



T cell receptor fingerprinting enables in-depth characterization of the interactions governing recognition of peptide-MHC complexes

Bentzen, Amalie Kai; Such, Lina; Jensen, Kamilla Kjærgaard; Marquard, Andrea Marion; Jessen, Leon Eyrych; Miller, Natalie J.; Church, Candice D; Lyngaa, Rikke Birgitte; Koelle, David M.; Becker, Jürgen C.

Total number of authors:
16

Published in:
Nature Biotechnology

Link to article, DOI:
[10.1038/nbt.4303](https://doi.org/10.1038/nbt.4303)

Publication date:
2018

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):

Bentzen, A. K., Such, L., Jensen, K. K., Marquard, A. M., Jessen, L. E., Miller, N. J., Church, C. D., Lyngaa, R. B., Koelle, D. M., Becker, J. C., Linnemann, C., Schumacher, T. N. M., Marcatili, P., Nghiem, P., Nielsen, M., & Hadrup, S. R. (2018). T cell receptor fingerprinting enables in-depth characterization of the interactions governing recognition of peptide-MHC complexes. *Nature Biotechnology*, 36(12), 1191-1196. <https://doi.org/10.1038/nbt.4303>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Title:

T cell receptor fingerprinting enables in-depth characterization of the patterns that govern peptide-MHC recognition

Authors: Amalie K. Bentzen^{1α}, Lina Such^{1α}, Kamilla K. Jensen², Andrea M. Marquard¹, Leon E. Jessen^{1,2}, Natalie J. Miller³, Candice D. Church³, Rikke Lyngaa^{1§}, David M. Koelle^{4,5}, Jürgen C. Becker⁶, Carsten Linnemann^{7#}, Ton N.M. Schumacher⁷, Paolo Marcatili², Paul Nghiem³, Morten Nielsen^{2,8}, Sine R. Hadrup^{1*}

¹Department of Micro and Nanotechnology, Technical University of Denmark, Lyngby, Denmark

²Department of Bio and Health Informatics, Technical University of Denmark, Lyngby, Denmark

³Department of Medicine, Divisions of Dermatology, University of Washington, Seattle, WA

⁴Department of Medicine, University of Washington, Seattle, WA

⁵Division of Vaccine and Infectious Diseases, Fred Hutchinson Cancer Research Center, Seattle, USA

⁶Translational Skin Cancer Research, University Hospital Essen and University of Duisburg-Essen, Germany

⁷Department of Molecular Oncology & Immunology, Oncode Institute, The Netherlands Cancer Institute, Amsterdam, The Netherlands

⁸Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, Buenos Aires, Argentina

[‡]These authors contributed equally

[#]Current affiliation: Kite Pharma Europe, Amsterdam

[§]Current affiliation: Gilead Sciences, Denmark

*Correspondence should be addressed to: sirha@dtu.dk

Abstract:

The promiscuous nature of T-cell receptors (TCRs) is fundamental for our ability to recognize a large range of pathogens; however, this feature makes it challenging to understand and control T-cell recognition¹. Existing technologies provide limited information about the key requirements for T-cell recognition and the ability of TCRs to cross-recognize structurally related elements^{2,3}. Herein we present a proof-of-concept of a novel ‘one-pot’ strategy to establish the patterns that govern TCR recognition of peptide-major histocompatibility complex (pMHC). We determine the affinity-based hierarchy of TCR interactions with MHC loaded with peptide variants, and apply this knowledge to understand the recognition motif, here termed the TCR fingerprint. The TCR fingerprints of 16 different TCRs were identified and used to predict and validate cross-recognized peptides from the human proteome. The identified fingerprints differed amongst TCRs recognizing the same epitope, demonstrating the value of this strategy for understanding T-cell interactions and assessing potential cross-recognition prior to selection of TCRs for clinical development.

Main:

The antigen specificity of T cells is conferred by the highly variable complementarity-determining regions (CDRs) of the TCR that interact with the pMHC⁴. Cellular immunity requires a pool of naïve T cells (the T-cell repertoire) that can recognize a multitude of potential pMHC antigens that originate from infections or cellular transformation. If a given TCR were only able to recognize a single combination of peptide and MHC, an individual would need more than 10^{15} CD8⁺ T cells to provide efficient coverage of all potential foreign peptides, whereas it is currently estimated that an individual only have around 10^7 - 10^8 different T cells⁵⁻⁷. The promiscuity of TCRs allows the recognition of numerous different pMHCs by each T cell, which serves to broaden the recognition space and ensure the effective recognition of most possible targets. It has been demonstrated that a single TCR can interact with more than 1 million different peptide-MHC combinations^{1,8,9}, however, due to technical limitations, little is known about the extent of promiscuity for any particular TCR or the pattern that govern the exact hierarchical avidities to various pMHCs. Peptide-MHC display libraries in yeast^{10,11} have very elegantly enabled interrogations of TCR cross-recognition and even identification of TCR targets without any preexisting knowledge of what that TCR recognize¹². Although such strategies theoretically offer an unbiased approach to determine the TCR recognition profile, not all possible positions are equally well represented and not all TCRs can be fully characterized¹². Additionally, so far, yeast display strategies has only been developed for the characterization of HLA-A*0201 restricted TCRs, and the technique is restricted to a few specialized laboratories. Another approach involves the functional interrogation of T cells exposed to peptides with only one fixed amino acid and a random composition for the rest of the sequence^{13,14}, however such approaches requires many TCR expressing cells, and do not provide an in-depth hierarchy of

binding interactions, as the TCR fingerprinting presented herein. The overall strategy is illustrated in **Supplementary Fig. 1** and the feasibility of this approach is presented in **Supplementary Fig. 2, Supplementary Table 1 and 2** and described in **Supplementary Note**.

Currently, the most widespread strategy for resolving potential cross-recognition of a TCR involves investigating the effect of single-position alanine substitutions on the reactivity of T cells². This strategy provides limited knowledge of a given TCRs recognition pattern, insufficient for describing the full TCR recognition profile (**Supplementary Fig. 2**). Here we present a novel approach that enables in-depth characterization of the TCR recognition patterns that are decisive for interaction with pMHC, which can be easily implemented in most immunology laboratories and is applicable across all foldable MHC molecule's. We leverage on the use of DNA barcode-labeled MHC multimers¹⁵ that uniquely allows a 'one-pot' strategy where the interaction of one clonal TCR with multiple related pMHC epitopes can be assessed simultaneously. This is possible 1) because the sequencing-based readout of the DNA barcode-based MHC multimer analysis allows direct quantification of the relative interactions of a given TCR to multiple pMHC variants, and 2) the high complexity of DNA barcodes that enables the generation of large libraries of differently labeled pMHC multimers. Based on such analysis, a hierarchy of pMHC interactions can be determined.

We first investigated the recognition pattern of two different TCRs isolated from individual Merkel cell carcinoma (MCC) patients, each recognizing a different Merkel cell polyomavirus (MCPyV)-derived peptide: 1) APNCYGNIPL restricted to HLA-B*0702 and 2) EWWRSGGFSF restricted to HLA-A*2402¹⁶. A collection of DNA barcode-labeled MHC multimers was produced for each TCR, containing all the peptides generated from sequentially

substituting every single amino acid of the two original 10-mer peptides with all naturally-occurring amino acids (n=191 for each library) (**Supplementary Table 3 and 4**). For the HLA-B*0702_{APN}-engaging TCR this data showed that the original amino acids asparagine (N) at position 3, tyrosine (Y) at position 5, and glycine (G) at position 6 were essential to maintain binding between the TCR and the MHC bound peptide. In contrast, there was some flexibility at positions 2, 7, 8, and 10 and the amino acids present at positions 1, 4, and 9 seemed to be the least critical for the TCR to recognize the MHC embedded peptide (**Fig. 1a-c**). For the HLA-A*2402_{EWV}-engaging TCR, the glycine (G) at position 7 and the phenylalanine (F) at position 8 were critical for maintaining the interaction between the TCR and the MHC-bound peptide. Position 2, 4, 5, 6, 9 and 10 was less restricted in terms of amino-acid requirements, but did display some selectivity. Barely any effect was seen when amino acids at positions 1 and 3 were substituted (**Fig. 1d-f**). Moreover, when applying the same two MHC multimer libraries to screen PBMCs from healthy donors, we found no noteworthy signal or weighted preference for any amino acids (**Fig. 1a, d and Supplementary Fig. 3 and Supplementary Fig. 4**).

Interestingly, results from screening of both TCRs showed that amino acid substitutions at the peptide-MHC anchor positions, which are predicted to impede peptide-MHC binding, can still allow the pMHC-TCR interaction (**Fig. 1g,h, Supplementary Fig. 5 and Supplementary Fig. 6**) This was experimentally validated using a APNCYGNIPL-based alanine substitution library to assess pMHC binding to HLA-B*0702 by MHC ELISA¹⁷, which showed that peptides with 40% reduced MHC binding capacity (compared to the original sequence) were still able to retain a similar level of pMHC-TCR interaction as the original MHC bound peptide (**Supplementary Fig. 5**). We moreover verified that all HLA-A*2402-EWVRSGGFSF peptide variants were stable during the course of the experiment, confirmed both by conformational dependent

enrichment of the pMHC library and direct MHC tetramer staining using peptide-MHC complexes with various binding affinities (**Supplementary Fig. 6**). Consequently the MHC anchor residues play a minor role in the TCR fingerprint (**Fig. 1c,f**), but will be important for the natural presentation of peptides.

