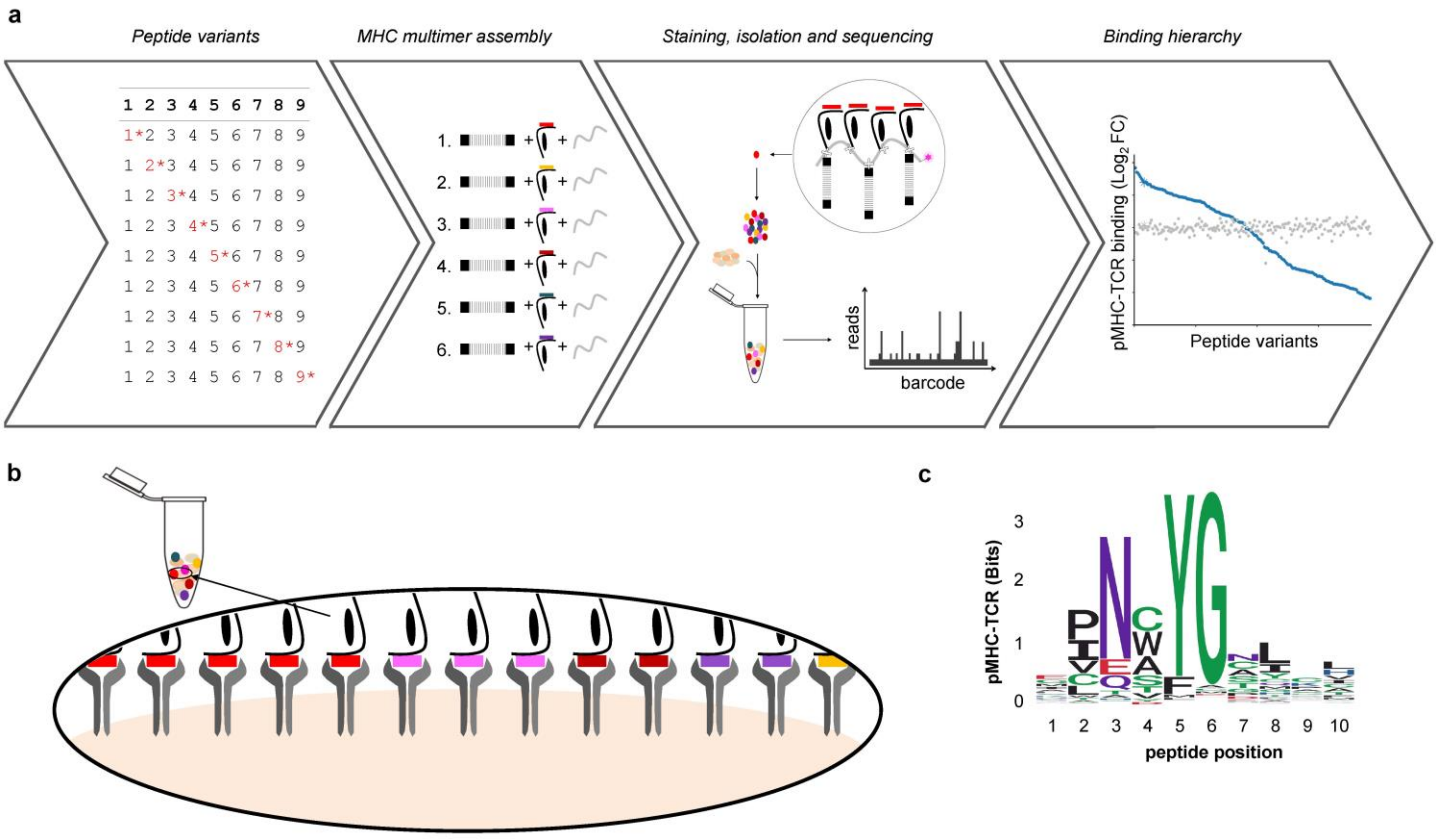
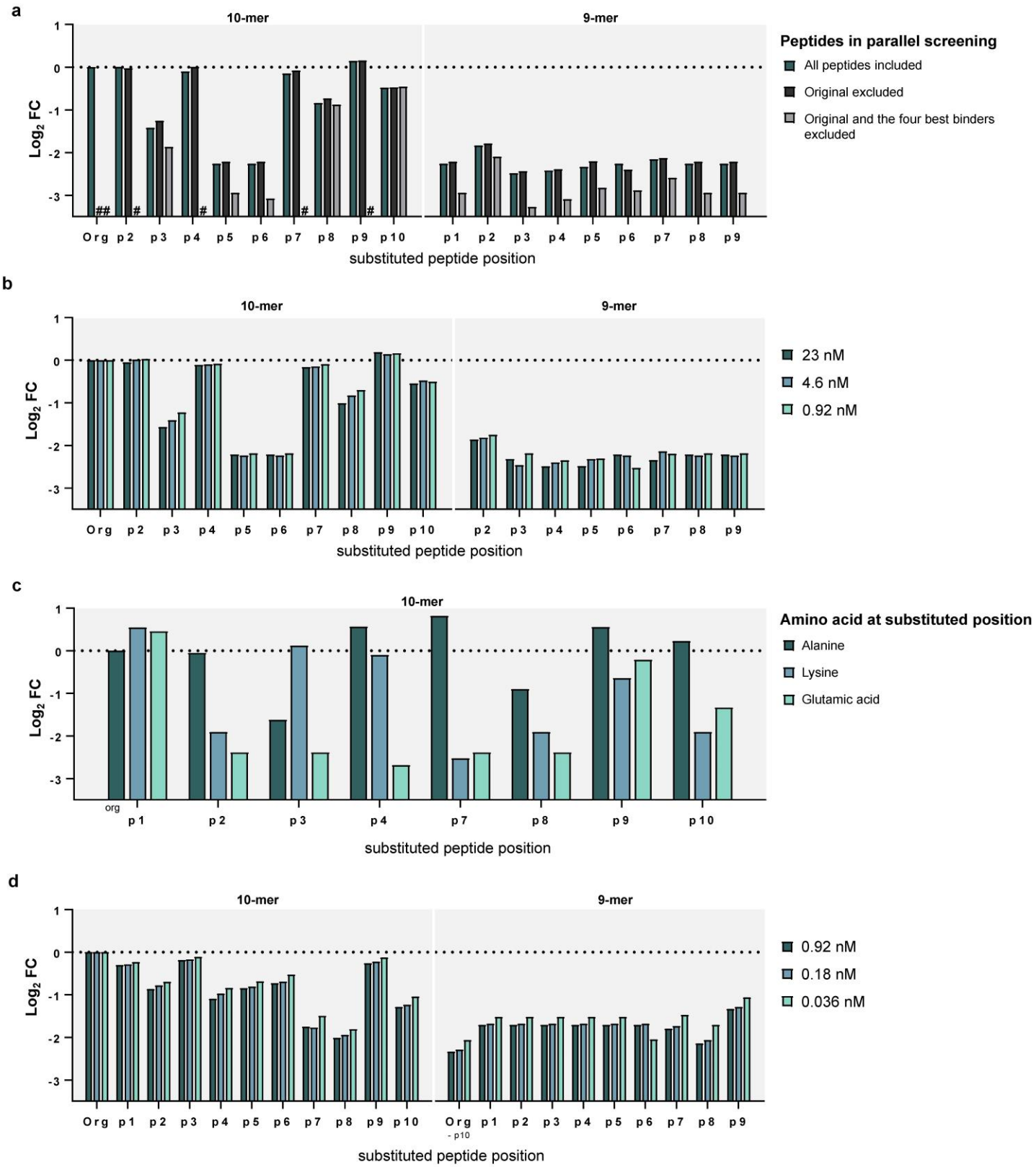


