

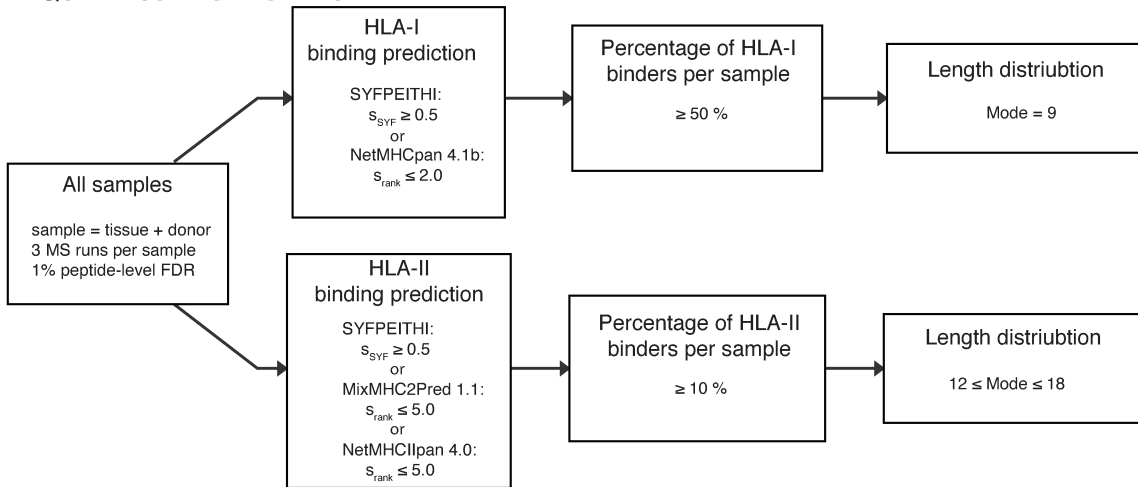
The HLA Ligand Atlas: A benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy

Ana Marcu[§], Leon Bichmann[§], Leon Kuchenbecker[§], Daniel Johannes Kowalewski, Lena Katharina Freudenmann, Linus Backert, Lena Mühlenbruch, András Szolek, Maren Lübke, Philipp Wagner, Tobias Engler, Sabine Matovina, Jian Wang, Mathias Hauri-Hohl, Roland Martin, Konstantina Kapolou, Juliane Sarah Walz, Julia Velz, Holger Moch, Luca Regli, Manuela Silginer, Michael Weller, Markus W. Löffler, Florian Erhard, Andreas Schlosser, Oliver Kohlbacher, Stefan Stevanović, Hans-Georg Rammensee, Marian Christoph Neidert

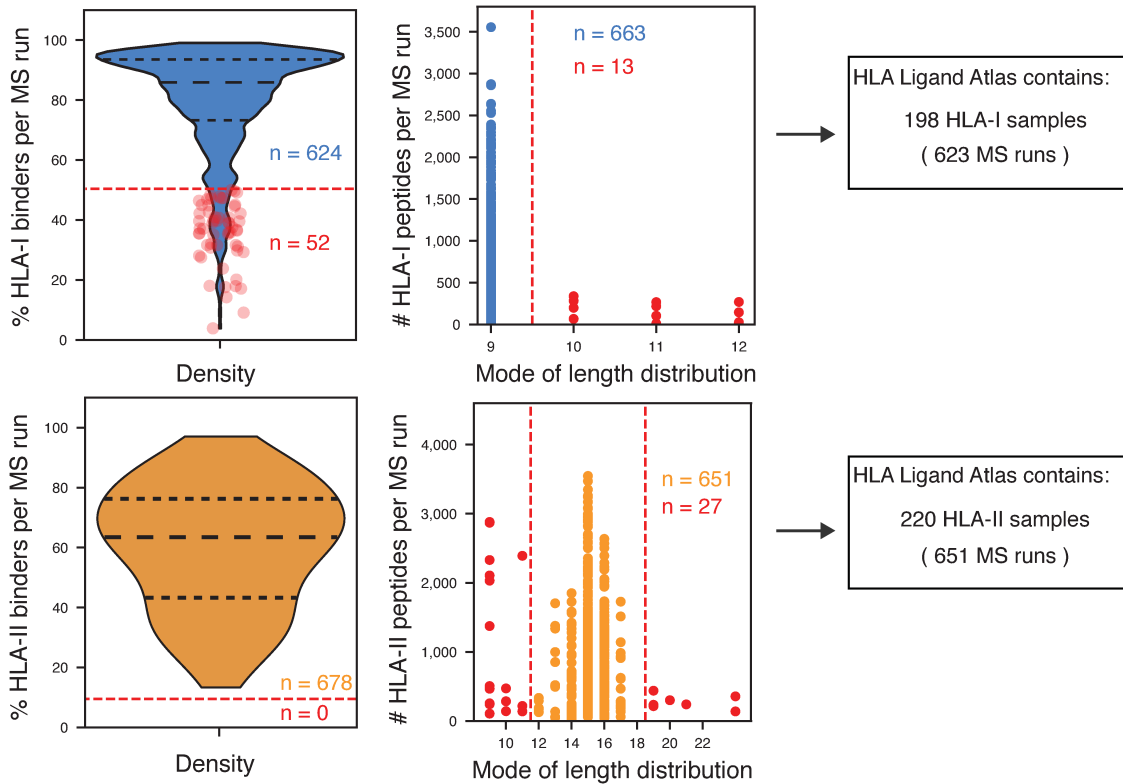
[§]these authors contributed equally to this work

