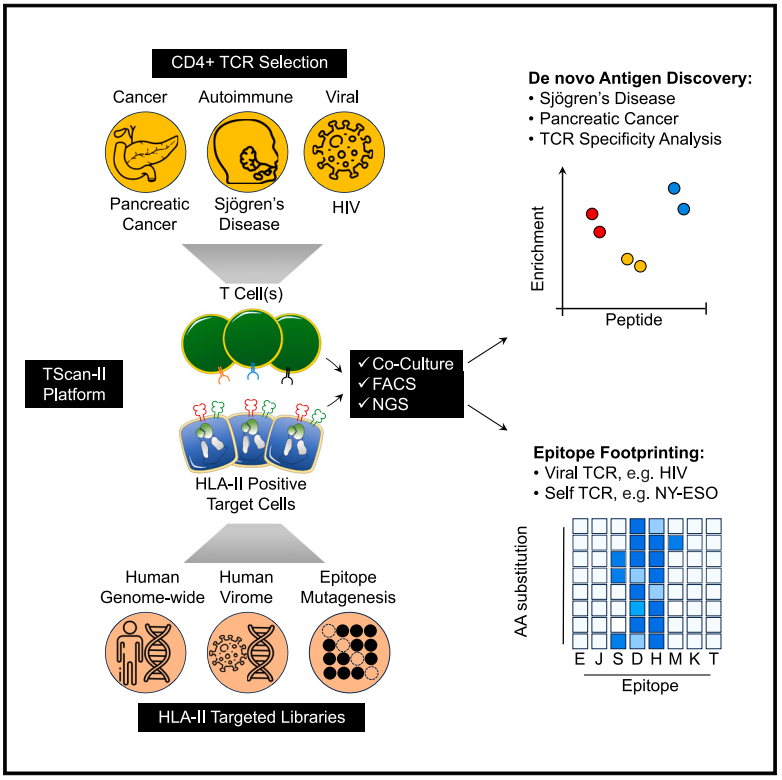
Cell

TScan-II: A genome-scale platform for the *de novo* identification of CD4$^+$ T cell epitopes

Graphical abstract



Highlights

- TScan-II finds the cognate and cross-reactive antigens of self and viral CD4$^+$ TCRs
- Saturated mutagenesis screens facilitate the mapping of the CD4$^+$ TCR epitope interface
- CD4$^+$ cells in Sjögren's disease recognize self-antigens on HLA-II-positive ductal cells
- TScan-II finds antigens within the non-canonical immunopeptidome for cancer TCRs

Authors

Mohammad H. Dezfulian, Tomasz Kula, Thomas Pranzatelli, ..., Senthil K. Muthuswamy, Blake M. Warner, Stephen J. Elledge

Correspondence

selledge@genetics.med.harvard.edu

In brief

TScan-II is a genome-scale platform that integrates endogenous antigen processing and T cell signaling, facilitating multiplex screens for the identification of physiologically relevant antigens recognized by CD4$^+$ T cells. The unbiased nature of the platform allows the identification of CD4$^+$ T cell antigens in Sjögren's disease and pancreatic cancer tumors.

Dezfulian et al., 2023, Cell *186*, 5569–5586
December 7, 2023
https://doi.org/10.1016/j.cell.2023.10.024

CellPress

Cell

CellPress
OPEN ACCESS

Resource

TScan-II: A genome-scale platform for the *de novo* identification of CD4$^+$ T cell epitopes

Mohammad H. Dezfulian,[1,2] Tomasz Kula,[1,2] Thomas Pranzatelli,[3] Nolan Kamitaki,[2,4,5] Qingda Meng,[6,13] Bhuwan Khatri,[7] Paola Perez,[8] Qikai Xu,[1,2] Aiquan Chang,[1,2] Ayano C. Kohlgruber,[1,2] Yumei Leng,[1,2] Ananth Aditya Jupudi,[9,10] Michelle L. Joachims,[7,9] John A. Chiorini,[3] Christopher J. Lessard,[7,11] A. Darise Farris,[9,10,11] Senthil K. Muthuswamy,[12] Blake M. Warner,[8] and Stephen J. Elledge[1,2,14,*]

[1]Division of Genetics, Department of Medicine, Howard Hughes Medical Institute, Brigham and Women's Hospital, Boston, MA, USA
[2]Department of Genetics, Harvard Medical School, Boston, MA, USA
[3]Adeno-Associated Virus Biology Section, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA
[4]Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA, USA
[5]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[6]Department of Medicine, Harvard Medical School, Boston, MA, USA
[7]Genes and Human Disease Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA
[8]Salivary Disorders Unit, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA
[9]Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA
[10]Departmentment of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
[11]Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
[12]Laboratory of Cancer Biology and Genetics, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
[13]Present address: College of Veterinary Medicine, Henan Agricultural University, Zhengzhou, China
[14]Lead contact
*Correspondence: selledge@genetics.med.harvard.edu
https://doi.org/10.1016/j.cell.2023.10.024

SUMMARY

CD4$^+$ T cells play fundamental roles in orchestrating immune responses and tissue homeostasis. However, our inability to associate peptide human leukocyte antigen class-II (HLA-II) complexes with their cognate T cell receptors (TCRs) in an unbiased manner has hampered our understanding of CD4$^+$ T cell function and role in pathologies. Here, we introduce TScan-II, a highly sensitive genome-scale CD4$^+$ antigen discovery platform. This platform seamlessly integrates the endogenous HLA-II antigen-processing machinery in synthetic antigen-presenting cells and TCR signaling in T cells, enabling the simultaneous screening of multiple HLAs and TCRs. Leveraging genome-scale human, virome, and epitope mutagenesis libraries, TScan-II facilitates *de novo* antigen discovery and deep exploration of TCR specificity. We demonstrate TScan-II's potential for basic and translational research by identifying a non-canonical antigen for a cancer-reactive CD4$^+$ T cell clone. Additionally, we identified two antigens for clonally expanded CD4$^+$ T cells in Sjögren's disease, which bind distinct HLAs and are expressed in HLA-II-positive ductal cells within affected salivary glands.

INTRODUCTION

The adaptive immune system generates a long-lived memory of the pathogenic antigens it encounters through a set of highly variable cell surface receptors found on B and T lymphocytes.[1] However, the fundamental question of which antigens are recognized remain incompletely understood. T cells play essential roles in tissue homeostasis and protect against pathogens and cancer. T cells recognize peptide antigens bound to human leukocyte antigens (HLAs) via their T cell receptors (TCRs) and are categorized into two subtypes, CD8$^+$ and CD4$^+$. CD8$^+$ T cells can be cytotoxic and directly destroy target cells following the recognition of human leukocyte antigen class-I (HLA-I)-bound antigens via the release of granzyme and perforin. The peptides that CD4$^+$ cells recognize are generated by the proteolysis of proteins in endosomes and lysosomes. This recognition stimulates cytokine secretion, activation, and differentiation into diverse T cell subsets with tailored functions.

Recent advances in single-cell sequencing technologies have illuminated T cells' cellular and molecular diversities in healthy tissues, sites of infection, tumors, and inflamed autoimmune organs. However, what T cells recognize in these contexts has remained elusive due to the lack of genome-scale high-throughput antigen discovery tools. Cell-based screening methods for CD8$^+$ T cells, such as TScan, have achieved proteome-wide success in identifying cognate antigens for TCRs from CD8$^+$ cells, but

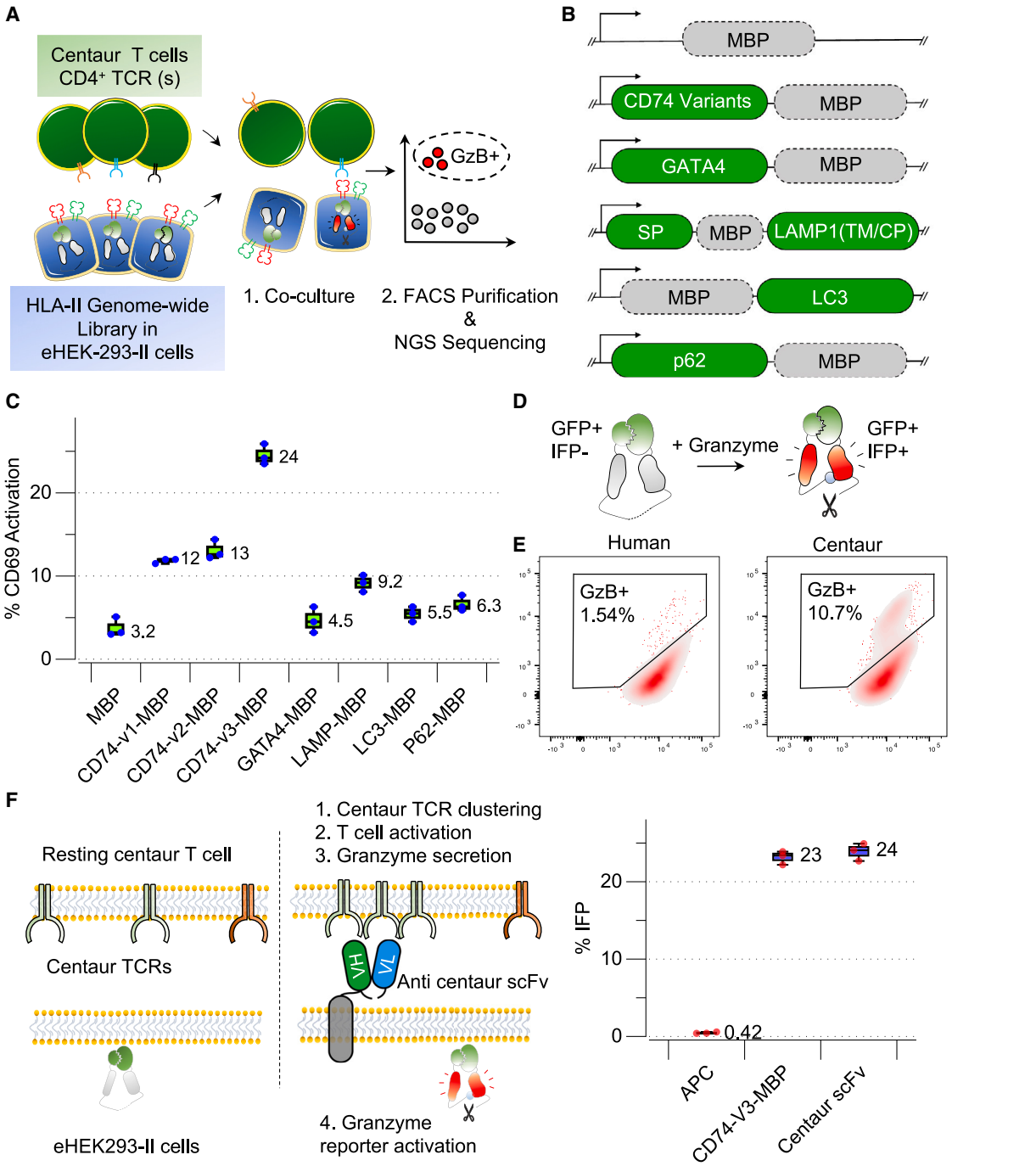


Figure 1. Development of the TScan-II platform

(A) Conceptual layout of the CD4+ antigen discovery platform.

(B) Schematic representation of fusion proteins developed to assess the presentation of full-length MBP fusions to CD74, GATA4, LAMP1 (SP, signal peptide; TM, transmembrane; CP, cytoplasmic), LC3, and p62 on HLA-DRB1*1501.

(C) Boxplots depicting the activation levels (CD69) of Ob1.A12 TCR-transduced SKW-3 cells following 6 h co-culture with eHEK-293-II cells with endogenously expressed full-length MBP fusions. The box is drawn around the inner quartile range, and the whiskers show minimum to maximum values for all plots shown here and below.

