low abundant (lowEThcD) precursors were selected first. AGC target was set to $2x10^5$ ions and a maximum injection time of 200 ms was allowed and the resulting MS/MS spectra were recorded in the Orbitrap with a resolution of 30,000. Total analysis time for the HCD/EThcD decision tree methods was 180 min with a gradient containing 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 and re-equilibration step to initial conditions. To further validate the presence of the candidates and to obtain reliable quantification data, InDel neoepitope candidates were measured from the same samples using a targeted MS2 method with the previously described settings. Precursor masses of targeted InDel neoepitope candidates are listed in Supplementary table 2 and were fragmented using both HCD and EThcD.

Generation of neoepitope databases

Frameshift peptide databases in FASTA format were constructed based on publicly available sequencing information from CCLE and COSMIC for HCT-116 using in-house

developed R scripts. Briefly, mutation information was obtained from public databases and used for *in silico* mutation of associated cDNA obtained from ENSEMBL using the biomaRt package (Durinck et al., 2009) and translated into protein sequences using the Biostrings package. Frameshifted protein sequences as well as 14 aa of the wild-type protein were included in the FASTA file. Published frameshift sequences originating from recurrent InDel mutations in MSI CRC were included in a separate database (Ballhausen et al., 2020). Previously published information on potential neoepitopes resulting from single nucleotide polymorphisms were included in a separate database (Scholtalbers *et al.*, 2015).

MS data analysis and identification of HLAp

Mass spectrometry raw data were analyzed using PEAKS Studio X (version 10.0, Bioinformatics Solutions). Raw files were subjected to the default data refinement before *de novo* sequencing and database search. The parent mass error tolerance was set to 10.0 ppm while the fragment mass error tolerance was set to 0.02 Da. Fragmentation mode was set either to "HCD" or "Mixed" (for EThcD and lowEThcD measurements with HCD fragmentation for single positive precursors). All raw files were first searched against the UniProt/SwissProt database (20659 entries, February 2019) and a database containing standard contaminants with oxidation of methionine (15.99 Da), carbamidomethylation of cysteine (57.02 Da), and acetylation of N-termini (42.01) 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 frameshift peptide databases based on CCLE (747 entries), COSMIC (1071 entries) and the SNP neoepitope datasets (1526 entries). All peptides identified at a peptide spectrum match FDR of 1% were exported and contaminants, as well as peptides shorter than 8 aa and longer than 15 aa, were filtered out. Unique peptides matching to entries of the UniProt/SwissProt database were reported as "endogenous", while peptides matching one of the frameshift peptide databases were reported as "InDel neoepitopes" and peptides matching the SNP neoepitope database were reported as "SNP neoepitopes".

In silico methods for validation of HLAp

Binding prediction to the expressed HLA allotypes was performed using NetMHCpan (version 4.0a)(Jurtz et al., 2017). The rank threshold for binders was set to <0.5% for strong binders (SB) and <2% for weak binders (WB). For peptides binding to more than one HLA allotype, only the best-ranked HLA allotype with its corresponding affinity in nM was reported. Peptide sequences were clustered using GibbsCluster (version 2.0)(Andreatta et al., 2017) with default parameters for MHC class I ligands and motifs were visualized with Seq2Logo (version 2.0)(Thomsen and Nielsen, 2012). Hydrophobicity indices of identified peptides were calculated using SSRcalc version Q (Krokhin, 2006) using the following parameters: 100Å C18 column, 0.1% Formic Acid (2015 model), no cysteine protection. Spectra recorded from synthetic peptides were compared to spectra measured from the samples. Briefly, all spectra recorded for a given peptide were compared to all spectra recorded from its synthetic counterpart. Raw data for matched fragment ions was exported from PEAKS Studio X (PSM-ions.txt),

normalized intensities of matched fragment ions were correlated and the match with the highest correlation was reported graphically. Moreover, retention time differences between the two peptides were calculated and reported.

