Immunochemical approaches to monitor and modulate the adaptive immune system

Luimstra, J.J.

7

Screening for neoantigen-specific CD8+ T cells using thermally-exchanged pMHCI multimers

Jolien J. Luimstra[1], Christina Heeke[2], Jitske van den Bulk[3], Brett Hos[4], Ferry Ossendorp[4], Noel F.C.C. de Miranda[3], Sine R. Hadrup[2], Jacques Neefjes[1], and Huib Ovaa[1]

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, The Netherlands
[2]Section for Immunology and Vaccinology, National Veterinary Institute, Technical University of Denmark, Copenhagen, Denmark
[3]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands
[4]Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands

Abstract

A role for T cell immunity in the clearance of tumor cells has well been established and is effectively put to use in therapies that selectively boost anti-tumor responses. It is proposed that the efficacy of checkpoint blockade immunotherapy largely relies on T cell reactivity to neoantigens: peptides derived from mutated (onco)proteins that are presented on major histocompatibility complex class I. Rapid detection of neoantigen-reactive T cells can support the development of specific immunotherapies as well as monitor response to therapy. In this study we aim to identify responses directed at HLA-A*02:01- or H-2K^{b}-binding neoantigens predicted from human colorectal cancer patients and from the murine colorectal cancer model MC38, respectively. We generated panels of DNA-barcoded major histocompatibility class I complexes loaded with predicted neoantigens using thermal MHC exchange technology and used them to identify T cell specificities from human and murine cell samples. We provide first evidence of the feasibility of the technique by confirming previously detected viral responses from healthy donors in two separate experiments. However, identification of true mouse or human neoantigens has proven challenging. Low cell counts and presumably the low neoantigen frequencies were bottlenecks in these first tests. Future work will focus on resolving the technical difficulties and on increasing the sensitivity of the technology.

Introduction

Oncogenesis is accompanied by the occurrence of somatic mutations in cancer cells. DNA nucleotide substitutions as well as insertions or deletions at protein-coding regions can result in the expression of mutated antigens[2]. Presentation of these so-called neoantigens on major histocompatibility complexes (MHCs) on the cell surface flags tumor cells for detection by neoantigen-specific T cells, potentially resulting in tumor clearance[3]. Due to their somatic origin, later in life, no central tolerance has been raised specifically against neoantigens supporting their potential as immunotherapeutic targets[4]. Unlike traditional tumor-associated antigens (TAAs) originating from overexpressed self-proteins, neoantigens are solely expressed on tumors and therefore neoantigen-based vaccines are expected to less frequently induce autoimmunity[5]. For this reason neoantigens are hot targets in the development of cancer therapeutics[6,7]. Vaccination with neoantigens through various delivery modes has resulted in therapeutic benefit in a number of preclinical and clinical studies[8-11].

Discovery of neoantigen-directed T cell reactivity in cancer patients is of high interest as it can support immunotherapeutic approaches. Unfortunately,

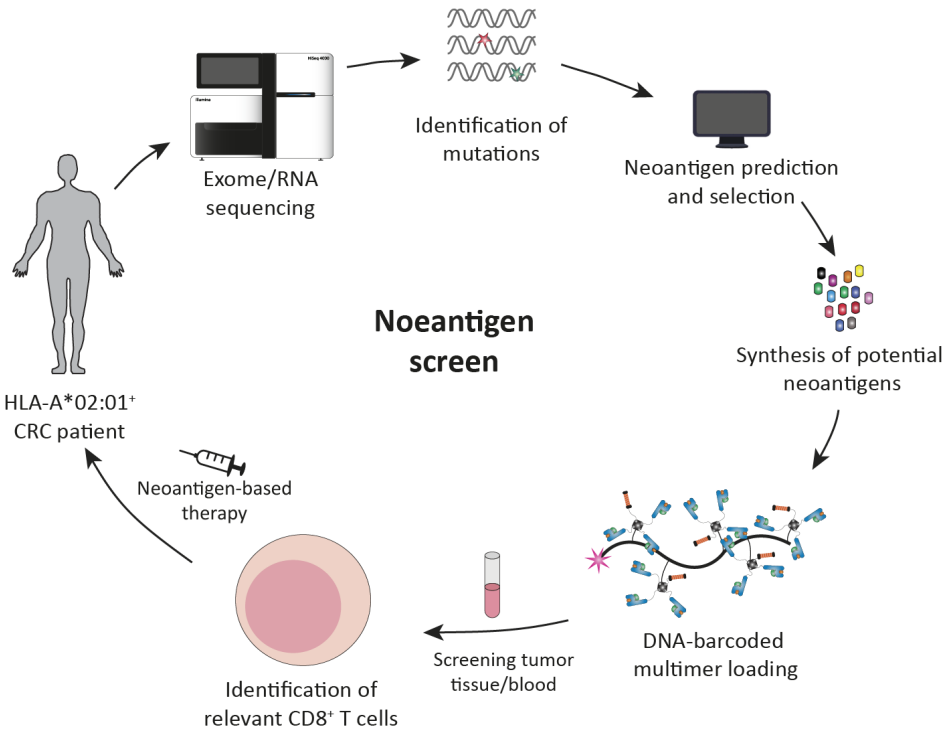

Figure 1. **Workflow of the anticipated neoantigen screen.** RNA and exomes from HLA-A*02:01+ colorectal cancer (CRC) patients are sequenced to identify mutations in expressed proteins. Potential neoantigens are predicted based on their HLA-A*02:01 binding motif. These peptides are then synthesized and loaded on conditional pMHCI multimers through thermal exchange. Consequently, these multimers can be used for screening of neoantigen-specific CD8+ T cells from patient material.

the detection of neoantigen-reactive T cells can prove challenging at several levels. Since neoantigens arise from patient-specific mutations they have to be identified on an individual basis[12-14]. Furthermore, most cancer mutations occur outside mutation "hotspots" and affect so-called "passenger" genes that are generally not analyzed in targeted, diagnostic procedures. By performing DNA sequencing of healthy and tumor tissues, somatic mutations can be identified and, in combination with RNA sequencing, transcribed putative neoantigens can be selected for a given cancer. The functional detection of neoantigen-reactive T cells often relies on the selection of mutated peptides predicted to bind a patient's MHC class I alleles[15,16]. Major efforts contribute to the development of advanced bioinformatic tools and algorithms, but these often fail to achieve accurate prediction[17-19].

Physical measurements, such as mass-spectrometric analyses of peptides eluted from tumor cells or tumor-infiltrating T cells (TILs) or functional assays to

measure cytokine secretion, are used to validate predicted neoantigens, but they require large numbers of cells and often lack sensitivity[20]. Another strategy is to characterize neoantigen-specific CD8+ T cells using pMHCI multimers. By using one fluorescent label per peptide, several multimers can be used simultaneously to identify specific T cells from a larger population using flow cytometry[21]. Conventional preparation of multimers required separate folding with each peptide for every specificity, but the development of several exchange techniques allows for the generation of multiple specificities in parallel[22-25]. We have recently described a novel exchange technology based on an increase in temperature, which is superior to preceding techniques in its potential for exchange on MHC

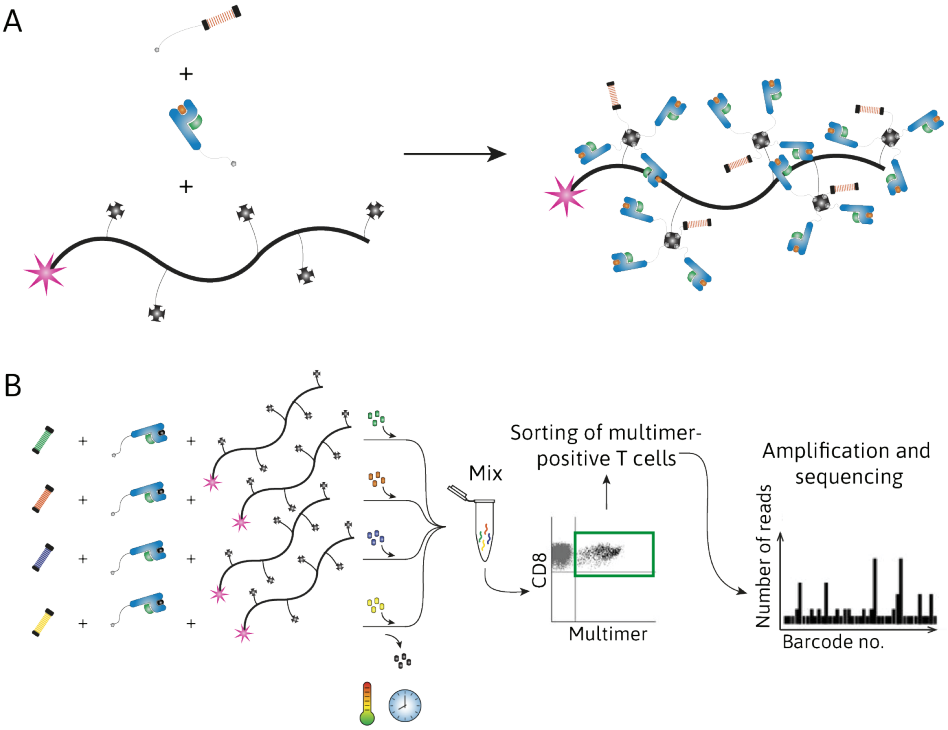

Figure 2. **Overview of generation and use of temperature-exchangeable DNA-barcoded MHCI multimers.** (A) DNA barcodes and MHCI monomers, both biotinylated, are added to fluorescently-labelled streptavidin-conjugated dextran backbones to form temperature-exchangeable DNA-barcoded MHCI multimers. (B) Dextran backbones and MHCI monomers with exchangeable peptide are combined with a specific DNA barcode per well. Each multimers is then loaded with a desired peptide and incubated at set exchange conditions. After exchange is complete, multimers are pooled, concentrated and added to a cellular suspension, from which multimer-positive CD8+ T cells are sorted on their fluorescent label. The DNA barcodes are then amplified using PCR and sequenced to identify the specific antigen-responsive T cells in the sample. (Picture adapted from Bentzen et al.[1]).

multimers, reducing pre-staining handling time even further[26].

In flow cytometry the number of parameters that can be measured simultaneously is limited by the number of fluorophores that can be detected simultaneously. Combinatorial coding has greatly increased the number of combinatorial parameters up to 63, but for large screens this number is insufficient[27,28]. Bentzen et al. devised a strategy to overcome this restriction by labelling with DNA barcodes[1]. Each multimer consists of a PE-labelled dextran backbone that accommodates multiple streptavidin moieties for conjugation of biotinylated MHC monomers and 25-oligonucleotide barcodes. Using this technology over 1000 T cell specificities can be detected from one sample. Labelled T cells are sorted using FACS based on their PE label, followed by amplification of the DNA barcodes using PCR and subsequent analysis with next-generation sequencing (NGS).

In this study, we combined our thermal exchange technology with DNA barcoding to screen for neoantigens in a colorectal cancer (CRC) mouse model (MC38) and in human HLA-A*02:01+ patients (Fig. 1). A group of CRC patients exhibit high microsatellite instability, giving rise to a high frequency of mutations and consequently this may result in a high number of neoantigens and increased responsiveness to immunotherapy[29]. We aim to identify bona fide neoantigens displayed by these patients in order to unravel therapeutic targets for immunotherapy.