Supplementary Figure 1

Using DNA barcode-labeled MHC multimers to generate TCR fingerprints

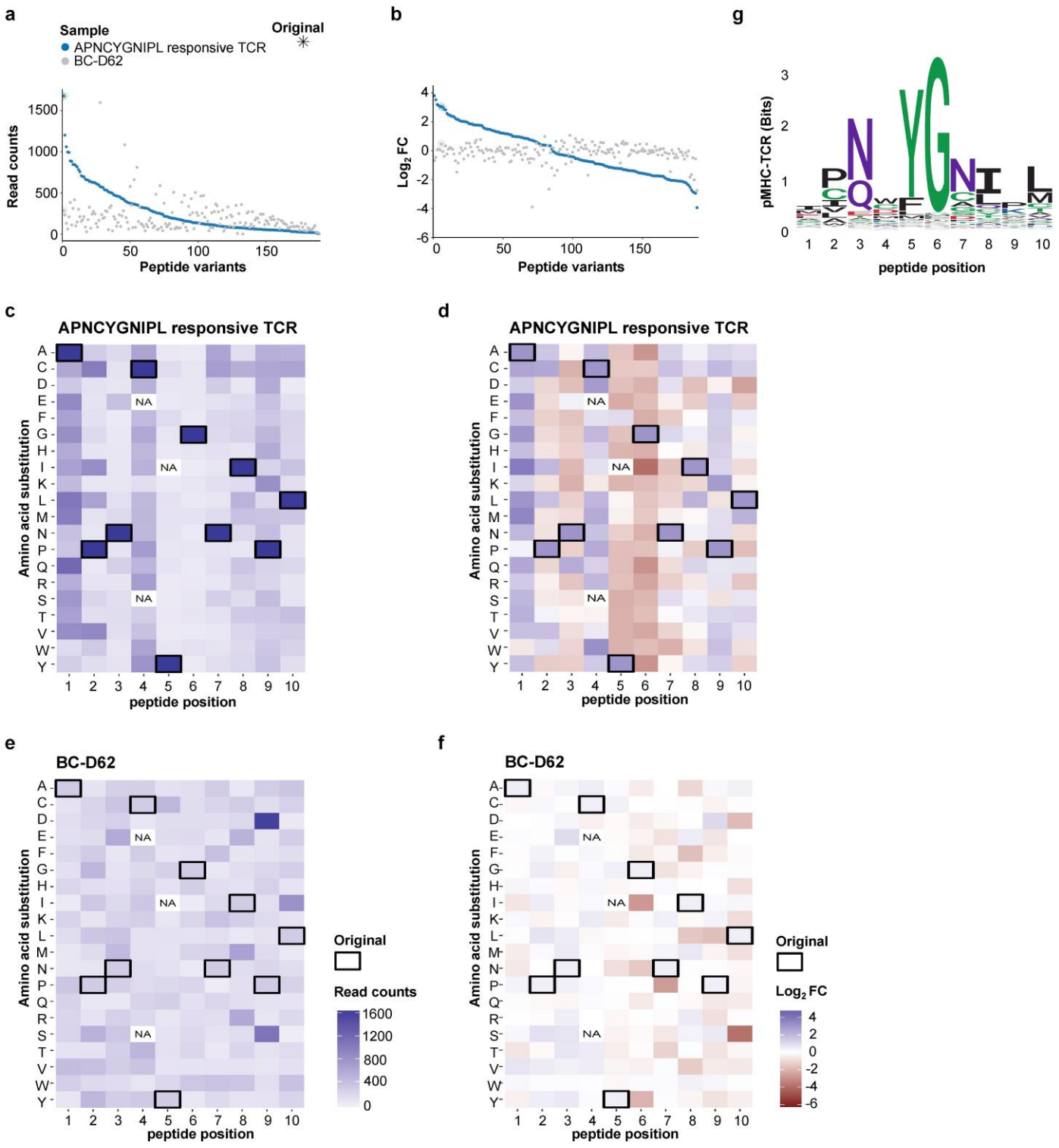
(**a**) Schematic overview of the workflow behind a TCR fingerprint. Each peptide position of the original peptide is substituted with all naturally occurring amino acids. The resultant peptide variants are assembled with MHC molecules, and individually multiplexed on a PE-labeled dextran backbone together 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 $\log_2$FC 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 fingerprint').



Supplementary Figure 2

Feasibility of using DNA barcode-labeled multimers to determine the binding hierarchy of multiple pMHCs in a 'one-pot' screen

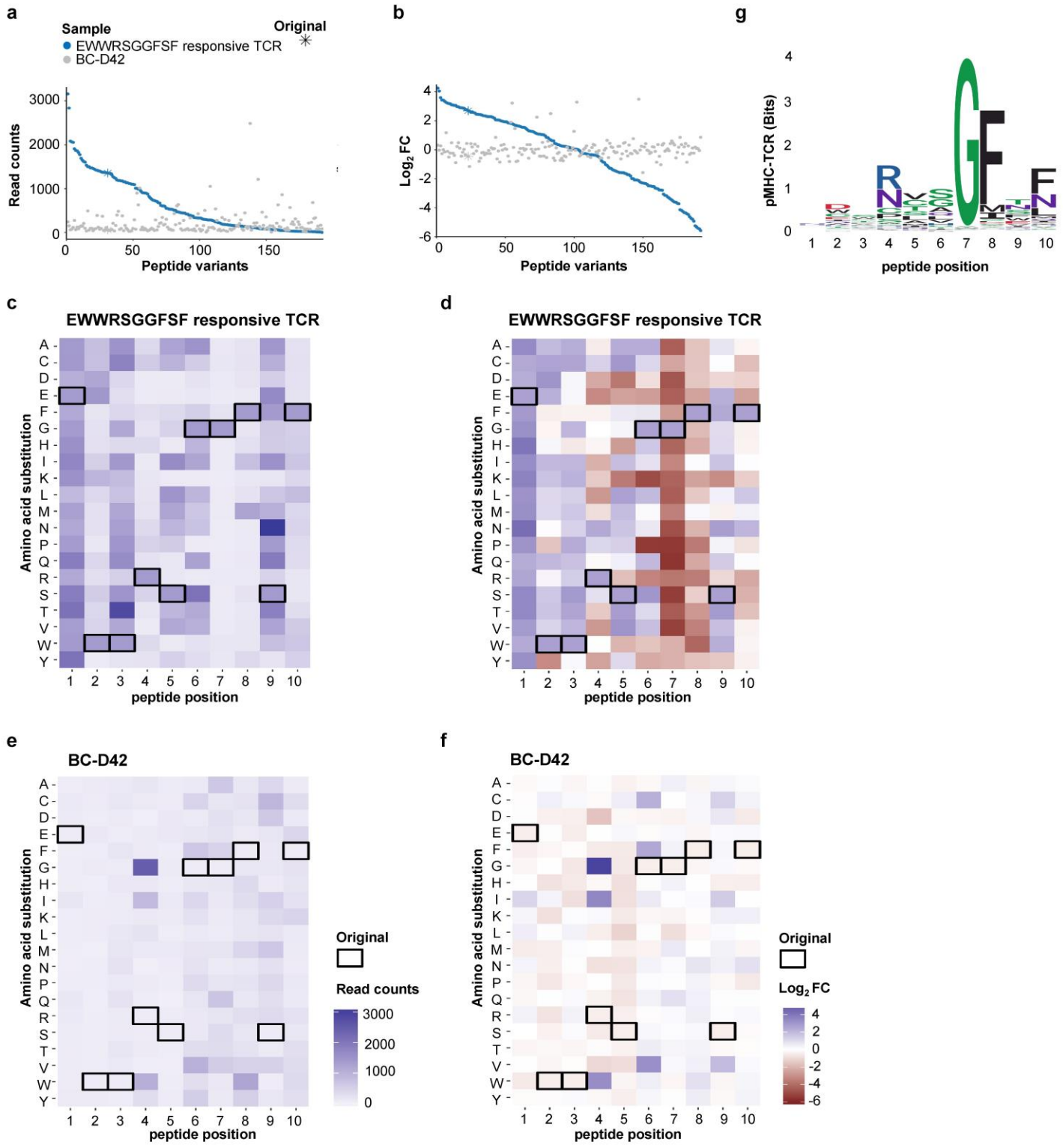
Bar plots of the relative log_2FC values from the DNA barcode-based analyses of PBMCs transduced with (**a-c**) a HLA-B*0702$_{APN}$ engaging TCR or (**d**) a HLA-A*2402$_{EWW}$ engaging TCR. Each bar represents a pMHC variant with an amino acid substitution at the given peptide position. Bars of the 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, n=1 for each variable). (**c**) Bar plot from a series of analyses where a range of 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, n=1 for each variable).