(D) The GzB reporter is visually represented in a schematic, illustrating how GzB mediates the cleavage of the substrate sequence (VGPD↓SGR) within $IFP^{GZB\text{-}Hi}$. Cleavage leads to IFP fluorescence (red).

(legend continued on next page)

methods for CD4$^+$ T cell antigen identification have not reached this level of sensitivity or throughput.[2] Initial attempts for CD4$^+$ antigen discovery have mainly relied upon single-chain peptide HLAs and, in instances where platforms can incorporate endogenous machinery, have been limited to a few thousand antigens, thus not being genome-scale or unbiased.[3–8]

Developing an antigen discovery platform for CD4$^+$ T cells is complicated by the lower affinity of CD4$^+$ TCRs for their cognate antigens, confinement of human leukocyte antigen class-II (HLA-II) expression to specialized cell types, and the complexity of HLA-II antigen-processing machinery. To circumvent this, we sought to exploit the successes of the TScan platform in identifying CD8$^+$ T cell antigens for CD4$^+$ T cell antigens. TScan utilizes the granzyme-mediated cytolytic ability of CD8$^+$ cells to activate a fluorescent reporter of granzyme B (GzB) protease activity. The target cells express a library of candidate antigens that are processed and presented endogenously on HLA-I molecules. Co-culture of the target cells with CD8$^+$ T cells of interest activates the GzB reporter in cells displaying cognate antigens, allowing the enrichment of rare cognate target cells with high sensitivity and specificity.[2] Unfortunately, the TScan system lacks the HLA-II antigen-processing and -presentation machinery required for CD4$^+$ T cell screening, and CD4$^+$ cells do not typically produce GzB. Here, we report the development of TScan-II, a highly engineered complementary genome-scale platform for systematically discovering antigens productively recognized by TCRs from CD4$^+$ T cells. We applied this platform for the *de novo* discovery of CD4$^+$ T cell antigens in pancreatic cancer and Sjögren's disease (SjD). The unbiased discovery of unsuspected specific antigens recognized by CD4$^+$ T cells helps unravel the mechanisms underlying cancer and autoimmunity.

RESULTS

Development of reporter cell lines capable of HLA-II antigen processing and delivery

To build a high-throughput CD4$^+$ antigen discovery platform capable of processing and presenting antigens at a genome scale that incorporates endogenous T cell signaling, we exploited a reporter of GzB activity. To do this, we had to overcome several technical challenges in both T cells and target cells. First, we had to overcome the need for HLA-II presentation machinery: unlike HLA-I,[9] the expression of HLA-II machinery is primarily limited to professional antigen-presenting cells (APCs),[10] which have low transduction rates and significantly lower brightness of the infrared GzB reporter, hampering utility for high-throughput screening. Second, we had to overcome the need to target endogenously expressed proteins of all types to HLA-II, which typically present peptides derived from endocytosed extracellular proteins undergoing proteolysis in lysosomal compartments.[10] This required the development of a system for efficient HLA-II targeting and processing of endogenously expressed antigens. Third, we had to overcome the need for granzyme secretion for reporter activation: unlike CD8$^+$ cells, CD4$^+$ cells, with some rare exceptions, do not typically possess GzB-mediated cytolytic activity; hence, we had to develop a platform for efficient CD4$^+$ TCR-dependent secretion of GzB into target cells (Figure 1A).

To develop a cell line capable of processing and presenting antigens on HLA-II cells like APCs, we engineered HEK293T cells. We mutated B2M via CRISPR-Cas9 to avoid potential cross-reactivity with HLA-I presented antigens (Figure S1A).[9] We engineered HEK293T cells to process and present antigens on HLA-II via the viral expression of the HLA-II transactivator (CIITA). CIITA plays an essential role in the transcription of classic HLA-II molecules such as HLA-DR, HLA-DP, and HLA-DQ and genes such as CD74, HLA-DM, and HLA-DO that are essential not only for regulating antigen processing but also for the effective loading of epitopes on HLA-II.[11] HEK293T cells expressing high HLA-II levels (CIITA-HEK239T) were sorted and used as APCs (Figure S1A). HLA-II alleles were assembled as a single construct, where the HLA-β and HLAα chains were separated by 2A self-cleaving peptides (β-P2A-α) and transduced into CIITA-HEK239T cells.

To optimize HLA-II antigen processing and presentation, we compared several strategies for lysosomal targeting, including fusions to LAMP1, CD74, LC3, p62, and GATA4 (Figure 1B). We used the low-affinity TCR (Ob.1A12), which recognizes a Myelin Basic Protein (MBP) epitope, for platform optimization.[12] Among these, a variant of CD74 was most efficient for MBP antigen trafficking and loading as measured by CD69 upregulation on cognate Ob.1A12 TCR+ T cells (Figure 1C). This segment of CD74 allows the proper folding and trafficking of HLA-II to the lysosome via dileucine-based endosomal targeting signals (LI, IL/ML).[13] We further enhanced antigen presentation by removing its arginine-based endoplasmic reticulum (ER) retention motif (CD74-v3) (Figure 1C). We also evaluated the fusions using a high-affinity human immunodeficiency virus (HIV) TCR (F24) antigen pair[14] and observed that CD74-v3 was the most efficient fusion for antigen presentation (Figure S1B). Furthermore, we showed that CD74-v3 fusion of long peptides antigens (90 aa or 56 aa) from different sources (CMV, BRAF^{V600E}, and NY-ESO) on various HLA-II alleles could activate SKW-3 T cells harboring the cognate TCRs (Figure S1C).

In the original TScan-I GzB reporter, the GzB cleavage amino acid sequence (VGPD$^{\downarrow}$FGR) was inserted into the scaffold of infrared fluorescent protein (IFP), which disrupts IFP (IFP^{GZB}) fluorescence,[2,15] which is restored upon GzB-mediated cleavage (Figure 1D). However, phenylalanine (F) exposure following GzB cleavage generates a potent N-degron that renders IFP^{GzB} unstable, limiting its fluorescence.[16] Thus, phenylalanine was mutated to serine (VGPD$^{\downarrow}$SGR), generating a stable GzB

(E) fluorescence-activated cell sorting (FACS) histograms depicting GzB reporter activation in eHEK-293-II following co-culture with centaur T cells. Target eHEK-293-II cells co-express the HLA-DRB1*1501 allele and full-length MBP fused to CD74-v3. Primary T cells express the unmodified (left panel) or centaur Ob1.A12 TCR and CD4 co-receptor (right panel).

(F) Frequency of eHEK-293-II cells that activate the GzB reporter in the presence of membrane-bound anti-centaur scFv or an endogenously processed full-length MBP fused to CD74-v3 following co-culture with Ob1.A12 centaur T cells. See Figure S1.

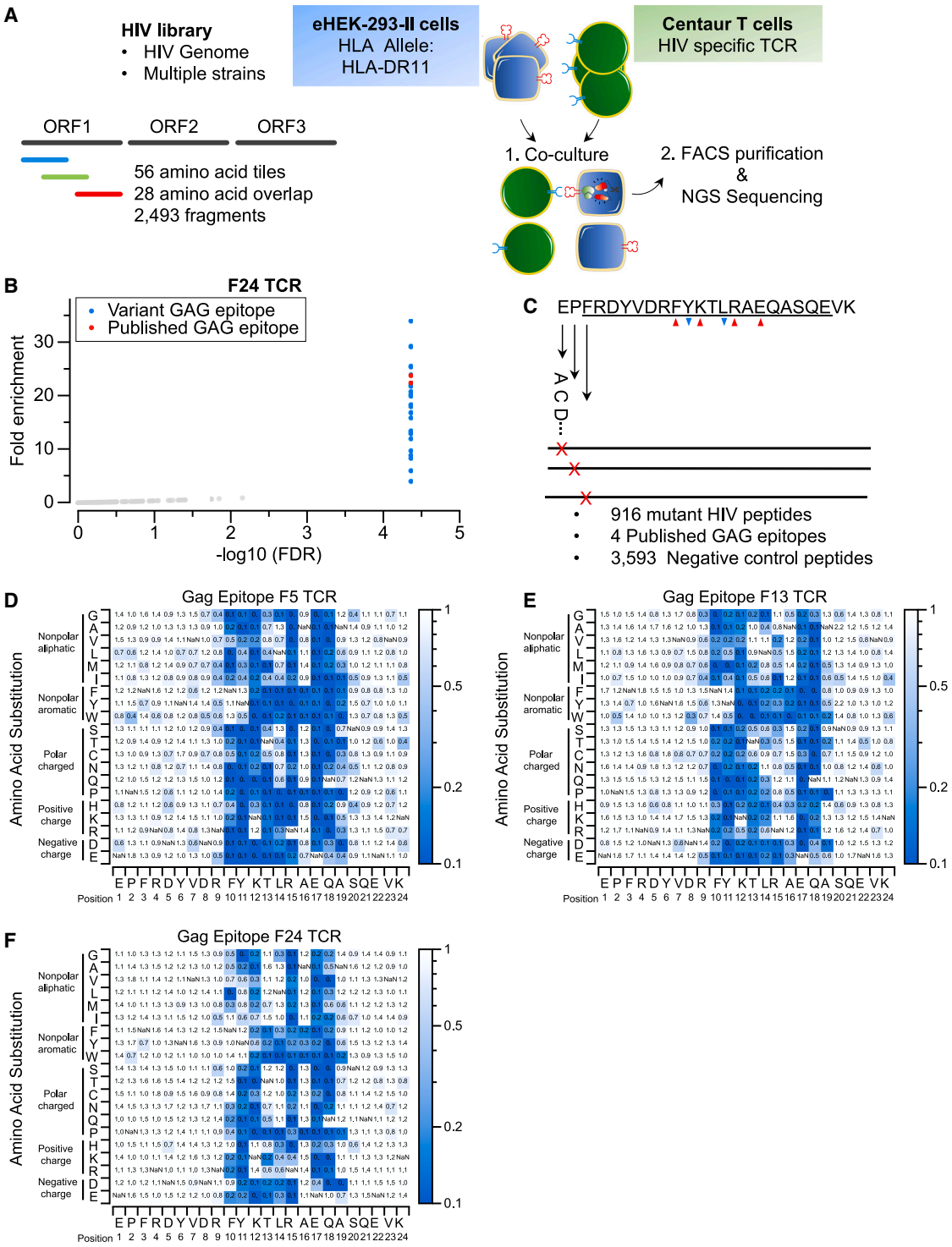


Figure 2. HIV genome-wide and mutagenesis TScan-II screens

(A) Schematic of TScan-II HIV genome-wide screen. The HIV library comprises 2,492 peptide fragments tiling across the genomes of multiple HIV strains in 56-aa steps with 28-aa overlap.

(B) TScan-II screen using an HIV elite controller TCR against the HIV library. Each dot represents one peptide. The y axis shows the geometric mean of the fold change across seven replicates for screens with HLA-DRB1*1101 and the Gag293-specific TCR (F24). Red dots represent peptides, with an epitope identical to the published Gag293 epitope for this TCR, and blue dots represent variant Gag293 epitopes. p values were calculated using permutation testing, and false discovery rate correction was done by the Benjamini-Hochberg method.

(legend continued on next page)

reporter with increased fluorescence, IFP$^{GzB\text{-}Hi}$ (Figure S1E). We expressed a caspase-resistant version of the inhibitor of caspase-activated DNase (ICAD^{CR}) to prevent genomic DNA fragmentation following granzyme-induced cell death in targeted HEK293T cells. We termed the B2M null target cells expressing CIITA, HLA allele, HLA-II-localized antigens, IFP$^{GzB\text{-}Hi}$, and ICAD^{CR} as "engineered HEK-293-II" (eHEK-293-II) cells (Figure S1K).