Results

HLA-A*02:01 proof-of-principle

As a proof-of-principle we used thermally-exchanged multimers to detect virus-specific CD8+ T cells in buffy coats from three healthy HLA-A*02:01+ donors. We selected eight common virus epitopes (see Table 1) originating from influenza A virus (IAV), Epstein-Barr virus (EBV), cytomegalovirus (CMV) or HIV, for which specific T cells were previously detected in one or more of the three buffy coats used in this experiment. Because the signal-to-noise ratio would be low when staining with only eight different DNA-barcoded pMHCI multimers, the selection of peptides was included in a larger panel consisting of 48 melanoma antigens (data not shown). This total of 56 peptides were loaded on DNA-barcoded HLA-A*02:01 multimers using thermal peptide exchange as depicted in Figure 2, A and B, and described in the Materials and Methods section. Next, these were used for staining of the buffy coats obtained from the HLA-A*02:01+ volunteers. CD8+multimer+ T cells were isolated using FACS and DNA barcodes in this population were identified by next-generation sequencing (NGS). In all three buffy coats, viral responses were detected in earlier studies. Using DNA-barcoded

Table 1. **Viral epitopes used for thermal exchange and MHCI multimer staining of CD8+ T cells in buffy coats from healthy volunteers.**

#	Sequence	Origin	BC83	BC104	BC112
V1	GILGFVFTL	IAV MP1	0.00%	**0.06%**	0.01%
V2	CLGGLLTMV	EBV LMP2	0.00%	**0.04%**	0.01%
V3	GLCTLVAML	EBV BMLF1	**0.81%**	**0.17%**	**1.64%**
V4	FLYALALLL	EBV LMP2	**0.09%**	**0.04%**	**0.02%**
V5	NLVPMVATV	CMV pp65	0.01%	0.00%	**7.48%**
V6	YVLDHLIVV	EBV BRLF1	**0.88%**	**0.11%**	**0.21%**
V7	VLEETSVML	CMV IE1	0.00%	0.00%	0.01%
V8	ILKEPVHGV	HIV Pol	0.00%	0.00%	0.01%

Frequencies indicate estimated percentages of antigen-specific T cells from total CD8+ T cells determined by sequencing of DNA barcodes. Values highlighted in bold face and green indicate specificities previously detected in that patient. BC, buffy coat; CMV, cytomegalovirus; EBV, Epstein-Barr virus; HIV, human immunodeficiency virus; IAV, influenza A virus.

thermally-exchanged HLA-A*02:01 multimers we detected all of them in similar frequencies (Table 1, highlighted in bold and marked in green), demonstrating the experimental feasibility of our approach.

Screening for neoantigens predicted from HLA-A*02:01+ CRC patients

After establishing proof-of-principle we set out to validate neoantigens predicted for five HLA-A*02:01-expressing CRC patients. Cancer exomes and transcriptomes were sequenced and compared to healthy tissue to reveal somatic mutations potentially giving rise to neoantigens. From these potential neoantigens, HLA-A*02:01-binding peptides of high and intermediate affinity were predicted using NetMHC, yielding 6, 13, 17, 136 and 336 sequences for patients P1 to P5, respectively (see Table S3 for sequences). The eight common viral epitopes used in the proof-of-principle (Table 1), were included as an experimental control, as well as a non-exchanged negative control. Multimers loaded with the predicted neoantigens or viral antigens were pooled and used to screen for reactive CD8+ T cells from a number of samples obtained from the tumor, peripheral blood or lymph node, as well as buffy coats from two healthy volunteers. Additional TIL subsets were included for patients P2, P3 and P4, based on a study by Duhen et al.[30] They described a unique subset of CD8+ TILs present in the tumor microenvironment, but not peripheral blood, that express both CD39 (a T cell exhaustion marker often co-expressed with PD-1) and CD103 (a cadherin involved in cytotoxic lysis)[31,32]. They found that this subset of T cells is enriched for tumor-reactive T cells. Therefore we included TILs selected for expression of both (double positive, DP), either (single positive, SP) or none (double negative, DN) of these two markers in an attempt to detect higher neoantigen-specific T cell frequencies in the DP subset.

As expected, barcodes corresponding to viral peptides were retrieved in one of the healthy controls and in some of the patient samples (Fig. 3). The peptides corresponding to the top 10 barcode reads increased compared to baseline are listed in Table S4. Consistent with the proof-of-principle experiment, CD8+ T cells specific for viral peptides V3, V4 and V6 were detected in buffy coat (BC) 83. Viral antigens were also detected in a number of the patient samples, but not in all of the sample types from the same patient. In patient P3, viral responses were

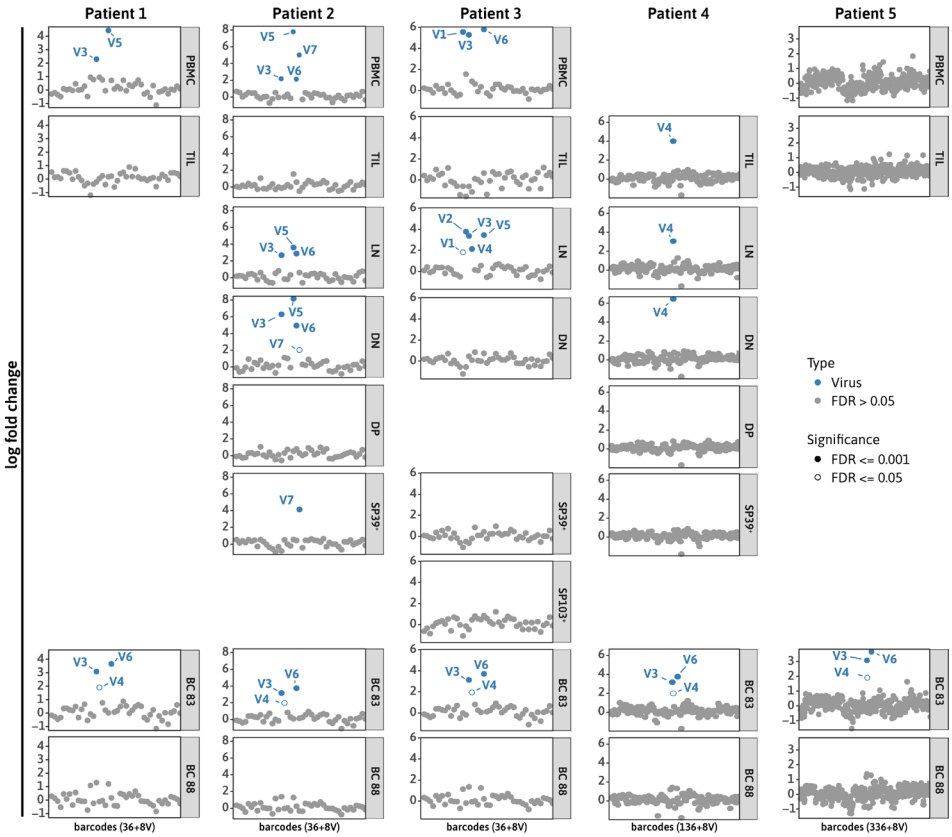

Figure 3. **Barcodes retrieved from the five human colorectal cancer (CRC) patients included in this study.** Patient samples were stained with a pool of HLA-A*02:01 multimers thermally exchanged for a selection of predicted neoantigens and eight common viral antigens (denoted with V, sequences listed in Table 1). Peptides predicted for patients P1, P2 and P3 were combined for a total of 36, while for patients P4 and P5, respectively, 136 and 336 neoantigens were predicted. Barcodes corresponding to viral antigens were detected in volunteer buffycoat 83 and in a number of patient samples. No significant neoantigen -specific responses were detected. FDR, false discovery rate; PBMCs, peripheral blood mononuclear cells; TIL, tumor-infiltrating lymphocyte; LN, lymph node; DN, double negative TIL subset expressing both CD39 and CD103; DP, double positive TIL subset; SP39+, single positive TIL subset expressing CD39; SP103+, single positive TIL subset expressing CD103.

detected in the lymph node and PBMC (peripheral blood mononuclear cell) samples, but not the TILs. TILs are likely more reactive against tumor antigens than against viral antigens, so this is not remarkable. In patient P2, virus-specific T cells were also detected in the double-negative (supposedly less tumor-reactive) TIL subset. Unfortunately, in none of the samples increased numbers of barcodes corresponding to predicted neoantigens were detected. This is not surprising, since neoantigen frequencies are generally low and hence do not give rise to high T cell numbers as viral antigens do. However, one of the peptides included in this set, TLVIYVARL (#268), was previously picked up in an activation assay using PBMCs from patient P4, but not in this screen. This hit will be validated in co-culture assays to determine if the result emanating from the first assay was a true or false positive.

Screening for MC38 neoantigen-specific CD8$^+$ T cells

To screen for neoantigens in the MC38 mouse model, a total of 1020 mutated H-2K^b binders were predicted based on expression profiles in tumor and healthy tissue. Neoantigen-specific responses were analyzed in naïve C57BL/6 mice that were untreated, vaccinated with irradiated MC38 tumor cells, or vaccinated with irradiated MC38 cells in combination with DMXAA (5,6-dimethylxanthenone-4-acetic acid). DMXAA is a murine STING (Stimulator of Interferon Genes) agonist that has demonstrated durable preclinical benefit by activating dendritic cells, anti-tumor CD8$^+$ T cells and inducing interferon (IFN-)β production[33,34]. Vaccination with irradiated tumor cells provides TAAs and danger signals (caused by radiation-induced immunogenic cell death) and consequently should result in priming and expansion of tumor-specific T cells. Mice were sacrificed one week after finishing a scheme of three vaccinations with two-week intervals.

In a first test of the experimental set-up, a small selection of 102 potential neoantigens were screened (see Table S5). These predicted neoantigens and OVA peptide SIINFEKL were loaded on DNA-barcoded H-2K^b multimers through thermal exchange and used to stain splenocytes isolated from vaccinated or untreated C57BL/6 mice. A no-peptide control was included as negative control and as experimental control OT-I T cells were spiked into the non-vaccinated sample (1% of total cells) for detection by SIINFEKL-loaded multimers.

After staining of the murine splenocytes, antigen-specific T cells were isolated using FACS. A clear population of multimer$^+$ CD8$^+$ T cells was visible in the non-vaccinated sample spiked with OT-I cells (Fig. 4, left), which was expected to consist of the SIINFEKL-specific OT-I T cells spiked in to the cell sample. However, even though the barcode corresponding to SIINFEKL (#104) was among the top 10 for the OT-I-spiked sample, only slightly more reads were detected in the sample compared to the baseline. This was the case for most barcodes: in all samples combined only two barcodes were detected above the significance threshold

Table 2. **Top 10 peptide specificities retrieved from the murine cell samples included in this study.**

		Used directly post exchange		
Sort count		204	221	573
	Sample	Non-vaccinated (+OT-I)	Vaccinated (MC38)	Vaccinated (MC38+DMXAA
Top 10 specificities	1	#24	#74	#17
	2	#89	#16	#16
	3	#63	#64	#64
	4	#60	#68	#63
	5	#87	#17	#28
	6	#70	#20	**#21**
	7	**#104**	#65	#24
	8	#12	#26	#78
	9	#72	#66	#20
	10	#74	#70	#15

Conditional H-2K^b multimers were exchanged for 102 predicted neoantigens and OVA peptide SIINFEKL. These were used to stain splenocytes isolated from non-vaccinated mice spiked with OT-I cells, mice vaccinated with irradiated MC38 tumor cells or from mice vaccinated with MC38 cells and DMXAA (5,6-dimethylxanthenone-4-acetic acid). Barcodes corresponding to peptides highlighted in red were detected at significantly increased levels (log fold change ≥2) compared to baseline reads. OVA peptide SIINFEKL is marked green in bold face; previously detected peptides are marked blue in bold face.