Supplementary Figure 3

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

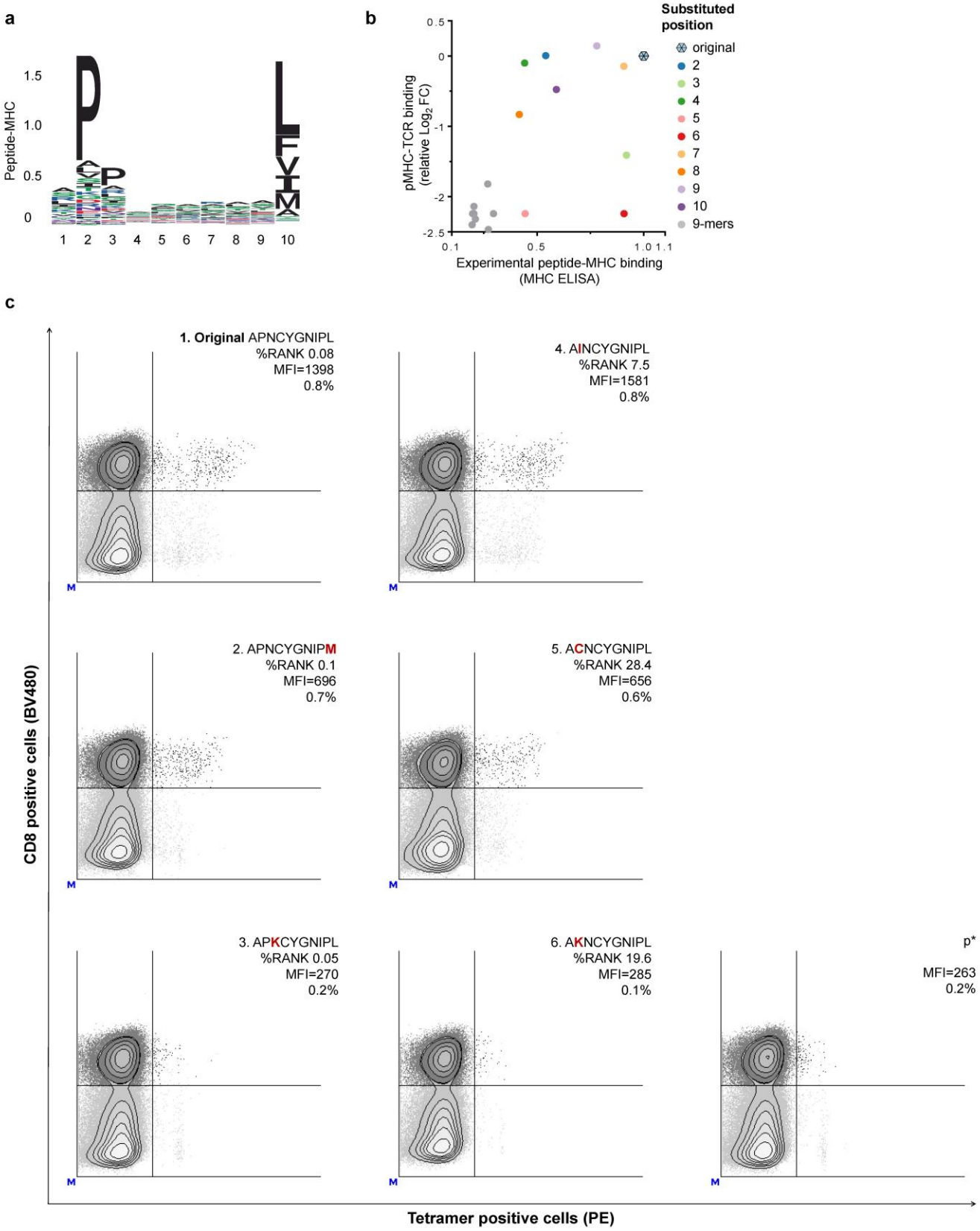
Results obtained from the DNA barcode-based analysis of T cells transduced with a TCR recognizing the HLA-B*0702 restricted peptide, APNCYGNIPL (replicate of the data shown in **Fig. 1a-c**). The analysis was performed with all possible variations of peptides created by single-position amino acid substitutions. The hierarchy of pMHC interactions expressed as (**a**) clonality reduced read counts or (**b**) log_2FC relative to a triplicate baseline sample (see Supplementary Note 1). A healthy donor PBMC sample (BC-D62) was screened with the same MHC multimer panel in parallel. The plotted order of read counts or log_2FC 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 log_2FC 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. **c** and **e** showing read counts, and **d** and **f** showing Log_2FC are colored according to the same keys respectively. (**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 Figure 4

Replicate screening of the HLA-A*2402$_{EWW}$-engaging TCR showing read counts and data from an irrelevant sample