Supplementary FIGURES

A QUALITY CONTROL WORKFLOW



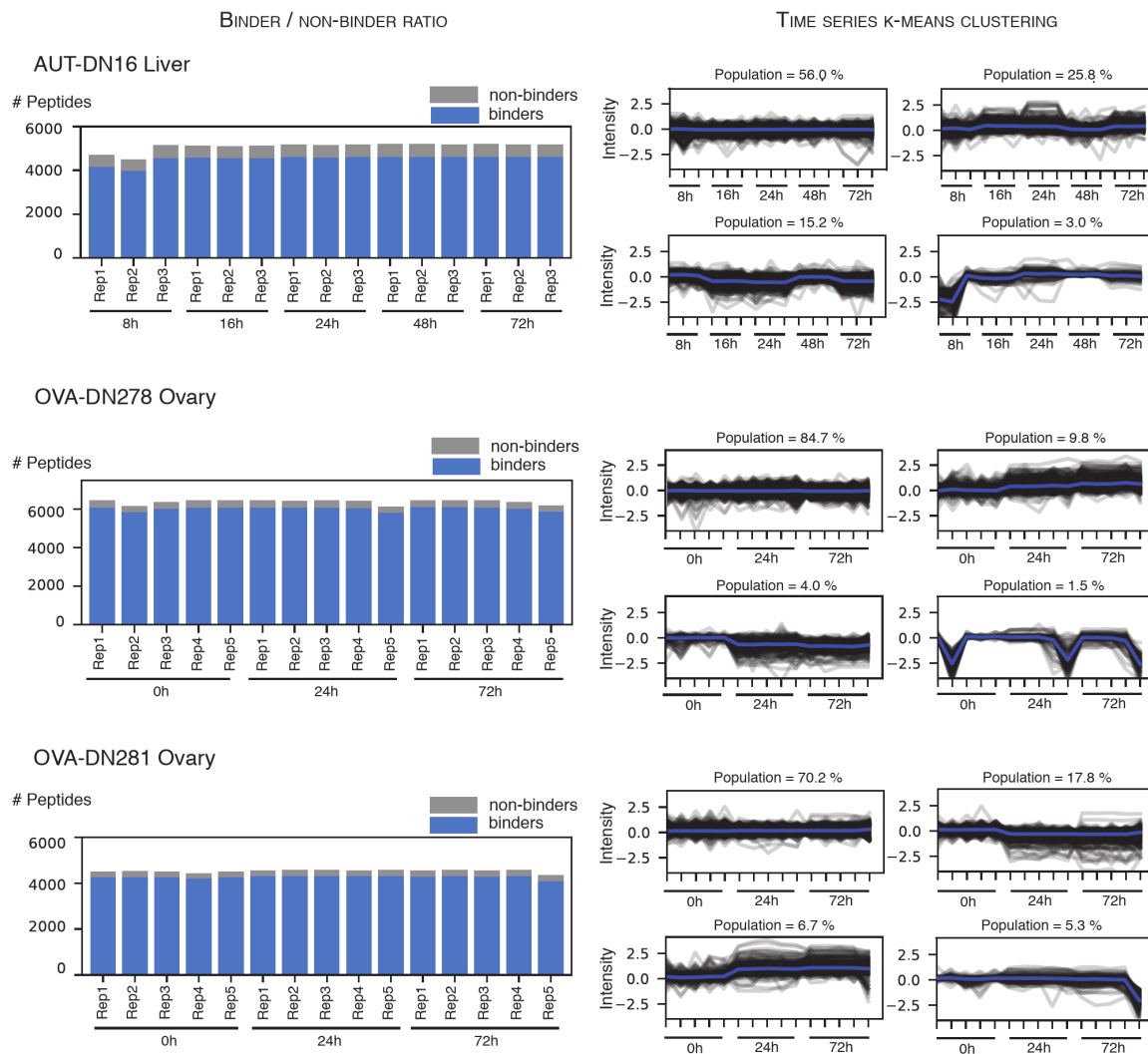
B OUTLIER REMOVAL AND TOTAL CONTENT OF THE HLA LIGAND ATLAS



Supplementary Figure S1: Computational workflow and quality control thresholds (related to Figure 1)

