Supplementary Data 1. Alanine substitutions and position variants of APNCYGNIPL. Applied in Supplementary Fig. 2

Substitution	Sequence
original	APNCYGNIPL
P1 -Ala	APNCYGNIPL
P2 -Ala	AANCYGNIPL
P3 -Ala	APACYGNIPL
P4 -Ala	APNAYGNIPL
P5 -Ala	APNCAGNIPL
P6 -Ala	APNCYANIPL
P7 -Ala	APNCYGAIPL
P8 -Ala	APNCYGNAPL
P9 -Ala	APNCYGNIAL
P10 -Ala	APNCYGNIPA
9-mer	APNCYGNIP
P1 -Ala	APNCYGNIP
P2 -Ala	AANCYGNIP
P3 -Ala	APACYGNIP
P4 -Ala	APNAYGNIP
P5 -Ala	APNCAGNIP
P6 -Ala	APNCYANIP
P7 -Ala	APNCYGAIP
P8 -Ala	APNCYGNAP
P9 -Ala	APNCYGNIA

Position variant	Sequence
original	APNCYGNIPL
-6 towards N-term	CKLLEIAPNC
-5 towards N-term	KLLEIAPNCY
-4 towards N-term	LLEIAPNCYG
-3 towards N-term	LEIAPNCYGN
-2 towards N-term	EIAPNCYGNI
-1 towards N-term	IAPNCYGNIP
+1 towards C-term	PNCYGNIPLM
+2 towards C-term	NCYGNIPLMK
+3 towards C-term	CYGNIPLMKA
+4 towards C-term	YGNIPLMKAA
+5 towards C-term	GNIPLMKAAF
+6 towards C-term	NIPLMKAAFK

Nature Biotechnology: doi:10.1038/nbt.4303

Supplementary Data 2. Alanine substitutions and position variants of EWWRSGGFSF. Applied in Supplementary Fig. 2

Substitution	Sequence
original	EWWRSGGFSF
P1 -Ala	AWWRSGGFSF
P2 -Ala	EAWRSGGFSF
P3 -Ala	EWARSGGFSF
P4 -Ala	EWWASGGFSF
P5 -Ala	EWWRAGGFSF
P6 -Ala	EWWRSAGFSF
P7 -Ala	EWWRSGAFSF
P8 -Ala	EWWRSGGASF
P9 -Ala	EWWRSGGFAF
P10 -Ala	EWWRSGGFSA
9-mer	EWWRSGGFS
P1 -Ala	AWWRSGGFS
P2 -Ala	EAWRSGGFS
P3 -Ala	EWARSGGFS
P4 -Ala	EWWASGGFS
P5 -Ala	EWWRAGGFS
P6 -Ala	EWWRSAGFS
P7 -Ala	EWWRSGAFS
P8 -Ala	EWWRSGGAS
P9 -Ala	EWWRSGGFA

Position variant	Sequence
original	EWWRSGGFSF
-6 towards N-term	GTTKFKEWWR
-5 towards N-term	TTKFKEWWRS
-4 towards N-term	TKFKEWWRSG
-3 towards N-term	KFKEWWRSGG
-2 towards N-term	FKEWWRSGGF
-1 towards N-term	KEWWRSGGFS
+1 towards C-term	WWRSGGFSFG
+2 towards C-term	WRSGGFSFGK
+3 towards C-term	RSGGFSFGKA
+4 towards C-term	SGGFSFGKAY
+5 towards C-term	GGFSFGKAYE
+6 towards C-term	GFSFGKAYEY

Nature Biotechnology: doi:10.1038/nbt.4303

Supplementary Data 3. APNCYGNIPL peptide variants applied in Fig. 1a-c,g and Supplementary Fig. 3 and Supplementary Fig. 5. See enclosed excel file.

Supplementary Data 4. EWWRSGGFSF peptide variants applied in Fig. 1d-f,h-k and Supplementary Fig. 4 and Supplementary Fig. 6. See enclosed excel file.

Supplementary Data 5. Information content of the structure-based model of APNCYGNIPL bound to HLA-B*0702. Related to Supplementary Fig. 7.

Position	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	0.88	3.27	1.09	0.22	0.25	4.24	0.71	0.26	0.41	2.09

Supplementary Data 6. Information content of the structure-based model of EWWRSGGFSF bound to HLA-A*2402. Related to Supplementary Fig. 8.

Position	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	0.35	1.32	0.66	0.21	0.17	3.87	3.44	0.26	0.30	2.41

Supplementary Data 7. EWWRSGGFSF double substitution variants applied in Fig. 1i-k and Supplementary Fig. 9. The MHC multimer panel comprised these peptides and the single substitution peptides listed in Supplementary Data 4. See enclosed excel file.