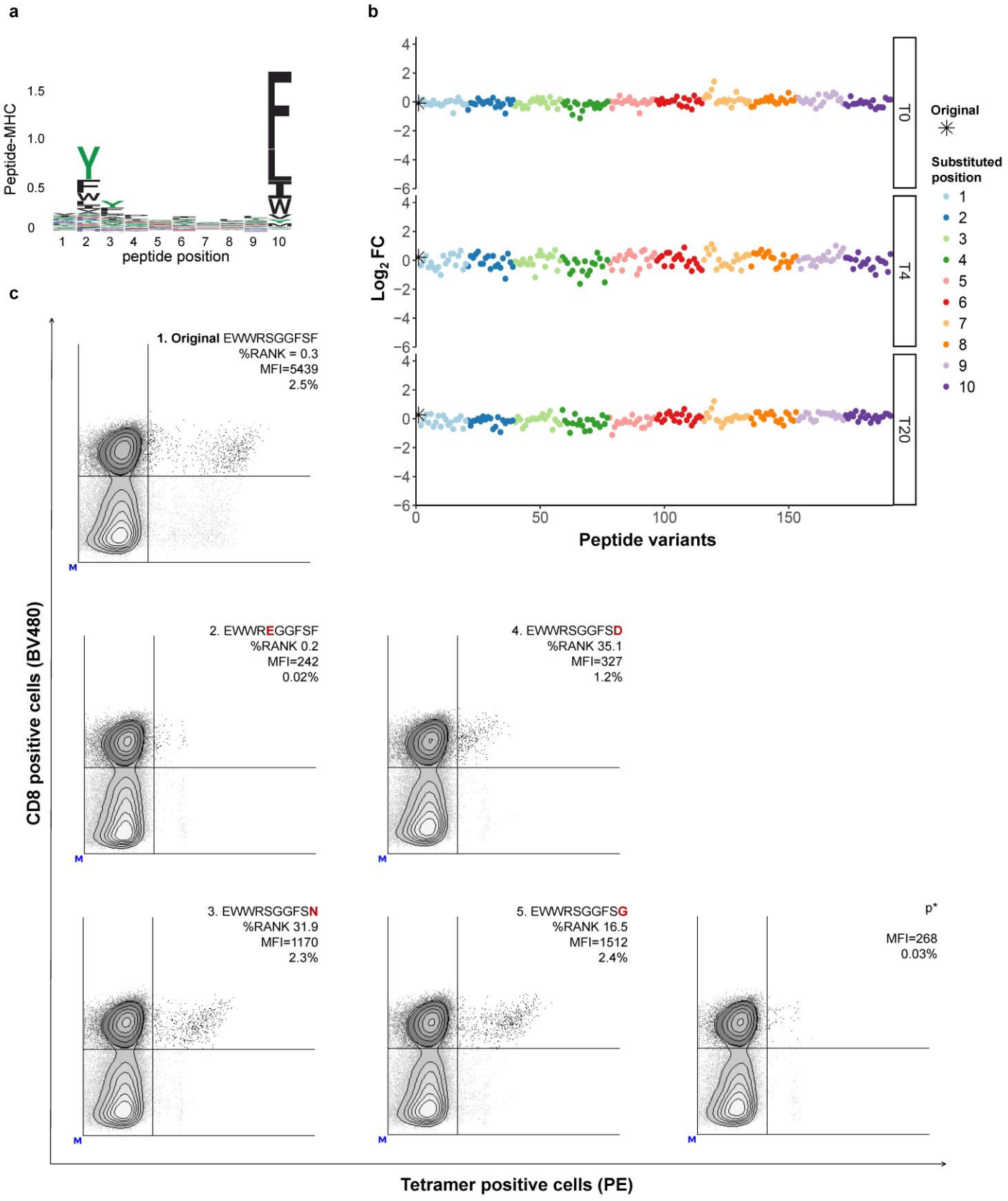
Results obtained from the DNA barcode-based analysis of T cells transduced with a TCR recognizing the HLA-A*2402 restricted peptide, EWWRSGGFSF (replicate of the data shown in **Fig. 1d-f**). The analysis was performed with all possible variations of peptides created by single-position amino acid substitutions. The hierarchy of pMHC interactions expressed as (**a**) clonality reduced read counts or (**b**) log_2FC 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_2FC of each pMHC-associated DNA barcode is determined by the hierarchy obtained from screening the HLA-A*2402$_{EWW}$ responsive TCR. (**c-f**) Heat map showing the clonality reduced read counts or log_2FC of (**c,d**) the HLA-A*2402$_{EWW}$ 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. **c** and **e** showing read counts, and **d** and **f** showing Log_2FC are colored according to the same keys respectively. (**g**) The replicate recognition pattern of the HLA-A*2402$_{EWW}$ interacting TCR, here visualized as a sequence logo based on the data from **a-d**.



Supplementary Figure 5

The effect of peptide–MHC interaction on analyzing the recognition pattern of the HLA-B*0702$_{APN}$-engaging TCR

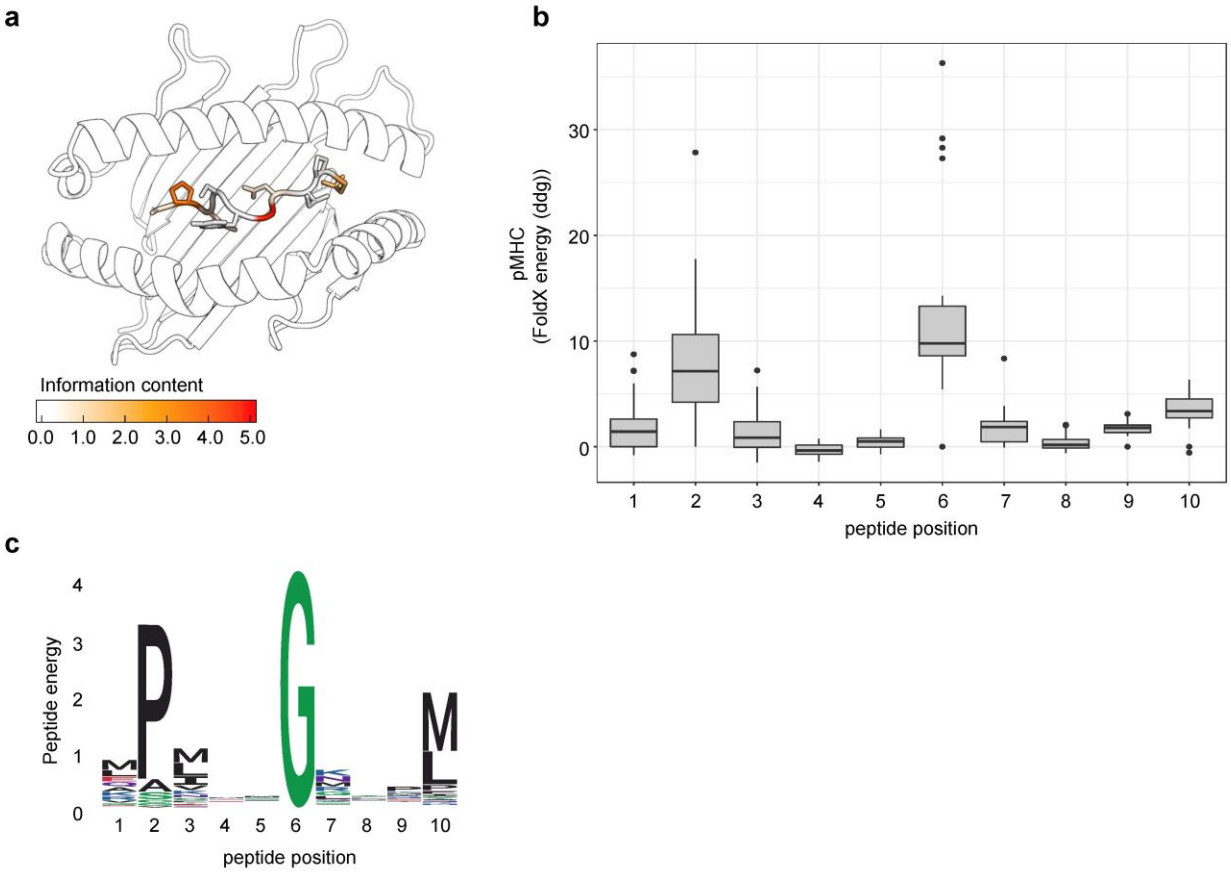
(**a**) Sequence logo showing the amino acid preferences for peptide binding to HLA-B*0702 constructed with NetMHCpan-4.0[49] by finding the binding cores of the top 1% strongest predicted binders among 100,000 natural random 10-mer peptides and visualizing these using R with the Shannon method in Seq2Logo[48]. (**b**) Scatter plot of the experimentally determined peptide binding to HLA-B*0702 (MHC ELISA, **Supplementary Note 3**), 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 twice.


a
Peptide-MHC
peptide position
b
Log2 FC
Peptide variants
T0
T4
T20
Original
Substituted position
c
1. Original EWWRSGGFSF
%RANK = 0.3
MFI=5439
2.5%
2. EWWREGGFSF
%RANK 0.2
MFI=242
0.02%
4. EWWRSGGFSD
%RANK 35.1
MFI=327
1.2%
3. EWWRSGGFSN
%RANK 31.9
MFI=1170
2.3%
5. EWWRSGGFSG
%RANK 16.5
MFI=1512
2.4%
p*
MFI=268
0.03%
CD8 positive cells (BV480)
Tetramer positive cells (PE)

Supplementary Figure 6

The effect of peptide–MHC interaction on analyzing the recognition pattern of the HLA-A*2402$_{EWW}$-engaging TCR