To gain a deeper understand of the pMHC structures, we next generated an *in silico* structure-based model of each of the original peptides: APNCYGNIPL bound to HLA-B*0702 and EWWRSGGFSF bound to HLA-A*2402 (**Supplementary Fig. 7 and Supplementary Fig. 8 and Supplementary Table 5 and 6**). This served to visualize the ‘best fit’ of the respective peptides in the MHC pocket, and could explain partially the amino acids essential for TCR recognition of pMHC, particularly the glycine (G) at position 6 and 7 of the HLA-B*0702 and A*2402 bound peptides respectively. An unfavorable change in energy, when substituting with any amino acid at this position, suggests that glycine at these positions may have an important role in maintaining the peptide in a conformation that promotes TCR engagement, but these glycines are not necessarily direct interaction points for the TCR. Thus the TCR fingerprint illustrates 1) the amino acids essential for direct interaction between the TCR and the pMHC and 2) the moieties crucial for maintaining a peptide conformation that favors such interactions.

A potential important limitation of the current technology is the lack of assessment for TCR interactions arising from mutual amino acid variations at several defined positions. To determine the potential impact of such mutual variations we designed a peptide library consisting of two independent alterations at position 4-8. To minimize the library size, only high-affinity HLA-A*2402 binding peptides were included (n=782, **Supplementary Table 7**). The majority of the mutual substituted peptides showed decreased TCR interaction properties compared to the single position amino-acid variations when screened in parallel (**Fig. 1i**). Several mutual beneficial

amino-acid combinations could be identified at position 4 and 5 (**Fig. 1j and Supplementary Fig. 9**), reflecting the preferences and flexibility determined by the original fingerprint at these positions (**Fig. 1f**). Interestingly, alanine (A) at position 6 specifically allowed isoleucine (I) and methionine (M) at position 8 (**Fig. 1k**), the only alternative residues tolerated at this position as determined in the original fingerprint. For the restricted positions 7 and 8 no alternative amino-acid combination were tolerated despite the mutual substitutions (**Supplementary Fig. 9**). TCR recognition of selected peptides with multiple substituted amino acids was confirmed by direct MHC tetramer staining (**Supplementary Fig. 9**). Thus, although the mutual substitutions did reveal favorable interactions and limit the combinatorial space for possible peptide interaction partners it did not resolve novel interaction requirements that were not accounted for in the original fingerprint.

Having established this, we next set out to resolve the recognition pattern of different TCRs recognizing the same target. We first investigated two murine TCRs, from the OT-1 and OT-3 TCR transgenic T cells, reported to have high and low functional avidity respectively to the H-2Kb restricted peptide SIINFEKL¹⁸. These T cells were screened with a library of 153 DNA barcode-labeled MHC multimers holding single amino acid substitutions of the SIINFEKL peptide (**Supplementary Table 8**). The hierarchy of pMHCs (**Fig. 2a,b and Supplementary Fig. 10**) were again used to generate individual TCR fingerprints to visualize the amino acids critical for TCR recognition (**Fig. 2c,d**). It is evident that both TCRs had a high dependence of the original amino acids at position 6 and 7 (E and K respectively). But, while the OT-1 TCR was more flexible at these positions, tolerating other amino acids with the same properties, it had a higher dependence of the original amino acids (S and N) at position 1 and 4 respectively, than the OT-3 TCR. T cell staining using fluorescent-labeled MHC multimers carrying one of six

SIINFEKL variants supported the different binding properties of the two TCRs (**Supplementary Fig. 11**).

We next examined 12 different TCRs derived from 4 MCC patients¹⁹ (**Supplementary Table 9**). They all recognized the same HLA-A*0201-restricted 9-mer peptide KLLLEIAPNC, derived from the common region of the oncogenic proteins, large- and small T antigen of MCPyV. The individual TCR recognition patterns were determined by screening each clone with the same library of 192 DNA barcode-labeled MHC multimers holding the KLLLEIAPNC peptide variants (**Supplementary Table 10**). Interestingly, we observed a substantial variance in the TCR fingerprints of these TCRs clones, even amongst those derived from the same patients. The most recurring pattern was a preference for the hydrophobic amino acids isoleucine (I), phenylalanine (F), or valine (V) at position 5 (**Fig. 2e**), though the influence of these amino acids varied between the TCRs, where a weaker restriction at position 5 seemed to be associated with a strong requirement for proline (P) at position 7, and in some cases asparagine (N) at position 8. Furthermore, several TCRs required glutamic acid (E) or methionine (M) at position 4; these positions were, in some cases, more important for TCR recognition than the hydrophobic amino acid at position 5. Clone 5 from patient w876 was special for having an additional preference for amino acids with a positive side chain, lysine (K) or arginine (R), at position 1. This pattern of similarity was clearly evident when clustering the 12 TCR fingerprints in a hierarchical manner (**Fig. 2e** and **Supplementary Fig. 12**). We validated two of the TCR fingerprints with distinct characteristics (clone 2 and 5 from patient w876), by assessing the functional capacity of the corresponding T-cell clones for recognition of alanine substitution variants of KLLLEIAPNC (**Fig. 2f** and **Supplementary Fig. 12**). 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 (**Supplementary Table 9**). The different TCR recognition patterns of the individual clones also imply a quantifiable difference in the number of potential peptide sequences that may be recognized by each TCR. Based on the fingerprint profile, we calculated, for each clone any possible substitution that resulted in a similar or enhanced pMHC-TCR interaction compared to the original peptide sequence. With this simplistic approach that do not take into account potential mutual substitution biases (e.g. unfavorable amino acid combinations), and defines a fixed number of MHC anchor residues; the 12 TCRs analyzed here may recognize a range of 12 to 28,080 different peptide sequences with similar or increased affinity compared to their respective original peptides (**Supplementary Table 11**). Interestingly, we observed that the number of potential targets for a given TCR inversely correlated to the functional efficacy of the T cells in question (**Fig. 2g**). This indicates that the TCR recognition pattern may not only be useful to further characterize TCRs with respect to their range of potential pMHC targets, but may also hint to the functional capacity.

To investigate the range of potential pMHC targets we next used the TCR fingerprints of the 12 HLA-A*0201 KLLLEIAPNC-specific T-cell clones (**Fig. 2e**) to predict peptides from the entire human proteome, that may potentially be cross-recognized by each TCR. We used the “Find Individual Motif Occurrences” (FIMO) software²⁰ to create a priority list based on the likelihood of cross-recognition between the TCR and the human proteome. Based on the top 1000 sequences from this priority list, for each TCR, we generated a correlation matrix illustrating the overlap of potential cross recognition between the different clones (**Supplementary Fig. 12**). This overlap follows the patterns of the TCR fingerprint, with clone 2 and 3 (from w678) sharing very similar profiles, but very distinct for the other clones, and clone 5 (from w876) with a

unique profile. From the priority list we applied the top 10 sequences of each TCR (**Supplementary Table 12**), along with the original peptide, to experimentally evaluate the cross-recognition, through one combined DNA barcode-labeled MHC multimer library (n=75) used to stain all 12 T-cell clones. We could confirm some level of recognition against 25 of the 75 peptides (**Fig. 3a**, and **Supplementary Table 12**). We again observe distinct cross-recognition properties of clone 2/3 w678 and clone 5 w876. For the most prominent hit, related to clone 5, w876 we confirmed the cross-recognition towards peptide (KTVGIYPNA) by conventional fluorescent-based tetramer staining (**Fig. 3b**) and by ICS after stimulation with peptide-pulsed HLA-A*0201-expressing cells (**Fig. 3c**). This peptide had a 44% sequence overlap with the original peptide. From the pool of cross-reactive peptides, the vast majority was predicted as non-binders (**Fig. 3a** and **Supplementary Table 12**), minimizing the actual risk for cross-recognition of clinical relevance. However, low-affinity peptides have been identified as targets for T cell recognition²¹. The exemplified cross-recognized peptide KTVGIYPNA is predicted as a low-affinity ligands (netMHCpan %Rank score of 3.2), but with a confirmed functional recognition. It is derived from ST6 N-acetylgalactosaminide (**Supplementary Table 13**) expressed at medium to high levels in myocytes. Consequently muscle tissue would be at direct risk of attack if applying such TCR in a clinical setting.

In summary, our data demonstrate the feasibility of a ‘one-pot’ approach for generation of TCR fingerprints, and the utility of those for characterizing potential TCR cross-recognition. This is valuable because the promiscuity of TCRs provides an intrinsic challenge to the use of TCRs in clinical applications and increases the risk of autoimmune reactions²². Although TCRs from patients have passed the thymic selection process, the strategies applied for TCR gene therapy may overrule the additional peripheral tolerance mechanisms that work to further shape the

characteristics of our TCR repertoire to avoid cross-recognition of healthy tissues. Consequently critical adverse events may arise from the use of natural TCRs²³. Additionally, many TCR transgenic strategies have involved the optimization of TCR affinity, for example through deep mutational scanning^{24,25}, without understanding the exact role of such modification in terms of pMHC recognition profiles.

Direct evidence for pathological effects from TCR cross-recognition stem from a clinical trial of adoptive cell therapy using such an affinity-optimized TCR for transduction of T cells^{3,26}. Severe adverse events was observed in this trial due to cross-recognition of healthy tissue, including two cases of fatal toxicity linked to T-cell cross-reactivity between the melanoma-associated antigen (MAGE-A3) and a titin-derived peptide expressed in healthy cardiac cells^{3,26}. With only 55% sequence overlap between those two peptides, this case highlights the great challenge facing current preclinical evaluations^{3,26,27}. The fingerprinting strategy presented herein thus provides an attractive screening tool for understanding the TCR recognition profile in greater detail prior to selection of TCRs intended for clinical development.