Development of T cells capable of GzB secretion following the recognition of cognate pHLA-II

We assessed the efficiency of GzB secretion into eHEK-293-II by introducing an intact CD4$^+$ TCR into CD8$^+$ cells. Although the expression of CD4$^+$ TCRs in CD8$^+$ cells permitted target cell activation, the signal-to-noise ratio was insufficient for a genome-scale screening. This problem could be due to (1) the absence of CD4 in CD8$^+$ T cells, (2) mispairing of exogenously introduced CD4$^+$ TCRs with endogenous TCRs, or (3) low affinities of CD4$^+$ TCR for cognate epitopes on HLA-II relative to CD8$^+$ TCRs for HLA-I. To circumvent these problems, we generated a hybrid TCR[17–19] that reduces endogenous mispairing. We also introduced CD4 with the engineered TCRs into CD8$^+$ cells for the efficient recruitment of Lck kinase to the TCR and peptide HLA-II (pHLA-II) synapse[20](Figure S1D). Co-culture of engineered CD8$^+$ cells expressing the enhanced TCR and CD4 molecule with eHEK-293-II cells expressing the cognate epitope resulted in a ~10-fold increase of GzB reporter activation (Figure 1E). We termed these engineered cells as "centaur" T cells (Figure S1K).

Mutating the endogenous TCR beta chains showed no significant improvement (Figures S1F and S1G), indicating that CD3 competition is not limiting. We also examined the cell surface levels of several T cell adhesion and co-stimulatory factors and found high levels of CD58 and CD155 on our eHEK-29T-II target cells but no CD80 expression. However, expressing CD80 yielded no improvement in the signal-to-noise ratio (Figure S1H). To assess whether the endogenous HEK-293T HLA-II alleles expressed following CIITA transduction impacted antigen detection, we mutated the endogenous HLA-II alleles (Figure S1I) but observed no discernable difference in the level of GzB reporter activity (Figure S1J).

To benchmark the relative efficiency of GzB delivery from centaur cells into target cells, we developed a membrane-bound centaur-specific scFv-expressing target cell that should maximally activate the target cells for comparison (Figure 1F). Strikingly, the percentage of target cells activating the GzB reporter was comparable in cells expressing the MBP antigen to that of target cells expressing the membrane-bound anti-centaur scFv, suggesting that our antigen delivery and presentation strategy results in efficient centaur TCR activation (Figure 1F). The genetic engineering performed to generate eHEK293-II target cells and centaur T cells is summarized schematically in Figure S1K.

Characterization and optimization of TScan-II using viral reactive TCRs

As an initial test assessing this antigen discovery platform, we focused on CD4$^+$ responses in a rare group of HIV-infected individuals known as "elite controllers" who maintain an undetectable HIV viral load in the absence of antiretroviral therapy.[21] To do this, we synthesized an HIV proteome-wide library, tiling across the entire proteomes of multiple different HIV strains, with each fragment barcoded twice via synonymous codons. We initially used a public CD4$^+$ TCR (F24) that recognizes the most immunoprevalent CD4$^+$ epitope (Gag293) in the HIV capsid protein (Figure 2A).[14] Our HIV library included two antigens that precisely matched the published epitope of this TCR. However, considering the library's composition of multiple strains, additional variants of this epitope are also present within this library for a total of 29 related Gag293 fragments. These variants exhibit 1–7 amino acid variations relative to the reported epitope. This TScan-II screen consistently observed reproducible enrichment of the two published antigens across all seven replicas (Figures 2B and S2A; Table S1). Of interest, all 29 fragments that were significantly enriched relative to the input library of 2,493 fragments were variants of the published Gag293 peptide, and the extent of their enrichment directly correlated to the number of amino acid differences relative to the published epitope (Figure S2C). Furthermore, there was a high level of concordance between both barcode fragments, affirming the reliability of the Gag293 epitope hits (Figure S2B). We confirmed the reactivity of F24 to these variants through co-culture experiments (Figure S2D). Thus, the TScan-II platform not only achieved 100% efficiency in identifying known TCR hits but also readily identified novel epitope variants from a complex library of antigens.

Mapping of CD4$^+$ TCR epitope interface using saturated mutagenesis screens

To assess whether TScan-II can enable finely detailed mapping of CD4$^+$ TCR epitopes, we generated a saturation mutagenesis library of the Gag293 epitope where each residue in the epitope is mutated to all other 19 amino acids (Figure 2C). Screens were performed using the three TCRs (F24, F3, and F5) reported in HIV elite controllers. As anticipated, only mutant Gag293 fragments enriched comparably to the wild-type (WT) Gag293 peptides and more strongly than all negative control peptides (Figures S2E–S2G; Table S1). The mutants of Gag293 fragments can be broadly categorized into two groups based on their behavior in the screen. The first group consists of enriched mutant fragments, where the substitution of amino acids had either no or minimal effect on TCR or HLA binding. The second fragment group dropped out of the screen, indicating that these mutations affected residues critical for TCR recognition or HLA binding. To identify critical residues, we compared the enrichment of each mutant peptide with the WT peptides (Figures 2D–2F). Notably, the crystal structure of the Gag293 epitope in complex with HLA-DR11 indicated that tyrosine and

(C) The Gag293 epitope saturation mutagenesis library. The library encodes 4 WT epitopes, 956 mutant Gag epitopes, and 3,593 negative controls. The downward blue arrows indicate anchor residues, whereas the upward red arrows indicate TCR-facing residues.
(D–F) Heatmap representation of comprehensive mutagenesis analysis of Gag293 epitope using 3 HIV elite controller TCRs, F5 (D), F13 (E), and F24 (F). The heatmap value represents this mutant's relative enrichment compared with the published (Gag) epitope. See Figure S2.

leucine (FRDYVDRFYKTLRAEQASQE) are anchor residues for HLA-DR11,[22] consistent with our studies in which tyrosine could only be replaced by aromatic residues such as phenylalanine or tryptophan, whereas leucine could be replaced primarily by residues with some hydrophobicity. This screen also revealed several non-anchor residues within the Gag293 epitope (FRDYVDRFYKTLRAEQASQE) as crucial for TCR activation. Based on structural data, these residues are involved in the binding of F24 and F13 TCRs. The correlation between the mutagenesis and the structural analysis further highlights the effectiveness of TScan-II in mapping the TCR and peptide-HLA interface. Interestingly, although each TCR had a unique epitope footprint, they each recognized all 29 naturally occurring epitope variants (Figure S2A). This flexibility in epitope recognition could contribute to the elite controller status of these individuals.

Virome-wide discovery of CD4$^+$ TCR antigens

To examine whether TScan-II can successfully identify cognate antigens from more complex libraries, we made a library tiled across the human virome (206 viral species and over 1,000 different strains) and screened using the F24 CD4$^+$ TCR (Figure 3A), which exhibited by far the most relaxed specificity (Figure 2F). Despite the 50-fold increase in library complexity, we observed enrichment of the antigens that precisely matched the published and variant epitopes described earlier (Figure 3B; Table S1). To assess whether the enriched fragments from the E2 protein of human papillomavirus type 49 (HPV) were due to cross-reactivity or noise, we generated a second library of the top enriched fragments from the primary screen with additional overlapping tiles. All enriched fragments from the second screen were either from the HIV Group-specific antigen (GAG) protein or the HPV protein (Figure S3B; Table S1). The minimal epitope recognized within E2/HPV antigens (Figure 3D) contained residues that were either identical to the core HIV Gag293 epitope mapped in Figure 2F or were permitted substitutions. We validated the recognition of the HPV epitope, further confirming that the original signals observed in the primary screen were valid hits and not noise (Figures 3D and S3A). These findings demonstrate the capability of TScan-II to successfully identify specific CD4$^+$ T cell epitopes within a complex virome-wide library.

Genome-scale discovery of self-reactive CD4$^+$ TCR targets

To assess TCR targets at genome scale, we examined a therapeutic TCR candidate for specificity. Cancer testis antigen 1B (CTAG1B/NY-ESO-1), a tumor-associated antigen, is expressed in various solid tumors but not in normal healthy tissues except for the thymus and testis. Given these unique characteristics, CTAG1B is considered an ideal target for adoptive cell therapy (ACT).[23,24] Several CTAG1B TCRs are actively being assessed in clinical studies. Although many therapeutic TCRs are thought to be specific initially, the complications observed in ACT patients associated with toxic off-targets reinforce the urgency of unbiased reactivity assessment of these TCRs.[25–27] Thus, we sought to assess the specificity of the HLA-DRB1*0401 (HLA-DRB4) restricted CTAG1B TCR (TCR-3598_2).[28]

We generated an HLA-II targeted library that tiled across the entire human proteome in 90-aa fragments with 45-aa overlap (Human Peptidome Version 2 [V2]). To assess whether the platform can be multiplexed, we also introduced HLA-DRB4, HLA-DRB1*03:01 (HLA-DR3), and HLA-DQB1*02:01_HLA-DQA1*05 (HLA-DQ2) into our target cells (Figure 4A). We performed a TScan-II screen and observed robust and reproducible enrichment of only six peptides in the library from 3 sets of open reading frames (ORFs) (Figure 4B; Table S2), two of which were the only fragments containing the cognate CTAG1B epitope. Additionally, three overlapping fragments belonging to Cancer testis antigen 2 (CTAG2/NY-ESO-2) ORF were enriched. The CTAG2 and CTAG1B proteins are paralogous proteins that share a 14-aa region of substantial identity and similarity spanning the original CTAG1B epitope (Figure 4C). Another strongly enriching antigen across all replicas was a peptide from Titin (TTN), a protein highly abundant in muscle tissues. This epitope was similar to the CATG1B epitope in a consecutive stretch of 5 amino acids and contained several conservative changes outside of that region (Figure 4C). In validation experiments, we observed reactivity as measured by granzyme secretion to each antigen but not to control antigens (Figure 4D). Interestingly, there was a direct correlation between the degree of TCR recognition and how tightly the peptide is predicted to bind HLA-DR4[29] (Figure S4A). These observations indicate that the platform can readily identify known and novel CD4$^+$ TCR self-antigens at genome scale, even when target cells express multiple HLA-II alleles simultaneously. Thus, the TScan-II platform is amenable to HLA multiplexing, which is especially important in instances where the TCR restriction across the patient's HLA alleles is unknown.