Mutation analysis

Underlying frameshift mutations of InDel neoepitopes and single nucleotide polymorphisms of SNP neoepitopes were verified by Sanger sequencing of PCR amplified fragments (Eurofins Genomics Germany GmbH, Köln) and analyzed using the Indigo webtool (Rausch et al., 2020). Mutation analysis of MSI CRC cell lines and patient samples was performed using fluorescently labeled primers for amplification. Fragments were visualized on an ABI3130xl (Applied Biosystems) genetic analyzer as described previously (Findeisen *et al.*, 2005). Primers are reported in Supplementary tables 3 – 5.

siRNA-mediated knockdown of UPF1

siRNA-mediated knockdown was performed with the previously reported siRNA targeting *UPF1* (AA-GAUGCAGUUCCGCUCCAUU-TT, Gehring et al. (2003)) using INTERFERin (Polyplus transfection®) for 48 h.

Quantitative real-time PCR

RNA was isolated using TriReagent (Sigma). 2 µg of RNA were reverse transcribed using the Revert-Aid™ H Minus Reverse Transcriptase Kit (Thermo Scientific) according to the manufacturer's protocol. qPCR was performed in technical triplicates on a StepOnePlus™ system (Applied Biosystems) using primaQuant CYBR qPCR Master Mix

(Steinbrenner Laborsysteme). Primers for *ATF3* (NMD-sensitive), *ATF4* (NMD-sensitive), *SC35A* (NMD-insensitive control), *SC35C* (NMD-sensitive), *SC35D* (NMD-sensitive), *UPP1* (NMD-sensitive), and the house-keeping gene *HPRT1* were reported elsewhere (Bhuvanagiri *et al.*, 2014; Cheruiyot *et al.*, 2018; Viegas *et al.*, 2007). qPCR primers for neoepitope candidates are reported in Supplementary table 6.

Immunization and ex vivo IFNγ ELISpot assay

HLA-A2.1/HLA-DR1-transgenic H-2 class I-/class II-knockout mice (Pajot *et al.*, 2004) were immunized weekly for three weeks with either a peptide pool consisting of 50 µg of *CKAP2*, *NFAT5*, *PSMC6*, and *TUBGCP3* peptide each or 100 µg of *CKAP2* or *TUBGCP3* peptides separately purchased from JPT Peptide Technologies (Berlin, Germany) or synthesized by the GMP & T Cell Therapy Unit at German Cancer Research Center (DKFZ; Heidelberg, Germany) and 20 µg CpG ODN 1826 (TIB MolBiol) suspended in 20 µl PBS. IFNγ ELISpot was performed *ex vivo* seven days after the last immunization with isolated splenocytes or CD8$^+$ T cells as described previously (Ballhausen *et al.*, 2020). CD8$^+$ T cells were isolated with the CD8a$^+$ T Cell Isolation Kit (Miltenyi Biotec) following the manufacturer's instruction using LS columns (Miltenyi Biotec). For ELISpot assay, MultiScreenHTS-IP plates (Merck) were activated with 70 µl of 70% ethanol per well for 5 min and washed five times with sterile PBS. 100 µl purified rat anti-mouse IFNγ antibody (1:200 in PBS; BD Biosciences) was added and incubated overnight at 4 °C. Plates were washed four times with PBS and blocked with 200 µl RPMI Medium 1640 supplemented with 10% FBS, 100 U/ml penicillin, and 0.1 mg/ml 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 1×10^6 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 4×10^5 splenocytes from naïve mice as antigen-presenting cells with 2 µg per well of the corresponding peptide and incubated for 4 h at 37 °C. Next, 1×10^5 CD8$^+$ T cells per 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 plus 0.01% tween, once with PBS and coated with 100 µl biotinylated rat anti-mouse IFNγ antibody (1:500 in PBS; BD Biosciences) for 1 h at room temperature. Next, plates were washed six times with PBS plus 0.01% tween and once with PBS. 100 µl AKP Streptavidin (1:500 in PBS; BD Biosciences) were added and incubated for 30 min at room temperature in the dark followed by three washing steps with PBS plus 0.01% tween and three times washing with PBS. BCTP/NBT substrate (100µl/ well; Thermo Fisher Scientific) was applied and incubated up to 45 min depending on the color development. The reaction was stopped with distilled water. Plates were dried overnight and analyzed using the CTL ImmunoSpot Reader. Concanavalin A (2 µg/well; Sigma) was used as an assay high control for IFNγ production in all performed experiments.