Acknowledgments: We would like to thank all Merkel Cell Carcinoma patients and healthy donors contributing material to this study; Thorsten Joeris, William Agace and Katharina Lahl for sharing of splenocytes from OT-1 and OT-3 transgenic mice, Bente Rotbøl and Tripti Tamhane for excellent technical assistance, and the group of SRH for scientific discussions. This research was funded in part through the European Research Council (ERC), StG 677268 NextDART, The Lundbeck Foundation Fellowship R190-2014-4178 (to SRH), the Danish Research Council (DFF – 4004-00422 and DFF - 7014-00055) (to SRH and MN), the NIH/NCI Cancer Center Support Grant P30 CA015704, K24-CA-139052, R01 CA176841, Kelsey Dickson Team Science Courage Research Team Award, Prostate Cancer Foundation Award #15CHAS04 and the UW MCC Patient Gift Fund (to PN).

Author contributions: AKB and LS conceived the concept, designed and performed experiments, analyzed data, made figures and wrote the manuscript; KKJ, AMM and LEJ conceived the bioinformatic processing, analyzed data, made figures and revised the manuscript; NJM and CDC, identified and established T cell clones; RL identified and characterized TCRs and TCR transduced T cells; DMK, JCB and PN provided patient material for TCR identification and supervised the studies; CL and TNS sequenced and expressed TCRs and revised the manuscript; PM and MN conceived the bioinformatic processing, supervised data analyses and revised the manuscript; SRH conceived the concept, supervised the study, designed experiments, analyzed data and wrote the manuscript.

Competing interests: AKB and SRH are co-inventors on a patent covering the use of DNA barcode-labeled MHC multimers (WO2015185067 and WO2015188839). NJM, CDC, DMK and PN are co-inventors on a patent application filed by their employer, University of Washington, concerning the HLA-A*0201 restricted TCR sequences.

Materials & Correspondence: TCR sequences and expression vectors must be obtained through an MTA. Correspondence to Sine Reker Hadrup.

References

1. Sewell, A. K. Why must T cells be cross-reactive? *Nat. Rev. Immunol.* **12**, 669–677 (2012).
2. Obenaus, M. *et al.* Identification of human T-cell receptors with optimal affinity to cancer antigens using antigen-negative humanized mice. *Nat. Biotechnol.* **33**, 402–7 (2015).
3. Cameron, B. J. *et al.* Identification of a Titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells. *Sci. Transl. Med.* **5**, 197ra103 (2013).
4. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 (1988).
5. Mason, D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol. Today* **19**, 395–404 (1998).
6. Arstila, T. P. *et al.* A direct estimate of the human alpha beta T cell receptor diversity. *Science* **286**, 958–961 (1999).
7. Robins, H. S. *et al.* Comprehensive assessment of T-cell receptor β -chain diversity in $\alpha\beta$ T cells. *Blood* **114**, 4099–4107 (2010).
8. Cornberg, M. & Wedemeyer, H. Hepatitis C virus infection from the perspective of heterologous immunity. *Curr. Opin. Virol.* **16**, 41–48 (2016).
9. Wooldridge, L. *et al.* A single autoimmune T cell receptor recognizes more than a million different peptides. *J. Biol. Chem.* **287**, 1168–77 (2012).
10. Birnbaum, M. E. *et al.* Deconstructing the peptide-MHC specificity of T cell recognition.

- Cell* **157**, 1073–87 (2014).
11. Adams, J. J. *et al.* Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity. *Nat. Immunol.* **17**, 87–94 (2016).
 12. Gee, M. H. *et al.* Antigen Identification for Orphan T Cell Receptors Expressed on Tumor-Infiltrating Lymphocytes. *Cell* **172**, 549–563.e16 (2018).
 13. Wooldridge, L. *et al.* CD8 Controls T Cell Cross-Reactivity. *J. Immunol.* **185**, 4625–4632 (2010).
 14. Schaubert, K. L. *et al.* Generation of robust CD8+ T-cell responses against subdominant epitopes in conserved regions of HIV-1 by repertoire mining with mimotopes. *Eur. J. Immunol.* **40**, 1950–1962 (2010).
 15. 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 (2016).
 16. Lyngaa, R. *et al.* T-cell responses to oncogenic merkel cell polyomavirus proteins distinguish patients with merkel cell carcinoma from healthy donors. *Clin. Cancer Res.* **20**, 1768–78 (2014).
 17. Rodenko, B. *et al.* Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. *Nat. Protoc.* **1**, 1120–32 (2006).
 18. Enouz, S., Carrié, L., Merkler, D., Bevan, M. J. & Zehn, D. Autoreactive T cells bypass negative selection and respond to self-antigen stimulation during infection. *J. Exp. Med.* **209**, 1769–1779 (2012).
 19. Miller, N. J. *et al.* Tumor-Infiltrating Merkel Cell Polyomavirus-Specific T Cells Are

- Diverse and Associated with Improved Patient Survival. *Cancer Immunol. Res.* **5**, 137–147 (2017).
20. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–8 (2011).
 21. Lee, P. P. *et al.* Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients. *Nat. Med.* **5**, 677–85 (1999).
 22. Albert, L. J. & Inman, R. D. Molecular mimicry and autoimmunity. *N. Engl. J. Med.* **341**, 2068–74 (1999).
 23. Van Den Berg, J. H. *et al.* Case Report of a Fatal Serious Adverse Event Upon Administration of T Cells Transduced with a MART-1-specific T-cell Receptor. *Mol. Ther.* **23**, 1541–1550 (2015).
 24. Fowler, D. M. & Fields, S. Deep mutational scanning: A new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
 25. Harris, D. T. *et al.* Deep mutational scans as a guide to engineering high affinity t cell receptor interactions with peptide-bound major histocompatibility complex. *J. Biol. Chem.* **291**, 24566–24578 (2016).
 26. Linette, G. P. *et al.* Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. *Blood* **122**, 863–71 (2013).
 27. Morgan, R. A. *et al.* Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy. *J. Immunother.* **36**, 133–51 (2013).

Methods

Ethical approval

All healthy donor material was collected under approval by the local Scientific Ethics Committee, and written informed consent was obtained according to the Declaration of Helsinki. Collection of MCC patient material was approved by the Fred Hutchinson Cancer Research Center Institutional Review Board and conducted according to the Declaration of Helsinki principles. Informed consent was received from all participants. Splenocytes from OT1 and OT3 transgenic mice were collected from collaborators, under regular approval from the national committee of animal health (approval no. M165-15, University of Lund, Sweden).

Cell samples

Peripheral blood mononuclear cells (PBMCs) from healthy donors were isolated from whole blood by density centrifugation on Lymphoprep (Axis-Shield PoC) and cryopreserved at -150°C in FCS (FCS; Gibco) + 10% DMSO. Mouse spleen suspensions were obtained by mashing the full spleen through a 70 μm cell strainer (Fischer Scientific). Red blood cells were lysed with RBC Lysis buffer (BioLegend) and used directly or cryopreserved at -150°C in FCS (FCS; Gibco) + 10% DMSO.

Generation of DNA barcodes and dextran conjugation

Attachment of 5' biotinylated AxBy DNA barcodes to PE- and streptavidin-conjugated dextran was performed as described in¹⁵. Oligonucleotides containing distinct 25-mer nucleotide sequences²⁸ were purchased from LGC Biosearch Technologies (Denmark) and PE- and streptavidin-conjugated dextran was provided by Immudex (Denmark) and FINA Biosolutions LCC (USA). All oligos carry a 6-nt unique molecular identifier (UMI)²⁹.

Peptide libraries

Most of the collections of peptide variants applied in this study were designed by sequentially substituting every single position of the full ‘original’ peptide sequence with all naturally occurring amino acids. Some libraries also included a number of length and position variants, extending the peptide either to the C- or N-terminal direction of the full protein sequence (**Supplementary Table 1, 2 and 10**). For the double substitution library applied in **Fig. 1i-k** and **Supplementary Fig. 9** the variants comprise of peptides with two independent amino acid substitutions at positions 4-8. These positions were substituted with all combinations of 12 different amino acids (T, R, P, N, M, K, I, G, F, E, C and A) to minimize the total library size. For the same purpose only high affinity peptides (%Rank<0.5) were synthesized and included in the MHC multimer analysis accumulating to 782 different peptides that were substituted at two positions simultaneously (**Supplementary Table 7**).

Peptides were purchased from Pepscan (Pepscan Presto) and dissolved to 10 mM in DMSO.

MHC monomer production

UV-sensitive ligands were synthesized as previously described^{17,30,31}. In brief, recombinant HLA-A*0201, HLA-A*2402, HLA-B*0702, and H-2Kb heavy chains and human or mouse β_2 microglobulin light chain were produced in *Escherichia coli*. HLA and H-2 heavy and light chains were refolded with UV-sensitive ligands and purified as described in³². Specific peptide-MHC complexes were generated by UV-mediated peptide exchange^{17,30,31,33}.

Generation of DNA barcode-labeled peptide-MHC multimer libraries

DNA barcode-labeled peptide-MHC multimers, all carrying a common fluorescent PE label, were generated as described in¹⁵. Immediately before staining, barcode-labeled MHC multimers were centrifuged for 5 min at 3,300 g and pooled (0.0036-2.3 pmol of each pMHC/sample) to enable parallel staining. An aliquot of ~5 μ L of the MHC multimer reagent pool was stored at -20 °C for baseline analysis.