(log fold change ≥2). This can be explained by the low cell numbers used for staining, and consequently the low number of sorted cells (Table 2). After thawing splenocyte counts were relatively low, so that only 100,000-500,000 cells per condition could be included, when in fact 1,000,000-2,000,000 are preferred. The lower limit of detection is about 20 copies of a specific CD8+ T cell and this number is challenging to reach with low numbers of PBMCs. Repeating this screen with more cells will likely yield more relevant and more reproducible data.

Discussion

The discovery of immune checkpoints as anti-cancer targets has sparked the field of cancer immunotherapy. Blocking of inhibitory molecules, such as CTLA-4 or the PD-1/PD-L1 interaction, results in reestablishment of pre-existing immune responses[35,36]. Accordingly, responses to checkpoint inhibition are highest in cancers with a high mutational burden where more neoantigens can be generated[37,38]. Vice versa, the efficacy of checkpoint inhibition can be further increased by priming of anti-tumor T cells through neoantigen-based

therapies[39]. The success of these therapies relies on neoantigens and hence it is important to know their identity. Identification of mutations has become relatively straightforward with the development of NGS, but predicting the immunogenicity of predicted neoantigens is less trivial. In this study we set out to validate predicted neoantigens by characterizing neoantigen-specific immune responses using conditional pMHCI multimers. Thermal exchange technology provides an easy method to generate large numbers of specific MHCI multimers in parallel. Using DNA barcode labeling up to 1000 peptides can be tested from a single sample, thus greatly reducing the sample volume required for analysis. Especially in the tumor field this provides huge benefit compared to conventional multimer staining. It has long been established that peptides with a high affinity for their cognate MHCI are not necessarily more immunogenic and hence potent immunogenic peptides may be missed by applying strict selection parameters[40,41]. Combining MHCI exchange technology with DNA barcoding allows broadening of the selection criteria to also include less obvious potential neoantigens that would otherwise not be included in screens. The experiments described here are

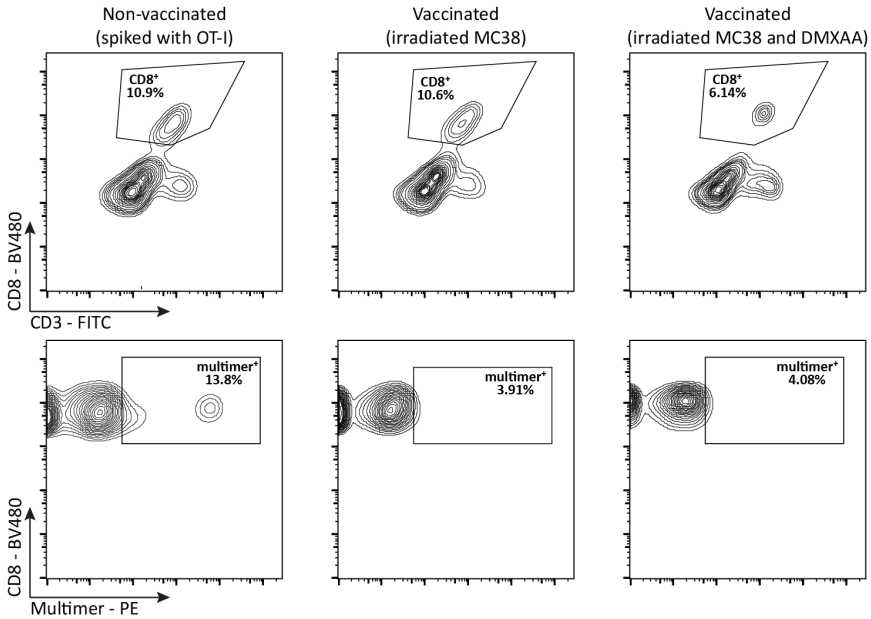

Figure 4. **FACS plots of MC38 splenocytes stained with a pool of 102 thermally-exchanged H-2K^b multimers (used immediately post exchange).** CD3$^+$CD8$^+$multimer$^+$ T cells were sorted based on fluorescence of the PE label conjugated to the multimerization backbone. Left: splenocytes isolated from non-vaccinated mice spiked with OT-I cells (1%). The population on the right likely shows OT-I cells stained with H-2K^b multimer exchanged for OVA peptide SIINFEKL. Center: splenocytes from mice vaccinated with irradiated MC38 tumor cells. Right: cells isolated from mice vaccinated with irradiated tumor cells and DMXAA (5,6-dimethylxanthenone-4-acetic acid).

a first step towards implementing DNA barcoding technology in combination with temperature-based MHC exchange technology to increase the throughput of the technology.

The screen performed with HLA-A*02:01 multimers exchanged for viral and melanoma epitopes served as a proof-of-principle, demonstrating that previously detected T cell responses against viral epitopes could be picked up using DNA-barcoded thermally-exchanged multimers. The fact that those same viral responses were detected in our human CRC patient screen was very promising, but in contrast no neoantigen barcodes were retrieved. Peptide #268 (TLVIYVARL) was shown to activate CD8+ T cells in PBMCs from patient P4, but in the pMHCI multimer screen we did not detect its corresponding barcode. This peptide is predicted to weakly bind HLA-A*02:01 and this may affect pMHCI multimer loading, although the predicted affinity of 235 nM is well beyond that of the template peptide IAKEPVHGV, which is 7,288 nM. Determining the exchange efficiency using HPLC can resolve whether impaired loading accounts for not detecting this peptide.

It is known that only few predicted peptides are bona fide neoantigens and that frequencies of neoantigen-specific T cells are low, hampering detection. Furthermore, in our study only a single MHCI allele was included, whereas each individual expresses up to six distinct HLA alleles. Inclusion of pMHCI multimers with additional patient-matched HLAs will undoubtedly increase the neoantigen discovery rate. Advancing thermal exchange technology will allow screening across the full range of HLA haplotypes expressed by each individual patient.

In an attempt to increase neoantigen-specific T cell frequencies, the mice used in our MC38 screen were vaccinated with irradiated MC38 tumor cells, with and without DMXAA. Despite FACS analysis clearly demonstrating a population in the OT-I spiked sample and the barcode corresponding to SIINFEKL turning up in the top 10 of elevated reads, no significant increase in reads compared to baseline was found. Due to the low cell count no T cell specificities were detected above the threshold and repetition of this screen with more cells will be necessary to demonstrate the potential of DNA-barcoded pMHCI screens for neoantigen discovery.

Materials and methods

Ethical approval

All animal experiments were approved by the animal ethics committee of the LUMC, which has been licensed by the Dutch Central Animal Experiments Committee. Experiments were performed by Federation of European Laboratory Animal Science Associations (FELASA)-accredited animal-handlers and monitored

by the animal welfare body according to the Dutch Act on animal experimentation (ex art. 14a, b, and c) and EU Directive 2010/63/EU ('On the protection of animals used for scientific purposes').

All patient material was collected under approval by the Medical Ethical Committee of the Leiden University Medical Centre (LUMC, protocol P15.282). Patient samples were anonymized and handled according to the medical ethical guidelines described in the Code of Conduct for Proper Secondary Use of Human Tissue of the Dutch Federation of Biomedical Scientific Societies. This research was conducted according to the recommendations outlined in the Helsinki declaration.

Human cell samples

PBMCs from patients were isolated from heparinized venous blood by use of Ficoll-Amidotrizoate (LUMC Pharmacy, Leiden, NL) density centrifugation. Tumor material and respective healthy colorectal and lymph node samples were obtained during surgery, cut into small fragments and digested using gentleMACS C tubes (Miltenyi Biotec), collagenase D (Roche) and DNAse I (Roche). The digested cells were incubated for 30 min at 37°C interrupted by three runs on the gentleMACS Dissociator (Miltenyi Biotec) and subsequently filtered by use of a 0.7 µm mesh filter. The tumor fragments and single cell digests were cryopreserved for analysis and culturing at later stages.

TIL collection was performed by culturing of tumor fragments in a 24-well plate with T cell medium (IMDM (Lonza BioWhittaker or Thermo Fisher), supplemented with 8% heat-inactivated pooled human serum, penicillin (100 IU/ml), streptomycin (100 µg/ml) and L-glutamine (4 mM)) and rIL-2 (1000 IU/ml). After 14-21 days of culturing, TILs were harvested and cryopreserved for later use. To increase the number of T cells available for screening, rapid expansion of TILs was performed by culturing with rIL-2 (3000 IU/ml), OKT3 (Miltenyi Biotec, 60 µg/ml) and irradiated (40 Gy) feeder cells (100-200×) for 4-5 days. Subsequently, culturing was continued up to two weeks in T cell medium supplemented with rIL-2 (3000 IU/ml).

A mixed lymphocyte tumor culture (MLTC) was performed by co-culturing PBMCs with lethally irradiated (100 Gy) tumor fragments in T cell medium. Recombinant human IL-4 was added at day 0 to prevent NK cell outgrowth. $PD1^+$ cell selection was performed after day 1 of co-culture. Cells were harvested and stained with PE-conjugated anti-PD1 antibodies (BD Biosciences). Subsequently, MACS was performed by use of magnetic anti-PE beads (Miltenyi Biotec) and magnetic separation (MS) columns (Miltenyi Biotec). $PD1^+$ cells and flow through were each cultured with irradiated (40 Gy) feeder cells (100-200×) and high-dose rIL-2 (3000 IU/ml). Culture medium containing rIL-2 was refreshed on alternate days. Cells were cryopreserved after a culturing period of two weeks.

CD39/CD103 double negative (DN), single positive (SN) and double positive (DP) TILs were isolated as described by Duhen et al.[30]. Briefly, cryopreserved PBMCs and TILs were thawed and enriched for T cells using a T cell enrichment kit (STEMCELL Technologies) and for TILs using EpCAM beads (STEMCELL Technologies). The enriched fractions were then labeled and sorted on a FACSAria II cell sorter (BD Biosciences). From the TILs memory T cell ($CD3^+CD4^-CD8^+CD45RA^-CR7^{+/-}$) subsets were sorted as $CD39^-CD103^-$ (DN), $CD39^-CD103^+$ (SP), and $CD39^+CD103^+$ (DP). For expansion sorted T cell subsets were cultured in complete RPMI 1640 medium supplemented with 2 mM glutamine, 1% (v/v) nonessential amino acids, 1% (v/v) sodium pyruvate, penicillin (50 U/ml), streptomycin (50 µg/ml), and 10% fetal bovine serum (Hyclone).

Sorted T cells were stimulated polyclonally with 1 µg/ml Phytohemagglutinin A (PHA, Sigma) in the presence of irradiated (40 Gy) allogeneic feeder cells (PBMC; 2×10^5 cells/well) and 10 ng/ml of IL-15 (BioLegend) in a 96-well round-bottom plate (Corning/Costar). T cell lines were maintained in complete medium with IL-15 for 2-3 weeks and then cryopreserved until analysis.

Murine cell samples

Female C57BL/6 mice of 8-10 weeks were purchased form Envigo, Harlan Laboratories and acclimatized for 1 week to the animal facility of the LUMC. The mice were housed in individually-ventilated-cage (IVC) systems in specific pathogen-free conditions and kept at room temperature. MC38 (murine colon carcinoma) cells were cultured in IMDM medium (Lonza) supplemented with 8% Fetal Calf Serum (FCS, Greiner), 100 IU/ml penicillin/streptomycin (Gibco), 2 mM glutamine (Gibco) and 25 mM 2-mercaptoethanol (culture medium). Cell lines were mycoplasma- and MAP-tested before injection. Mice were subcutaneously injected thrice in the right-flank with 5×10^6 irradiated (15,000 rads) MC38 cells in 200 µL of PBS with a two week interval, whereby one group of MC38 injections was adjuvanted with 100 µg of DMXAA (5,6-dimethylxanthenone-4-acetic acid, InvivoGen).