Label-free quantification of HLAp

Quantification of HLA-presented peptides was performed using the raw output (peptides.csv) from PEAKS and a custom script in the R programming language (ISBN 3-900051-07-0). First, measured intensities from InDel neoepitope candidates using parallel reaction monitoring were combined with measured intensities from endogenous peptides, and data were filtered to include only peptides which were measured in at least

two out of three replicates per condition and MS method used. Subsequent steps were performed for each MS method separately. 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 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 was performed for hits with the PANTHER classification system (Mi et al., 2019) 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. Results were visualized using the ggplot2 package (ISBN 978-3-319-24277-4).

Statistical analysis

Luciferase assay and qPCR data are presented as mean ± SD 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 using a two-sided, unpaired t-test with correction for multiple hypothesis testing. $p \leq 0.05$ was considered significant.

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-HLA-ABC antibody (W6/32)	Barnstable et al. (1978); Parham et al. (1979)	N/A
Rabbit monoclonal anti-β-2-microglobulin (EP2978Y)	abcam	Cat# ab75853; RRID:AB_1523204
Mouse monoclonal anti-α-Tubulin	Sigma-Aldrich	Cat# T5168; RRID:AB_477579
Rabbit-anti-mouse IgG (HRP Conjugate)	Sigma-Aldrich	Cat# A9044; RRID:AB_258431
Goat-anti-rabbit IgG (HRP Conjugate)	Sigma-Aldrich	Cat# A0545; RRID:AB_257896
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
Bacterial and Virus Strains		
Biological Samples		
Human tissues (tumor DNA)	Tissue bank of the German Collaborative Group on HNPCC (Heidelberg, Germany)	N/A
Colorectal cancer cell lines (genomic DNA)	Michalak *et al.* (2020); Woerner *et al.* (2001); Woerner *et al.* (2007)	N/A
Chemicals, Peptides, and Recombinant Proteins		
Dimethyl pimelimidate	Thermo Fisher Scientific Pierce	Cat# 21666
5-Azacytidine (5AZA)	Sigma-Aldrich	Cat# A2385
synthetic peptides for validation and *in vivo* immunization; see Supplementary table 2	JPT Peptide Technologies (Berlin, Germany) and GMP & T Cell Therapy Unit at German Cancer Research Center (Heidelberg, Germany)	N/A
Critical Commercial Assays		
Dual-Luciferase Reporter Assay System	Promega	Cat# E1910
primaQuant CYBR qPCR Master Mix	Steinbrenner Laborsysteme	Cat# SL-9902
CD8a+ T Cell Isolation Kit	Miltenyi Biotec	Cat# 130-104-075
1-Step BCIP/NBT Substrate Solution	Thermo Fisher Scientific	Cat# 34042
Deposited Data		
Mass spectrometry raw data and search results	This paper	PRIDE: PXD021755