Staining with DNA barcode-labeled multimers

Cryopreserved cells were thawed and washed in RPMI + 10% FCS, then washed in barcode-cytometry buffer (PBS + 0.5% BSA + 100 μ g/mL herring DNA + 2 mM EDTA) and incubated for 30 min at 37 °C in the presence of 50 nM dasatinib. $0.5-2 \times 10^6$ cells were incubated for 15 min at 37 °C with the DNA-barcoded MHC multimer pool in a total volume of 80 μ L (final concentration of each distinct pMHC: 0.036-23 nM). Next, a 5x antibody mix composed of CD8-PerCP (Invitrogen MHCD0831) (final dilution 1/50) or BV510 (BD 563256, clone RPA-T8) (final dilution 1/25) or BV480 (BD 566121, clone RPA-T8) (final dilution 1/50), dump channel antibodies (CD4-FITC (BD 345768) (final dilution 1/80), CD14-FITC (BD 345784) (final dilution 1/32), CD19-FITC (BD 345776) (final dilution 1/16), CD40-FITC (Serotech MCA1590F) (final dilution 1/40), and CD16-FITC (BD 335035) (final dilution 1/64)) and a dead cell marker (LIVE/DEAD Fixable Near-IR; Invitrogen L10119) (final dilution 1/1000) was added and incubated for 30 min at 4 °C. Cells were washed three times in barcode-cytometry buffer and fixed in 1% paraformaldehyde (PFA). If the cells were not acquired within 24 h, they were washed twice and resuspended in barcode-cytometry buffer. Cells were acquired within a week after multimer staining. For staining for mouse splenocytes, OT1 and OT3 T cells, the

following antibodies were used: CD8a-BV480 (BD 566096, clone 53-6.7) and CD3-FITC (BioLegend 100206, clone 145-2C11).

Cell sorting

Cells were sorted on a FACS Aria Fusion (BD) into tubes containing 100 μ L of barcode-cytometry buffer (tubes were saturated with PBS + 2% BSA in advance). Using FACSDiva software, we gated on single, live CD8 positive and ‘dump’ (CD4, 14, 16, 19, and 40) negative lymphocytes and sorted either all multimer-positive (PE) cells within this population. The sorted cells were centrifuged for 10 min at 5,000g and the buffer was removed. The cell pellet was stored at -80 °C in a minimal amount of residual buffer (<20 μ L). Gating strategy exemplified in **(Supplementary Fig. 13)**.

DNA barcode amplification

DNA barcode amplification was performed as described in Bentzen et al., 2016¹⁵. PCR amplification was conducted on isolated cells (in <20 μ L of buffer) or on the stored aliquot of the MHC multimer reagent pool (diluted 50,000 \times in the final PCR), which was used as the baseline to determine the number of DNA barcode reads within a non-processed MHC multimer reagent library. PCR products were purified with QIAquick PCR Purification kit (Qiagen, 28104). The amplified DNA barcodes were sequenced at Sequetech (USA) using Ion Torrent PGM 316 or 318 chip (Life Technologies).

Processing of sequencing data derived from multimer-associated DNA barcodes

Sequencing data were processed by the software package ‘Barracoda,’ available online at (<http://www.cbs.dtu.dk/services/barracoda>). This tool identifies the barcodes used in a given experiment, assigns sampleID and pMHC specificity to each barcode, and counts the total number of reads and clonally reduced reads for each pMHC-associated DNA barcode. Furthermore, it accounts for barcode enrichment, expressed as \log_2FC , based on methods designed for the analysis of RNA-seq data. See details in *Statistical analyses*.

Normalization of \log_2FC values relative to the original peptide-MHC

To compare the \log_2FC values of the original peptide with those obtained from the peptides with an amino acid substitution, all \log_2FC s were normalized using the formula: $z = \frac{x-\omega}{\sigma}$, where z is the normalized \log_2FC , x is the \log_2FC of the peptide variation, σ is the standard deviation of all \log_2FC s and ω is the \log_2FC of the original peptide. Applied in **Fig. 1g,h** and **Supplementary Figure 2 and 5**.

Generation of fluorescently-labeled MHC tetramers

MHC tetramers were assembled on PE-conjugated streptavidin (BioLegend, Nordic Biosite, Denmark) as previously described^{34,35} and acquired on a BD LSR Fortessa. Gating strategy exemplified in **Supplementary Fig. 13**.

MHC- ELISA

The MHC-ELISA was performed as described in¹⁷. 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 μ g/mL herring DNA + 2 mM EDTA) and were subsequently incubated with 1 μ M 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 μ mol 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 μ g/mL herring DNA + 2 mM EDTA) before sorting all beads in a FSC-SSC plot. The sorted beads were centrifuged for 10 min at 5,000g and the buffer was removed. The pellet was stored at -80 °C in a minimal amount of residual buffer (<20 μ L). The pMHC associated DNA barcodes were amplified and sequenced as described previously, 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**).

T-cell functional assays

EC₅₀ determination

EC₅₀ values for MCC CD8⁺ T cell clones were determined in¹⁹. Briefly, secreted IFN- γ was measured after incubating 2×10^4 clonal T cells with 5×10^4 peptide-pulsed T2 cells (peptide sequence KLLLEIAPNC). Peptide was used at log₁₀ 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 EC₅₀ 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 μ M 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% CO₂. 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-TNF α antibody (BioLegend, 502930), APC-conjugated anti-IFN γ 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.

TCR gene capture

*HLA-B*0702_{APN} and HLA-A*2402_{EWV}-responsive CD8⁺ T cells:* TCR gene capturing was performed as described in ³⁶. Briefly, DNA isolated from MCC CD8⁺ T cells, was sheared to fragments of 500–600 bp with a Covaris system (S-series, D10%, I5, C/b 200, 30 s) and the resulting DNA fragments were purified with SPRI beads (Agencourt). Sequence library preparation was performed using the TruSeq DNA Sample Preparation kit (Illumina) with the adaptation of only seven cycles for the final library amplification. Illumina TruSeq 6-bp indexes (as designed by the manufacturer) were used for multiplexing. Multiplexed TCR captures were performed using a custom-designed Agilent SureSelect bait library with the following adaptations: pools of six to eight DNA libraries were captured with 1:10 of a bait reaction and block 3 in the hybridization mixture was replaced with a custom NKI block 3. The NKI block 3 consisted of equal amounts of two DNA oligos (IDT-DNA, Iowa, USA) at 16.6 $\mu\text{g } \mu\text{l}^{-1}$: NKI 3.1 5'AGATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTC TTCTGCTTG/3'ddC/-3' and NKI 3.2 5'CAAGCAGAAGACGGCATAACGAGATNNNNNNGTGACTGGAGTTCAGAC

GTGTGCTCTTCCGATCT/3'ddC/-3'. Captured library fragments were split into two fractions and PCR enrichment (15 cycles) was performed using the Illumina P5 and P7 oligonucleotides (IDT-DNA, Iowa, USA). P5 primer 5'-AATGATACGGCGACCACCGAGATCT-3' and P7 primer 5'CAAGCAGAAGACGGCATACGAG-3'. Quantification of PCR reactions were validated on a BioAnalyzer DNA Chip (Agilent) and were combined in equal amounts and diluted (10 nM) afterwards. Paired-end sequencing was performed using the Illumina HiSeq2000 platform with a read length of 75-100 bp. CDR3 TCR sequences were identified from the sequencing data as previously reported³⁷.

KLL-specific CD8⁺ T cell clones: Simultaneous sequencing of TCR α and TCR β repertoires was performed as described in³⁸.

TCR transduction

Retroviral transduction was performed as previously described in³⁹. Briefly, Phoenix-A cells were used as packaging cells and were transfected with 10 μ g of retroviral plasmid DNA. Virus supernatant was harvested 2 days after transfection and was either used immediately or was snap-frozen and stored at -80 °C. PBMCs were activated by CD3/CD28 beads (Human T cell expander; Invitrogen/Dynal) in a 1:2 ratio (cell/bead). After 30 min of incubation at RT, non-CD3⁺ cells were removed by magnetic separation. Cells were incubated in RPMI media + 10% human serum containing IL-15 (Peprotech; 5 ng/mL) and rh-IL-2 (Novartis; 100 IU/mL). For bead-based transduction, beads were incubated with retronectin overnight at 4 °C. Beads were washed with PBS and blocked with 2% BSA following a 2-h incubation with virus supernatant (1-2 x 10⁷ beads/ mL virus supernatant). PMBCs were incubated with virus-coated beads at a

1:10 ratio (cell/bead) for 24 h at 37 °C. The transduction efficiency was determined 5 days after transfection and was always over 60%.

Affinity predictions

The binding affinity of each of the amino acid substitutions of the original peptide sequences was predicted using NetMHCpan 4.0⁴⁰. The %rank is the rank of the predicted binding affinity compared to a set of random natural peptides. This measure is not affected by inherent bias of certain MHC molecules towards higher or lower means predicted affinities. Strong binders are defined as having %rank < 0.5, and weak binders with %rank < 2.

Modeling of MHC-bound peptides

The *in silico* structure-based pMHC model of HLA-B*0702_{APN} and HLA-A*2402_{EWV} with the original peptide embedded in the MHC binding pocket was made using MODELLER⁴¹ and the conformation of the original peptide was then optimized using the robotics-based kinematic closure (KIC)⁴² protocol from Rosetta⁴³.

The program FoldX⁴⁴ was used to model all single amino acid substitutions of the original peptide, and to predict their effect on the interaction as the difference between the predicted binding energy of the MHC to the mutated and original peptide, respectively: $\Delta\Delta G = \Delta G_{\text{amino acid substitution}} - \Delta G_{\text{original}}$. A $\Delta\Delta G > 0$ indicates that a given substitution has destabilizing properties, and a $\Delta\Delta G < 0$ indicates that a given substitution has stabilizing properties.