We mapped the TCR-CTAG1B interface using saturation mutagenesis (Figures S4B and S4C; Table S2), generating a heatmap of critical residues within the epitope's 10-aa core (L7-N16) (Figure 4E). By comparing critical residues with the predicted TTN epitope, we found that TTN contains an identical central core of residues (K9-V13) and differs from CTAG1B only by five amino acids in its outer core, each of which (L7V, L8R, S14T, G15S, and N16L) is permissive to different degrees. Likewise, CTAG2 differs by only two residues with tolerated changes (L7V and E10D) explaining its cross-reactivity. An interesting takeaway is that mutagenesis to alanine in 3 key positions, V13, S14, and G15, had minimal effects, suggesting that alanine scanning maps may be unreliable for predicting cross-reactivity. Unlike CTAG1B, CTAG2 and TTN have a broad expression profile across various healthy tissues (Figure 4F), which could be problematic in a clinical setting using this TCR for ACT.[27]

Characterization of pancreatic-cancer-reactive CD4$^+$ TCRs

In a cancer setting, identifying the neo-antigens recognized by T cells is approachable, as they can be identified by DNA sequencing and tested, but self-antigens pose a much more difficult problem, limiting their widespread applicability for TCR-based immunotherapy. To explore TScan-II in this capacity, we extended our previous study of co-culturing HLA-II-positive pancreatic cancer tumor organoids with autologous patient primary blood mononuclear cells (PBMCs)[30] (Figure 5A). This approach resulted in the expansion of organoid-primed T cell

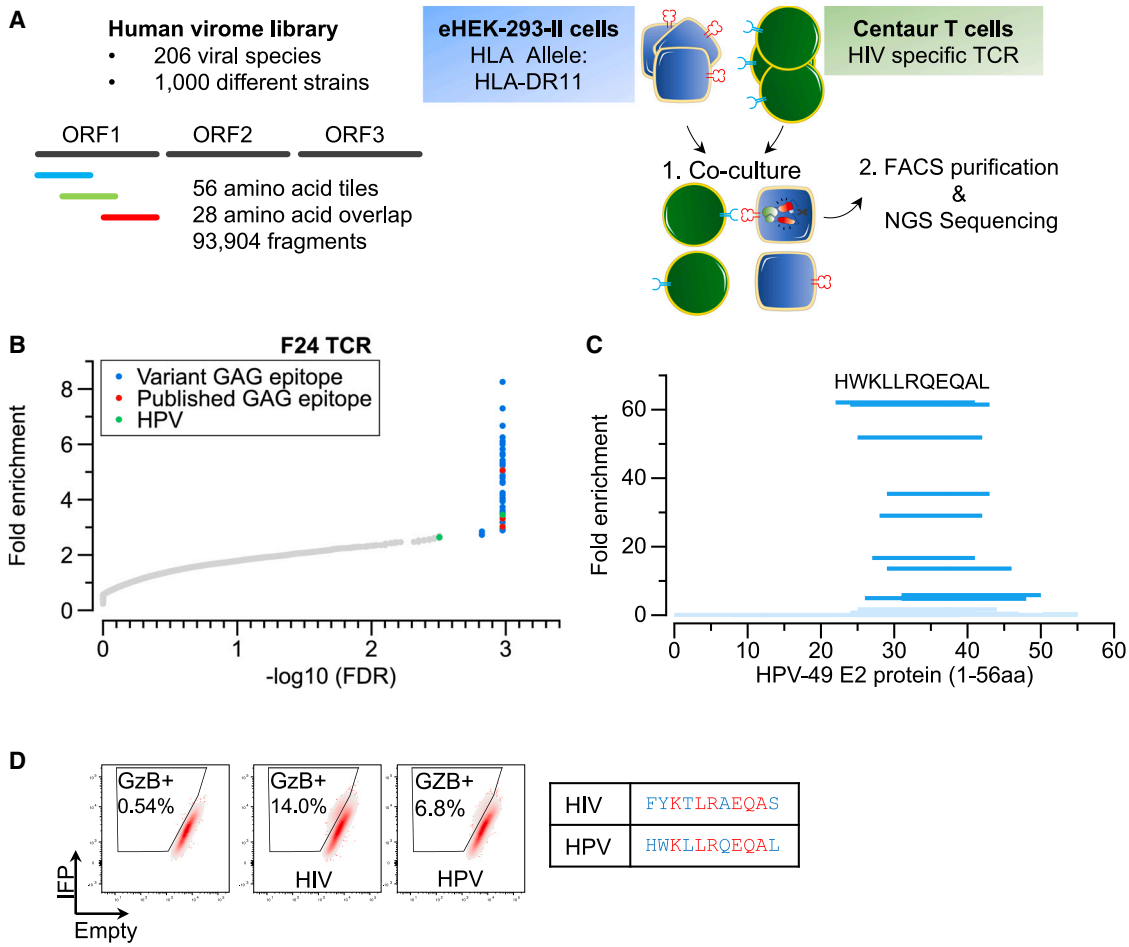


Figure 3. Virome-wide TScan-II screen

(A) Schematic of the TScan-II human virome screen. The human virome library comprises 93,904 56-aa fragments collectively tiled across the human virome (206 viral species and >1,000 strains), including multiple HIV strains.

(B) TScan-II screen of the HIV elite controller F24 TCR with the virome library. Each dot represents one peptide, with the y axis plotting the geometric mean of the fold change across eight replicates against the HLA-DRB1*1101 allele. Red dots represent peptides with WT Gag293 (Gag) epitope for this TCR, blue dots represent variant Gag293 epitopes, and green dot represents a peptide from the HPV type 49. p values were calculated by permutation testing, and false discovery rate correction was done by the Benjamini-Hochberg method.

(C) TScan-II screen of a library of 12-, 15-, 18-, and 20-aa fragments tiled at 1-aa intervals across the 56mer antigenic segments of HPV. The length of each dark blue indicates the scoring segment of HPV. Fragments in light blue did not score. The sequence depicted represents the minimal epitope.

(D) FACS depicting reporter activation after a 6-h co-culture with F24 T cells. eHEK-293-II cells express the HLA-DRB1*1101 allele and a control peptide (left), HIV (middle), or HPV (right). Identical residues shared between the minimal HPV epitope and core Gag293 epitope are colored in red, whereas blue residues represent permitted amino acid substitutions based on Figure 2F. See Figure S3.

(opT) clones, of which approximately 90% of the TCRs originated from a single Th1 CD4$^+$ T cell clone (CXCR3$^+$, CCR4$^-$, CD69$^+$, and CD103$^+$).[30] Intriguingly, this CD4$^+$ T cell clone possessed a unique TCRβ chain and two distinct TCRα(α,α′) chains. Thus, we synthesized two TCRs with a unique TCRβ chain and either one of the TCRα chains (DHM6 TCR and DHM7 TCR). Notably, DHM6 TCR recognized only the matched patient-derived tumor organoid, but not a tumor organoid derived from a different patient (unmatched) (Figure 5SA).

To identify the target of DHM6, we generated a patient-specific neo-antigen library by performing whole-exome sequencing of matched tumor and normal cell DNA from the pancreatic cancer patient and identified 54 nonsynonymous somatic mutations within tumor cells (Figure S5B; Table S3). We screened the TCR against the neo-antigen library and did not observe a reproducible enrichment of any of the peptide fragments, suggesting that DHM6 likely does not recognize a neo-antigen (Figures S5C and S5D).

To examine whether DHM6 TCR might recognize a self-antigen, we designed an expanded human peptidome library (Human Peptidome Version 3 [V3]), which encompassed not only the annotated human proteome as used in the CTAG1B screen but also fragments derived from splicing defects such as intron retention and exon junction, alternative translation products, endogenous retroviruses, antisense strands, and hypothetical proteins covering 586,167 fragments (Figure 5C).[31–37] Using

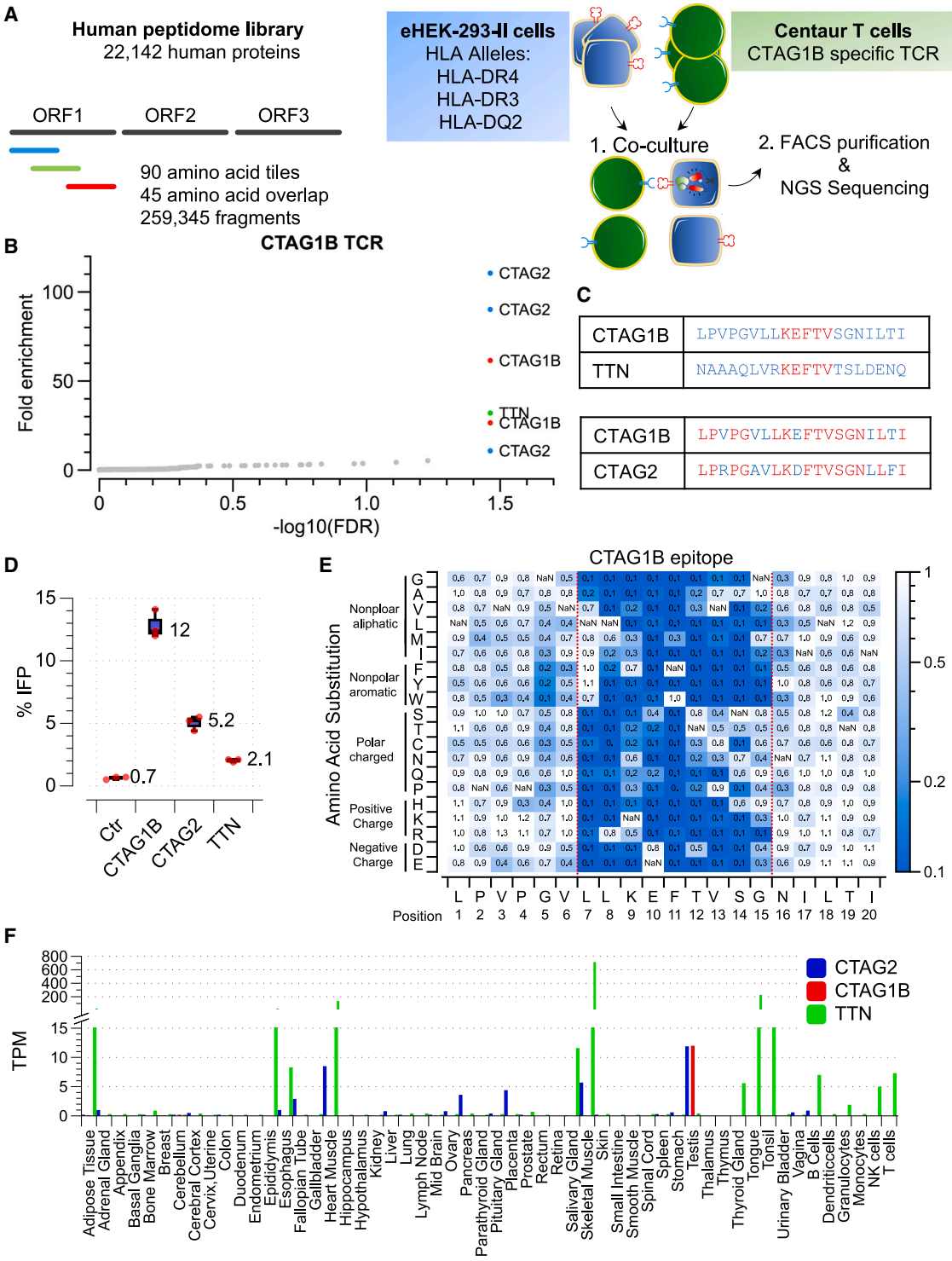


Figure 4. Human genome-wide screening using NY-ESO-specific CD4⁺ TCR

(A) Schematic of TScan-II human genome-wide screen. The Human Genome-Wide V2 library comprises 259,345 antigens that tile across the entire human proteome in 90-aa fragments with 45-aa overlap.

(B) TScan-II screen of CTAG1B CD4⁺ TCR (3598-2) against the Human Genome-Wide V2 library. The eHEK-293-II cells express the HLA-DRB1*04, HLA-DRB1*03:01, and HLA-DQB1*02_HLA-DQA1*05 alleles. p values were calculated by permutation testing, and false discovery rate correction was done by the Benjamini-Hochberg method.

(C) Predicted 20-aa epitopes from the enriched 90-aa peptides.

(legend continued on next page)

this library, we performed a TScan-II screen using DHM6 T cells (Figure 5C) and observed a robust enrichment of a peptide fragment from a hypothetical protein (LOC100131311) residing in the antisense strand of induced myeloid leukemia cell differentiation protein (MCL1) (Figure 5D; Table S3).[38] Utilizing strand-specific gene expression analysis, we detected transcripts spanning the length of the antigenic fragment in 33% of the assessed organoids, including the matching patient tumor (Figure S5F). Using a set of overlapping tiles, we mapped the epitope to a 15-aa segment (Figure S5E). We deconvoluted the HLA-DP restriction of this antigen using target cells expressing each HLA-DP allele heterodimer (Figure S5G). Notably, these alleles can be found in up to 25% of individuals of African descent, particularly among Black South Africans.[39]

Tissue-infiltrated CD4$^+$ T cells in SjD recognize self-antigens

In almost all autoimmune diseases, there is a pressing need to understand the antigenic landscape of tissue-infiltrating T cells,[40] a key prerequisite for understanding the underlying etiology and developing immune-modulatory therapies. One such autoimmune disease is SjD, a chronic systemic archetypal and heterogeneous autoimmune disease with special targeting of the exocrine glands.[41] Its etiology is unknown, but it appears to engage the adaptive immune system, as evidenced by the focal lymphocytic infiltration of the salivary gland (SG) and lacrimal gland. The lymphocytic foci in the SG of SjD patients are predominantly CD4$^+$ T cells[42–44] with an antigen-experienced phenotype.[45–48] Additionally, ductal cells of the SG in SjD patients express high levels of HLA-II and are surrounded by CD4$^+$ T cells and B cells.[46,48–52] In response to interferon γ (IFNγ) exposure, the SG ductal cells express T cell co-stimulatory molecules such as CD80, CD86, CD40, and Inducible T cell co-stimulator ligand (ICOSL).[46,53–56] Moreover, historically, an association between the expanded C*07:01–B*08:01–DRB1*03:01–DQB1*02:01 haplotype and the development of SjD was suggested in patients of European ancestry.[57–61] Collectively, these observations raise the question of whether any of the clonally expanded infiltrating T cells are indeed self-reactive.