(A) Data from three LC-MS/MS runs (technical replicates) per sample (one tissue from one subject) were processed with MHCquant with a local peptide-level FDR of 1%. Identified peptides were categorized into peptides predicted as strong, weak, or non-binders. HLA-I peptides were predicted by SYFPEITHI and NetMHCpan 4.0, employing a cut-off threshold of $s_{\text{SYF}} \geq 0.5$ and $s_{\text{rank}} \leq 2.0$, respectively. HLA-II peptides were predicted with MixMHCPred, NetMHCIIpan 4.0 and SYFPEITHI with equal thresholds. Samples with less than 50% and 10% predicted binders per technical replicate for HLA-I and HLA-II immunopeptidomes were excluded from further analysis and are not included in the HLA Ligand Atlas database release. Finally, we employed a stringent cut-off concerning the mode of the peptide length distribution for HLA-I (mode equal to 9) and -II (mode between 12 to 18) immunopeptidomes.

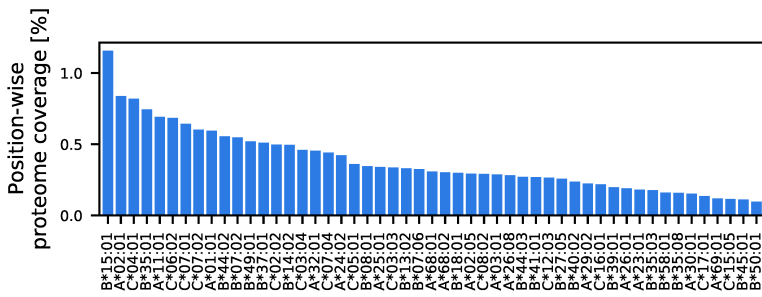
(B) LC-MS/MS runs not pertaining to the aforementioned QC thresholds (dashed red lines) concerning the percentage of predicted HLA-binders and the peptide length distribution were not included in the HLA Ligand Atlas database release: Violin plots (left) depict the percentage of peptides predicted to be HLA-binders per LC-MS/MS run and the quality control cut-off for LC-MS/MS runs employed for HLA-I and -II immunopeptidomes. Dot plots (center) depict the mode of the peptide length distribution encountered per LC-MS/MS run and the quality control cut-off employed for HLA-I and -II. The number of LC-MS/MS runs (HLA-I - blue and HLA-II orange) failing the QC thresholds is indicated by red dots. The final number of LC-MS/MS runs and samples selected for inclusion into the HLA Ligand Atlas database release and data analysis are indicated on the right and differ for HLA-I and -II experiments.



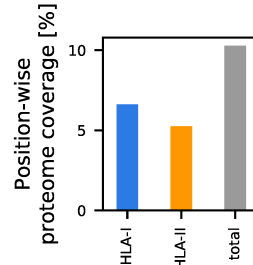
Supplementary Figure S2: Time-Series Experiment (related to Figure 1)

The time series experiment was carried out on three biological samples from three subjects. Bar plots (left hand side) indicate the number of identified HLA-I predicted binders (blue) and predicted non-binders (grey) across technical replicates and timepoints for each timeseries. Time series (right hand side) indicate individual clusters (K-Means using four seeds) of trajectories of quantified MS1 intensity across technical replicates and timepoints. Intermediate trend lines for each cluster are indicated in blue and the percentage of trajectories populating a given cluster is annotated at the top edge of each plot. The trajectories of all timepoints were set in relation to the initial timepoint. The analysis reveals that the number of identified peptides and their percentage of predicted HLA-binders is constant and that most trajectories do not vary much across time points.