Spleens from the mice were obtained one week after the final injection and mashed on single cell strainers with the blunt end of a 5 ml syringe and washed with culture medium. Cellular precipitates after centrifugation were treated with 5 ml of lysis buffer for 3 minutes at room temperature and subsequently washed with culture medium. Splenocytes were frozen in 10% DMSO in FCS at a concentration of 10×10^6 cells/ml and stored in liquid nitrogen. Similarly, OT-I/Thy1.1/CD45.2 cells were obtained from the spleens of in-house bred transgenic mice, although samples were enriched for $CD8^+$ lymphocytes (Mouse CD8 T Lymphocyte Enrichment Set, BD IMag) and frozen at a concentration of 4×10^6 cells/mL.

Peptide prediction and synthesis

For exome sequencing, reads were mapped against the human reference genome (hg38) using the Burrows-Wheeler Aligner (BWA-mem version 0.7.15) algorithm with default parameters[42]. Duplicate reads were removed using Picard Tools (http://picard.sourceforge.net). Genome Analysis Toolkit (GATK version 3.8; Broad Institute) was used for base quality recalibration[43]. Subsequently, single-nucleotide variants and indels were called using a combination of three popular software tools: muTect 2, varScan 2 and Strelka[44-46]. The resulting vcf files were combined into a single file using GATK CombineVariants. Variants were then functionally annotated using the ensembl Variant Effect Predictor (VEP)[47]. Variants annotated as protein disrupting or altering were further investigated if at least one read with the alternative allele was present in the RNAseq data. Reads generated by RNAseq were mapped against the same hg38 genome build using gsnap[48]. Integrative Genomics Viewer (IGV) was used for visually inspecting variants[49,50]. Manual review of aligned reads was used to reduce the risk of false positives and incorrect calls[51]. Prediction of binding to HLA-A*02:01 was performed for 8-12 amino acid peptide sequences using NetMHC and NetMHCpan. All strong and weak binders were selected for multimer screening. Murine MC38 neoantigen prediction was performed as described by Hos et al.[52]

Peptides were synthesized in our lab using standard solid-phase peptide synthesis or ordered from Pepscan. Synthesis was performed using Syro I and Syro II synthesizers using *N*-methyl-2-pyrrolidone as solvent. Resins and amino acids were purchased from Nova Biochem. Peptides were purified by reversed-phase HPLC over a preparative Waters X-bridge C18 column in a Waters HPLC system using water/acetonitrile mixtures containing 0.1% TFA. Peptide purity and composition were analyzed sample-wise by LC-MS using a Micromass LCT Premier mass spectrometer (Waters) equipped with a 2795 separation module (Alliance HT) and 2996 photodiode array detector (Waters). The samples were separated using a water/acetonitrile gradient over a Kinetix C18 column (Phenomenex). Analysis was performed using MassLynx 4.1 software (Waters Chromatography).

Multimer preparation

Temperature-exchangeable HLA-A*02:01-IAKEPVHGV and H-2K^{b}-FAPGNAPAL complexes were expressed and folded essentially as described previously[24,26,53], with minor alterations. Folded complexes were concentrated using a 30 kDa MWCO PES Vivaflow 200 protein concentrator system (Sartorius), driven by a Masterflex L/S peristaltic pump. Consequently the buffer was exchanged for 300 mM NaCl and 20 mM Tris•Cl, pH 8 using a NAP-10 column. Samples were filtered using a Spin-X column and biotinylated overnight using BirA ligase, supplemented with ATP, biotin and protease inhibitors. The following day samples were concentrated using Amicon Ultra-15 30 kDa MWCO centrifugal filter units (Merck Millipore)

and purified by gel filtration size exclusion chromatography (300 mM NaCl and 20 mM Tris•Cl, pH 8; Superdex 75 16/600 column, GE Healthcare) on an NGC system (Bio-Rad). After another round of concentration to 2-4 mg/ml using Amicon Ultra-15 30 kDa filters, they were snap-frozen and stored at -80°C in the same buffer supplemented with 15% glycerol. Biotinylation was verified by incubation of biotinylated MHCI monomers with streptavidin, followed by gel filtration chromatography on a Shimadzu Prominence system equipped with a 300 × 7.8 mm BioSep SEC-s3000 column (Phenomenex) using PBS as mobile phase. Data were analyzed using Shimadzu LabSolutions software (version 5.85).

Multimers were assembled as described by Bentzen et al.[1]. Briefly, PE- and streptavidin-conjugated dextran backbones (Fina Biosolutions, final concentration 6.92×10^{-8} M) were added to 5'-biotinylated AxBy DNA barcodes (DNA Technology, sequences in Table S1 and Table S2), which were titrated per batch of dextran. After incubating for 30 min at 4°C MHCI monomers were added at a final concentration of 30 µg/ml, followed by another 30-min incubation at 4°C. To each well a different peptide was added at a final concentration of 60 µM, and plates were incubated at previously described exchange temperatures (5 minutes at room temperature for H-2K^{b} and 3 hours at 32°C for HLA-A*02:01). For stability and to saturate unoccupied streptavidin binding sites a solution containing 500 µM D-biotin, 100 µg/ml herring DNA, 0.5% BSA, 2 mM EDTA and 5% glycerol (in PBS) was added and incubated for 20 min on ice.

Barcode-labelled exchangeable multimers were centrifuged at 3300 × *g* for 5 min at 4°C to sediment aggregates and then pooled at 0.043 µg pMHC per sample. Pools were collected in reservoirs that were pre-saturated for at least 2 hours with 2% BSA to prevent sticking. Using (also pre-saturated) Vivaspin6 or Vivaspin20 centrifugal concentrators (100 kDa MWCO, Sartorius) to a volume of ~80 µl per sample. Concentrated pools were centrifuged for 5 min at 3300 × *g* before adding to cell suspension. A 5 µl aliquot was stored at -20°C for later use as baseline sample.

pMHCI multimer staining and sorting

Cryopreserved cell suspensions were thawed in and washed with RPMI supplemented with 10% FCS, and subsequently washed with barcode cytometry buffer (BCB; PBS with 0.5% BSA, 100 µg/ml herring DNA, 2 mM EDTA) and incubated with 50 nM dasatinib for 30 min at 37°C. For human samples 2×10^{6} cells and for murine samples ~4×10^{5} cells were stained with pooled DNA-barcoded multimers in 100 µl BCB total volume for 15 min at 37°C. Human samples were stained with antibody mix composed of anti-CD8-V510, dump channel FITC-conjugated antibodies against CD4, CD14, CD16, CD19 and CD40 (all BD Biosciences), and near-IR viability dye (Invitrogen), for 30 min at 4°C. Murine samples were stained with antibody mix composed of anti-CD3-FITC (BioLegend), anti-CD8-BV480

(BD Horizon) and near-IR viability dye (Invitrogen), for 30 min at 4°C. Cells were washed three times with BCB, filtered and fixed in 1% PFA for overnight storage.

Stained cells were washed three times in BCB prior to sorting on a FACSAria or FACSMelody (BD Biosciences) into pre-saturated tubes containing 200 µl BCB. From human samples the population of single, live $CD8^+$, $dump^-$, PE $(multimer)^+$ lymphocytes was sorted using FACSDiva (BD Biosciences) software. From murine samples single, live $CD8^+$, $CD3^+$, PE $(multimer)^+$ lymphocytes were sorted. Further analysis was performed using FACSDiva or FlowJo (FlowJo, LLC) software.

DNA barcode amplification and analysis

DNA barcodes were amplified using a Taq PCR Master Mix Kit with 0.3 µM appropriate forward- (with a distinct sample ID embedded) and reverse primers comprising Ion Torrent PGM 5' and 3' adaptors. Sorted cells (in less than 20 µl buffer) and the stored baseline aliquot (diluted 10,000× in H_2O) were amplified using a PCR program with the following conditions: 95°C 10 min; 36 cycles: 95°C 30 s, 60°C 45 s, 72°C 30 s and 72°C 4 min. PCR products were analyzed using gel electrophoresis (E-Gel, Invitrogen), pooled at similar concentrations according to visual inspection and then purified using the QIAquick PCR Purification kit (Qiagen) according to standard procedure. The amplified barcodes were sequenced at Sequetech (USA) or the LUMC Sequence Analysis Support Core (SASC, NL). Sequencing data were analyzed and visualized using the online tool 'Barracoda' (http://www.cbs.dtu.dk/services/barracoda) developed at DTU.

Acknowledgements

The authors thank Cami Talavera Ormeño and Paul Hekking (Department of Cell and Chemical Biology, LUMC, Leiden, Netherlands) for synthesis of peptides. We also thank Dina Ruano (Department of Pathology, LUMC, Leiden, Netherlands), for sequencing of the human CRC screen and assistance with analysis.

This work was supported by a grant from the Institute for Chemical Immunology (ICI, to H. Ovaa) and is part of Oncode Institute, which is partly financed by the Dutch Cancer Society.

Author contributions: J.J. Luimstra, J. Neefjes, and H. Ovaa conceived and designed the study with input from N. de Miranda and F. Ossendorp. J. van den Bulk prepared human cell samples and B. Hos prepared murine cell samples. N.F. de Miranda predicted peptides. J.J. Luimstra and C. Heeke generated DNA-barcoded multimers and performed T cell staining experiments. J.J. Luimstra wrote the manuscript with input from all authors.

References

1. Bentzen, A. K. *et al.* Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol* **34**, 1037-1045, doi:10.1038/nbt.3662 (2016).
2. Schumacher, T. N. & Hacohen, N. Neoantigens encoded in the cancer genome. *Curr Opin Immunol* **41**, 98-103, doi:10.1016/j.coi.2016.07.005 (2016).
3. Tran, E. *et al.* Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* **350**, 1387-1390, doi:10.1126/science.aad1253 (2015).
4. Corthay, A. Does the immune system naturally protect against cancer? *Front Immunol* **5**, 197, doi:10.3389/fimmu.2014.00197 (2014).
5. Kreiter, S., Castle, J. C., Tureci, O. & Sahin, U. Targeting the tumor mutanome for personalized vaccination therapy. *Oncoimmunology* **1**, 768-769, doi:10.4161/onci.19727 (2012).
6. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69-74, doi:10.1126/science.aaa4971 (2015).
7. Aldous, A. R. & Dong, J. Z. Personalized neoantigen vaccines: A new approach to cancer immunotherapy. *Bioorg Med Chem* **26**, 2842-2849, doi:10.1016/j.bmc.2017.10.021 (2018).
8. Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547**, 222-226, doi:10.1038/nature23003 (2017).
9. Bezu, L. *et al.* Trial watch: Peptide-based vaccines in anticancer therapy. *Oncoimmunology* **7**, e1511506, doi:10.1080/2162402X.2018.1511506 (2018).
10. Keskin, D. B. *et al.* Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* **565**, 234-239, doi:10.1038/s41586-018-0792-9 (2019).
11. Ott, P. A. *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217-221, doi:10.1038/nature22991 (2017).
12. Vita, R. *et al.* The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* **43**, D405-412, doi:10.1093/nar/gku938 (2015).
13. Lundegaard, C. *et al.* NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res* **36**, W509-512, doi:10.1093/nar/gkn202 (2008).
14. Capietto, A. H., Jhunjhunwala, S. & Delamarre, L. Characterizing neoantigens for personalized cancer immunotherapy. *Curr Opin Immunol* **46**, 58-65, doi:10.1016/j.coi.2017.04.007 (2017).
15. Karasaki, T. *et al.* Prediction and prioritization of neoantigens: integration of RNA sequencing data with whole-exome sequencing. *Cancer Sci* **108**, 170-177, doi:10.1111/cas.13131 (2017).
16. Wilson, E. A. & Anderson, K. S. Lost in the crowd: identifying targetable MHC class I neoepitopes for cancer immunotherapy. *Expert Rev Proteomics* **15**, 1065-1077, doi:10.1080/14789450.2018.1545578 (2018).
17. Altmann, D. M. New tools for MHC research from machine learning and predictive algorithms to the tumour immunopeptidome. *Immunology* **154**, 329-330, doi:10.1111/imm.12956 (2018).
18. Blaha, D. T. *et al.* High-Throughput Stability Screening of Neoantigen/HLA Complexes Improves Immunogenicity Predictions. *Cancer Immunol Res* **7**, 50-61, doi:10.1158/2326-6066.CIR-18-0395 (2019).
19. Bjerregaard, A. M. *et al.* An Analysis of Natural T Cell Responses to Predicted Tumor Neoepitopes. *Front Immunol* **8**, 1566, doi:10.3389/fimmu.2017.01566 (2017).
20. Bassani-Sternberg, M. *et al.* Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun* **7**, 13404, doi:10.1038/ncomms13404 (2016).
21. Altman, J. D. *et al.* Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**, 94-96 (1996).
22. Saini, S. K. *et al.* Dipeptides catalyze rapid peptide exchange on MHC class I molecules. *Proc Natl Acad Sci U S A* **112**, 202-207, doi:10.1073/pnas.1418690112 (2015).
23. Amore, A. *et al.* Development of a hypersensitive periodate-cleavable amino acid that is methionine- and disulfide-compatible and its application in MHC exchange reagents for T cell characterisation. *Chembiochem* **14**, 123-131, doi:10.1002/cbic.201200540 (2013).
24. Rodenko, B. *et al.* Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. *Nat Protoc* **1**, 1120-1132, doi:10.1038/nprot.2006.121 (2006).
25. Toebes, M., Rodenko, B., Ovaa, H. & Schumacher, T. N. Generation of peptide MHC class I mo-