We employed TScan-II to assess whether SG-expanded CD4$^+$ T cells can recognize a self-antigen present in our human genome-wide library (Figure 6A) by profiling a large cohort of published SG-infiltrating CD4$^+$ lymphocytes from 15 SjD patients who shared the HLA-DR3 and HLA-DQB1*02:01 alleles (Table S4).[47] Twenty TCR pairs across five SjD patients were ultimately chosen for subsequent genome-scale TScan-II screening according to the strategy detailed in the STAR Methods (Table S4). Interestingly, although in normal T cells, up to one-third express two distinct TCRα chains, approximately 50% of the SjD T cells did.[60,62–64] In most cases, following the transduction of TCRs that were sourced from T cells that expressed two TCRα chains (α or α'), we could only detect a high level of expression for one of the two TCR pairs (TCRβ-α or TCRβ-α') on the cell surface of primary T cells (Figure S6A).

We performed the first genome-wide screen using a mixture of 5 centaur T cell lines pooled, each expressing a unique TCR (4A, 7A, 7G, 9B, and 10C). We observed the enrichment of two adjacent peptide fragments from Mitogen-Activated Protein Kinase Kinase Kinase 4 (MAP3K4/MTK1) (Figure 6B; Tables S5 and S6). Deconvolution identified TCR4A and HLA-DR3 restriction (Figures 6F and S6B). This TCR was obtainted from a subject 4 that was postive for anti–Sjögren's-syndrome-related antigen A (SSA+) and negative for Sjögren's-syndrome-related antigen B (SSB−). We performed a second genome-wide TScan-II screen using only 4A T cells (Figure 6C; Tables S5 and S6) and identified the same two MAP3K4 fragments. Henceforth, the remaining genome-wide screens were performed using pools of 5 TCRs. We identified a second antigen reactivity from the additional three genome-wide screens, this time to DNA-damage-induced apoptosis suppressor (DDIAS), using a collection of 5 TCRs (3A, 10A1, 10D, 10E, and 10F) (Figure 6D; Tables S5 and S6). Deconvolution revealed TCR3A (subject 3, SSA−/SSB−) and HLA-DQ2 restriction (Figures 6F and S6C). To further assess the multiplexing capability, we performed a screen with 3A and 4A within a pool of six total T cell clones and detected both MAP3K4 and DDIAS, thereby demonstrating the system's ability to identify specific TCR-antigen interactions even in the presence of multiple active T cell clones (Figure 6E; Tables S5 and S6). Further experiments confirmed the original two antigens, and none of the other 18 TCRs identified additional antigens even when all of the patient-specific HLA-II alleles were included in the screens (Figures S6D and S6E; Tables S5 and S6).

Identification of a cross-reactive *Bacteroidales* peptide

Little is known about how tolerance is broken in autoimmune diseases. One hypothesis proposes that cross-reactive T cells, capable of recognizing microbial and self-peptides, play a role in the development of autoimmune disorders.[65–67] Thus, we took advantage of the knowledge of the minimal core residues sufficient to activate both TCRs (Figures S6F–S6I) to bioinformatically impute viral and bacterial epitopes that may activate the 3A and 4A TCRs. We identified a peptide fragment from the TonB-dependent receptor protein of *Bacteroidales* that can readily activate TCR 4A (Figure 6G). Intriguingly, an increase in the abundance of *Bacteroidales* has been reported in the oral microbial flora of SjD patients compared with healthy controls, although the clinical significance of this finding is not known.[68] Further investigation is needed to assess the significance of this finding to SjD pathogenesis.

(D) Validation of the identified antigens in target cells following 6-h co-culture with cells expressing CTAG1B TCR and eHEK-293-II cells expressing HLA-DRB1*04. The box is drawn around the inner quartile range, and the whiskers show minimum to maximum values.
(E) Heatmap representation of comprehensive mutagenesis analysis of the CTAG1B epitope in HLA-DRB1*04 eHEK-293-II cells using CTAG1B-specific TCR (3598-2). The mutant epitopes are expressed in the context of a 56-aa fragment. Heatmaps are as in Figure 2.
(F) Transcript expression levels for CTAG2, CTAG1B, and TTN from 61 human tissues based on data from three sources: HPA, GTEx, and FANTOM5. Bar graph with the y axis plotting consensus normalized expression value (TPM, transcripts per kilobase million) and x axis representing the multiple human tissues and blood cell types. See Figure S4.

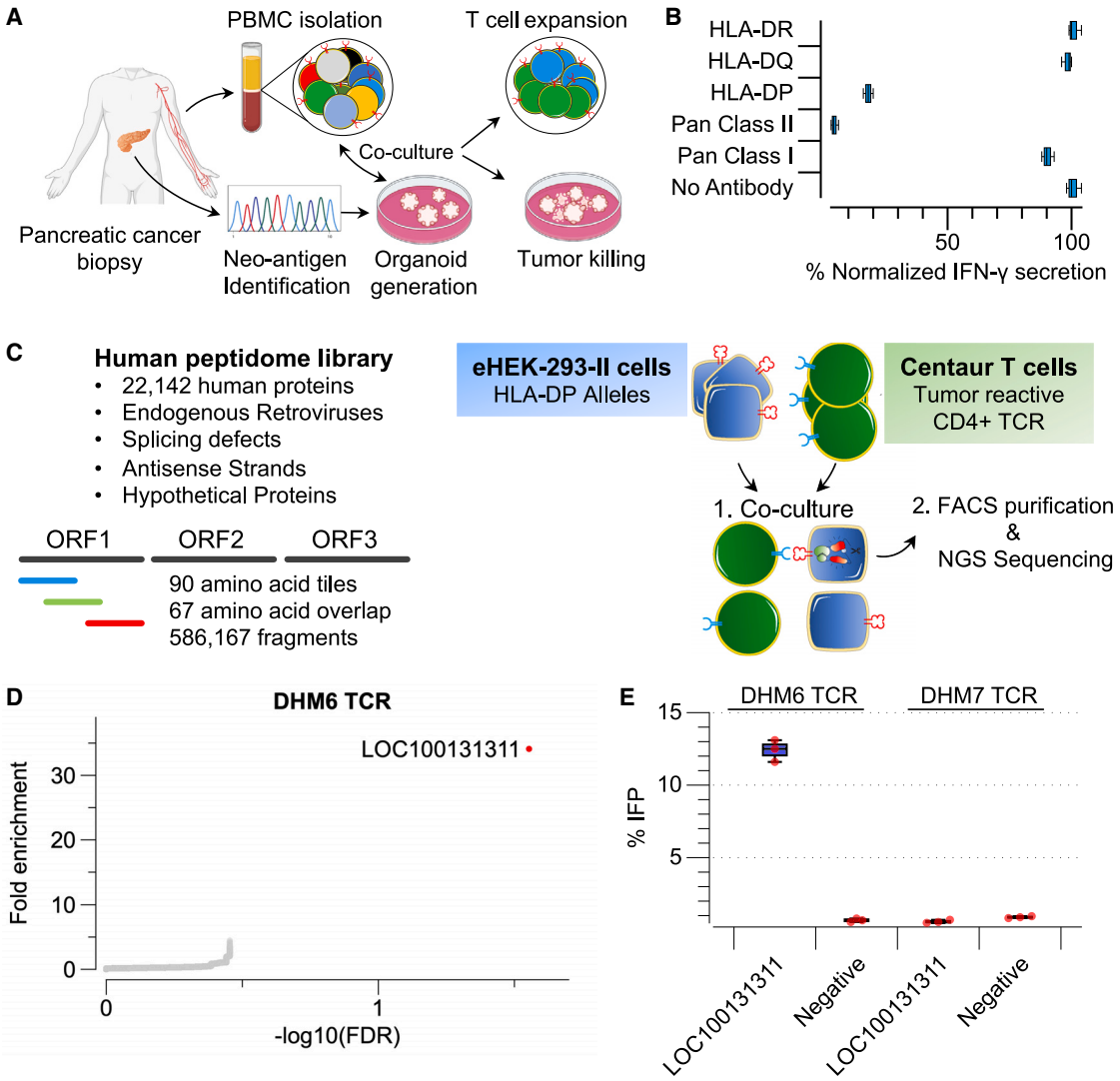


Figure 5. Characterization of pancreatic-cancer-reactive TCRs

(A) Schematic of the experimental workflow to derive and characterize pancreatic-cancer-tumor-reactive T cell clones.

(B) Characterizing tumor reactivity and HLA-II allele restriction of tumor-reactive TCRs using patient-derived organoids. Primary centaur T cells were co-cultured with patient-derived organoids. IFN-γ secretion was used as a readout of T cell activity. Pan-HLA class-I antibody (W6/32), Pan-HLA class-II antibody (Tu39), HLA-DR (L243), HLA-DP (B7/21), and HLA-DQ (SPVL3) were used for blocking experiments. The box is drawn around the inner quartile range, and the whiskers show minimum to maximum values for all plots shown here and below.

(C) Schematic of TScan-II human genome-wide screen. The Human Peptidome V3 genome-wide library comprises 586,167 antigens that tile across the entire human proteome, alternative translational products, endogenous retroviruses, splicing defects, antisense strands, and hypothetical proteins in 90-aa fragments with 67-aa overlap.

(D) TScan-II screen of $CD4^+$ TCR (DHM6) against the Human Peptidome V3 library and HLA-DP expressing library eHEK-293-II cells. p values were calculated by permutation testing, and false discovery rate correction was done by the Benjamini-Hochberg method.

(E) Validation of the identified antigens in target cells using the GzB reporter following 6-h co-culture with DHM6-transduced T cells and indicated 90 aa antigens. See Figure S5.