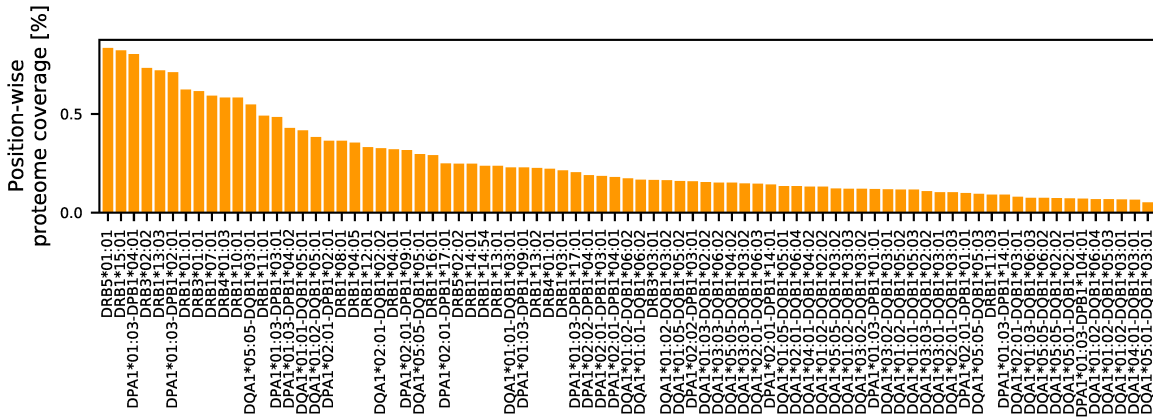
A POSITION-WISE PROTEOME COVERAGE HLA-I



C POSITION-WISE PROTEOME COVERAGE INCLUDING NON-BINDERS



B POSITION-WISE PROTEOME COVERAGE HLA-II

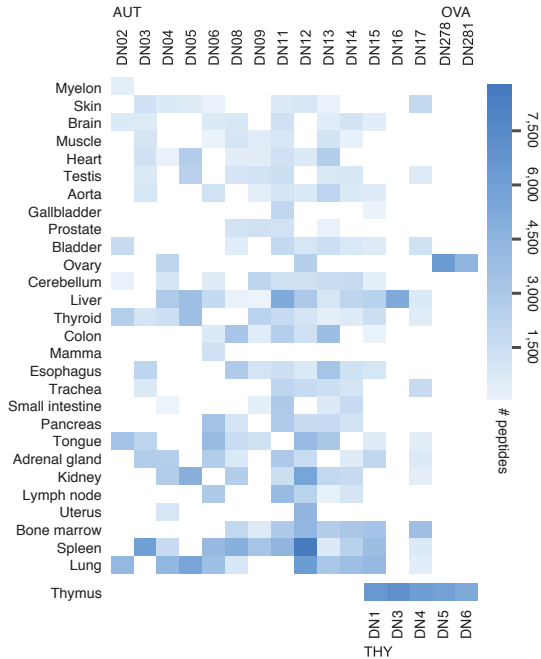


Supplementary Figure S3: Position-wise proteome coverage (related to Figure 1C, Figure 2B and C)

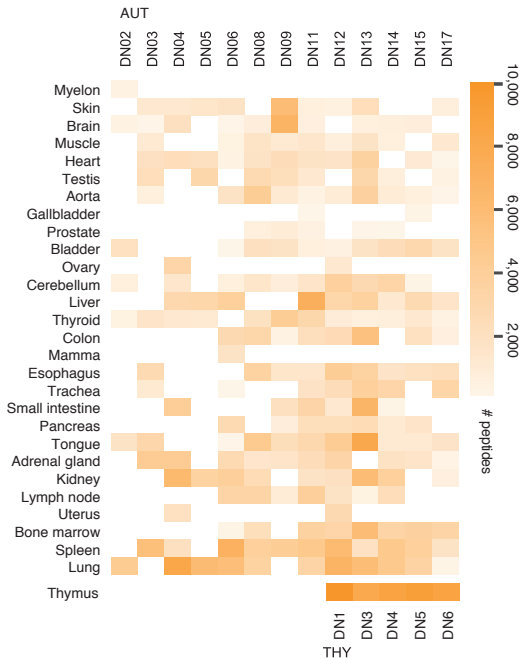
(A-B) Analysis of the different propensities of HLA-I (blue) and II (orange), allotypes concerning the presentation of peptides. In turn different percentages of the source proteome are position-wise covered. For example, HLA-B*15:01 ligands cover over 1% of the HLA-I source proteome (see Figure 2B, C).

(C) Global analysis of the position-wise coverage of all proteins presented on either HLA-I (blue), HLA-II (orange), or both independent of specific allotypes (see Figure 1C).