nomers and multimers through ligand exchange. *Curr Protoc Immunol* **Chapter 18**, Unit 18 16, doi:10.1002/0471142735.im1816s87 (2009).
26. Luimstra, J. J. *et al.* A flexible MHC class I multimer loading system for large-scale detection of antigen-specific T cells. *J Exp Med* **215**, 1493-1504, doi:10.1084/jem.20180156 (2018).
27. Newell, E. W., Klein, L. O., Yu, W. & Davis, M. M. Simultaneous detection of many T-cell specificities using combinatorial tetramer staining. *Nat Methods* **6**, 497-499, doi:10.1038/nmeth.1344 (2009).
28. Hadrup, S. R. *et al.* Parallel detection of antigen-specific T-cell responses by multidimensional encoding of MHC multimers. *Nat Methods* **6**, 520-526, doi:10.1038/nmeth.1345 (2009).
29. Dudley, J. C., Lin, M. T., Le, D. T. & Eshleman, J. R. Microsatellite Instability as a Biomarker for PD-1 Blockade. *Clin Cancer Res* **22**, 813-820, doi:10.1158/1078-0432.CCR-15-1678 (2016).
30. Duhen, T. *et al.* Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors. *Nat Commun* **9**, 2724, doi:10.1038/s41467-018-05072-0 (2018).
31. Gupta, P. K. *et al.* CD39 Expression Identifies Terminally Exhausted CD8+ T Cells. *PLoS Pathog* **11**, e1005177, doi:10.1371/journal.ppat.1005177 (2015).
32. Djenidi, F. *et al.* CD8+CD103+ tumor-infiltrating lymphocytes are tumor-specific tissue-resident memory T cells and a prognostic factor for survival in lung cancer patients. *J Immunol* **194**, 3475-3486, doi:10.4049/jimmunol.1402711 (2015).
33. Corrales, L. *et al.* Direct Activation of STING in the Tumor Microenvironment Leads to Potent and Systemic Tumor Regression and Immunity. *Cell Rep* **11**, 1018-1030, doi:10.1016/j.celrep.2015.04.031 (2015).
34. Woo, S. R. *et al.* STING-dependent cytosolic DNA sensing mediates innate immune recognition of immunogenic tumors. *Immunity* **41**, 830-842, doi:10.1016/j.immuni.2014.10.017 (2014).
35. Blank, C., Gajewski, T. F. & Mackensen, A. Interaction of PD-L1 on tumor cells with PD-1 on tumor-specific T cells as a mechanism of immune evasion: implications for tumor immunotherapy. *Cancer Immunol Immunother* **54**, 307-314, doi:10.1007/s00262-004-0593-x (2005).
36. Iwai, Y. *et al.* Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. *Proc Natl Acad Sci U S A* **99**, 12293-12297, doi:10.1073/pnas.192461099 (2002).
37. Hellmann, M. D. *et al.* Tumor Mutational Burden and Efficacy of Nivolumab Monotherapy and in Combination with Ipilimumab in Small-Cell Lung Cancer. *Cancer Cell* **33**, 853-861 e854, doi:10.1016/j.ccell.2018.04.001 (2018).
38. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463-1469, doi:10.1126/science.aaf1490 (2016).
39. van den Bulk, J., Verdegaal, E. M. & de Miranda, N. F. Cancer immunotherapy: broadening the scope of targetable tumours. *Open Biol* **8**, doi:10.1098/rsob.180037 (2018).
40. vanderBurg, S. H., Visseren, M. J. W., Brandt, R. M. P., Kast, W. M. & Melief, C. J. M. Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability. *Journal of Immunology* **156**, 3308-3314 (1996).
41. Rosendahl Huber, S. K. *et al.* Chemical Modification of Influenza CD8+ T-Cell Epitopes Enhances Their Immunogenicity Regardless of Immunodominance. *PLoS One* **11**, e0156462, doi:10.1371/journal.pone.0156462 (2016).
42. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).
43. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).
44. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219, doi:10.1038/nbt.2514 (2013).
45. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22**, 568-576, doi:10.1101/gr.129684.111 (2012).
46. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811-1817, doi:10.1093/bioinformatics/bts271 (2012).
47. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, doi:10.1186/s13059-016-0974-4 (2016).
48. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875, doi:10.1093/bioinformatics/bti310 (2005).

49. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).
50. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).
51. Robinson, J. T., Thorvaldsdottir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant Review with the Integrative Genomics Viewer. *Cancer Res* **77**, e31-e34, doi:10.1158/0008-5472.CAN-17-0337 (2017).
52. Hos, B.J. *et al.* Identification of a neo-epitope dominating endogenous CD8 T cell responses to MC-38 colorectal cancer. *OncoImmunology* doi:10.1080/2162402X.2019.1673125 (2019).
53. Toebes, M. *et al.* Design and use of conditional MHC class I ligands. *Nat Med* **12**, 246-251, doi:10.1038/nm1360 (2006).

Supplementary Data

Supplementary Table 1. **Oligo A sequences used in this study**

Oligo A #	Barcode Sequence
A1	CGAGGGCAATGGTTAACTGACACGT
A2	CAGAAAGCAGTCTCGTCGGTTCGAA
A3	TAAGTAGCGGGCATAATGTACGCTC
A5	GGGCTGCGGAGCGTTTACTCTGTAT
A6	AAACGTATGTGCTTTGTCGGATGCC
A7	ATATCATCATAGGCTTAGCGACGTA
A8	AGGAAAATCTGCTACCGCCAATGAT
A9	CTGATTGACTGCATGGAGGCTATAC
A10	GTGGCGACTTCACGATTATCTGAAC
A11	CCTGTATTGAAGGTTCAGTCCTGTT
A12	GGCTCTATAAGGTTTCCTCAAAGGT
A14	AGAGAATATGTCGCTCCCGTTATGT
A15	GCAGTTAGATATGCAGTTACCTGAC
A16	CTTCACCCGAACATGCAGTGTTATT
A17	AAAGCCGTTGCAGTATCGTCTGAGC
A18	GCTGGATGTTAATAACTGCGGTCCG
A19	ACGAGTTGACATGGACGGATCCCTC
A20	TTCATCACTCATTGTTCTGAGTAGG
A21	ATGTTTAATCTAACTTGATGCCTCC
A22	TAATACGCCTGAGGTGTTGGGTTGC
A23	AGTCGGCATTGCTACCATAACTGTT
A24	CCGGACCGCTATTAACCTTGTACTG
A25	CTAGATGCTGCGAACGGAAGCTGTC
A26	TGTTCCAAGCGGTTGAACGATTAGC

Supplementary Table 2. **Oligo B sequences used in this study**

Oligo B #	Barcode Sequence
B61	GTTAGGTCGGCAGGTCAGTATGACC
B62	CGGGAGTTGGATCTGCGTAGAGTCC
B63	CCGGTTTTATACCCTCGTTCCCCGA
B64	CAGAACTACAGGCTGGCATGGATGC
B65	ATTCTGATGGGTAGAAACCGTTCCC
B66	GAGCGTGAGTTCCATGGAAAATTAC
B67	AGTAAAGGCTCACTGCTATCGCACT
B68	ATTTATTCGCACAATCGCCGAGTGC
B69	TACTCAACGACGTGGGGTAGGATCC
B70	GATATTCGGATCTTGGCTCGGACTG
B71	TTTCCTTGTTCGGATCGGTCGAGAA
B72	TGGAAACGACTGGTGTATGCATTCC
B73	GCTGTCAGTAGCGCCAGTACAATTT
B74	CTTTATGGGATAGCAAGACCTCTCC
B75	CATATGGATTTGTTGCATCCTGATG
B76	TGCAATATGGGTCGCGTTCAGTCGT

Supplementary Table 2 (continued). **Oligo B sequences used in this study**

Oligo B #	Barcode Sequence
B77	TAATTGCCTTGCGGTGCGTGTTGAT
B78	GGCAGGCTAGCTTAGTTGTAGCGGT
B79	CACATACTCAGACTCCCTGTCATAG
B80	TTGATCACAGCACGAATACGTTTCC
B81	TTTGAATAACCTTTCGCCTCTCGTG
B82	TGATTGCTTTGCCCTATAGCTACGT
B83	AGTGAAGTTACCTGGTGTTCCCTTC
B84	GCGGGATGTGCATTGCCAAGTTACC
B85	TTAAGTTGCCCAATTATTGTCCGCC
B86	GACAATGTAGGGGCCGTCTCAAGTA
B87	GCCATAGAGTCTACCGTCACCTCCG
B88	GTGCTACCATCGAGCGGAGGTATTT
B89	GAGTCCGATTGCTTTATCTGCTACC
B90	AATGGGCCTGCTACTCGCCATTATT
B111	CGTTGAAATAGTCGCATCTCTCACG
B113	TAAATGGCCTACATTCGAACGGTTG
B115	GTTATCCGTAGAGCGGTGCAAGTCC
B116	TCTTCGACATCTGGCATCACGACCT
B117	TTGGTTTATGATCACTGAAATGCCC
B118	ATGGGGATAGCCATCAGTTGGGCTA
B120	AATCGTAGTCGTCAGCCGCTAATAA
B121	GTCTTGTCGTGGGACGCATGTATCC
B122	GAAGTGAGGCGCTAACGCTCTAGGG
B124	GGTCATTCTAGTGAACTAATCCCCT
B200	GCTGAGGCGTTCCACTTGGATCGTT
B201	ATTGGGGACTTCCCTTTGCATTCTT
B202	AAATGGGACCGACACACTCTTAGCA
B203	GGCTTTACGGAACCCCGTGACTAGA
B204	AAGGTACTGGGCCGGTCCAATACAG
B205	TAGTAATACATACGCCCAGGCGGTA
B206	TACATGTTCGTTCTGCGTTACTCAC
B207	CTAATCAATGGTCCACGTTCTAGGG
B208	GCTACTACACCCGAGGTCGAGAGGA
B209	GTTCTGCAATTTCAATTCCCGGTCC
B210	AAGGCTTTCTAGCCACGTATGCGAA
B211	TGTACTGAGGGAGTAACTGCCCGTT
B212	ATTGTCAGATGTCGTAGGTCGCCGC
B213	AAATCCATTTATCGGTCTGTCGTGA
B214	AGTCATATAGTTGTATTCTCCCTGC
B215	ATTTGGACAGCATATTGACGTCGGA
B216	CGCATTCCGACAATATCGTGTTGTA
B217	CCCAGAACGTGAGTCAGTGTTCGCA
B218	TCCTTAGTTTTCCGGCTAAATGAGA
B219	GATGTTATTGCTTGCACAACGGCTG