Generation of sequence logos

TCR interaction with pMHC

The TCR fingerprints were created based on the \log_2 FCs calculated by Barracoda. The amino acid substitution setup that we applied allowed us to assign a single \log_2 FC value to each amino acid residue at each position in the peptide sequence. The \log_2 FC of the original peptide was assigned to each of the respective amino acid residues at each position. This way a position-specific scoring matrix (PSSM) for each clone was calculated with the number of rows corresponding to the number of positions in the peptide and number of columns corresponding to the number of naturally occurring amino acids. We then used softmax to normalize each row of the PSSMs to sum to 1 and the resulting position-specific frequency matrices were then converted into Shannon logos.

Energy-based sequence logos of peptide-MHC

The sequence logo of the structurally predicted peptide-MHC binding preference for HLA-B*0702_{APN} and HLA-A*2402_{EWV} was made using the energies from the FoldX analysis of each amino acid substitution to generate a PSSM, in which each row represents positions in the peptide and each column represent an amino acid substitution. We then normalized each row in the PSSM. All FoldX energies were normalized using the following formula:

$$N_{ij} = \frac{e^{-A_{ij}}}{\sum_j (e^{-A_{ij}})}$$

where N_{ij} is the normalized FoldX energy for each position in the normalized PSSM and A_{ij} is the FoldX energies in the i th row and j th column of the PSSM.

The PSSMs were normalized so that each position summed to 1 and the sequence logos were then generated with the Shannon method in Seq2Logo⁴⁵. Furthermore, we calculated the information content I for each position in the peptide using the following formula, $I =$

$\log(20) + \sum_a p_a \cdot \log(p_a)$, where p_a is the normalized FoldX energy for each amino acid substitution. This information content is shown in this structural pMHC model of HLA-B*0702_{APN} and HLA-A*2402_{EWV} in Supplementary **Fig. 7 and 8**, and **Supplementary Table 5 and 6**.

Predicted peptide binding affinities

The predicted peptide binding affinity of each amino acid substitution of the original peptide sequence (shown in **Fig 1g,h**) was found using NetMHCpan-4.0⁴⁶.

Principal component analysis

In order to visualize the inter Shannon logo distances, we flattened each PSSM to a vector with elements corresponding to the number of naturally occurring amino acids (n=20) times the number of peptide positions (n=9). We then stacked the flattened PSSMs to form a matrix with the number of rows corresponding to the number of clones and number of columns as before. On this combined matrix, we could then perform a PCA and visualize by standard methods. Using the PC1-PC12, hierarchical clustering was performed and we visualized the distances using a dendrogram. Applied in **Fig. 2e** and **Supplementary Fig. 12**.

Estimating the TCR promiscuity

For estimating the TCR promiscuity, i.e. the estimated number of cross-recognized peptides by a given TCR, the log-ratios were calculated by transforming the log₂FCs given by Barracoda for each TCR to fold change using base 2 and then dividing each fold change with the fold change of the original peptide. The calculated fold change ratios were then applied to estimate the number of possible cross-recognized peptides calculated per TCR using standard combinatorics on the

positional set of log ratios larger than zero. In this calculation the anchor positions (2 and ∞) were included with the numbers 2 and 3 respectively, reflecting the selectivity of HLA-A*0201 (determined from the ‘naturally presented ligands’ given by NetMHCpan-4.0⁴⁶). The calculations for H-2Kb followed the same rationale, with the anchor positions (3, 5 and 8) included with the numbers 4, 4 and 3 respectively. The obtained values for TCR promiscuity were plotted against the experimentally obtained EC₅₀ values⁴⁷ (**Fig. 2g**) and are listed in **Supplementary Table 11**.

Prediction of cross-reactive peptides related to MCC clones

From each Shannon logo cross-reactive peptides were predicted from the corresponding PSSMs using the “Find Individual Motif Occurrences” (FIMO) software²⁰, which searches the human proteome for sequences that match each logo. For each MCC clone, the 10 peptides with the highest likelihood for cross recognition to the given TCR (lowest p-value) were synthesized.

Signature of potential cross-recognition

The correlation matrix (**Supplementary Fig. 12**) was constructed based on the output from the FIMO database. For each clone, the corresponding top 1,000 peptides (based on p-values) were retrieved. The total pool of top peptides from all clones was then, per peptide, scored against the Shannon PSSMs using the sum of positional scores. The set of scores per clone were then correlated all-against-all using Pearson’s correlation coefficient. Lastly, the set of correlations per clone was clustered using hierarchical clustering.

Statistical analyses

The statistical processing of DNA barcode reads was developed in¹⁵ and is based on methods designed for the analysis of RNA-seq data, implemented in the R package edgeR⁴⁹. Fold

changes in read counts mapped to a given sample relative to mean read counts mapped to triplicate baseline samples are estimated using normalization factors determined by the trimmed mean of M-values method⁴⁸. P values (applied in **Fig. 3a**) were calculated by comparing each experiment individually to the mean baseline sample reads using a negative binomial distribution with a fixed dispersion parameter set to 0.1. False-discovery rates (FDRs) were estimated using the Benjamini–Hochberg method. Specific barcodes with an FDR < 0.1% were defined as significant. At least 1/1,000 reads associated with a given DNA barcode relative to the total number of DNA barcode reads in that given sample was set as threshold avoid false-positive detection of T-cell populations due to low number of reads in the baseline samples.

The statistical analyses in **Fig. 2g** were conducted using GraphPad Prism 7. Normal distribution was assessed using a Shapiro-Wilk normality test, suggesting log transformation of the data. The statistical test and Pearson correlation (two-tailed) is therefore performed on log transformed data.

Data and code availability

All relevant data and codes are available from the authors. TCR sequences and expression vectors must be obtained through an MTA. For DNA barcode analysis the tool Barracoda is available online at (<http://www.cbs.dtu.dk/services/barracoda>).

References (methods only):

28. Xu, Q., Schlabach, M. R., Hannon, G. J. & Elledge, S. J. Design of 240,000 orthogonal 25mer DNA barcode probes. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2289–94 (2009).
29. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–4 (2011).
30. Toebe, M. *et al.* Design and use of conditional MHC class I ligands. *Nat. Med.* **12**, 246–51 (2006).
31. Chang, C. X. L. *et al.* Conditional ligands for Asian HLA variants facilitate the definition of CD8+ T-cell responses in acute and chronic viral diseases. *Eur. J. Immunol.* **43**, 1109–20 (2013).
32. Hadrup, S. R. *et al.* High-throughput T-cell epitope discovery through MHC peptide exchange. *Methods Mol. Biol.* **524**, 383–405 (2009).
33. Frøsig, T. M. *et al.* Design and validation of conditional ligands for HLA-B*08:01, HLA-B*15:01, HLA-B*35:01, and HLA-B*44:05. *Cytom. Part A* **87**, 967–975 (2015).
34. Hadrup, S. R. *et al.* Parallel detection of antigen-specific T-cell responses by multidimensional encoding of MHC multimers. *Nat. Methods* **6**, (2009).
35. Andersen, R. S. *et al.* Parallel detection of antigen-specific T cell responses by combinatorial encoding of MHC multimers. *Nat. Protoc.* **7**, 891–902 (2012).
36. Linnemann, C. *et al.* High-throughput identification of antigen-specific TCRs by TCR gene capture. *Nat. Med.* **19**, 1534–41 (2013).
37. Bolotin, D. A. *et al.* Next generation sequencing for TCR repertoire profiling: platform-

- specific features and correction algorithms. *Eur. J. Immunol.* **42**, 3073–83 (2012).
38. Han, A., Glanville, J., Hansmann, L. & Davis, M. M. Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat. Biotechnol.* **32**, 684–692 (2014).
 39. Kühlcke, K. *et al.* Retroviral transduction of T lymphocytes for suicide gene therapy in allogeneic stem cell transplantation. *Bone Marrow Transplant.* **25 Suppl 2**, S96-8 (2000).
 40. Nielsen, M. & Andreatta, M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med.* **8**, 33 (2016).
 41. Fiser, A. & Sali, A. ModLoop: Automated modeling of loops in protein structures. *Bioinformatics* **19**, 2500–2501 (2003).
 42. Stein, A. & Kortemme, T. Improvements to Robotics-Inspired Conformational Sampling in Rosetta. *PLoS One* **8**, (2013).
 43. Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H. & Meiler, J. Practically useful: What the Rosetta protein modeling suite can do for you. *Biochemistry* **49**, 2987–2998 (2010).
 44. Schymkowitz, J. *et al.* The FoldX web server: An online force field. *Nucleic Acids Res.* **33**, 382–388 (2005).
 45. Thomsen, M. C. F. & Nielsen, M. 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**, 281–287 (2012).
 46. Jurtz, V. *et al.* NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions

Integrating Eluted Ligand and Peptide Binding Affinity Data. *J. Immunol.* j11700893
(2017). doi:10.4049/jimmunol.1700893

47. Miller, N. J. *et al.* Tumor-Infiltrating Merkel Cell Polyomavirus-Specific T Cells Are Diverse and Associated with Improved Patient Survival. *Cancer Immunol. Res.* **5**, 137–147 (2017).
48. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
49. Andersen, R. S. *et al.* High frequency of T cells specific for cryptic epitopes in melanoma patients. *Oncoimmunology* **2**, e25374 (2013).