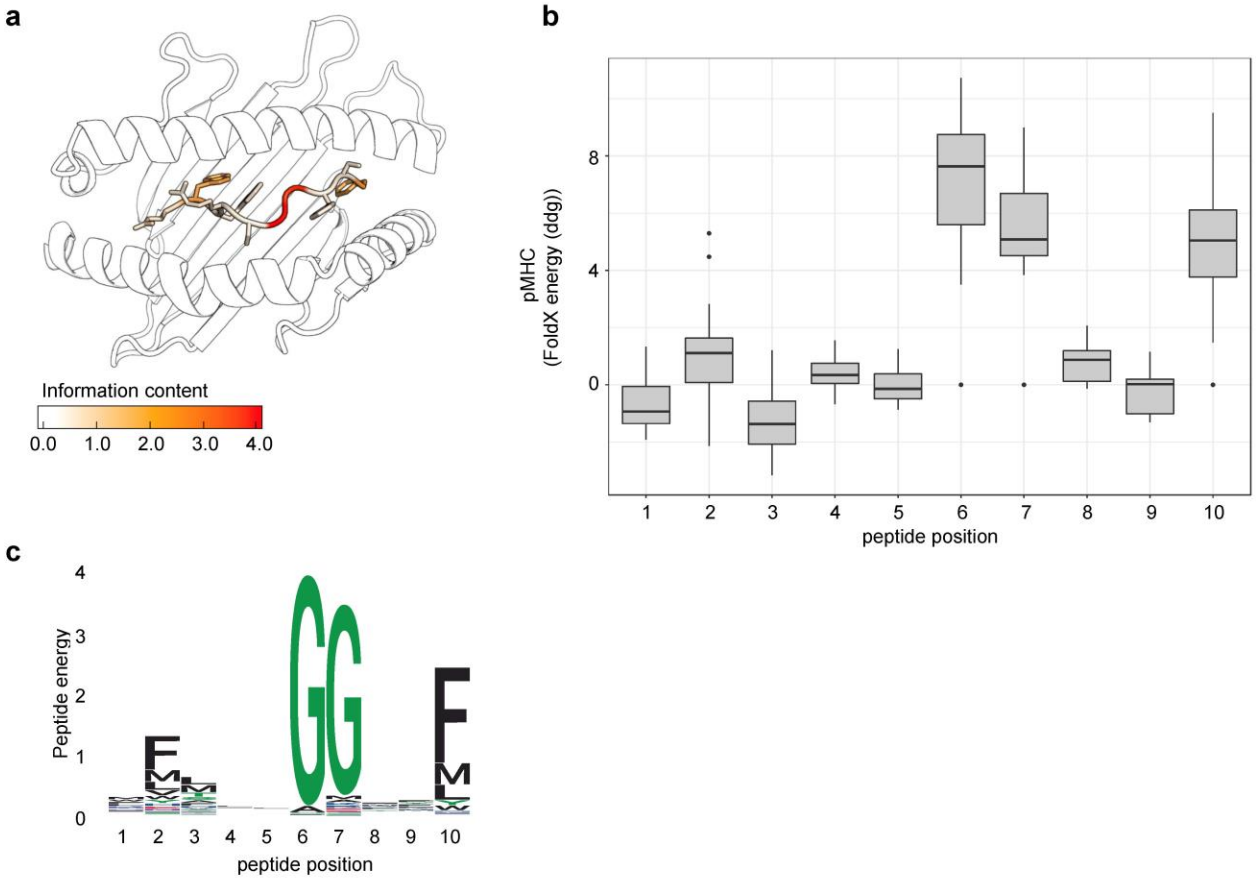
(**a**) Sequence logo showing the amino acid preferences for peptide binding to HLA-A*2402 constructed with NetMHCpan-4.0[49] by finding the binding cores of the top 1% strongest predicted binders among 100,000 natural random 10-mer peptides and visualizing these using R with the Shannon method in Seq2Logo[48]. (**b**) Shows the sequencing output expressed in log_2FC after enrichment of only the correctly folded pMHC amongst the full DNA barcode-labeled MHC multimer library comprising all 191 single substitution EWWRSGGFSF variants (**Supplementary Data 4** and **Supplementary Note 3**). 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$_{EWW}$ engaging TCR with fluorescently labeled MHC multimers carrying one of five variations of EWWRSGGFSF 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 twice.



Supplementary Figure 7

***In silico* modeling of the APNCYGNIPL peptide conformation when bound to HLA-B*0702**

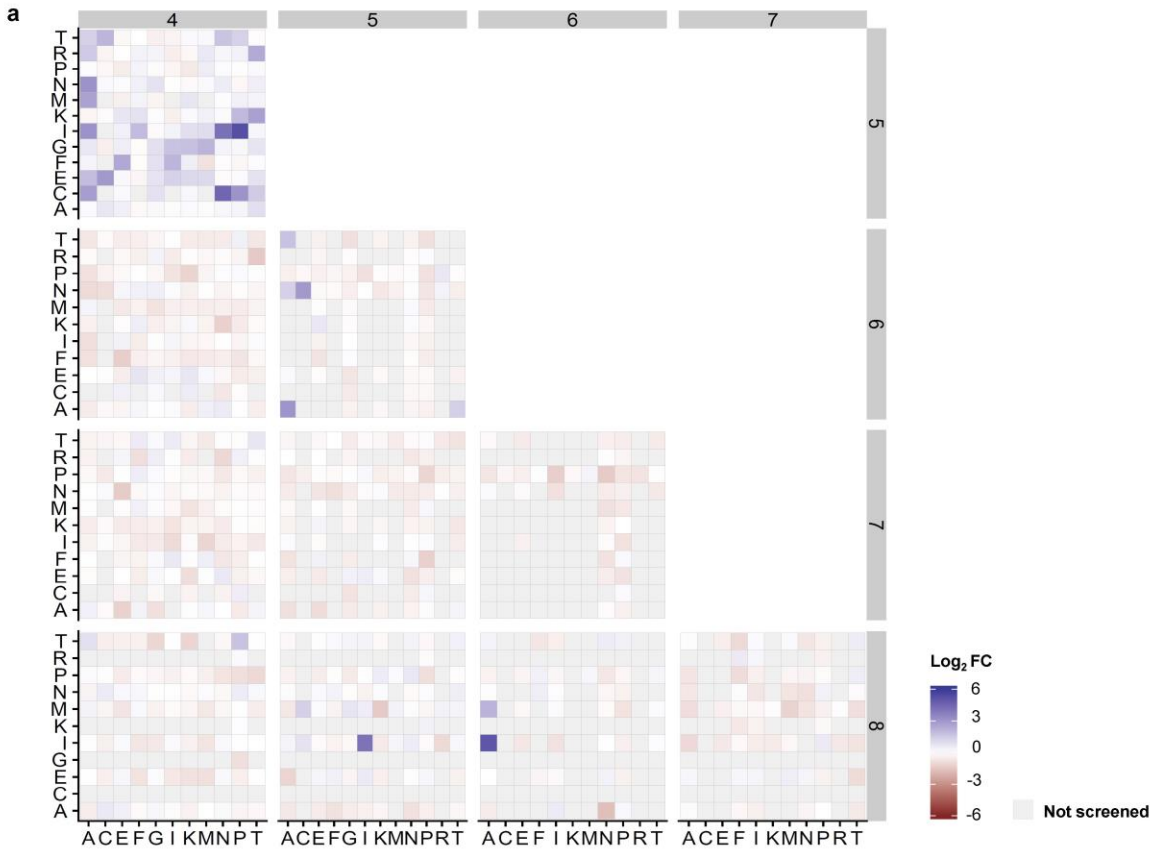
(**a**) Structural model of the original peptide, APNCYGNIPL, binding to HLA-B*0702. Color scaling indicates the importance of a given amino acid for retaining the conformation of the original peptide when bound to the MHC (the information content is listed in **Supplementary Data 5**). (**b**) Bar plot 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 peptide position (x-axis) with all naturally occurring amino acids (n=19). 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**.