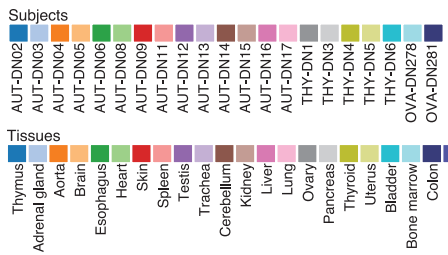
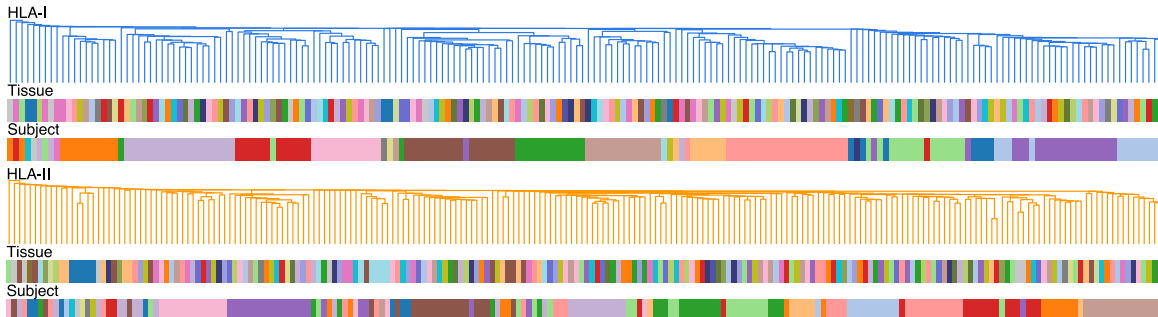
A NUMBER OF HLA-I PEPTIDES PER DONOR