Supplementary Table 3. **HLA-A*02:01 neoantigen sequences predicted from five colorectal cancer (CRC) patients, with DNA barcode annotations**

#	Sequence	Patient	Barcode
1	ALAAAQCSA	P5	A1B61
2	ALALVVAMA	P5	A1B62
3	ALGSTAPPA	P4	A1B63
4	ALQDMSSTA	P5	A1B64
5	CLAAWVPAPA	P5	A1B65
6	FHIFYQLLGA	P5	A1B66
7	FLELSLRA	P5	A1B67
8	FLPSSCSLA	P4	A1B68
9	FLPSSCSLAPA	P4	A1B69
10	FLRTKLCFA	P5	A1B70
11	FQECHIPFPA	P5	A1B71
12	GLACLGLSWYA	P5	A1B72
13	GLDDRSPQA	P5	A2B61
14	GLGGTHHMA	P4	A2B62
15	GLNHGNFFA	P2	A2B63
16	HIWPKGFEA	P4	A2B64
17	HLYDTLHWA	P3	A2B65
18	ILGSGTSFA	P5	A2B66
19	KLAYFSLSA	P4	A2B67
20	KLDLKVPKA	P5	A2B68
21	KLVFLGLDNA	P4	A2B69
22	KMPEMSIKA	P4	A2B70
23	KVMLTAPPA	P5	A2B71
24	LIISEYFTA	P4	A2B72
25	LLFLGPLAPA	P5	A3B61
26	LLISQGLKA	P4	A3B62
27	LLSPPEPQA	P5	A3B63
28	LMAPLSPGA	P5	A3B64
29	MLGQLSAEA	P5	A3B65
30	MLILGKDTA	P5	A3B66
31	MLLPPRPAA	P4	A3B67
32	MLSASIMYA	P5	A3B68
33	MMMGQFERDA	P4	A3B69
34	QLLLLLPRA	P5	A3B70
35	RLDLSAGPTA	P5	A3B71
36	RLYHPDTHHA	P5	A3B72
37	RMFIPAAAA	P5	A5B61
38	RMWVSMCPA	P5	A5B62
39	SLAEPSPPA	P2	A5B63
40	SLAQAPIPA	P5	A5B64
41	SLFDSVYGA	P5	A5B65
42	SLLGGVLRRA	P5	A5B66
43	SLQPPTLGA	P4	A5B67
44	TLAIRFISA	P5	A5B68
45	VIAASVPRA	P5	A5B69
46	WLCGWTSSA	P5	A5B70
47	WLLGLLMPFRA	P2	A5B71
48	WVLPSLPMA	P4	A5B72
49	YISRCAPPA	P4	A6B61
50	YLNLTVLA	P5	A6B62
51	YMDLILASA	P4	A6B63
52	YMQWVWGA	P4	A6B64
53	YTDRALAFYA	P4	A6B65
54	LLLPEGIRC	P4	A6B66
55	LLDDNQAPF	P5	A6B67
56	SLDDIIRHDF	P5	A6B68
57	YLQKLSVEF	P4	A6B69
58	ALAPRSATI	P5	A6B70
59	CLFEFLTGI	P5	A6B71
60	CMGGMNWRPI	P2	A6B72
61	FAHNRNWYI	P4	A7B61
62	FLEENCADI	P5	A7B62
63	FLNGATPYEKGI	P3	A7B63
64	FLPPRLKKI	P5	A7B64
65	FMMPPEETI	P4	A7B65
66	FTEKNLWLI	P5	A7B66
67	FVAPLVPLPI	P5	A7B67
68	GLHRQLLYI	P4	A7B68
69	GLHSVGAYI	P5	A7B69
70	GMEEATVAI	P5	A7B70
71	GMIHMLDGI	P5	A7B71
72	GMWAQLPCI	P4	A7B72
73	GVYPVEGFEI	P5	A8B61
74	HLPAGWGEALQI	P5	A8B62
75	ILDPSYHI	P4	A8B63
76	ILIYSWCRI	P5	A8B64
77	ILLGTFLAI	P5	A8B65
78	ILLKMEIQI	P5	A8B66
79	ILSDPENNI	P5	A8B67
80	IRPPLLPYIVNI	P4	A8B68
81	IVLGSCVYI	P5	A8B69
82	KIAFHIKSI	P5	A8B70
83	KLCQGMHQI	P3	A8B71
84	KLINPDKKI	P4	A8B72
85	KLLHTQKVVYI	P5	A9B61
86	KLNGQTMEI	P5	A9B62
87	KMEIQIFKI	P5	A9B63
88	KQLAVSICI	P5	A9B64
89	KTGMEILLWI	P5	A9B65
90	LIMVYLFSI	P4	A9B66
91	LLAVVIQFQI	P5	A9B67
92	LLIDLMEQEI	P5	A9B68
93	LLILCVHAKI	P5	A9B69
94	LLPPPTEWLI	P5	A9B70
95	LLPPPTEWLIPI	P5	A9B71
96	LLWILLKMEI	P5	A9B72
97	MLHRGLLLI	P5	A1B73
98	MMLATKLTI	P4	A1B74
99	RLFGTWINKI	P4	A1B75
100	SLKDSQFSI	P5	A1B76
101	SLLLLPEGI	P4	A1B77
102	SLPTTPLYFI	P5	A1B78
103	SLSHILTCGI	P4	A1B79
104	SLYYDYEPPI	P5	A1B80
105	SMSSTPLTI	P5	A1B81
106	STAAEVVAI	P5	A1B82
107	TLLSRLPAI	P4	A1B83
108	TLSPAITSI	P5	A1B84
109	TMQPWPCSI	P5	A2B73
110	VIAGGIWHI	P2	A2B74
111	VLNPYVKHSI	P5	A2B75
112	VLVEEVAEKCI	P5	A2B76
113	WLGPGLRMGI	P5	A2B77
114	WLSRSAFYCI	P5	A2B78
115	YITAFFCWI	P4	A2B79
116	YLDLYLIHWPI	P5	A2B80

Supplementary Table 3 (continued). **HLA-A*02:01 neoantigen sequences predicted from five colorectal cancer (CRC) patients, with DNA barcode annotations**

#	Sequence	Patient	Barcode
117	YLLKVCERI	P1	A2B81
118	YLQKLSVEFQI	P4	A2B82
119	YLVASDQRPI	P1	A2B83
120	YPQLKALPPI	P5	A2B84
121	ALGPAASAL	P5	A3B73
122	ALLPPPTEWL	P5	A3B74
123	ALNPSAPSL	P5	A3B75
124	ALPRLPVPL	P5	A3B76
125	ALSLDTQNL	P5	A3B77
126	ALVKSSEEL	P4	A3B78
127	ALWSAVTLL	P5	A3B79
128	AMAAALGVL	P5	A3B80
129	AMAQVTHPL	P4	A3B81
130	AMVAVPMVL	P5	A3B82
131	ASLQNLLFKL	P4	A3B83
132	AVFGHHFSL	P5	A3B84
133	CLAAEITRL	P4	A5B73
134	CMADGSTAL	P5	A5B74
135	FGMSVCSWPL	P5	A5B75
136	FIQQMVHAL	P5	A5B76
137	FLARPLPWPL	P4	A5B77
138	FLFSDKLGEL	P3	A5B78
139	FLLAGAHL	P5	A5B79
140	FLLKNIIFL	P4	A5B80
141	FLLPGKKIL	P5	A5B81
142	FLMHLYLEL	P5	A5B82
143	FLMVLVWLPL	P5	A5B83
144	FLNKPSIIL	P2	A5B84
145	FLPGSTPSL	P5	A6B73
146	FLPPLLLLLL	P4	A6B74
147	FLSHYLQKL	P4	A6B75
148	FLSLPETAL	P5	A6B76
149	FLSTLPHL	P2	A6B77
150	FLTTSSLML	P4	A6B78
151	FLVQNIHTLAGL	P4	A6B79
152	FLYNNLVESL	P5	A6B80
153	FMFRTWGRML	P5	A6B81
154	FMRWIIGL	P4	A6B82
155	FPMPNYQAAL	P4	A6B83
156	FSWSNTTLL	P1	A6B84
157	GLEVSGAFPQL	P5	A7B73
158	GLFSEDGATL	P4	A7B74
159	GLHLHPSPAL	P5	A7B75
160	GLIDGMHMHL	P5	A7B76
161	GLLPQTKTL	P5	A7B77
162	GLNLGPQVAL	P5	A7B78
163	GLPPEPEVPPAL	P5	A7B79
164	GLQDQEPSL	P5	A7B80
165	GLQKEIAEL	P4	A7B81
166	GLRMGIGLNL	P5	A7B82
167	GLRTEAPPTL	P4	A7B83
168	GLSAQHVPSL	P5	A7B84
169	GLSSFGQSL	P5	A8B73
170	GLTEPVLIWL	P4	A8B74
171	GMGGSTITL	P5	A8B75
172	GMHSRLSSL	P4	A8B76
173	GMLTVIGQGL	P5	A8B77
174	GVHPSLAPL	P5	A8B78
175	HLDTFHLSL	P5	A8B79
176	HLHESCMLSL	P4	A8B80
177	HLISQCEQL	P5	A8B81
178	HLQIRWPNLPRL	P5	A8B82
179	HLTHLEAAL	P4	A8B83
180	IIATVLYGPL	P5	A8B84
181	ILANTVKPFL	P4	A9B73
182	ILDPSYHIPPL	P4	A9B74
183	ILKLWLGPGL	P5	A9B75
184	ILPMKIPRQL	P5	A9B76
185	ILTHIIECL	P5	A9B77
186	IMPMNILYL	P2	A9B78
187	IMYACVFCL	P5	A9B79
188	ITLGFGWML	P5	A9B80
189	KISFENLHL	P5	A9B81
190	KISHCPHLL	P4	A26B115
191	KLFEMAYKRWHL	P4	A9B83
192	KLIYQGHLL	P2	A9B84
193	KLMKNIQFPL	P5	A10B61
194	KLMPWNCCL	P5	A10B62
195	KLQAETEEL	P4	A10B63
196	KLTTEYLSL	P4	A10B64
197	KPLLSYPLVL	P5	A10B65
198	LLAACPLHL	P5	A10B66
199	LLAEVDVPKL	P3	A10B67
200	LLAPSGHLL	P5	A10B68
201	LLFGLKGEL	P5	A10B69
202	LLIQQINFHL	P4	A10B70
203	LLKKIASTFYL	P4	A10B71
204	LLKYVRTPTL	P5	A10B72
205	LLLALPHEL	P5	A11B61
206	LLLCVQALL	P4	A11B62
207	LLLPGWCRL	P5	A11B63
208	LLLQQPPPL	P5	A11B64
209	LLSQICSHL	P3	A11B65
210	LLVDKHKYFL	P5	A11B66
211	LLVGSNQWEL	P5	A11B67
212	LMLSAQLCL	P4	A11B68
213	LMPGGSCWRL	P5	A11B69
214	LMPIFSPEL	P4	A11B70
215	LQAECDQYL	P4	A11B71
216	MEKLADIVTEL	P5	A11B72
217	MLAPPFPPPL	P4	A12B61
218	MLLNTPFTL	P5	A12B62
219	MLNTQDSSILPL	P4	A12B63
220	MLTAPPASL	P5	A12B64
221	MLVPGGTRVCQL	P5	A12B65
222	MMEAGLSEL	P5	A12B66
223	NLELDPIFL	P5	A12B67
224	NLGPSLVGL	P2	A12B68
225	NMLCFNFKL	P5	A12B69
226	NMSKVETGL	P5	A12B70
227	NVLSSLWYL	P5	A12B71
228	PLSFVLHFL	P5	A12B72
229	QLGKEDLGL	P4	A14B61
230	QLQIIFLEL	P5	A14B62
231	RLHTWSQGL	P4	A14B63
232	RLLDSEEPL	P5	A14B64