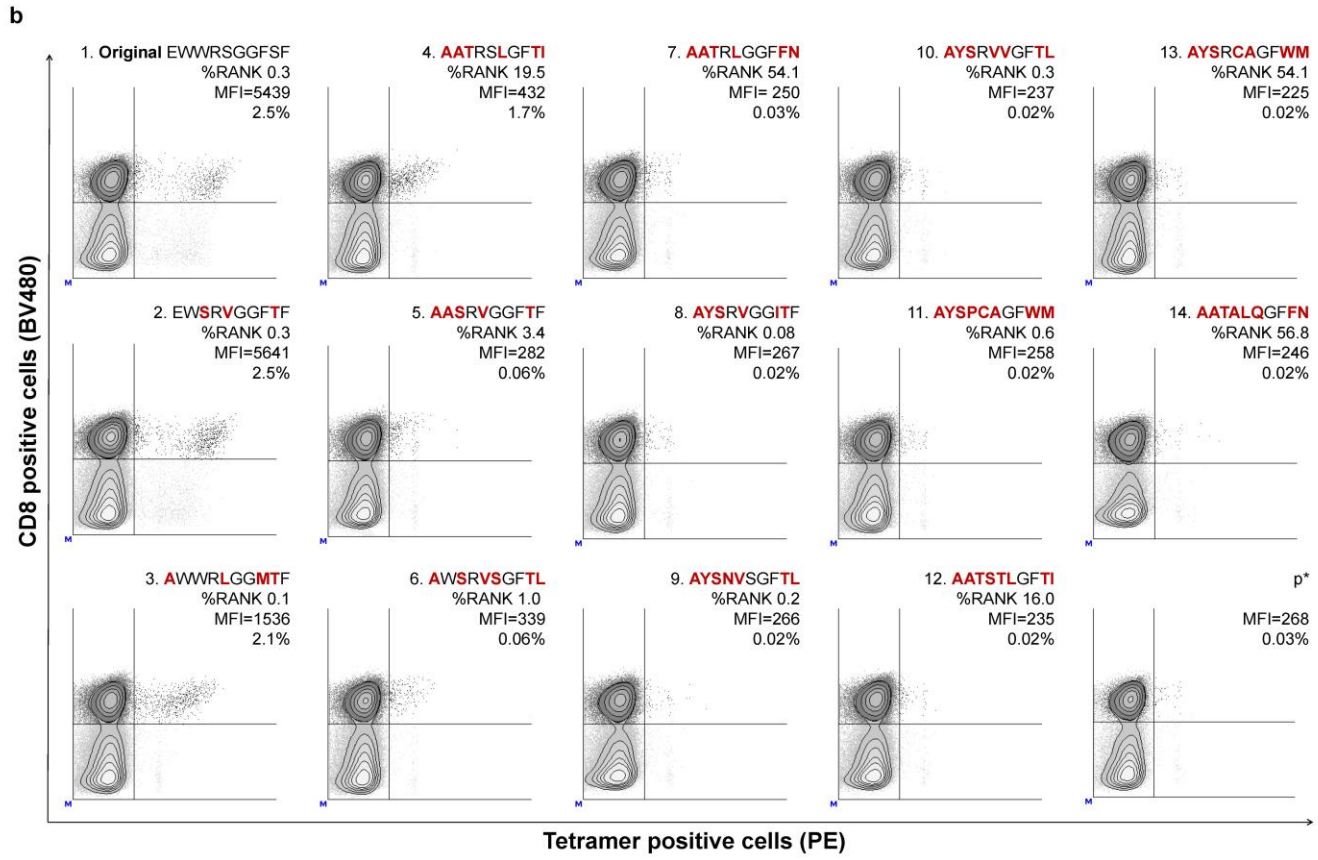


Supplementary Figure 8

***In silico* modeling 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 Data 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 peptide position (x-axis) with all naturally occurring amino acids (n=19). 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$_{EWW}$ based on the energy change ($\Delta\Delta$) calculated in **b**.

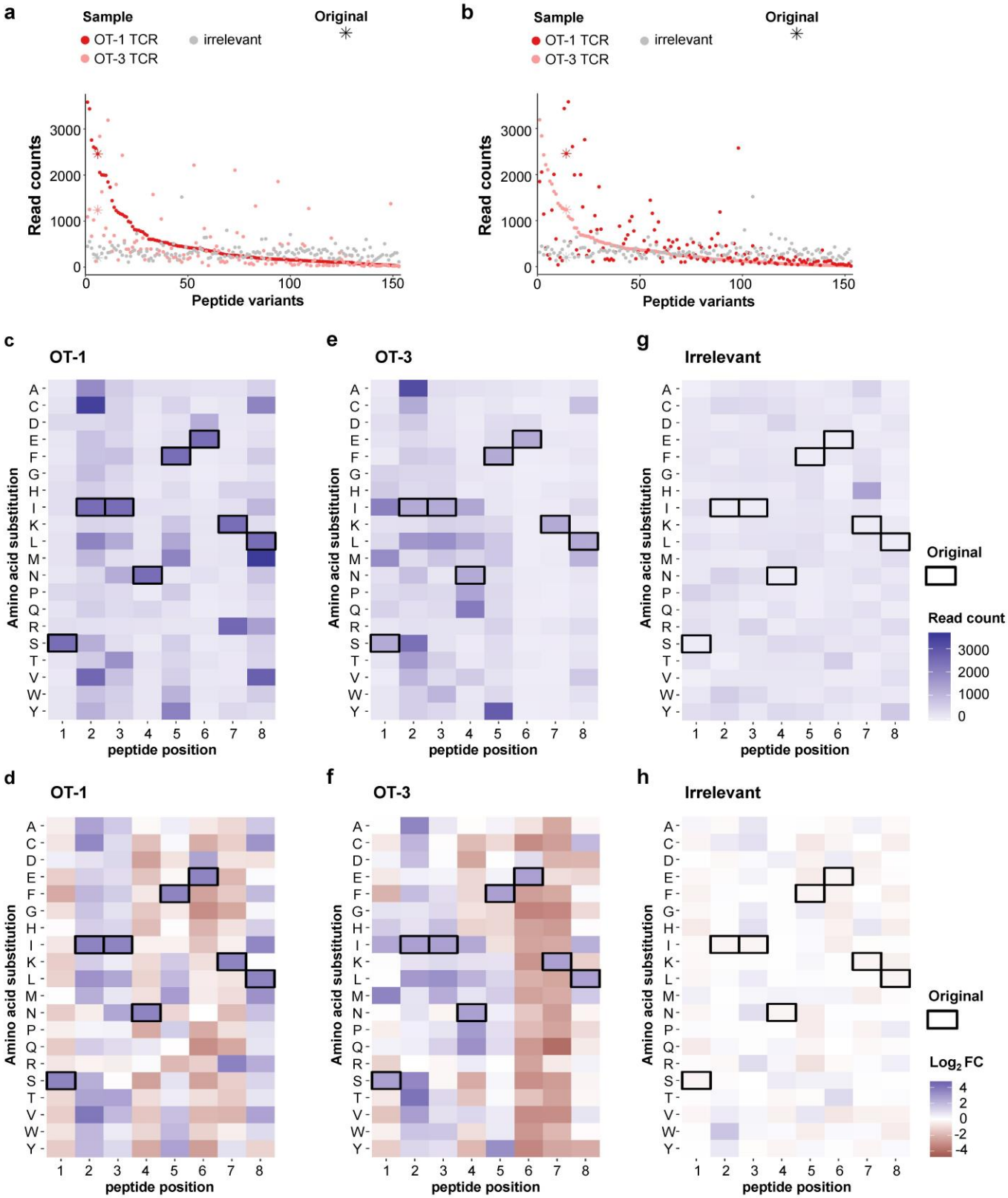

a
4
5
6
7
5
6
7
8
Log2 FC
6
3
0
-3
-6
Not screened
A C E F G I K M N P T
A C E F G I K M N P R T
A C E F I K M N P R T
A C E F I K M N P R T