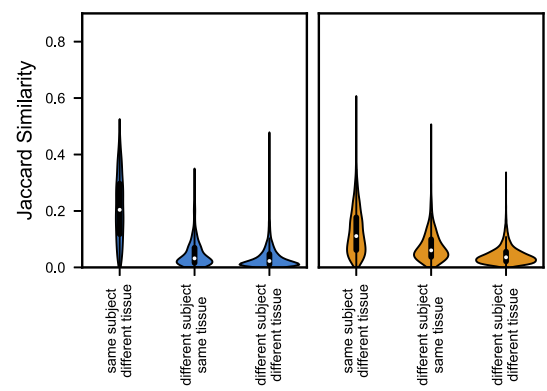
B NUMBER OF HLA-II PEPTIDES PER DONOR



C HIERARCHICAL CLUSTERING OF SAMPLES BASED ON PEPTIDE JACCARD SIMILARITY



D DISTRIBUTION OF PAIRWISE PEPTIDE JACCARD SIMILARITIES

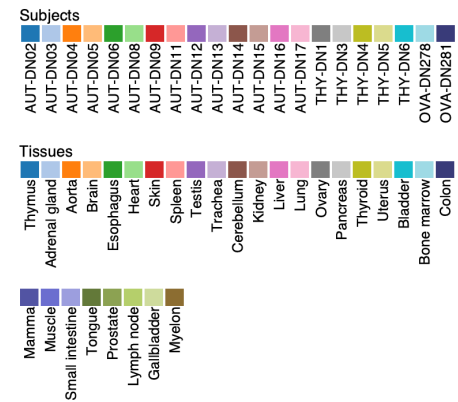
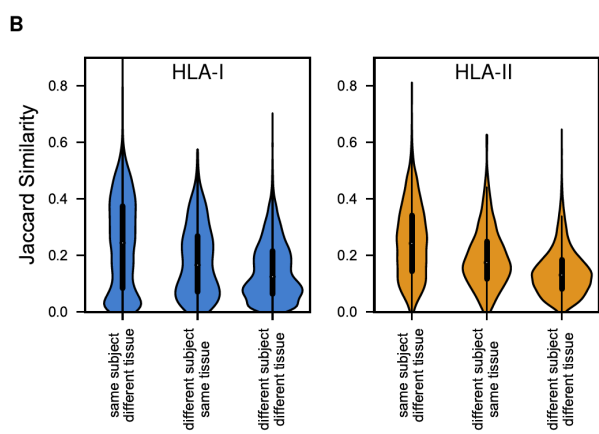
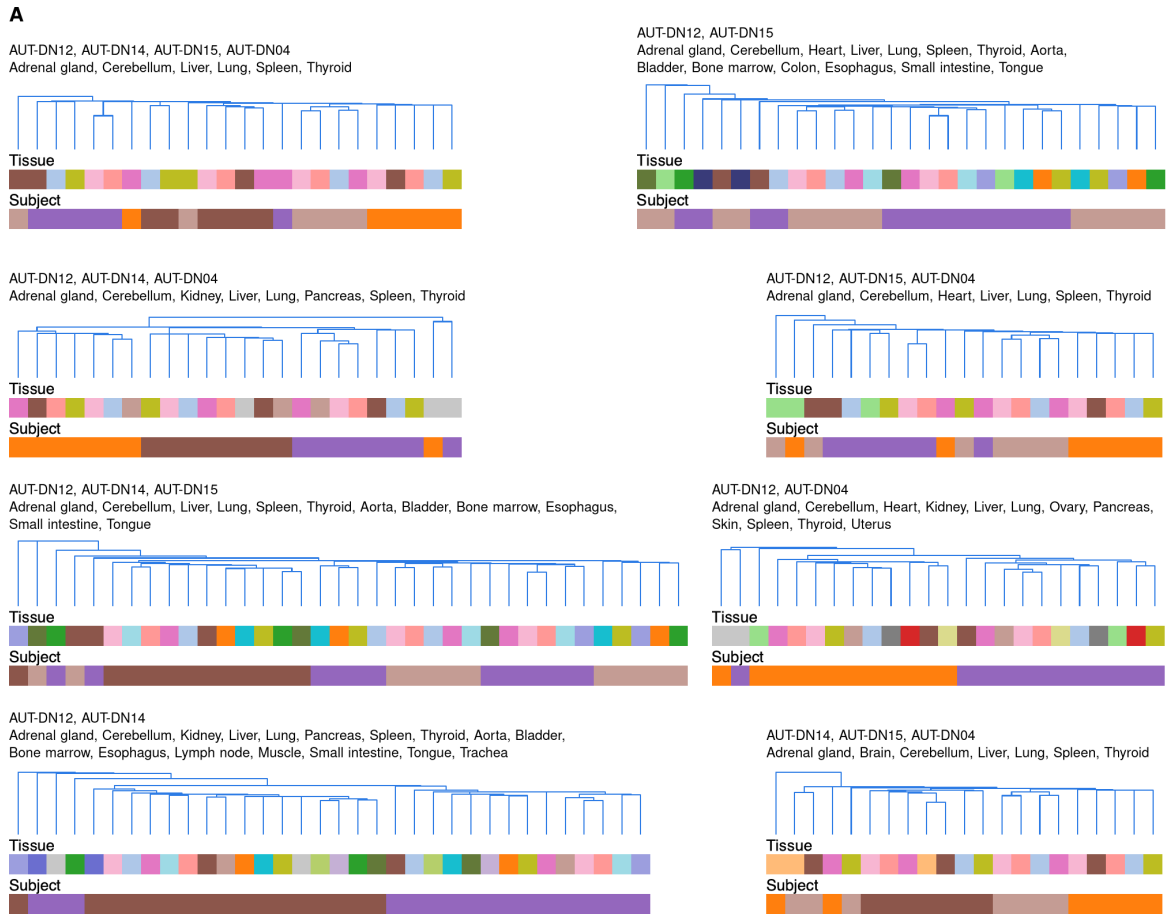


Supplementary Figure S4: Distribution of peptide identifications and yields across tissues and subjects and peptide-based hierarchical sample clustering (related to Figure 2D, E, and Figure 4A)

(A-B) HLA-I peptide yields per tissue and subject are illustrated in a heatmap for HLA-I and -II. The color range is in accordance with the number of peptides identified in each sample as indicated in the legend on the right (HLA-I – blue, HLA-II – orange).

(C) Pairwise hierarchical clustering of samples based on the Jaccard similarity between HLA-I (blue) and HLA-II (orange) peptides. The dendrogram illustrates the nearest neighbor based on the similarity between tissues and subjects. See Figure 2D.

(D) Violin plots illustrate the distribution of the peptide Jaccard similarity index for each pairwise comparison between the same subject - different tissues; different subjects - the same tissue, and different subject - different tissues. Results confirmed for HLA-I and -II source proteins Figure 2D and E, respectively.

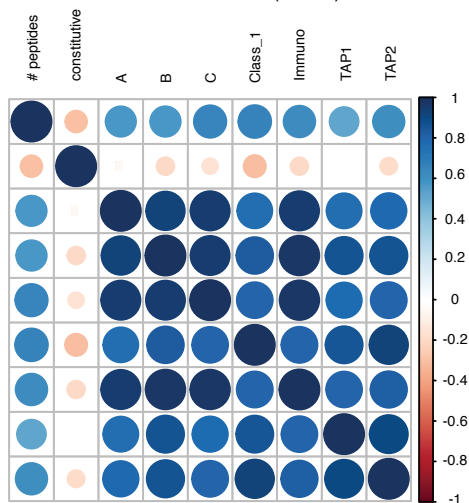


Supplementary Figure S5: Hierarchical sample clustering for HLA and tissue subset matched sample groups (related to Figure 2D, E)