Supplementary Table 3 (continued). **HLA-A*02:01 neoantigen sequences predicted from five colorectal cancer (CRC) patients, with DNA barcode annotations**

#	Sequence	Patient	Barcode
233	RLMTHYCAML	P5	A14B65
234	RLPRPAHAL	P5	A14B66
235	RLSSSELSPL	P5	A14B67
236	RLWPVLDPCCPL	P5	A14B68
237	RMAATRTSL	P5	A14B69
238	RMQGLGFLL	P5	A14B70
239	RPLLALVNSL	P5	A14B71
240	RVYPRPRVAL	P5	A14B72
241	SELSPLTPRL	P5	A15B61
242	SIYPPPRAL	P4	A15B62
243	SLHFLCWSL	P5	A15B63
244	SLIFGLILKL	P5	A15B64
245	SLLEKVSKKRL	P5	A15B65
246	SLLPSCCAL	P5	A15B66
247	SLLRQPVQL	P5	A15B67
248	SLLSCPPFL	P4	A15B68
249	SLNWPEALPHL	P5	A15B69
250	SLPAGPSAL	P5	A15B70
251	SLPCTPLWL	P5	A15B71
252	SLPSTQLPL	P5	A15B72
253	SLQNLLFKL	P4	A16B61
254	SLVGTQTLL	P5	A16B62
255	SLVPRGTPL	P5	A16B63
256	SLVSFLMHL	P5	A16B64
257	SLYPQNMTL	P4	A16B65
258	SMAPTQTCL	P4	A16B66
259	SMCPAGTWCL	P5	A16B67
260	SQSEQGLLL	P5	A16B68
261	SQTPVPPGL	P5	A16B69
262	STFGMSVCSWPL	P5	A16B70
263	TLAGLHVHL	P5	A16B71
264	TLAQPELFL	P2	A16B72
265	TLGFGWMLIL	P5	A17B61
266	TLLLKAPTL	P5	A17B62
267	TLMDCHWQPL	P4	A17B63
268	TLVIYVARL	P4	A17B64
269	TLWPSLPSSTL	P5	A17B65
270	TMGGYCGYL	P4	A17B66
271	TMNDSKHKL	P5	A17B67
272	TMVDFIKSTL	P4	A17B68
273	TMVQGPAGL	P4	A17B69
274	TVLSQGWEL	P5	A17B70
275	VELMKHFAWL	P4	A17B71
276	VIFSGALLGL	P4	A17B72
277	VLDPEGIRGL	P5	A18B61
278	VLILGTKRL	P5	A18B62
279	VLQKKRILL	P5	A18B63
280	VLREKVPCL	P5	A18B64
281	VLSMTTRIFL	P5	A18B65
282	VLSSLWYLNL	P5	A18B66
283	VLTQLVLNL	P5	A18B67
284	VLVESKLRGL	P5	A18B68
285	WLCSPAPWL	P5	A18B69
286	WLGMALIPL	P4	A18B70
287	WLGTLWPSL	P5	A18B71
288	WLLQKSPQL	P4	A18B72
289	WLNTKMKFFL	P5	A10B73
290	YCLFAASLLL	P4	A10B74
291	YIFTILSSL	P4	A10B75
292	YLDSIGNLPFEL	P5	A10B76
293	YLIFKPDVML	P4	A10B77
294	YLKIKHLLL	P3	A10B78
295	YLNTNPVCGL	P4	A10B79
296	YQAQLRISL	P5	A10B80
297	YTLVPSTVAL	P3	A10B81
298	YTYLTIFDL	P5	A10B82
299	YVLPRALSL	P2	A10B83
300	YYLMTVMERL	P5	A10B84
301	ALIPLPGM	P4	A11B73
302	FCLLVVVVLM	P5	A11B74
303	FLAKDPHPM	P5	A11B75
304	FLFTVPIEDM	P5	A11B76
305	FLTGIPLSM	P5	A11B77
306	FLVLSMPAM	P4	A11B78
307	ILLWILLKM	P5	A11B79
308	ILNSLPSSM	P4	A11B80
309	ILSMEKIPPM	P4	A11B81
310	KMICRGMTM	P4	A11B82
311	RLHRLILPM	P5	A11B83
312	RLQPMISVRM	P4	A11B84
313	RLYQSVLSM	P5	A12B73
314	RMLETVLRM	P4	A12B74
315	RMVWVELEM	P5	A12B75
316	SLDDWSLIYM	P3	A12B76
317	SLRMLVQPEM	P5	A12B77
318	TQMSNLVNM	P4	A12B78
319	YVQAFQVGM	P5	A12B79
320	TLIDFFCEDKKP	P5	A12B80
321	FLMRKRWPS	P5	A12B81
322	FLQSHVPKS	P5	A12B82
323	RPLWCLLPPS	P4	A12B83
324	AIIDFSVWT	P5	A12B84
325	ALLWDTETT	P5	A14B73
326	ALWPYGPLT	P5	A14B74
327	FLFKEEEFTT	P5	A14B75
328	FLPNADMET	P4	A14B76
329	FPLVALLWDT	P5	A14B77
330	HLHHHLPTT	P5	A14B78
331	ILAPKLLST	P5	A14B79
332	KMPPFVSTT	P5	A14B80
333	VLLKKIAST	P4	A14B81
334	VVMGVCPFT	P5	A14B82
335	WLIPIAMAT	P5	A14B83
336	YLAFFPPT	P5	A14B84
337	YLLARYYYT	P5	A15B73
338	YLTIFDLLET	P5	A15B74
339	YPLPPWPWST	P5	A15B75
340	AINDVLWACV	P5	A15B76
341	ALDLYHVLV	P5	A15B77
342	ALGLDVIDQV	P1	A15B78
343	ALKIPQGQRV	P4	A15B79
344	ALKQYACTV	P5	A15B80
345	ALMPPSPLPSRV	P5	A15B81
346	ALPHLLLLV	P5	A15B82
347	ALPQLEHQV	P5	A15B83
348	ALRPHPAAV	P5	A15B84

Supplementary Table 3 (continued). **HLA-A*02:01 neoantigen sequences predicted from five colorectal cancer (CRC) patients, with DNA barcode annotations**

#	Sequence	Patient	Barcode
349	ALSEALWVV	P4	A16B73
350	ALSKHLTNPFLV	P5	A16B74
351	ALSWRNVPV	P5	A16B75
352	ALTEELHQKV	P5	A16B76
353	ALVWLKDPV	P4	A16B77
354	AMIVEQPEV	P5	A16B78
355	APLDLRLAWV	P5	A16B79
356	AQHCLLLLV	P5	A16B80
357	AQIQRPIQV	P5	A16B81
358	AQSGPLSFV	P5	A16B82
359	CLSPMGLGV	P5	A16B83
360	CMNEHWMPV	P5	A16B84
361	ELWAVDHLQV	P5	A17B73
362	FHLCSVATRV	P4	A17B74
363	FLCEKEQIV	P4	A17B75
364	FLGNMFHV	P5	A17B76
365	FLINTFEGV	P5	A17B77
366	FLKFLQGV	P5	A17B78
367	FLLGMATV	P4	A17B79
368	FLQGVALAV	P5	A17B80
369	FLRGCAPSWV	P5	A17B81
370	FLRLQVEGV	P4	A17B82
371	FLSHYLQKLSV	P4	A17B83
372	FQNRGEAAV	P5	A17B84
373	FQVLVRILPV	P5	A18B73
374	FSAPPNSLV	P5	A18B74
375	GLAVTYGV	P5	A18B75
376	GLDFFWKQEV	P5	A18B76
377	GLGEPKQPV	P4	A18B77
378	GLGSFVVGV	P5	A18B78
379	GLLPLASTV	P5	A18B79
380	GLPLAMAQV	P4	A18B80
381	GLPPRHHGV	P5	A18B81
382	GLVEEPMEDV	P5	A18B82
383	GMATVNNCV	P4	A18B83
384	GMEHFSTPV	P5	A18B84
385	GMKLLGITLV	P5	A1B111
386	GVVTSGPGV	P5	A1B113
387	HLAPPRYSPQV	P4	A1B115
388	HLIKERPLV	P5	A1B116
389	HLLWRLPAPV	P4	A1B117
390	HLSEKALEV	P4	A1B118
391	HTYSSIPVV	P1	A1B120
392	IIAGGASLV	P5	A1B121
393	ILASGFIDV	P5	A1B122
394	ILGEGRAEAV	P5	A1B124
395	ILLVNSLKV	P5	A26B116
396	ILSAITQPV	P5	A19B113
397	ILSALRVSPV	P5	A2B111
398	ILYPDEVACMV	P5	A2B113
399	IMGKMEADPEV	P3	A2B115
400	ITITFVTAV	P3	A2B116
401	ITSAAIYHV	P5	A2B117
402	IVAAGVASGV	P5	A2B118
403	KAFFGPVYYV	P5	A2B120
404	KIVAYMYLV	P1	A2B121
405	KLAIHVGLAV	P5	A2B122
406	KLESPALKQV	P5	A2B124
407	KLQGAVCVV	P5	A20B111
408	KLSLFIVCTV	P5	A20B113
409	KLSVEFQIV	P4	A3B111
410	KMDLDGMLTV	P5	A3B113
411	KQVMLQLYV	P5	A3B115
412	KVGDILQAV	P5	A3B116
413	KVSGTLLTV	P4	A3B117
414	LFSYMQWV	P4	A3B118
415	LLAAWAAPSGV	P5	A3B120
416	LLGSPDPEGV	P5	A3B121
417	LLHILSFVV	P5	A3B122
418	LLLAVRSFV	P5	A3B124
419	LLLKFTASV	P5	A21B111
420	LLSEHAVIV	P5	A21B113
421	LLSRVEILPV	P4	A5B111
422	LLTDLTSWGV	P5	A5B113
423	LLVIEKNLMV	P5	A5B115
424	LMTAKIVGNV	P5	A5B116
425	LMVDPSHEV	P5	A5B117
426	LMVSAGRGLWAV	P3	A5B118
427	LTNAGMLEV	P5	A5B120
428	LVMKGQIPV	P5	A5B121
429	MIISRHLASV	P4	A5B122
430	MLDVDLDEV	P4	A5B124
431	MLHKSIPV	P5	A22B111
432	MLNVNLDPPV	P5	A22B113
433	MLVPGGTRV	P5	A6B111
434	MMMRNQENV	P5	A6B113
435	NLEEPPSSV	P5	A6B115
436	NLQAMSLYV	P5	A6B116
437	NLYGMSKVAV	P4	A6B117
438	PLVHITEEV	P4	A6B118
439	QIFPITPPV	P5	A6B120
440	QLAGKRIGV	P5	A6B121
441	QLAIQVLLV	P5	A6B122
442	QLHPQLLLPV	P3	A6B124
443	QLILLILCV	P5	A23B111
444	QLPGSATYPV	P4	A23B113
445	RLEFIAHV	P5	A7B111
446	RLFICISGV	P5	A7B113
447	RLLGQTDMAV	P4	A7B115
448	RLMAGQQQV	P4	A7B116
449	RLQPMISV	P4	A7B117
450	RLTQMSNLV	P4	A7B118
451	RMKRLPVAV	P5	A7B120
452	RMQCVAVFAV	P5	A7B121
453	RQLPQMSKV	P5	A7B122
454	RSFDEVEGV	P5	A7B124
455	RTGPHILIV	P5	A24B111
456	SLAECGARGV	P4	A24B113
457	SLASWDVPV	P5	A8B111
458	SLCRLWPV	P5	A8B113
459	SLHGHVAAV	P4	A8B115
460	SLIEFDTLV	P5	A8B116
461	SLILSFQRV	P4	A8B117
462	SLLHTTFPHRQV	P5	A8B118
463	SLPMIATV	P5	A8B120
464	SLPSSMEIAV	P4	A8B121