Autoimmune antigens in SjD are expressed in HLA-II-expressing SG cells

To assess the expression of antigens in the SG of SjD patients, we used single-cell RNA sequencing (scRNA-seq) data of the minor SG biopsies collected from SjD patients fulfilling the classification criteria for SjD (Figure 7A). As expected, we observed a clear increase in the abundance of B cell and T cell infiltration within the SGs of SjD patients compared with non-SjD controls (Figure 7B). In the epithelial compartment, MAP3K4 was expressed in the acinar, ductal, and myoepithelial cells within the SG (Figure 7C). DDIAS expression was mostly limited to the ductal cells of the SG (Figure 7C). Both MAP3K4 and DDIAS are expressed at higher levels in the ducal cells of SjD patients (Figure S7A), consistent with studies showing that the

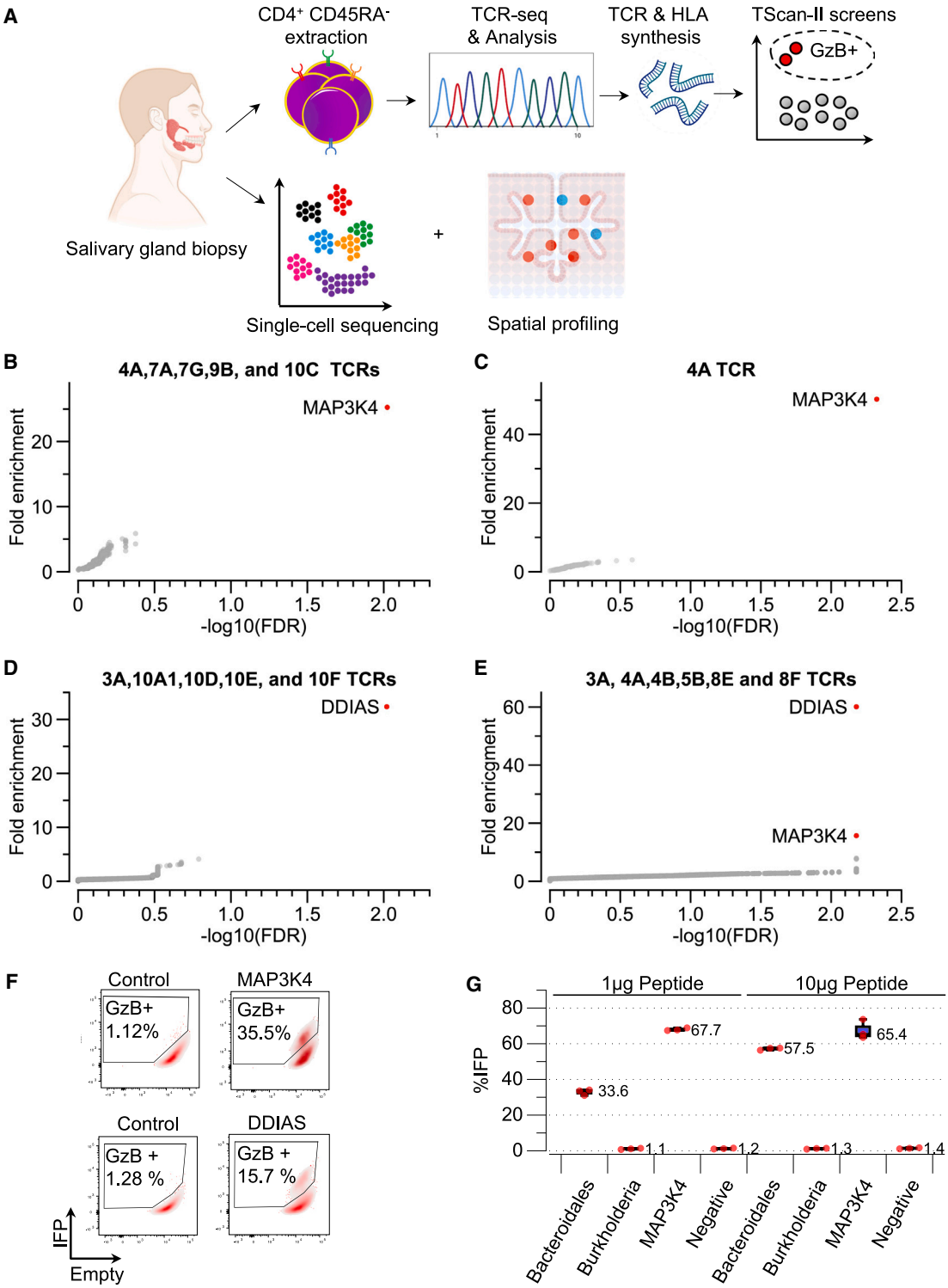


Figure 6. Human genome-wide TScan-II screens of clonally expanded CD4⁺ T cells in SjD

(A) Schematic representation of the experimental workflow to characterize clonally expanded CD4⁺ T cell clones in SjD patients.

(B–E) TScan-II screens were performed against the Human Genome-Wide V2 library with eHEK-293-II cells expressing HLA-DRB1*04, HLA-DRB1*03:01, and HLA-DQB1*02_HLA-DQA1*05. p values were calculated by permutation testing of the geometric mean across enrichment values. False discovery rate correction was done by the Benjamini-Hochberg method.

(B) Pooled TScan-II screen of 5 TCRs, 4A, 7A, 7G, 9B, and 10C.

(legend continued on next page)

expression of DDIAS is upregulated in the SG of SjD patients.[69,70] To assess the spatial expression relative to the ductal cells and lymphocytic infiltrates, we used the Visium platform (10× Genomics). Visium provides unbiased, coarse-grained, spatially resolved (~55 μm) transcriptomic information but cannot resolve individual cells. However, in direct support of our scRNA-seq results, we observed a partial co-location of the expression of *MAP3K4* and *DDIAS* with markers of ductal cells (i.e., *S100A2* and *CFTR*).[71] Furthermore, we detected transcripts for both *DDIAS* and *MAP3K4* in areas of lymphocytic inflammation (*arrows*, Figure S7B).

Given the expansion of $CD4^+$ cells in the SjD SG, a central question is where the autoimmune targets encounter HLA-II expression. Is it in the associated APCs in the inflammatory infiltrates, or could it also be on the SG tissues themselves? In healthy SGs, we found HLA-II expression on B cells, macrophages, and dendritic cells, as expected. However, HLA-II was also expressed within the SG's mucous acini and ductal cells (Figure 7C) and is markedly further increased in the ductal cells of SjD patients (Figure S7A). A main characteristic of focal lymphocytic infiltration in SjD patients is the periductal arrangement of $CD4^+$ cells.[52,72] Importantly, ductal cells express both MAP3K4 and DDIAS. Hence, multiple cell types within the SG are likely competent to present HLA-II antigens during disease progression. We also detected IFNGR1 and ISG15 gene expression within the ductal cells; therefore, the observed upregulation of HLA-II within the ductal cells could likely result from chronic SG inflammation and interferon (Figures 7A and S7A). Collectively, we observed that both MAP3K4 and DDIAS are expressed within specific populations of cells in the SG that are most likely capable of processing and presenting HLA-II antigens, which could contribute to SjD.

DISCUSSION

The immune system's primary function is to maintain tissue homeostasis by recognizing and eliminating pathogens and damaged cells. Uncovering which antigens are recognized has remained largely elusive, partially due to the absence of high-throughput tools to probe T cell specificity. Here, we describe TScan-II, a genome-scale platform for the *de novo* identification of $CD4^+$ TCR antigens. TScan-II employs endogenous HLA-II-processing machinery to process antigens from full-length ORFs or long peptide fragments, thus greatly expanding the complexity of antigenic libraries. The coupling of these complex libraries with a reporter of T cell activity having a high signal-to-noise ratio resulted in the enrichment of scarce HLA-II antigens. We applied TScan-II to decipher the antigen specificity of disease-relevant $CD4^+$ TCRs in infectious disease, cancer, and autoimmunity. In our proof-of-principle screens, TScan-II was capable of identifying cognate antigens from viral and self-reactive $CD4^+$ TCRs and provided specificity insights for both. Interestingly, in both HIV and CTAG1B screens, the identified cross-reactive epitopes had minimal homology to the original epitope, underscoring the difficulty of predicting TCR specificities in the absence of unbiased genome-scale approaches. The ease with which the platform can be multiplexed at both T cell and HLA levels allows this tool to be highly customizable to various basic and translational applications, enabling genome-scale antigen discovery for $CD4^+$ T cells.

The unbiased nature of TScan-II makes it an ideal tool for the dissection of the antigenic landscape associated with $CD4^+$ cells within the tumor microenviroment (TME). In this regard, we employed patient-derived organoids to assess the reactivity of $CD4^+$ T cell clones toward HLA-II-positive pancreatic cancer tumors. Taking an unbiased approach with a library covering unconventional transcripts and translation products in addition to the entire proteome allowed us to discover that this tumor-reactive TCR recognized an epitope originating from an uncharacterized gene expressed within matching patient tumor organoids. The ability of TScan-II to find different classes of antigens within the non-canonical immunopeptidome showcases its utility in cancer research.

In the autoimmune context, we applied TScan-II to SjD and examined the antigen specificity of clonally expanded $CD4^+$ T cells within the SG patients. By screening 20 expanded clones from 5 SjD patients, we identified self-antigens that can readily and potently activate two of the most expanded clones from two individuals. A key question concerns how HLA-II self-antigens are presented to $CD4^+$ cells in SjD. There are two possibilities: (1) antigens can be phagocytosed and presented by APC cells within the SG, or (2) they can be presented directly to $CD4^+$ T cells by HLA-II-positive SG cells. Ductal epithelial cells within the SG have been circumstantially implicated in presenting self-antigens to $CD4^+$ cells. Using single-cell data, we detected HLA-II expression on ductal cells within the SG of healthy individuals. This expression is considerably upregulated within the inflamed SG of SjD subjects relative to non-SjD subjects. Importantly, both DDIAS and MAP3K4 were expressed by the HLA-II-positive ductal cells, consistent with the possibility that both antigens can be processed and presented by ductal epithelial cells directly to $CD4^+$ T cells. Spatial transcriptomics showed that both antigens were expressed in areas within the SG with high lymphocyte density. Importantly, peptides spanning the length of DDIAS and MAP3K4 proteins have been previously identified in HLA-II ligandome datasets from the bone marrow, kidney, and lungs.[73,74] Collectively, not only can DDIAS and MAP3K4 be processed and presented on HLA-II alleles, but they are also expressed in cells expressing HLA-II and reside in the neighborhood of periductal immune infiltrates within the SG. How cells directly present peptides from their cytoplasmic proteins on HLA-II molecules is unknown, but this could be an internal form of presentation employing bulk autophagy, or

(C) TScan-II screen of 4A TCR.
(D) Pooled TScan-II screen using 5 TCRs, 3A, 10A1, 10D, 10E, and 10F.
(E) Pooled TScan-II using 6 TCRs, 3A, 4A, 9C, 8A, 8B, and 5B.
(F) FACS plots depicting activation of GzB reporter following a 6-h co-culture experiment with 3A and 4A centaur T cells toward the indicated cognate antigens.
(G) Relative activation of the GzB reporter in cells pulsed with the indicated peptide fragment co-cultured with 4A T cells. The box is drawn around the inner quartile range, and the whiskers show minimum to maximum values for all plots shown. See Figure S6.