Figure legends:

Fig. 1. The recognition pattern of two different TCRs that recognize MCC-derived peptides restricted to HLA-B*0702 or HLA-A*2402

(a-c) Results obtained from the DNA barcode-based analysis of T cells transduced with a TCR recognizing the HLA-B*0702 restricted peptide, APNCYGNIPL. The analysis was performed with all possible variations of peptides created by single-position amino acid substitutions. **(a)** Shows the hierarchy of pMHC interactions expressed as \log_2FC of read counts relative to a triplicate baseline sample (see **Supplementary Note**). A healthy donor PBMC sample (BC-D42) was screened with the same MHC multimer panel in parallel. For both samples the plotted order of \log_2FC s of each pMHC-associated DNA barcode is determined by the hierarchy obtained from screening the HLA-B*0702_{APN} responsive TCR. **(b)** Heat map of amino acid preferences of the HLA-B*0702_{APN} responsive TCR based on data from **(a)**. Each row represents a given amino acid and each column a position in the peptide sequence. The amino acids of the original peptide target are marked in black boxes. **(c)** Recognition pattern of the HLA-B*0702_{APN} interacting TCR, here visualized as a sequence logo based on the data from **a** and **b**. **(d-f)** Results obtained from the DNA barcode-based analysis of T cells transduced with a TCR recognizing the HLA-A*2402 restricted peptide, EWWRSGGFSF. The analysis was performed with all possible variations of peptides created by single-position amino acid substitutions. Visualization of data corresponding to **(a-c)**. **(g-h)** Scatter plot of the predicted peptide binding, %Rank (x-axis) of all naturally occurring amino acid substitutions of **(g)** APNCYGNIPL to HLA-B*0702 or **(h)** EWWRSGGFSF to HLA-A*2402, in relation to the experimentally obtained TCR-pHLA interaction (y-axis). The color indicates the position of the amino acid substitution. %Rank<2 (dotted line) marks the recommended cut-off of peptides that are considered as binders to MHC.

(i-k) Results from a parallel MHC multimer analysis of the TCR recognizing the HLA-A*2402 restricted peptide, EWWRSGGFSF, with a MHC multimer library composed of peptides with single amino acid substitutions corresponding to the one used in **(d-f)** as well as double amino acid substitutions covering 12 naturally occurring amino acids, where positions 4-8 are substituted two amino acids at a time (n=973, see full list in **Supplementary Table 4 and 7**). **(i)** Shows the obtained \log_2FC values grouped according to the number of substitutions within the peptide sequence. **(j-k)** Heatmaps showing the \log_2FC obtained for peptides with amino acids substituted at **(j)** position 4 and 5 simultaneously, or **(k)** position 6 and 8 simultaneously. Each row and column represents a given amino acid substitution (see heat map of all screened substitutions in **Supplementary Fig. 9**). The original amino acids are marked in bold and peptides in the same row or column are only substituted at one position, indicated with the one-letter code. All data is representative of duplicate analyses.

Fig. 2. Diverse recognition patterns of TCRs recognizing the same pMHC epitope

(a-b) The TCR-pMHC interaction hierarchy obtained from DNA barcode-based analysis of mouse OT-1 and OT-3 T cells, both recognizing the H-2Kb restricted peptide SIINFEKL. The analysis was performed with all possible variations of peptides created by single-position amino acid substitutions. \log_2FC of read counts relative to a triplicate baseline (see **Supplementary Note**) is plotted according to the hierarchy obtained from **(a)** the OT-1 T cells, or **(b)** the OT-3 T cells, compared to the signal obtained using T cells from wild-type C57BL/6 mice (irrelevant), screened with the same MHC multimer panel. **(c,d)** The different TCR fingerprints obtained from the screening of **(c)** OT-1 and **(d)** OT-3 derived T cells. The OT-1 and OT-3 T cells were screened once. See **Supplementary Fig. 10** for the read counts of the corresponding data, and **Supplementary Fig. 11** for single fluorescent-based MHC multimer stainings of a range of

SIINFEKL variants. (e) TCR fingerprints of 12 MCC clones all originally identified for their recognition of the HLA-A*0201 restricted peptide KLLLEIAPNC. The fingerprints are clustered according to the similarity of their recognition pattern. Data is representative of duplicate analyses. (f) Bar plots showing the cytokine secretion after stimulating the clonal T cells with peptides containing alanine substitutions at the indicated positions compared to the obtained TCR fingerprints (from e) of clone 2 and clone 5, w876. The grey bars indicate the original peptide that has an alanine at position 6. Cytokine secretion was determined once (individual frequencies are shown in **Supplementary Fig. 12**). (g) Correlation between the number of targets estimated for each TCR, based on data from e (x axis), and the obtained EC_{50} values of each clone (y axis). Each dot represents one T cell clone. Dots of the same color indicate clones derived from one patient. R^2 is based on Pearson's r on the log-transformed values..

Fig. 3. Cross-reactivity of HLA-A*0201_{KLL} responsive TCRs

(a) Screening for T cell recognition of 75 peptides that are potentially cross-recognized by one or more of the 12 clonal T cells that has the HLA-A*0201 restricted KLLLEIAPNC peptide as original target. For each clone the top 10 potential cross-reactive peptides were synthesized and used to screen for TCR cross recognitions using DNA barcode-labeled MHC multimers. Total library size was 75 peptides (**Supplementary Table 12**). The p-value resulting from the DNA barcode-based screen of all 75 pMHC multimers and all 12 clones are plotted (y-axis) according to %Rank score (x-axis). Dotted line at $y = 3$ represent the selected threshold of $FDR < 0.1\%$. Dotted line at $x=2$ marks the recommended cut-off of peptides that are considered as binders to MHC. The closed symbol indicate a response that were also stained with fluorescently labeled MHC tetramers. Data is mean of duplicate analyses. (b) Contour plots from the fluorescently-

based tetramer screening of three clones that all recognize the original HLA-A*0201 restricted KLEIAPNC peptide. Tetramers are generated from either the original peptide target (KLEIAPNC), a peptide (KTVGIYPNA) that was cross-recognized by the TCR of clone 5, w876, in **a**, or a peptide (RTCEIQGWC) that was not recognized by any of the clones in **a**. The clones were spiked into a healthy donor PBMC sample (BC) in equal amounts. The percentage out of total CD8⁺ T cells is indicated within the contour plots. (c) Bar plot depicting the frequency of cytokine producing cells of CD8⁺ T cells after stimulating clone 5, w876, with HLA-A*0201 expressing cells pulsed with the indicated 9-mer peptide. Tetramer staining and cytokine secretion was determined once.

Supplementary material

Supplementary Note

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 ¹⁵.

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 ¹⁵, is used to retrieve the number of read counts for each pMHC-associated DNA barcode, and to calculate the log₂FC 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^{24,25} 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 selection¹⁵.

When assessing the TCR fingerprint, the position of a given pMHC within the binding hierarchy is defined based on the \log_2FC 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.

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 MCPyV-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 Table 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 reported⁴⁹. 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**).

Figure legends for supplementary figures:

Supplementary Fig. 1. Using DNA barcode-labeled MHC multimers to generate TCR fingerprints

(a) Schematic overview of the workflow behind a TCR fingerprint. Each peptide position of the original peptide is substituted with all naturally occurring amino acids. The resultant peptide variants are assembled with MHC molecules, and individually multiplexed on a PE-labeled dextran backbone together with individual DNA barcodes. Thus, a given DNA barcode forms a tag for the corresponding pMHC variant. T cells recognizing the original pMHC are incubated with all MHC multimer variants simultaneously. The MHC multimer binding cells are isolated based on the shared PE-label and their associated DNA barcodes are amplified and sequenced. The hierarchy of pMHC interactions is revealed from the relative number of DNA-barcode reads specifically associated with a given MHC multimer. DNA barcode reads are typically transformed to \log_2FC values (data in the example corresponds to results from Fig. 1a). (b) A clonal TCR will bind a pool of pMHC variants in a hierarchical manner that is governed by the avidity of the given pMHC and the TCR. (c) The hierarchy of TCR-pMHC interactions is used to determine the recognition pattern of that given TCR, visualized as a Shannon plot (here determined 'TCR fingerprints')

Supplementary Fig. 2. Feasibility of using DNA barcode-labeled multimers to determine the binding hierarchy of multiple pMHCs in a ‘one-pot’ screen

Bar plots of the relative \log_2FC values from the DNA barcode-based analyses of PBMCs transduced with **(a-c)** a HLA-B*0702_{APN} engaging TCR or **(d)** a **HLA-A*2402_{EWV}** engaging TCR. Each bar represents a pMHC variant with an amino acid substitution at the given peptide position. Bars of the same color indicate peptides that were included in a parallel screening. The dotted line at zero indicates the normalized \log_2FC of the original pMHC. **(a-b)** The MHC multimer libraries applied in the analyses included all possible 9- and 10-mer variants of APNCYGNIPL generated from substituting with alanine. **(a)** Shows a series of three analyses where # indicates a peptide variant that was excluded in a given screening. Data is representative of triplicate analyses. **(b)** Bar plot from a series of analyses using the same MHC multimer library of APNCYGNIPL variants. In these analyses a varying amount of DNA barcode-labeled MHC multimers were used, i.e. a titration from 23 nM to 0.92 nM in respect to each pMHC (corresponding to 100%, 20% or 4% of the amount originally applied). The analysis was performed once. **(c)** Bar plot from a series of analyses where a range of 9- and 10-mer peptides produced from substituting with alanine, lysine or glutamic acid were applied to generate three different libraries of DNA barcode-labeled MHC multimers that were used to stain the HLA-B*0702_{APN} engaging TCR. **(d)** The MHC multimer libraries applied in these analyses included all possible 9- and 10-mer variants of EWWRSGGFSF generated from substituting with alanine. The series of analyses are equivalent to that shown in **(b)** only conducted on a different TCR and at lower MHC multimer concentrations, i.e. a titration from 0.92 nM to 0.036 nM in respect to

each pMHC (corresponding to 4%, 0.8% or 0.16% of the amount originally applied). The analysis was performed once.