b
CD8 positive cells (BV480)
Tetramer positive cells (PE)
1. Original EWWRSGGFSF
%RANK 0.3
MFI=5439
2.5%
4. AATRSLGFTI
%RANK 19.5
MFI=432
1.7%
7. AATRLGGFFN
%RANK 54.1
MFI= 250
0.03%
10. AYSRVVGFTL
%RANK 0.3
MFI=237
0.02%
13. AYSRCAGFWM
%RANK 54.1
MFI=225
0.02%
2. EWSRVGGFTF
%RANK 0.3
MFI=5641
2.5%
5. AASRVGGFTF
%RANK 3.4
MFI=282
0.06%
8. AYSRVGGITF
%RANK 0.08
MFI=267
0.02%
11. AYSPCAGFWM
%RANK 0.6
MFI=258
0.02%
14. AATALQGFFN
%RANK 56.8
MFI=246
0.02%
3. AWWRLGGMTF
%RANK 0.1
MFI=1536
2.1%
6. AWSRVSGFTL
%RANK 1.0
MFI=339
0.06%
9. AYSNVSGFTL
%RANK 0.2
MFI=266
0.02%
12. AATSTLGFTI
%RANK 16.0
MFI=235
0.02%
p*
MFI=268
0.03%

Supplementary Figure 9

The HLA-A*2402$_{EWW}$-engaging TCRs recognize peptides with multiple amino acid substitutions

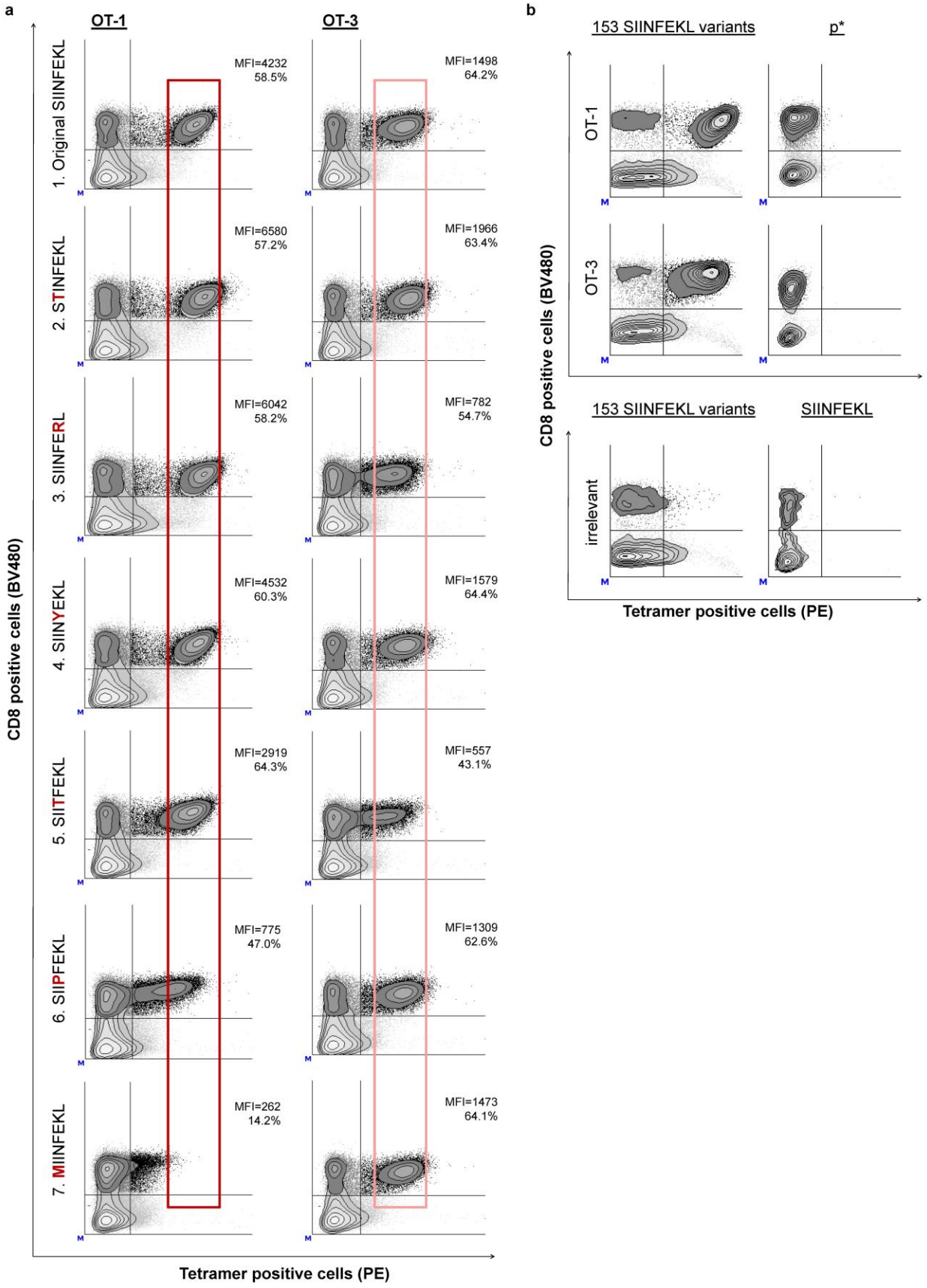
(**a**) Heat map showing the amino acid preferences of the TCR recognizing the HLA-A*2402 restricted peptide, EWWRSGGFSF, expressed as log_2FC of read counts relative to a triplicate baseline sample (see Supplementary Note). The TCR transduced clones were stained with a MHC multimer library composed 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=776, **Supplementary Data 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$_{EWW}$ engaging TCR with fluorescently labeled MHC multimers carrying one of 14 variations of EWWRSGGFSF 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 fluorescent-based MHC multimer stainings were performed once.



Supplementary Figure 10

The DNA barcode-based MHC multimer analysis of OT-1 and OT-3 T cells, showing read counts, log_2FC 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 log_2FC 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.