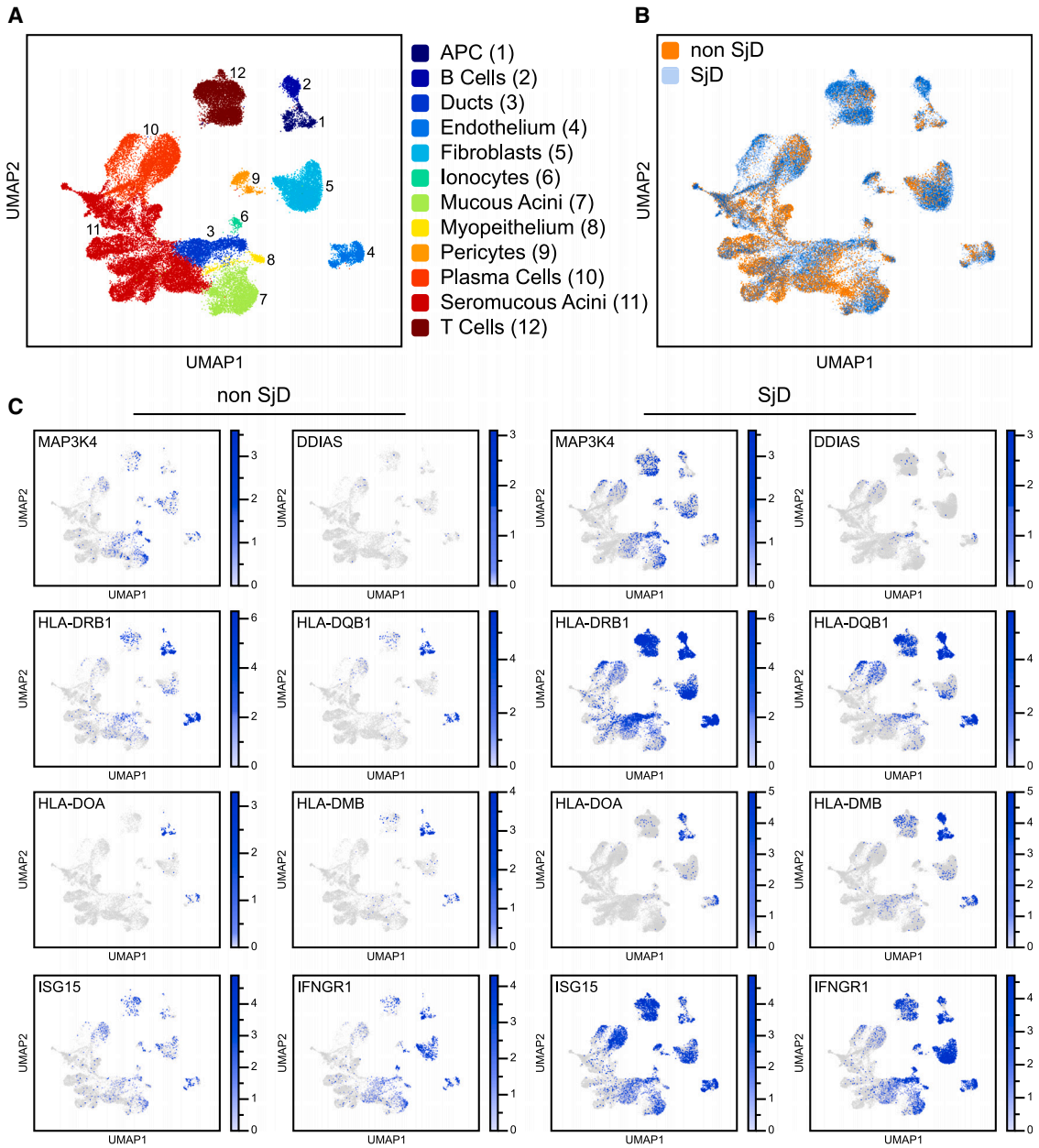


Figure 7. Single-cell expression profiling of CD4$^+$ antigens and HLA-II machinery within the salivary gland

(A) Uniform manifold approximation and projection (UMAP) visualization of scRNA-seq data from dissociated SGs of healthy individuals and SjD patients. Cells from minor SG from n = 6 subjects who did not fulfill ACR 2016 classification criteria and were otherwise healthy and n = 7 SjD subjects who fulfilled American College of Rheumatology Classification Criteria were integrated into a single object containing 12 clusters. Selective clusters are highlighted by their functional annotation.

(B) UMAP plot depicting the abundance of SG cells in non-SjD (healthy) and SjD subjects overlaid on the UMAP plot depicted in (A).

(C) Expression of various HLA-II marker genes and MAP3K4 and DDIAS among non-SjD and SjD patients overlaid on the UMAP plot depicted in (A). See Figure S7.

peptides transported into the ER may be able to load onto the HLA-II molecule prior to transport to the lysosomes.[10] These specific TCRs may aid in the further testing of the HLA-II antigen presentation on SG in the future.

For several autoimmune diseases, such as rheumatoid arthritis (RA) and type 1 diabetes, a subset of patients display a strong association with certain HLA-II alleles.[75–78] Historically, studies in SjD have identified associations with the HLA-II locus, which were later corroborated by genome-wide association studies (GWASs) among SjD patients.[57–61,79,80] In Europeans, the C*07:01–B*08:01–DRB1*03:01–DQB1*02:01 haplotype exhibits the strongest genetic association with SjD.[60,61,79] In this

study, we identified self-antigens restricted to both of these HLA-II alleles. Is this an unusual case where two different HLA-II alleles that happen to be linked can each independently contribute to SjD pathogenesis, or could another mechanism also be involved? During the course of our studies, several publications suggested that the causative genetic element in the C*07:01–B*08:01–DRB1*03:01–DQB1*02:01 haplotype may actually be due to the *C4* gene.[81–83] This extended haplotype is in strong linkage disequilibrium (LD) with a *C4A* null allele in populations of European origin.[81,84,85] The same haplotype has also been linked to systemic lupus erythematosus (SLE).[57–60,86] Recombinants that separate the HLA from the *C4* locus have shown that *C4* is the likely cause of SLE.[81,87,88] In addition, heterozygous *C2* deficiency—when present in combination with a low *C4A* copy number—substantially increased the risk of SLE and primary SjD by 10-fold compared with individuals with 2 *C4A* copies and normal *C2*.[83] The complement system plays a critical role in the clearance of cellular debris, such as apoptotic cell fragments and bacterial pathogens. The impaired clearance of cellular debris, pathogens, and immune complexes due to the absence or lower levels of complement could, in turn, result in the accumulation of antigen fragments (clearance hypothesis).[89–91] Lupus-prone mice expressing the human C4A protein have an improved ability to clear apoptotic cells and exhibit lower levels of self-reactive B cells and decreased levels of autoantibodies targeting SSA, relative to mice expressing human C4B, providing supporting evidence for the clearance hypothesis.[92] This hypothesis would be consistent with our finding on the cross-reactive epitope in *Bacteroidales*. Furthermore, GWASs have shown that a subset of SjD and SLE patients of European descent exhibit HLA-II associations independent of C4 allele status.[93,94] In particular, variations in the intergenic sequences between HLA-DRB1 and HLA-DQA1 (XL9) region in different haplotypes contribute to an increase in HLA-II gene expression.[95–97] Assuming the accuracy of these findings, the association between disease susceptibility and either lower C4 protein levels in one haplotype or the increase in HLA-II expression in others would result in a pro-immunogenic increased amount of self-antigen presentation rather than a common antigen specificity among affected SjD individuals.[95] As a result, multiple self-antigens expressed in the exocrine glands, each with varying HLA-II restrictions, are likely to contribute to the $CD4^+$ T cell-mediated pathogenesis observed in SjD. The divergent HLA-II restriction of the two self-antigens identified here is consistent with the above hypothesis. The simultaneous involvement of complement and $CD4^+$ T cells presents a complex scenario where the chronic interplay of innate and adaptive immune systems synergizes to initiate and advance SjD, culminating in the heterogeneous clinical presentations witnessed among patients. Future studies probing the antigenic landscape of expanded T cell clones across a large cohort of patients with SjD and multiple autoimmune diseases will further shed light on commonalities and differences among antigen-governing pathogenesis.

Collectively, through the various applications described here, we provide evidence that the TScan-II platform can facilitate the identification of TCR-epitope relationships in infectious diseases, cancer, and autoimmunity and will serve as an avenue for the dissection of many more such relationships in the future.

Limitations of the study

TScan-II has several limitations. First, it relies on the genetic encoding of candidate antigens, which does not include certain antigens, such as CD1-bound lipids. Second, in its current form, it is unlikely to present peptides with post-translational modifications (PTMs). One potential approach to address this limitation is the use of full-length proteins or engineering modifiers, such as protein kinases and citrullination enzymes, to reside in the ER. Finally, this unbiased screen revealed that many TCRs do not recognize a self-antigen, an observation that may not actually be a limitation of the platform per se but possibly a feature of the biology of T cell infiltration in SjD. Two of the possible explanations are that the self-antigen target is absent from our library or that the TCRs are not self-reactive but instead recognize epitopes derived from bacterial, fungal, viral, or food sources, given the exposure of the SG to the environment.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human Participants
  - Cell lines
  - Primary T cells
- METHOD DETAILS
  - T-cell extraction and expansion
  - T-cell transduction
  - Lentiviral production
  - Knockout of endogenous B2M gene
  - Mutation of endogenous HLA-II genes
  - Knockout of endogenous TCRβ gene
  - Generation of CIITA and IFP^{GzB-Hi} $ICAD^{CR}$ constructs
  - Generation of membrane bound anti-CD3 scFv
  - Generation of membrane-bound anti-centaur scFv
  - HLA allele design and cloning
  - $CD4^+$ TCR design and cloning
  - Validation experiments
  - Library design, cloning, and transduction
  - TScan-II NGS sequencing
  - TScan-II data analysis
  - Characterization of organoid-reactive T-cells
  - Pancreatic cancer neo-antigen screen
  - Characterization of SjD T-cells
  - SjD TScan-II screens
  - Salivary gland single-cell data tissue acquisition
  - Salivary gland scRNA-seq
  - Salivary gland scRNA-seq data processing

- Salivary gland scRNA-seq data analysis
- 10X Visium tissue harvest
- 10X Visium slides preparation, and sequencing
- 10X Visium data analysis

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2023.10.024.

ACKNOWLEDGMENTS

This research was supported by the Sjögren's Foundation (M.H.D.); Howard Hughes Medical Institute Investigator (S.J.E.); Department of Defense BC171184 (S.J.E.), U24AI118633 (S.J.E.), 5R01 AR073855 (C.J.L.), and 5R01 AR074310 (A.D.F.); NIDCR Division of Intramural Research Award DE000704 (B.M.W.); the NIDCR Sjogren's Clinical Investigations Team; the NIDCR/NIDCD Genomics and Computational Biology Core; and NIH training grant T32 HG002295 (N.K.). S.K.M. was supported by NIH Intramural Research Program and Kim and Judy Davis. The authors would also like to thank David Owen for comments on the manuscript.

AUTHOR CONTRIBUTIONS

M.H.D. and S.J.E. conceived and designed the experiments. M.H.D. performed the TScan experiments. M.H.D, T.K., and N. K. analyzed the TScan-II screen data. A.J. and M.L.J. provided SjD TCR sequences. B.M.W., P.P., T.P., and J.A.C. performed the single-cell experiments. B.M.W., T.J.P., and M.H.D. analyzed the single-cell experiments. C.J.L., B.K., and A.D.F. performed the Visium spatial transcriptomics. C.J.L., B.K., and M.H.D. analyzed the Visium spatial transcriptomics. Q.M. and S.K.M. performed the pancreatic cancer organoid co-culture experiments. M.H.D. and S.J.E. wrote the paper, and B.M.W., A.D.F., and C.J.L. edited the paper. All the authors reviewed and approved the final manuscript.

DECLARATION OF INTERESTS

M.H.D., T.K., and S.J.E. are patent holders of the TScan-I and TScan-II technologies. T.K. is a co-founder of TScan Therapeutics and ImmuneID and serves on the Scientific Advisory Board of TScan Therapeutics. S.J.E. is a founder of TScan Therapeutics, ImmuneID, MAZE Therapeutics, and Mirimus. S.J.E. serves on the Scientific Advisory Board of TScan Therapeutics and Maze Therapeutics. In accordance with Partners HealthCare's conflict of interest policies, the Partners Office for Interactions with Industry has reviewed S.J.E.'s financial interest in TScan and determined that it creates no significant risk to the welfare of the participants in this study or to the integrity of the research.