Supplementary Fig 3. Replicate screening of the HLA-B*0702_{APN} engaging TCR showing read counts and data from an irrelevant sample

Results obtained from the DNA barcode-based analysis of T cells transduced with a TCR recognizing the HLA-B*0702 restricted peptide, APNCYGNIPL (replicate of the data shown in **Fig. 1a-c**). The analysis was performed with all possible variations of peptides created by single-position amino acid substitutions. The hierarchy of pMHC interactions expressed as **(a)** clonality reduced read counts or **(b)** log₂FC relative to a triplicate baseline sample (see Supplementary Note). A healthy donor PBMC sample (BC-D62) was screened with the same MHC multimer panel in parallel. The plotted order of read counts or log₂FC of each pMHC-associated DNA barcode is determined by the hierarchy obtained from screening the HLA-B*0702_{APN} responsive TCR. **(c-f)** Heat map showing the clonality reduced read counts or log₂FC of **(c,d)** the HLA-B*0702_{APN} responsive TCR or **(e,f)** the healthy donor BC-D62, respectively. Each row represents a given amino acid and each column a position in the peptide sequence. The amino acids of the original peptide target are marked with black borders. **(g)** The replicate recognition pattern of the HLA-B*0702_{APN} interacting TCR, here visualized as a sequence logo based on the data from **a-d**.

Supplementary Fig. 4. Replicate screening of the HLA-A*2402_{EWV} engaging TCR showing read counts and data from an irrelevant sample

Results obtained from the DNA barcode-based analysis of T cells transduced with a TCR recognizing the HLA-A*2402 restricted peptide, EWWRS GGFSF (replicate of the data shown in **Fig. 1d-f**). The analysis was performed with all possible variations of peptides created by single-position amino acid substitutions. The hierarchy of pMHC interactions expressed as **(a)** clonality reduced read counts or **(b)** \log_2 FC relative to a triplicate baseline sample (see Supplementary Note). A healthy donor PBMC sample (BC-D42) was screened with the same MHC multimer panel in parallel. The plotted order of read counts or \log_2 FC of each pMHC-associated DNA barcode is determined by the hierarchy obtained from screening the HLA-A*2402_{EW} responsive TCR. **(c-f)** Heat map showing the clonality reduced read counts or \log_2 FC of **(c,d)** the HLA-A*2402_{EW} responsive TCR or **(e,f)** the healthy donor BC-D42, respectively. Each row represents a given amino acid and each column a position in the peptide sequence. The amino acids of the original peptide target are marked with black borders. **(g)** The replicate recognition pattern of the HLA-A*2402_{EW} interacting TCR, here visualized as a sequence logo based on the data from **a-d**.

Supplementary Fig. 5. The effect of peptide-MHC interaction on analyzing the recognition pattern of the HLA-B*0702_{APN} engaging TCR

(a) Sequence logo showing the amino acid preferences for peptide binding to HLA-B*0702 constructed with NetMHCpan-4.0⁴⁶ by finding the binding cores of the top 1% strongest predicted binders among 100,000 natural random 10-mer peptides and visualizing these using R with the Shannon method in Seq2Logo⁴⁵. **(b)** Scatter plot of the experimentally determined peptide binding to HLA-B*0702 (MHC ELISA), given as the relative quantity of a pMHC variant (only alanine substitutions) after UV-exchange, compared to that of the original peptide,

in correlation to the experimentally obtained recognition properties of the given pHLA-B*0702 interacting with the TCR. The MHC ELISA data is average of 5 experiments (n=2). Each dot in **b** represents one peptide-MHC variation. The coloring indicates the position of a given amino acid substitution and the asterisks indicates the original peptide. **(c)** Dot plots from staining the HLA-B*0702_{APN} engaging TCR with fluorescently labeled MHC multimers carrying one of six variations of APNCYGNIPL and an irrelevant peptide (p*, the UV conditional peptide). 1-3 are all examples of peptides predicted as strong binders to HLA-B*0702 (%Rank<0.5), here only 1 and 2 are recognized by the TCR. 4-6 are examples of peptides that are predicted as poor binders to MHC (%Rank>2), but the TCR is still able to recognize 4 and 5. The respective peptide sequences (substitutions in red), %Rank, MFI and percentages out of total CD8⁺ T cells are indicated within the contour plots. The fluorescent based multimer stainings were performed once.

Supplementary Fig. 6. The effect of peptide-MHC interaction on analyzing the recognition pattern of the HLA-A*2402_{EWV} engaging TCR

(a) Sequence logo showing the amino acid preferences for peptide binding to HLA-A*2402 constructed with NetMHCpan-4.0⁴⁶ by finding the binding cores of the top 1% strongest predicted binders among 100,000 natural random 10-mer peptides and visualizing these using R with the Shannon method in Seq2Logo⁴⁵. **(b)** Shows the sequencing output expressed in log₂FC after enrichment of only the correctly folded pMHC amongst the full DNA barcode-labeled MHC multimer library comprising all 191 single substitution EWVRSGGFSF variants (**Supplementary Table 4**). Correctly folded pMHC were captured on W6/32 beads that were subsequently sorted. The pMHC associated DNA barcodes were amplified and sequenced. This

was performed at T0, which was just after collecting and reducing the volume of the MHC multimer panel, T4, after four hours incubation at RT, and at T20, after 20 hours incubation at RT. Each dot represents one peptide-MHC variation. The coloring indicates the position of a given amino acid substitution and the asterisks indicates the original peptide. (c) Dot plots from staining the HLA-A*2402_{EWV} engaging TCR with fluorescently labeled MHC multimers carrying one of five variations of EWWRS GGFSF and an irrelevant peptide (p*, the UV conditional peptide). 1 and 2 are examples of peptides predicted as strong binders to HLA-A*2402 MHC (%Rank<0.5), only peptide 1 is recognized by the TCR. 3-5 are examples of peptides that are predicted as poor binders to MHC (%Rank>2), but the TCR is still able to recognize all three peptides. The respective peptide sequences (substitutions in red), %Rank, MFI and percentages out of total CD8⁺ T cells are indicated within the contour plots. The experiments were performed once.

Supplementary Fig. 7. *In silico* modelling of the APNCYGNIPL peptide conformation when bound to HLA-B*0702

(a) Structural model of the original peptide, APNCYGNIPL, binding to HLA-B*0702. Color scaling indicates the importance of a given amino acid for retaining the conformation of the original peptide when bound to the MHC (the information content is listed in **Supplementary Table 5**). (b) Bar plot of the FoldX derived $\Delta\Delta G$ energy between the original peptide (APNCYGNIPL) and HLA-B*0702, as well as all the peptide variants created by substituting each position with all naturally occurring amino acids. If the $\Delta\Delta G$ is larger than zero it indicates that the given amino acid substitution has destabilized the peptide-MHC interaction while a $\Delta\Delta G$ smaller than zero indicates that the substitution has stabilized the peptide-MHC interaction. (c)

Sequence logo showing the structurally predicted peptide-MHC binding preference for HLA-B*0702_{APN} based on the energy change ($\Delta\Delta$) calculated in **b**.

Supplementary Fig. 8. *In silico* modelling of the EWWRSGGFSF peptide conformation when bound to HLA-A*2402

(a) Structural model of original peptide, EWWRSGGFSF, binding to HLA-A*2402. Color scaling indicates the importance of a given amino acid for retaining the conformation of the peptide when bound to the MHC (the information content is listed in **Supplementary table 6**).

(b) Bar plot of the FoldX derived $\Delta\Delta G$ energy between the original peptide (EWWRSGGFSF) and HLA-A*2402, as well as all the peptide variants created by substituting each position with all naturally occurring amino acids. If the $\Delta\Delta G$ is larger than zero it indicates that the given amino acid substitution has destabilized the peptide-MHC interaction while a $\Delta\Delta G$ smaller than zero indicates that the substitution has stabilized the peptide-MHC interaction. (c) Sequence logo showing the structurally predicted peptide-MHC binding preference for HLA-A*2402_{EWV} based on the energy change ($\Delta\Delta$) calculated in **b**.

Supplementary Fig. 9. The HLA-A*2402_{EWV} engaging TCR recognize peptides with multiple amino acid substitutions.

(a) Heat map showing the amino acid preferences of the TCR recognizing the HLA-A*2402 restricted peptide, EWWRSGGFSF, expressed as \log_2FC of read counts relative to a triplicate baseline sample (see Supplementary Note). The TCR transduced clones were stained with a MHC multimer library composed of peptides with all single amino acid substitutions of the original EWWRSGGFSF peptide, as well as double amino acid substitutions covering 12

naturally occurring amino acids, where positions 4-8 were substituted two amino acids at a time. The heat map includes only double amino acid substitution variants (the effect of single amino acid substitutions can be seen in **Fig. 1** and **Supplementary Fig. 4**). Double substitution variants resulting in peptides with a %Rank>0.5 were excluded from the analysis (total number of double substitution peptide variants, n=782, **Supplementary Table 7**). The data largely confirm the findings when screening only with the single substitution library, that there is some flexibility of the amino acids at positions 4 and 5, while almost no other double substitutions are tolerated. The barcode-based analysis of double substitutions is representative of duplicate analyses. **(b)** Dot plots from staining the HLA-A*2402_{EW} engaging TCR with fluorescently labeled MHC multimers carrying one of 14 variations of EWWRS_{GGFSF} and an irrelevant peptide (p*, the UV conditional peptide). The variants are comprised of peptides with 3 to 8 amino acid substitutions compared to the original peptide sequence (1). The substitutions were either at positions that had no preference for certain amino acids for TCR recognition (positions 1, 3 or 9, as visualized in the TCR fingerprint, **Fig. 1f**) or substituted with an amino acid that was prominent in the TCR fingerprint (positions 2, 4, 5, 6, 8 and 10). The MHC anchor at positions 2 were in some cases instead substituted with tyrosine (Y), which is the preferred amino acid for the peptide-MHC interaction (peptides 8, 9, 10, 11, 13, see **Supplementary Fig. 6**). We saw that the TCR were able to recognize peptide variants with up to 6 amino acids substituted compared to the original sequence (peptide 4 and 6) when the substituted amino acids were represented in the TCR fingerprint. The respective peptide sequences (substitutions in red), %Rank, MFI and percentages out of total CD8⁺ T cells are indicated within the contour plots. The experiments were performed once.