Supplementary Data 8. SIINFEKL peptide variants applied in Fig. 2a-d, Supplementary Fig. 10 and Supplementary Fig. 11. See enclosed excel file.

Nature Biotechnology: doi:10.1038/nbt.4303

Supplementary Data 9. Sequence information on the TCRs of the 12 HLA-A*0201$_{KLL}$ responsive clones. Related to Fig. 2e-g and Supplementary Fig. 12

Patient	Clone	TRA sequence*	TRB sequence*
w678	1	CVLNNNDMRF	CAIRQFDANTGELFF
	1.5	unknown alpha	CAIRQFDANTGELFF
	2	CAFRVSHDMRF	CASSIIAGSSYNEQFF
	3	CVVATYSGGGADGLTF	CASSIIAGSSYNEQFF
	4	CVVATYSGGGADGLTF	CASSSGNPSTDTQYF
w876	2	CVVTEYSGGGADGLTF	CASRGQNTGELFF
	5	CAYNQGGKLIF	CASSVLNTGELFF
	5.5	unknown alpha	CASSVLNTGELFF
	6	CVVPLYSSASKIIF	CASSDTPDLNTEAFF
	9	CVGNNNDMRF	CAIRRQDQNTGELFF
w830	4	CVLNNNDMRF	CASSILGASNQPQHF
w683	1	CVVALYSGGGADGLTF	CASRSQNYYGYTF

***** Although, the TCR recognition pattern is ultimately determined by the TCR sequence, the specific involvement of alpha and beta variable regions did not explain all the differences observed between the TCR fingerprints (**Fig. 2e**). Some TCRs with shared beta sequences (TRB) showed very similar fingerprints, e.g. clone 2 and clone 3 of patient w678, while other TCRs also sharing the same beta domain were substantially different in their recognition pattern, e.g. clone 5 and 5.5 of patient 876. Hence, the sequence of the alpha and beta variable regions were not sufficient to predict the TCR recognition similarities across the T cell clones.

Supplementary Data 10. KLLEIAPNC peptide variants applied in Fig. 2 and Supplementary Fig. 12. See enclosed excel file.

Supplementary Data 11. The estimated promiscuity of the 12 MCC clones screened in Fig. 2e-g and Fig. 3.

Clonal TCR	# of potential targets (adjusted for pMHC anchors)	# of potential targets (not adjusted)
w678_1	432	1,296
w678_1.5	810	3,645
w678_2	28,080	425,880
w678_3	25,740	360,360
w678_4	264	1,408
w876_2	24	24
w876_5	48	56
w876_5.5	216	288
w876_6	360	960
w876_9	384	1,344
w683_1	12	18
w830_4	264	528
OT-1	13,056	864,688
OT-3	62,208	1,749,600

Supplementary Data 12. The 75 peptides that comprise the top-10 of the FIMO-based priority list that best match each of the TCR fingerprints of the 12 HLA-A*0201 MCCKLL engaging TCRs. Including the original KLLEIAPNC target. Related to Fig. 3 and Supplementary Fig. 12. See enclosed excel file.

Supplementary Data 13. Expression characteristics of ST6 N-acetylgalactosaminide. Related to Fig. 3.

Gene	ST6GALNAC3		
Protein	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3		
Organism	Human		
Function	Sialyltransferase		
Pathway	Protein glycosylation		
Expression	**Tissue**	**Cell types**	**Stain**
	Muscle tissue	Myocytes	Medium /
	Endocrine tissue	Glandular cells	Low
	Gastrointestinal tract	Glandular cells	Low
	Breasts and female reproductive	Squamous epithelial cells	Low
	Brain	Neuronal cells	Low

Supplementary Note 1. Concept of establishing TCR recognition patterns

We examine the requirements for TCR interaction with pMHC by the use of collections of peptide variants. The variants are based on the peptide that was originally described as the target of the given TCR, and includes peptides generated from sequentially substituting one or more positions of the originally identified target with all naturally occurring amino acids (**Supplementary Fig. 1**). All peptide variants are loaded onto MHC molecules and multimerized on a PE-labeled dextran backbone that is also individually labeled with different DNA barcodes, such that each pMHC variation is specifically associated with a unique DNA barcode [16]. Subsequently, T cells expressing a given TCR (either TCR-transduced T cells or T-cell clones) are stained with the whole pool of DNA barcode-labeled MHC multimers generated from the given collection of peptide variants (**Supplementary Fig. 1**). TCR clonal T cells will bind several different pMHCs in a hierarchy-based manner that is determined by the affinity between the pMHC and the given TCR (**Supplementary Fig. 1**). Upon selection of clonal MHC multimer-binding T cells, the MHC multimer-associated DNA barcodes can be retrieved and the number of DNA barcode reads reflects the relative binding hierarchy amongst the different pMHCs included in the analyses. 'Barracoda' (http://www.cbs.dtu.dk/services/barracoda), the software developed to analyze sequencing data from DNA barcode-labeled MHC multimers [16], is used to retrieve the number of read counts for each pMHC-associated DNA barcode, and to calculate the $\log_2$FC in read counts relative to the mean read counts of three baseline samples. The baseline samples consist of unprocessed aliquots of the full library of DNA barcode-labeled MHC multimers. Unprocessed in the meaning that the multimers are not mixed with cell samples, but the DNA barcodes within each aliquot are directly amplified and sequenced. Thus, the baseline samples enables normalization of the DNA barcode reads of the samples relative to the actual input of DNA barcodes from the MHC multimer panel. The methodology has some similarities to deep mutational scanning[26,27]