Supplementary Table 3 (continued). **HLA-A*02:01 neoantigen sequences predicted from five colorectal cancer (CRC) patients, with DNA barcode annotations.** V, viral epitope

#	Sequence	Patient	Barcode
465	SLPVTSLSSV	P5	A8B122
466	SLQWPLKSRV	P4	A8B124
467	SLRTDCLLAV	P5	A25B111
468	SLVPEREKMLV	P5	A25B113
469	SLWYLNLTV	P5	A9B111
470	SMFVGSDTV	P5	A9B113
471	SMRECALHTV	P5	A9B115
472	SMTCKVMTSWAV	P5	A9B116
473	SQMDGLEV	P5	A9B117
474	SQVQLAIQV	P5	A9B118
475	SVFPNILNV	P4	A9B120
476	TALDVLANV	P5	A9B121
477	TETFALILYV	P5	A9B122
478	TLEERTSSV	P5	A9B124
479	TLGAALPPWPV	P4	A26B111
480	TLGAMDLGV	P5	A26B113
481	TLRQTTSVPV	P5	A19B115
482	TLSVIRDYLV	P5	A19B116
483	TLVEELITV	P5	A19B117
484	VIGAVVATV	P4	A19B118
485	VISAISEAV	P5	A19B120
486	VLAENVNMCV	P5	A19B121
487	VLKPCLIPV	P5	A19B122
488	VLKPFLFTV	P5	A19B124
489	VLLQSESGTAPV	P5	A22B115
490	VLLVLVLAV	P4	A22B116
491	VLMGCWLEV	P5	A22B117
492	VLSKGEIVV	P5	A22B118
493	WLASGRPCV	P5	A20B115
494	WLIVLTQLV	P5	A20B116
495	WLPKMPPFV	P5	A20B117
496	WLRELTSIV	P3	A20B118
497	WMTMDHLLV	P5	A20B120
498	WVLAALLAV	P5	A20B121
499	YLAHTVNAYKLV	P5	A20B122
500	YLEQLKMTV	P3	A20B124
501	YLGDILLAV	P5	A23B115
502	YLPPGFMFKV	P3	A23B116
503	YLPRTMDFGINV	P5	A23B117
504	YLSGRQKFWV	P2	A23B118
505	YMACKDEGCKLV	P3	A21B115
506	YQSAGITGV	P5	A21B116
507	YTWLGAMPV	P4	A21B117
508	SLWGNPTQY	P5	A21B118

#	Sequence	Origin	Barcode
V1	GILGFVFTL	FLU MP1	A21B120
V2	CLGGLLTMV	EBV LMP2	A21B121
V3	GLCTLVAML	EBV BMLF1	A21B122
V4	FLYALALLL	EBV LMP2	A21B124
V5	NLVPMVATV	CMV pp65	A24B115
V6	YVLDHLIVV	EBV BRLF1	A24B116
V7	VLEETSVML	CMV IE1	A24B117
V8	ILKEPVHGV	HIV Pol	A24B118
X	No peptide control		A25B115

Supplementary Table 4. **Top 10 peptide specificities retrieved from the five human colorectal cancer (CRC) patients included in this study.**

	Patient 1		Patient 2						Patient 3						Patient 4					Patient 5		BC 83	BC 88
						TIL subsets						TIL subsets					TIL subsets						
Sample ->	TILs	PBMCs	TILs	PBMCs	LN	DP	DN	SP39	TILs	PBMCs	LN	DN	SP39	SP103	TILs	LN	DP	DN	SP39	TILs	PBMCs		
Top 10 specificities 1	#15	V5	V5	V5	V5	**#264**	V5	V7	**#63**	V6	V2	**#502**	**#400**	**#400**	V4	V4	**#52**	V4	**#115**	**#128**	**#457**	V6	V3
2	#110	V3	**#264**	V7	V6	**#110**	V3	#117	#264	V1	V6	#117	**#502**	V6	**#115**	V6	V4	**#19**	V6	**#474**	**#418**	V3	V5
3	#209	V1	#400	V3	V3	#400	V6	#404	**#209**	V3	V3	V6	#264	**#502**	V6	**#115**	**#90**	**#276**	**#490**	**#64**	**#114**	V4	V1
4	#264	V4	#117	V6	V2	#119	V7	V5	#117	V2	V4	#264	#117	#110	**#31**	**#343**	**#19**	**#115**	**#328**	**#480**	**#494**		
5	#224	V2	#404	#199	**#186**	V6	V4	**#186**	**#399**	V4	V1	**#400**	#186	#186	**#19**	**#490**	**#343**	**#9**	**#24**	**#410**	**#398**		
6	#199	#502	**#149**	**#186**	#83	#502	#502	**#264**	#110	**#502**	#404	#186	**#316**	**#496**	**#490**	**#363**	**#353**	**#31**	**#19**	**#446**	**#64**		
7	#400	**#117**	V6	#294	**#264**	**#186**	#209	#199	#149	#404	**#400**	V3	#119	V7	**#52**	**#449**	**#115**	**#490**	**#367**	**#398**	**#109**		
8	#83	#400	**#186**	#209	V7	#404	#400	#209	#186	#264	V5	**#316**	V6	#117	**#448**	**#306**	**#438**	**#118**	**#118**	**#5**	**#256**		
9	#63	V6	#119	V1	V1	V5	**#264**	V4	V6	**#400**	#110	#119	**#209**	**#500**	**#53**	**#19**	**#367**	**#52**	**#414**	**#79**	**#319**		
10	#316	#110	#209	#399	**#15**	#117	#119	**#149**	**#316**	**#209**	#186	#404	**#138**	#504	**#191**	**#215**	**#51**	**#448**	**#438**	**#354**	**#436**		

Various samples from five patients were stained with a pool of HLA-A*02:01 multimers thermally exchanged for a selection of predicted neoantigens and common viral antigens (denoted with V, sequences listed in Table 1 and Table S3). Barcodes corresponding to peptides highlighted in red were detected at significantly increased levels (log fold change ≥2) compared to baseline reads; peptides in yellow at slightly increased levels (1≤ log fold change ≤2). Underlined peptides in bold are peptides predicted specifically for that patient. TIL, tumor-infiltrating lymphocyte; PBMC, peripheral blood mononuclear cell; LN, lymph node; DP, double positive TIL subset expressing both CD39 and CD103; DN, double negative; $SP39^{+}$, single positive expressing CD39; $SP103^{+}$, single positive expressing CD103.

Supplementary Table 5. **H2-K^b neoantigen sequences predicted from MC38 mice, with DNA barcode annotations**

#	Sequence	Barcode
1	SIIVFNLL	A1B212
2	IVFNLLEL	A1B213
3	FVIDFKPL	A1B200
4	KFNFKTAL	A1B201
5	ALPVRFSL	A1B202
6	MARPWGLL	A1B203
7	RHCWYLAL	A1B205
8	FSLQFALL	A1B204
9	SSMVPSAL	A1B206
10	FALLMGTL	A1B207
11	RNRRIFAL	A1B208
12	SPFLLITL	A1B209
13	LAPIKFAL	A1B210
14	VNSIHALL	A1B211
15	ASASLSRL	A2B200
16	VSLWPDLL	A2B201
17	TAIELGTL	A2B202
18	STPQLLPL	A2B203
19	VLESYLNL	A2B205
20	VGPRYDFL	A2B204
21	KILTFDRL	A2B206
22	LTFDRLAL	A2B207
23	EAERFANL	A2B208
24	VTVFVNNL	A2B209
25	PMFLFKTL	A2B210
26	SASRYALL	A2B211
27	SSIKVVGL	A3B200
28	VNMDGASL	A3B201
29	TVVGLSNL	A3B202
30	TGSVFGEL	A3B203
31	ANVLFFGL	A3B205
32	QILVFLIL	A3B204
33	LGVLFSQL	A3B206
34	YMYVPTAL	A3B207
35	LGSIFSTL	A3B208
36	RSVLHGCL	A3B209
37	FINLYGLL	A3B210
38	IHPVMSTL	A3B211
39	AALSPASL	A5B200
40	VQFMSCNL	A5B201
41	GAFVLQLL	A5B202
42	SSFVPVGL	A5B203
43	TSIGMLYL	A5B205
44	RLYETFNL	A5B204
45	FTPSHPPL	A5B206
46	RTLCVGNL	A5B207
47	LAIMTQHL	A5B208
48	AWVPFGGL	A5B209
49	FSYIVELL	A5B210
50	LTFFHSGL	A5B211
51	FTFLFFAL	A6B200
52	VVWFFTFL	A6B201
53	TFLFFALL	A6B202
54	RGLIRYRL	A6B203
55	AVIGYSLL	A6B205
56	VTLKPPFL	A6B204
57	SSSTAAAL	A6B206
58	SNHVLGHL	A6B207
59	LSVEPFRL	A6B208
60	LGYSSVGL	A6B209
61	VGLPWVTL	A6B210
62	VSRHHRAL	A6B211
63	RTVLRLLSL	A7B200
64	VVIAIFIIL	A7B201
65	YSMGKDAGL	A7B202
66	VAVLPVLSL	A7B203
67	IGACKAMNL	A7B205
68	VGQSVWLGL	A7B204
69	VMIAGKVAL	A7B206
70	CVVPFTDLL	A7B207
71	IVGHFYGGL	A7B208
72	VAQTPHGFL	A7B209
73	LDFGFWHEL	A7B210
74	IANFQLCPL	A7B211
75	ISRDLASML	A8B200
76	VKRTRFLRL	A8B201
77	SHPRRHRRL	A8B202
78	STQMHRALL	A8B203
79	TQMHRALLL	A8B205
80	VQKKFSRNL	A8B204
81	ISYDPDTCL	A8B206
82	SMPSAKVSL	A8B207
83	LGVCMYGML	A8B208
84	LAQKIHQNL	A8B209
85	LYLSSRSLL	A8B210
86	CSYLPELPL	A8B211
87	QVFKVIGNL	A9B200
88	ASLLPSMPL	A9B201
89	WNCPFSQRL	A9B202
90	QLYLLCCQL	A9B203
91	VQLASRSLL	A9B205
92	MSYFLQGTL	A9B204
93	SSPYSLHYL	A9B206
94	STSFNFNSL	A9B207
95	QVVKYHRVL	A9B208
96	VVKYHRVLL	A9B209
97	GSWAYCRAL	A9B210
98	KLYTRYAFL	A9B211
99	ASIIVFNLL	A1B214
100	IIVFNLLEL	A1B215
101	GKILTFDRL	A1B216
102	GGKILTFDRL	A1B217
103	No peptide control	A1B219
104	SIINFEKL	A1B218