Supplementary Fig. 10. The DNA barcode-based MHC multimer analysis of OT-1 and OT-3 T cells showing read counts, log₂FC and data from an irrelevant sample.

The underlying data from the DNA barcode-based analysis of splenocytes derived from OT-1 or OT-3 transgenic mice or T cells from wild-type C57BL/6 mice, respectively, used for **Fig. 2a-d**. The analysis was performed with all possible variations created by single-position amino acid substitutions of the H-2Kb restricted peptide, SIINFEKL. **(a,b)** Shows the clonality reduced read counts of each pMHC-associated DNA barcode plotted based on the hierarchy of counts derived from **(a)** screening the OT-1 T cells or **(b)** screening the OT-3 T cells. The obtained read counts from screening splenocytes from wild-type C57BL/6 mice are also plotted. **(c-h)** Heat map showing the read counts or log₂FC derived from screening of **(c,d)** the OT-1 T cells, **(e,f)** the OT-3 T cells or **(g,h)** the wild-type C57BL/6 splenocytes respectively. Each row represents a given amino acid and each column a position in the peptide sequence. The amino acids of the original peptide target are marked with black borders. The analysis was performed once.

Supplementary Fig. 11. Fluorescent-based MHC multimer analyses of OT-1 and OT-3 T cells

(a) Dot plots from the fluorescent-based MHC multimer staining of splenocytes derived from OT-1 or OT-3 mice. MHC tetramers carrying one of seven variations of SIINFEKL were applied in single color MHC multimer analyses. When comparing the staining of OT-1 and OT-3 in relation to their binding to the original peptide (1), the most variation between the TCRs is seen when staining with MHC multimers carrying peptide 3, 6 or 7. Peptide 3, which has a K to R substitution at position 7, leads to a higher MFI of the OT-1 T cells but a lower MFI of the OT-3 T cells. The opposite is seen for peptides 6 and 7, with N to P or S to M substitutions at position

4 and 1 respectively. When staining with the corresponding tetramers, the OT-1 T cells has a marked decreased in MFI, whereas the MFI of the OT-3 T cells is retained. These binding properties reflects the amino acid preferences evident from the individual TCR fingerprints (**Fig. 2c,d**). The respective peptide sequences, MFI and percentages out of total CD8⁺ T cells are indicated adjacent to or within the contour plots. All peptide variants are predicted to be strong binders of H-2Kb (%Rank<0.5). **b**) Dot plots from staining OT-1, OT-3 or wild-type C57BL/6 splenocytes with the full panel of 153 DNA barcode-labeled MHC multimer variants or a single irrelevant MHC multimer carrying the UV-conditional peptide (p*) or the original peptide SIINFEKL. The MFI and percentages out of total CD8⁺ T cells are indicated within the contour plots. The MHC multimer stainings were performed once.

Supplementary Fig. 12. Detailed information on the 12 HLA-A*0201_{KLL} engaging TCRs

(a) Principal component analysis (PCA) to visualize the diversity of the 12 TCR fingerprints of the MCC-derived T cells from **Fig. 2e**. Each dot represents one T cell clone. **(b)** Bar plots of the cytokine secretion of clone 2 and clone 5, w876, after stimulating the clonal T cells with peptides containing alanine substitutions at the indicated positions. The bars correspond to the individual stainings of the data shown in **Fig. 2f**. The grey bars indicate the original peptide that has an alanine at position 6. Cytokine secretion was determined once. **(c)** A correlation matrix showing the interrelationship between each of the 1000 peptides (predicted from the human proteome) that match best the individual TCR fingerprints of the 12 MCC clones (n=12,000) of **Fig. 2e**. Similarity is depicted on both color grading and size of dots. **(d)** Screening for T cell recognition of 75 peptides that are potentially cross-recognized by one or more of the 12 clonal T cells that

has the HLA-A* A0201 restricted KLEIAPNC peptide as original target. A library including 75 potentially cross-recognized peptides were applied in a DNA barcode-based MHC multimer analysis (**Fig. 3a** and **Supplementary Table 12**). Results from all 75 pMHC multimers and all 12 clones are plotted as Clonality reduced read counts resulting from the DNA barcode-based screen of all 75 pMHC multimers and all 12 clones are plotted (y-axis) according to %Rank score (x-axis). Dotted line at $x=2$ marks the recommended cut-off of peptides that are considered as binders to MHC. The closed symbol indicate a response that were also stained with fluorescently labeled MHC tetramers. Data is mean of duplicate analyses.

Supplementary Fig. 13. General gating strategy

(a) Shows an example of the initial gating of CD8⁺ T cells. (b) Shows an example of the multimer positive population of total CD8⁺ T cells, which was sorted in a DNA barcode-based analysis.

Supplementary Tables:

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

Substitution	Sequence	Position variant	Sequence
original	APNCYGNIPL	original	APNCYGNIPL
P1 -Ala	APNCYGNIPL	-6 towards N-term	CKLLEIAPNC
P2 -Ala	AANCYGNIPL	-5 towards N-term	KLLEIAPNCY
P3 -Ala	APACYGNIPL	-4 towards N-term	LLEIAPNCYG
P4 -Ala	APNAYGNIPL	-3 towards N-term	LEIAPNCYGN
P5 -Ala	APNCAGNIPL	-2 towards N-term	EIAPNCYGNI
P6 -Ala	APNCYANIPL	-1 towards N-term	IAPNCYGNIP
P7 -Ala	APNCYGAIPL	+1 towards C-term	PNCYGNIPLM
P8 -Ala	APNCYGNAPL	+2 towards C-term	NCYGNIPLMK
P9 -Ala	APNCYGNIAL	+3 towards C-term	CYGNIPLMKA
P10 -Ala	APNCYGNIPA	+4 towards C-term	YGNIPLMKAA
9-mer	APNCYGNIP	+5 towards C-term	GNIPLMKAAF
P1 -Ala	APNCYGNIP	+6 towards C-term	NIPLMKAAF ^K
P2 -Ala	AANCYGNIP		
P3 -Ala	APACYGNIP		
P4 -Ala	APNAYGNIP		
P5 -Ala	APNCAGNIP		
P6 -Ala	APNCYANIP		
P7 -Ala	APNCYGAIP		
P8 -Ala	APNCYGNAP		
P9 -Ala	APNCYGNIA		

Supplementary Table 2. Alanine substitutions and position variants of EWWRSGGFSF.

Applied in Supplementary Fig. 2

Substitution	Sequence	Position variant	Sequence
original	EWWRSGGFSF	original	EWWRSGGFSF
P1 -Ala	AWWRSGGFSF	-6 towards N-term	GTTKFKEWWR
P2 -Ala	EAWRSGGFSF	-5 towards N-term	TTKFKEWWR
P3 -Ala	EWARSGGFSF	-4 towards N-term	TKKFKEWWRSG
P4 -Ala	EWWRSGGFSF	-3 towards N-term	KKFKEWWRSGG
P5 -Ala	EWWRAGGFSF	-2 towards N-term	FKEWWRSGGF
P6 -Ala	EWWRSGGFSF	-1 towards N-term	KEWWRSGGFS
P7 -Ala	EWWRSGAFSF	+1 towards C-term	WWRSGGFSFG
P8 -Ala	EWWRSGGASF	+2 towards C-term	WRSGGFSFGK
P9 -Ala	EWWRSGGFAF	+3 towards C-term	RSGGFSFGKA
P10 -Ala	EWWRSGGFSFA	+4 towards C-term	SGGFSFGKAY
9-mer	EWWRSGGFS	+5 towards C-term	GGFSFGKAYE
P1 -Ala	AWWRSGGFS	+6 towards C-term	GFSFGKAYEY
P2 -Ala	EAWRSGGFS		
P3 -Ala	EWARSGGFS		
P4 -Ala	EWWRSGGFS		
P5 -Ala	EWWRAGGFS		
P6 -Ala	EWWRSGGFS		
P7 -Ala	EWWRSGAFS		
P8 -Ala	EWWRSGGAS		
P9 -Ala	EWWRSGGFA		

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

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

Supplementary Table 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 Table 6. Information content of the structure-based model of EWWRS GGFSF

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 Table 7. EWWRS GGFSF 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 Table 4. See enclosed excel file.

Supplementary Table 8. SIINFEKL peptide variants applied in Fig. 2a-d, Supplementary Fig.

10 and Supplementary Fig. 11. See enclosed excel file.

Supplementary Table 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	CASSGNPSTDTQYF
w876	2	CVVTEYSGGGADGLTF	CASRGQNTGELFF
	5	CAYNQGGKLIF	CASSVLNTGELFF
	5.5	unknown alpha	CASSVLNTGELFF
	6	CVVPLYSSASKIIF	CASSDTPDLNTEAFF
	9	CVGNNDMRF	CAIRRQDQNTGELFF
w830	4	CVLNNNDMRF	CASSILGASNQPQHF
w683	1	CVVALYSGGGADGLTF	CASRSQNYGYTF

* 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 Table 10. KLLLEIAPNC peptide variants applied in Fig. 2 and Supplementary Fig. 12. See enclosed excel file.

Supplementary Table 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 Table 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 MCKKLL engaging TCRs. Including the original KLEIAPNC target. Related to Fig. 3 and Supplementary Fig. 12. See enclosed excel file.

Supplementary Table 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