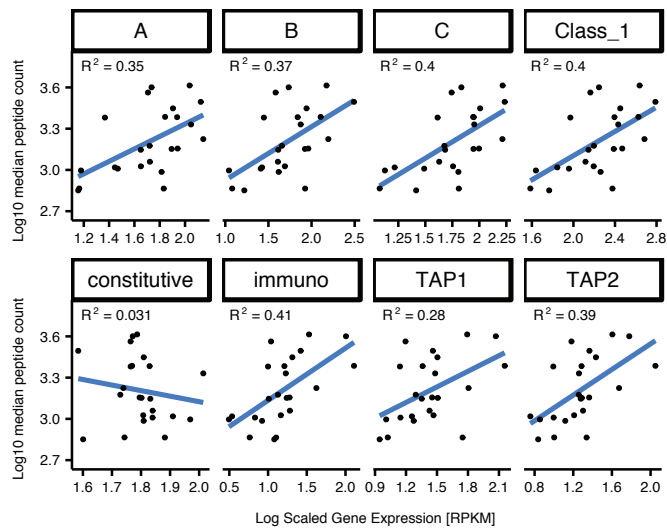
(A) Hierarchical clustering of predicted HLA-A*02:01 binding peptide sets based on their Jaccard similarity. Samples were selected from donors carrying the HLA-A*02:01 allele and subsequently filtered for predicted binders against that allele. The samples were then divided into subsets of donors and tissues such that every tissue was measured in every donor. The resulting subsets of samples were then clustered as described in the Methods section.

(B) The distributions of Jaccard similarity scores for equivalently HLA allele matched sample pairs over all HLA allotypes present in the dataset.

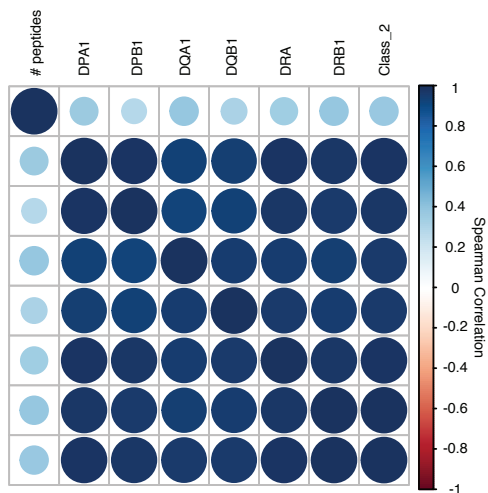
A CROSS CORRELATION MATRIX (HLA-I)



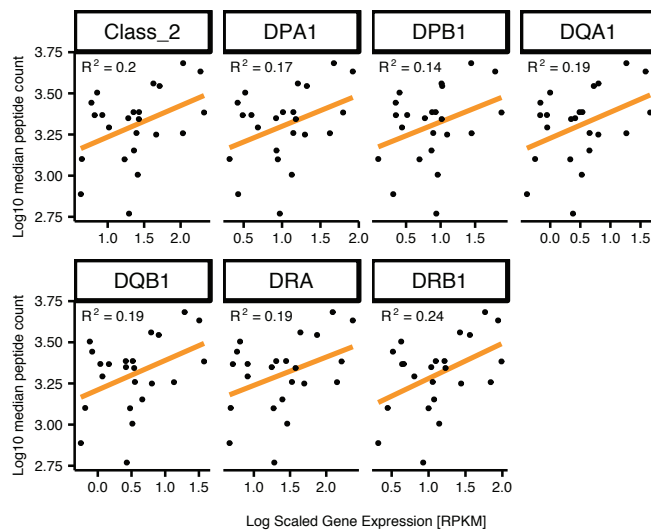
B LINEAR MODELS (HLA-I)



C CROSS CORRELATION MATRIX (HLA-II)