that determines the impact of certain mutations on a phenotypical or functional readout. Using the DNA barcode-tagged MHC multimers we can select very precisely for T cells binding to the pMHC complexes, and through a flow cytometry based cell sorting procedure, retrieve only the DNA barcodes associated to pMHC complexes of relevance for T cell binding, out of a large pool of irrelevant pMHC associated DNA barcodes, even after only a single round of selection[16].

When assessing the TCR fingerprint, the position of a given pMHC within the binding hierarchy is defined based on the $\log_2$FC value given by 'Barracoda'. The contribution of each individual amino acid at a given position can be determined based on the presence of a given peptide in the binding hierarchy. Such contributions can be visualized as a heat map or translated into a sequence logo, referred to as the TCR fingerprint, which serves to visualize the TCR recognition pattern, with the size of the one-letter amino acid code reflecting the selectivity of that TCR at the given position (**Supplementary Fig. 1**). The TCR fingerprinting strategy can be applied to any TCR restricted to a foldable MHC class I molecule, however it does required the generation of a peptide amino acid substitution library unique for each TCR specificity, and consequently may be advantageous specifically for collections of TCRs raised for the same specificity.

Supplementary Note 2: Feasibility and sensitivity for resolving the binding hierarchy of TCR-pMHC interactions

To prove the feasibility of our strategy for resolving the binding hierarchy of multiple pMHCs that are recognized by the same TCR, we generated a collection of DNA barcode-labeled MHC multimers that contained variants of a MCP*y*V-derived 10-mer HLA-B*0702-restricted peptide, APNCYGNIPL. T cells transduced with a TCR recognizing this target were stained with a smaller panel of DNA barcode-labeled MHC multimers generated from single amino acid substitutions using only alanine (A) and

length (9-mer) or position variants of the original APNCYGNIPL peptide, resulting in 34 peptide variants (**Supplementary Fig. 2** and **Supplementary Data 1**). Length (9-mer) or position variants were included to ensure that TCR recognition was only feasible at the determined position of the protein, and the original length of the peptide. Although recognition of position and length variants are less likely to occur, compared to single amino acid substitution within the original peptide, examples of TCRs recognizing both 9 and 10mer variants of the same peptide has been reported[51]. To ensure that low-avidity interactions were not outcompeted by high-avidity interactions in a parallel screening, we first used this panel as a basis in a set of three consecutive analyses where we 1) included the full collection of MHC multimer variants, 2) excluded the multimers carrying the original peptide and 3) excluded the multimers carrying the original peptide and the additional four peptide variants with the strongest pMHC-TCR interactions (**Supplementary Fig. 2**). For the same purpose we repeated the analyses, either applying 5- or 25-fold reduced amounts of the multimer reagents, compared to what was used in the initial analysis (i.e., 4.6 nM and 0.92 nM of each pMHC, **Supplementary Fig. 2**). The results were comparable, and showed that any potential competition between the different pMHCs did not limit our detection of low-affinity binders and that our platform for determining the binding hierarchy was sensitive enough to detect both high-avidity and low-avidity interactions simultaneously. Next, we applied collections of MHC multimers were the peptide variants were generated from substituting with either lysine (K), which has a positively-charged side chain, or glutamic acid (E), which has a negatively-charged side chain. These parallel staining's verified that alanine substitutions alone are insufficient to describe the critical pMHC-TCR interaction points (**Supplementary Fig. 2)**.

Supplementary Note 3. Investigation of peptide-MHC stability

MHC- ELISA

The MHC-ELISA was performed as previously described[19]. In brief, specific peptide-MHC complexes were generated by UV-mediated peptide exchange (2.5 nmol/L, based on 100% rescue) and incubated on streptavidin-coated plates (2 mg/mL) for 1 h at 37 °C. Subsequently the peptide-MHCs were incubated with horseradish peroxidase-conjugated β_2-m antibody (1.4 mg/mL; Acris GmbH), which bind only to correctly folded MHC molecules, for 1 h at 37 °C. After washing, a colorimetric reaction was initiated by the addition of tetramethylbenzidine peroxidase substrate. Absorbance was determined (450 nm wavelength) by an ELISA reader.