ReFrame database of recurrent InDel mutations in MSI CRC	Ballhausen *et al.* (2020); https://github.com/atb-data/neoantigen-landscape-msi	N/A
TRON Cell Line Portal (database of neoepitopes originating from single nucleotide polymorphisms)	Scholtalbers *et al.* (2015); http://celllines.tron-mainz.de/	N/A
UniProt/SwissProt database	https://www.uniprot.org/	RRID:SCR_002380
Experimental Models: Cell Lines		
Homo sapiens: HCT-116 cells	Department of Applied Tumor Biology, Heidelberg University (Heidelberg, Germany)	DSMZ Cat# ACC 581; RRID:CVCL_0291
Mus musculus: HB95 hybridoma cells	Department of Oncology, Ludwig Institute for Cancer Research (Lausanne, Switzerland)	ATCC Cat# HB-95; RRID:CVCL_7872
Experimental Models: Organisms/Strains		
HLA-A2.1/HLA-DR1-transgenic H-2 class I-/class II-knockout mice	Institute Pasteur (Paris, France); Pajot *et al.* (2004)	N/A
Oligonucleotides		
Primers for mutation analysis by Sanger sequencing; see Supplementary table 4 & 5	This paper	N/A
Fluorescently labeled primers for mutation analysis by capillary gel electrophoresis; see Supplementary table 6	This paper	N/A
Primers for qPCR; see Supplementary table 7	This paper; Bhuvanagiri *et al.* (2014); Cheruiyot *et al.* (2018); Viegas *et al.* (2007)	N/A
siRNA targeting UPF1 (AA-GAUGCAGUUCCGCUCCAUU-TT)	Gehring *et al.* (2003)	N/A
CpG ODN 1826	TIB MolBiol	N/A
Recombinant DNA		
Renilla-HBB WT/*Renilla-HBB* NS39 plasmids, firefly plasmid (NMD efficiency assay)	Boelz *et al.* (2006)	N/A
Software and Algorithms		
RStudio (version 1.3.1093)	https://rstudio.com/	RRID:SCR_000432
biomaRt	Durinck *et al.* (2009)	RRID:SCR_002987
Biostrings	https://bioconductor.org/packages/release/bioc/html/Biostrings.html	RRID:SCR_016949
PEAKS Studio X (version 10.0)	Bioinformatics Solutions	N/A
NetMHCpan (version 4.0a)	Jurtz *et al.* (2017)	RRID:SCR_018182
GibbsCluster (version 2.0)	Andreatta *et al.* (2017)	N/A

Seq2Logo (version 2.0)	Thomsen and Nielsen (2012)	N/A
SSRcalc (version Q)	Krokhin (2006)	N/A
Indigo webtool for rapid InDel discovery in Sanger chromatograms	(Rausch *et al.*, 2020); https://www.gear-genomics.com/	N/A
limma	Ritchie *et al.* (2015)	RRID: SCR_010943
vsn	Huber *et al.* (2002)	RRID:SCR_001459
Msnbase	Gatto and Lilley (2012)	N/A
Panther GO-term analysis	Mi *et al.* (2019)	N/A
ggplot2	https://cran.r-project.org/web/packages/ggplot2/index.html	RRID:SCR_014601
Other		
jetPRIME (DNA transfection)	Polyplus transfection	Cat# 114-15
INTERFERin (siRNA transfection)	Polyplus transfection	Cat# 409-10
Passive Lysis 5X Buffer	Promega	Cat# E1941
Poly-Prep Chromatography Columns (9 ml)	Bio-Rad Laboratories	Cat# 7311550
Sepharose-Protein A conjugate 4B beads	Invitrogen	Cat# 101142
Positive Pressure-96 Processor	Waters	Cat# 186006961
96-well filter plate with 3 µm glass fiber and 10 µm polypropylene	Agilent	Cat# 201017-100
Sep-Pak tC18 plate 100 mg sorbent	Waters	Cat# 186002321
LS Columns	Miltenyi Biotec	Cat# 130-042-401
MultiScreenHTS-IP plates for ELISpot	Merck	Cat# MSIPN4W50
AKP Strepavidin	BD Biosciences	Cat# 554065
Concanavalin A from *Canavalia ensiformis*	Sigma-Aldrich	Cat# C5275

Supplemental References

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

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. Nature Communications *11*, 4740. 10.1038/s41467-020-18514-5.

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.

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.

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.

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.

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.

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.

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. Molecular & Cellular Proteomics *17*, 533-548. 10.1074/mcp.TIR117.000383.

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.

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.

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., 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.

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.

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.

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. Analytical Chemistry *78*, 7785-7795. 10.1021/ac060777w.

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.

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.

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.

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.

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.

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.

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.

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.

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.