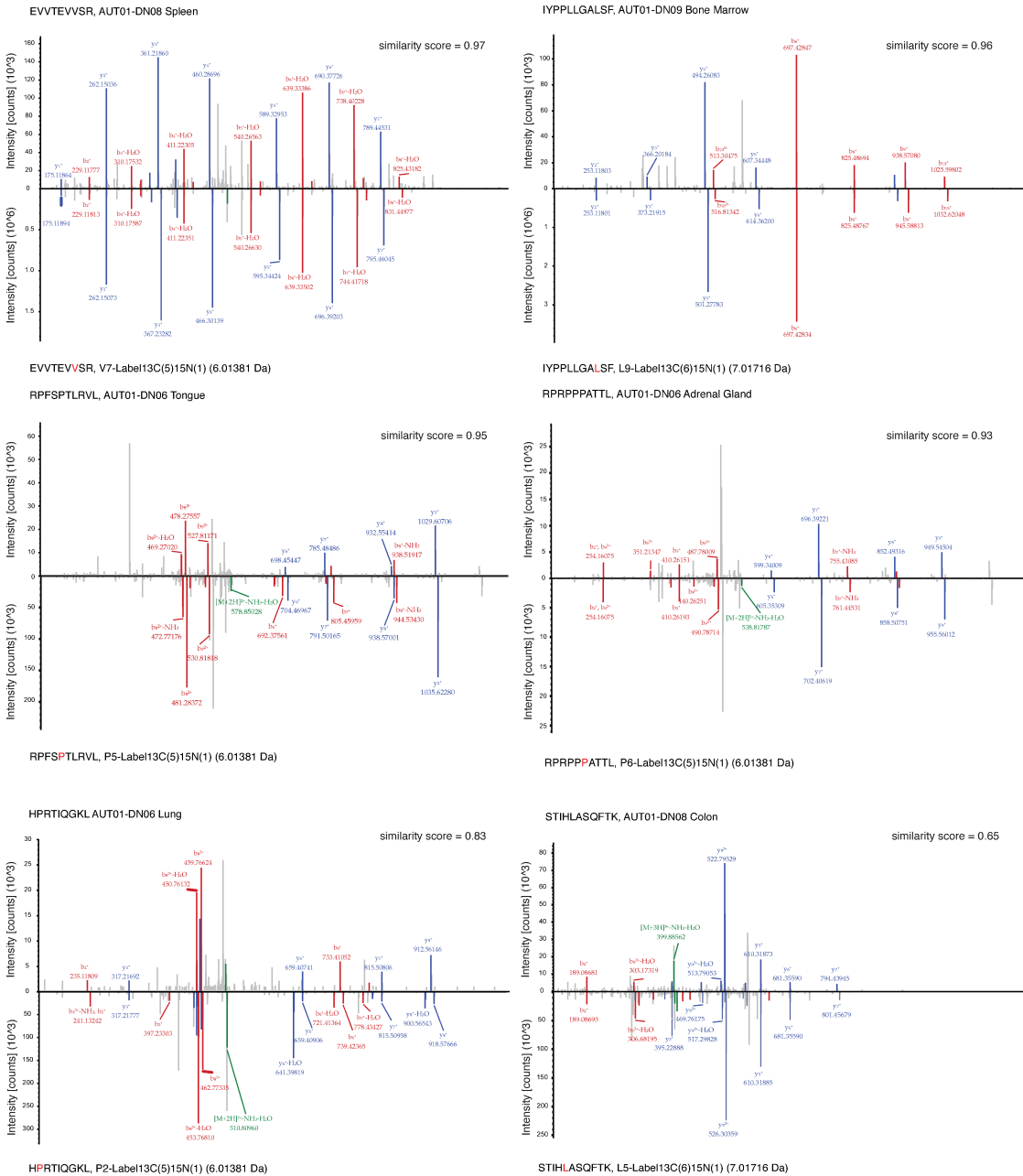
D LINEAR MODELS (HLA-II)



Supplementary Figure S6: Linear models of gene expression across tissues (related to Figure 3). The correlation of HLA-I and -II peptide yields across tissues was assessed with respect to the expression levels (RPKM) of multiple immune related genes of the HLA-I and -II antigen presentation pathway. (RNA expression values were taken from a different publication (Boegel et al., 2018)).

(A,C) Symmetrical cross-correlation matrices were generated, illustrating the spearman correlation coefficients (ρ) of the number of identified HLA-I and -II ligands with expression levels (RPKM) of relevant genes in the HLA-I and -II antigen presentation pathways and among each other. The color-coded dots and their size represent the degree (Spearman ρ) of positive (blue) or negative correlation (red).

(B,D) In addition, linear models illustrate HLA-I and -II immunopeptidome yields correlated to log scaled gene expression values (RPKM) of genes involved in the HLA-I and -II antigen processing pathway, respectively. The R2 correlation coefficient is depicted in the top left corner for each comparison.



Supplementary Figure S7: Manual validation of cryptic peptide identifications with isotope labeled synthetic peptides (related to Figure 6). Six exemplary spectral comparisons depict cryptic peptides identified in diverse tissues and subjects (upper spectrum) related to their synthetic isotope labeled counterpart (lower spectrum). The isotope labeled amino acid is highlighted in red and its corresponding label and mass are annotated. Matching b- (red), y-ions (blue) and neutral losses (green) and their corresponding fragment masses are annotated to each peak in the spectra. The spectral similarity score of the given comparison is annotated in the top right.