Bead capture of correctly folded MHC

The stability of the HLA-A*2402-EWWRSGGFSF peptide variations were measured by capturing all correctly folded pMHCs within the combined MHC multimer panel with beads coated with anti-HLA-ABC monoclonal antibodies (clone W6/32, Biolegend 311334) (**Supplementary Fig. 6**). The antibody coated beads were produced from QuantumPlexM Streptavidin beads (Bangs Laboratories) by first washing according to manufactures instructions and incubating with antibody (10,000 beads:25 ng W6/32 antibody) for 30 min at 4 °C. The beads were washed twice in barcode-cytometry buffer (PBS + 0.5% BSA + 100 μg/mL herring DNA + 2 mM EDTA) and were subsequently incubated with 1 uM biotin (Avidity, Bio200) for 30 min at 4 °C. For capture of correctly folded MHC, the full DNA barcode-labeled MHC multimer library (1.15 pmol of each pMHC =0.22 μmol total MHC) were incubated with 10,000 W6/32 coated beads for 30 min at 4 °C, then washed three times in barcode-cytometry buffer (PBS + 0.5% BSA + 100 μg/mL herring DNA + 2 mM EDTA) before sorting all

beads in a FSC-SSC plot. The sorted beads were centrifuged for 10 min at 5,000*g* and the buffer was removed. The pellet was stored at −80 °C in a minimal amount of residual buffer (<20 µL). The pMHC associated DNA barcodes were amplified and sequenced as described previously[16], and the presence of correctly folded pMHCs within the MHC multimer panel could be verified through the presence of the corresponding DNA barcode (**Supplementary Fig. 6**).

Supplementary Note 4. T-cell functional assays

EC_{50} determination

The EC_{50} values for MCC $CD8^+$ T cell clones were determined in a previous study[22]. Briefly, secreted IFN-γ was measured after incubating 2 x 10^4 clonal T cells with 5 x 10^4 peptide-pulsed T2 cells (peptide sequence KLLEIAPNC). Peptide was used at $\log_{10}$ dilutions to a final concentration from 10^6 to 10^{12} molar in 200 mL of T-cell medium for 36 hours. IFN-γ in cell culture supernatants was assayed by ELISA according to the manufacturer's recommendations (Human IFN gamma ELISA Ready-SET-Go Kit; Affymetrix). The amount of IFN-γ secretion by each T-cell clone was used to estimate the EC_{50} of each clone, i.e., the amount of peptide leading to 50% of the maximum level of IFN-γ secretion.

Intracellular cytokine staining

MCC $CD8^+$ T-cell clones and T2 cells were thawed and washed in RPMI + 10% FCS. T2 cells were resuspended in X-vivo + 5% human serum, pulsed with 5 µM peptide, and left at room temperature for 2 h. 150,000 T cells were stimulated with pulsed T2 cells at a 1:1 ratio. GolgiPlug (BD, 555029) (1/250 dilution) and PE-conjugated anti-CD107 (BD, 555801) (1/10 dilution) were added before incubation for 4 h at 37 °C and 5% CO_2. Cells were washed twice in PBS + 10% FCS and stained with surface

antibodies: FITC-conjugated anti-CD3 antibody (BD, 345763) and BV-510-conjugated anti-CD8 antibody (BD, 563256 (final dilution 1/10 of each antibody) and dead cell marker (LIVE/DEAD Fixable Near-IR; Invitrogen L10119) (final dilution 1/1,000) and incubated for 30 min at 4 °C. Cells were washed twice in PBS + 10% FCS and incubated overnight at 4 °C in fixation buffer (1:4, eBioscience 00-5123-43, to diluent eBioscience 00-5223-56). The following day, cells were washed and resuspended in permeabilization buffer (1:10 buffer to water, eBioscience 00-8333-56) and stained with intracellular antibodies PE-Cy7-conjugated anti-TNFα antibody (BioLegend, 502930), APC-conjugated anti-IFNγ antibody (BD 341117), and Brilliant Violet 421 anti-IL-2 antibody (BioLegend, 500328) (final dilution 1/10 of each antibody), for 30 min at 4 °C. Cells were washed in permeabilization buffer and resuspended in PBS + 10% FCS.

51. Andersen, R. S. *et al.* High frequency of T cells specific for cryptic epitopes in melanoma patients. *Oncoimmunology* **2,** e25374 (2013).