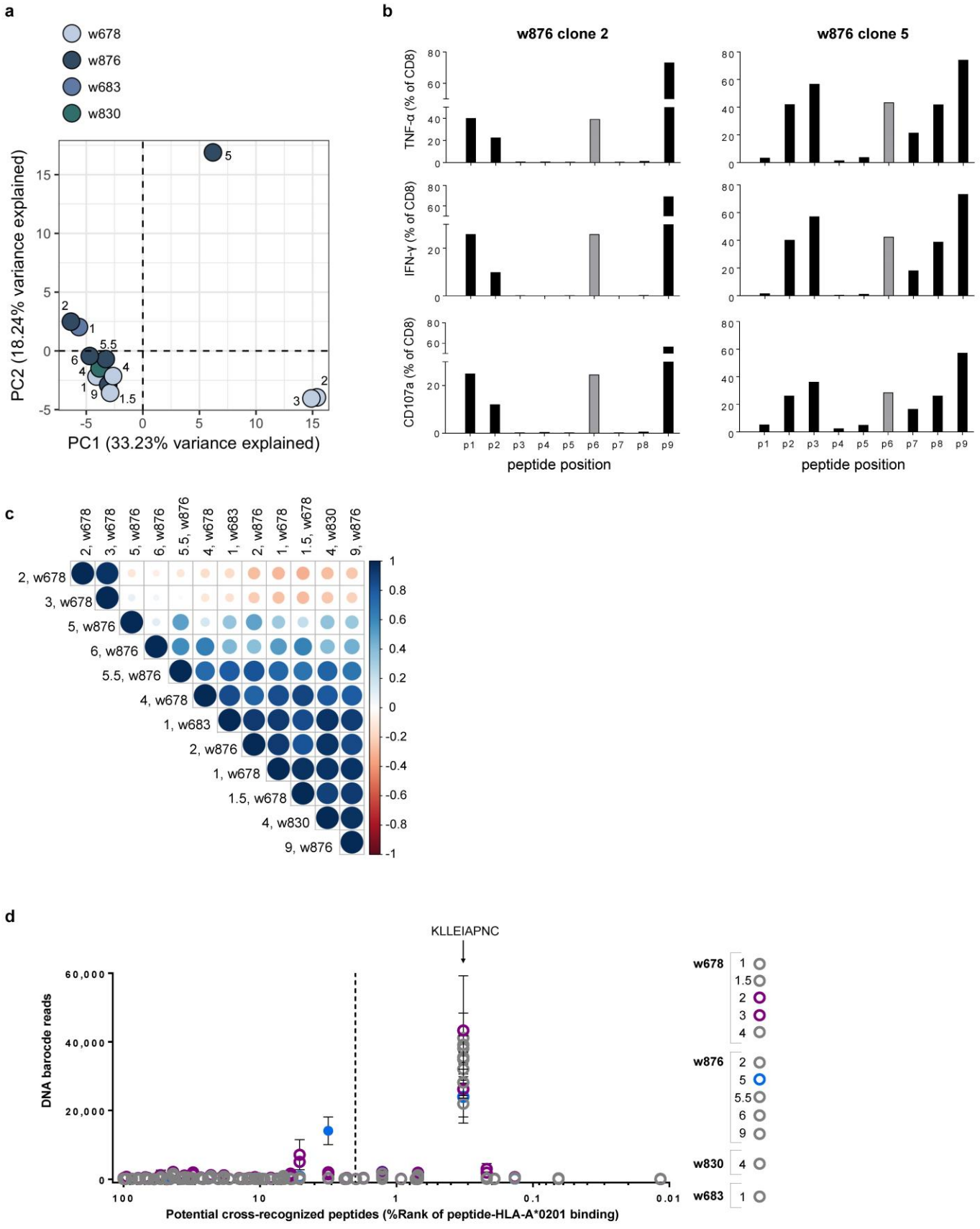


Nature Biotechnology: doi:10.1038/nbt.4303

Supplementary Figure 11

Fluorescence-based MHC multimer analyses of OT-1 and OT-3 T cells

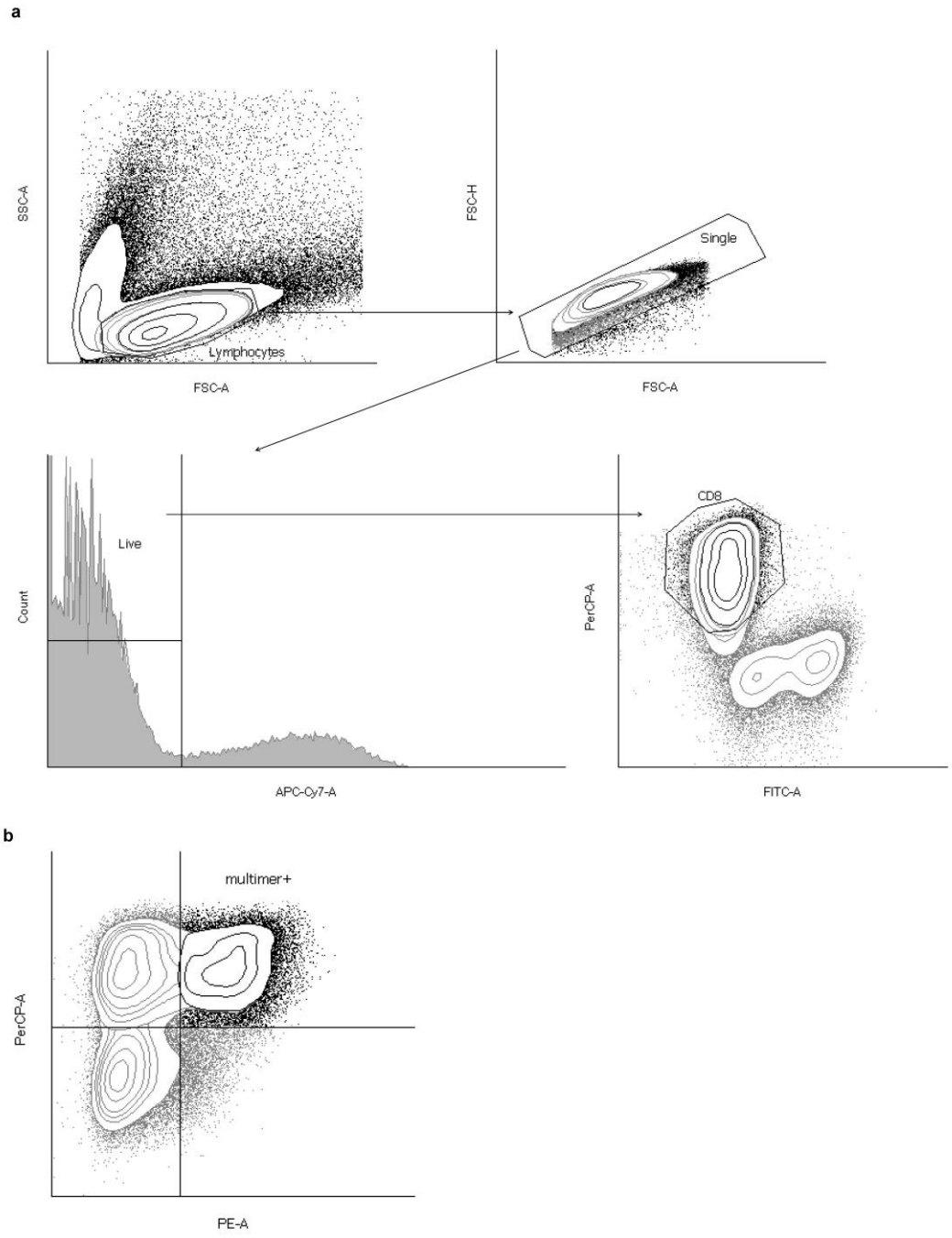
(**a**) Dot plots from the fluorescent-based MHC multimer staining of splenocytes derived from OT-1 or OT-3 mice. MHC tetramers carrying one of seven variations of SIINFEKL were applied in single color MHC multimer analyses. When comparing the staining of OT-1 and OT-3 in relation to their binding to 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, 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.


a
w678
w876
w683
w830
PC2 (18.24% variance explained)
PC1 (33.23% variance explained)
b
w876 clone 2
w876 clone 5
TNF-α (% of CD8)
IFN-γ (% of CD8)
CD107a (% of CD8)
peptide position
c
d
KLLEIAPNC
DNA barocde reads
Potential cross-recognized peptides (%Rank of peptide-HLA-A*0201 binding)

Supplementary Figure 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 (n=12 individual T cell clones). (**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**. Correlation was determined using Pearson's correlation coefficient. 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 KLLEIAPNC 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 Data 12)**. The 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 indicates a response that were also stained with fluorescently labeled MHC tetramers. The plotted data is mean of duplicate analyses and error bars indicate range of duplicates.



Supplementary Figure 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. This gating strategy was applied for all MHC multimer analyses.