Received: April 10, 2023
Revised: July 12, 2023
Accepted: October 25, 2023
Published: November 27, 2023

REFERENCES

1. Flajnik, M.F., and Kasahara, M. (2010). Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat. Rev. Genet. *11*, 47–59. https://doi.org/10.1038/nrg2703.
2. Kula, T., Dezfulian, M.H., Wang, C.I., Abdelfattah, N.S., Hartman, Z.C., Wucherpfennig, K.W., Lyerly, H.K., and Elledge, S.J. (2019). T-scan: A genome-wide method for the systematic discovery of T cell epitopes. Cell *178*, 1016–1028.e13. https://doi.org/10.1016/j.cell.2019.07.009.
3. Birnbaum, M.E., Mendoza, J.L., Sethi, D.K., Dong, S., Glanville, J., Dobbins, J., Ozkan, E., Davis, M.M., Wucherpfennig, K.W., and Garcia, K.C. (2014). Deconstructing the peptide-MHC specificity of T cell recognition. Cell *157*, 1073–1087. https://doi.org/10.1016/j.cell.2014.03.047.
4. Joglekar, A.V., and Li, G. (2021). T cell antigen discovery. Nat. Methods *18*, 873–880. https://doi.org/10.1038/s41592-020-0867-z.
5. Sharma, G., Rive, C.M., and Holt, R.A. (2019). Rapid selection and identification of functional CD8+ T cell epitopes from large peptide-coding libraries. Nat. Commun. *10*, 4553. https://doi.org/10.1038/s41467-019-12444-7.
6. Li, G., Bethune, M.T., Wong, S., Joglekar, A.V., Leonard, M.T., Wang, J.K., Kim, J.T., Cheng, D., Peng, S., Zaretsky, J.M., et al. (2019). T cell antigen discovery via trogocytosis. Nat. Methods *16*, 183–190. https://doi.org/10.1038/s41592-018-0305-7.
7. Kisielow, J., Obermair, F.-J., and Kopf, M. (2019). Deciphering CD4+ T cell specificity using novel MHC-TCR chimeric receptors. Nat. Immunol. *20*, 652–662. https://doi.org/10.1038/s41590-019-0335-z.
8. Lee, M.N., and Meyerson, M. (2021). Antigen identification for HLA class I- and HLA class II-restricted T cell receptors using cytokine-capturing antigen-presenting cells. Sci. Immunol. *6*, eabf4001. https://doi.org/10.1126/sciimmunol.abf4001.
9. Blum, J.S., Wearsch, P.A., and Cresswell, P. (2013). Pathways of antigen processing. Annu. Rev. Immunol. *31*, 443–473. https://doi.org/10.1146/annurev-immunol-032712-095910.
10. Roche, P.A., and Furuta, K. (2015). The ins and outs of MHC class II-mediated antigen processing and presentation. Nat. Rev. Immunol. *15*, 203–216. https://doi.org/10.1038/nri3818.
11. Reith, W., LeibundGut-Landmann, S., and Waldburger, J.-M. (2005). Regulation of MHC class II gene expression by the class II transactivator. Nat. Rev. Immunol. *5*, 793–806. https://doi.org/10.1038/nri1708.
12. Wucherpfennig, K.W., and Hafler, D.A. (1995). A review of T-cell receptors in multiple sclerosis: clonal expansion and persistence of human T-cells specific for an immunodominant myelin basic protein peptide. Ann. N. Y. Acad. Sci. *756*, 241–258. https://doi.org/10.1111/j.1749-6632.1995.tb44522.x.
13. Schröder, B. (2016). The multifaceted roles of the invariant chain CD74 — more than just a chaperone. Biochim. Biophys. Acta *1863*, 1269–1281. https://doi.org/10.1016/j.bbamcr.2016.03.026.
14. Benati, D., Galperin, M., Lambotte, O., Gras, S., Lim, A., Mukhopadhyay, M., Nouël, A., Campbell, K.-A., Lemercier, B., Claireaux, M., et al. (2016). Public T cell receptors confer high-avidity CD4 responses to HIV controllers. J. Clin. Invest. *126*, 2093–2108. https://doi.org/10.1172/JCI83792.
15. To, T.-L., Piggott, B.J., Makhijani, K., Yu, D., Jan, Y.N., and Shu, X. (2015). Rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis in vivo. Proc. Natl. Acad. Sci. USA *112*, 3338–3343. https://doi.org/10.1073/pnas.1502857112.
16. Timms, R.T., Zhang, Z., Rhee, D.Y., Harper, J.W., Koren, I., and Elledge, S.J. (2019). A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. Science *365*, eaaw4912. https://doi.org/10.1126/science.aaw4912.
17. Soetandyo, N., Wang, Q., Ye, Y., and Li, L. (2010). Role of intramembrane charged residues in the quality control of unassembled T-cell receptor α-chains at the endoplasmic reticulum. J. Cell Sci. *123*, 1031–1038. https://doi.org/10.1242/jcs.059758.
18. Cohen, C.J., Zhao, Y., Zheng, Z., Rosenberg, S.A., and Morgan, R.A. (2006). Enhanced antitumor activity of murine-human hybrid T-cell receptor (TCR) in human lymphocytes is associated with improved pairing and TCR/CD3 stability. Cancer Res. *66*, 8878–8886. https://doi.org/10.1158/0008-5472.CAN-06-1450.
19. Cohen, C.J., Li, Y.F., El-Gamil, M., Robbins, P.F., Rosenberg, S.A., and Morgan, R.A. (2007). Enhanced antitumor activity of T cells engineered to express T-cell receptors with a second disulfide bond. Cancer Res. *67*, 3898–3903. https://doi.org/10.1158/0008-5472.CAN-06-3986.

20. Courtney, A.H., Lo, W.-L., and Weiss, A. (2018). TCR signaling: mechanisms of initiation and propagation. Trends Biochem. Sci. *43*, 108–123. https://doi.org/10.1016/j.tibs.2017.11.008.

21. Lambotte, O., Boufassa, F., Madec, Y., Nguyen, A., Goujard, C., Meyer, L., Rouzioux, C., Venet, A., and Delfraissy, J.-F.; SEROCO-HEMOCO Study Group (2005). HIV controllers: A homogeneous group of HIV-1-infected patients with spontaneous control of viral replication. Clin. Infect. Dis. *41*, 1053–1056. https://doi.org/10.1086/433188.

22. Galperin, M., Farenc, C., Mukhopadhyay, M., Jayasinghe, D., Decroos, A., Benati, D., Tan, L.L., Ciacchi, L., Reid, H.H., Rossjohn, J., et al. (2018). CD4+ T cell–mediated HLA class II cross-restriction in HIV controllers. Sci. Immunol. *3*, eaat0687. https://doi.org/10.1126/sciimmunol.aat0687.

23. Thomas, R., Al-Khadairi, G., Roelands, J., Hendrickx, W., Dermime, S., Bedognetti, D., and Decock, J. (2018). NY-ESO-1 based immunotherapy of cancer: current perspectives. Front. Immunol. *9*, 947. https://doi.org/10.3389/fimmu.2018.00947.

24. Park, T.S., Groh, E.M., Patel, K., Kerkar, S.P., Lee, C.-C.R., and Rosenberg, S.A. (2016). Expression of MAGE-A and NY-ESO-1 in primary and metastatic cancers. J. Immunother. *39*, 1–7. https://doi.org/10.1097/CJI.0000000000000101.

25. Cameron, B.J., Gerry, A.B., Dukes, J., Harper, J.V., Kannan, V., Bianchi, F.C., Grand, F., Brewer, J.E., Gupta, M., Plesa, G., et al. (2013). 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. https://doi.org/10.1126/scitranslmed.3006034.

26. Morgan, R.A., Chinnasamy, N., Abate-Daga, D., Gros, A., Robbins, P.F., Zheng, Z., Dudley, M.E., Feldman, S.A., Yang, J.C., Sherry, R.M., et al. (2013). Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy. J. Immunother. *36*, 133–151. https://doi.org/10.1097/CJI.0b013e3182829903.

27. Linette, G.P., Stadtmauer, E.A., Maus, M.V., Rapoport, A.P., Levine, B.L., Emery, L., Litzky, L., Bagg, A., Carreno, B.M., Cimino, P.J., et al. (2013). Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. Blood *122*, 863–871. https://doi.org/10.1182/blood-2013-03-490565.

28. Poncette, L., Chen, X., Lorenz, F.K., and Blankenstein, T. (2019). Effective NY-ESO-1-specific MHC II-restricted T cell receptors from antigen-negative hosts enhance tumor regression. J. Clin. Invest. *129*, 324–335. https://doi.org/10.1172/JCI120391.

29. Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. *48*, W449–W454. https://doi.org/10.1093/nar/gkaa379.

30. Meng, Q., Xie, S., Gray, G.K., Dezfulian, M.H., Li, W., Huang, L., Akshinthala, D., Ferrer, E., Conahan, C., Perea Del Pino, S., et al. (2021). Empirical identification and validation of tumor-targeting T cell receptors from circulation using autologous pancreatic tumor organoids. J. Immunother. Cancer *9*, e003213. https://doi.org/10.1136/jitc-2021-003213.

31. Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B.J., et al. (2020). Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. Nat. Commun. *11*, 1293. https://doi.org/10.1038/s41467-020-14968-9.

32. Zhao, Q., Laverdure, J.-P., Lanoix, J., Durette, C., Côté, C., Bonneil, É., Laumont, C.M., Gendron, P., Vincent, K., Courcelles, M., et al. (2020). Proteogenomics uncovers a vast repertoire of shared tumor-specific antigens in ovarian cancer. Cancer Immunol. Res. *8*, 544–555. https://doi.org/10.1158/2326-6066.CIR-19-0541.

33. Ouspenskaia, T., Law, T., Clauser, K.R., Klaeger, S., Sarkizova, S., Aguet, F., Li, B., Christian, E., Knisbacher, B.A., Le, P.M., et al. (2022). Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. Nat. Biotechnol. *40*, 209–217. https://doi.org/10.1038/s41587-021-01021-3.

34. Xiang, R., Ma, L., Yang, M., Zheng, Z., Chen, X., Jia, F., Xie, F., Zhou, Y., Li, F., Wu, K., and Zhu, Y. (2021). Increased expression of peptides from non-coding genes in cancer proteomics datasets suggests potential tumor neoantigens. Commun. Biol. *4*, 496. https://doi.org/10.1038/s42003-021-02007-2.

35. Starck, S.R., and Shastri, N. (2011). Non-conventional sources of peptides presented by MHC class I. Cell. Mol. Life Sci. *68*, 1471–1479. https://doi.org/10.1007/s00018-011-0655-0.

36. Wang, S., Mao, C., and Liu, S. (2019). Peptides encoded by noncoding genes: challenges and perspectives. Signal Transduct. Target. Ther. *4*, 57. https://doi.org/10.1038/s41392-019-0092-3.

37. Lv, D., Chang, Z., Cai, Y., Li, J., Wang, L., Jiang, Q., Xu, K., Ding, N., Li, X., Xu, J., and Li, Y. (2022). TransLnc: a comprehensive resource for translatable lncRNAs extends immunopeptidome. Nucleic Acids Res. *50*, D413–D420. https://doi.org/10.1093/nar/gkab847.

38. Adinarayana, K.P.S., Sravani, T.S., and Hareesh, C. (2011). A database of six eukaryotic hypothetical genes and proteins. Bioinformation *6*, 128–130. https://doi.org/10.6026/97320630006128.

39. Gonzalez-Galarza, F.F., McCabe, A., dos Santos, E.J.M.D., Jones, J., Takeshita, L., Ortega-Rivera, N.D., Cid-Pavon, G.M.D., Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., et al. (2020). Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res. *48*, D783–D788. https://doi.org/10.1093/nar/gkz1029.

40. Kamradt, T., and Mitchison, N.A. (2001). Tolerance and autoimmunity. N. Engl. J. Med. *344*, 655–664. https://doi.org/10.1056/NEJM200103013440907.

41. Bloch, K.J., Buchanan, W.W., Wohl, M.J., and Bunim, J.J. (1965). Sjoegren's syndrome. A clinical, pathological, and serological study of sixty-two cases. Med. (Baltim.) *44*, 187–231.

42. Adamson, T.C., 3rd, Fox, R.I., Frisman, D.M., and Howell, F.V. (1983). Immunohistologic analysis of lymphoid infiltrates in primary Sjogren's syndrome using monoclonal antibodies. J. Immunol. *130*, 203–208. https://doi.org/10.4049/jimmunol.130.1.203.

43. Christodoulou, M.I., Kapsogeorgou, E.K., and Moutsopoulos, H.M. (2010). Characteristics of the minor salivary gland infiltrates in Sjögren's syndrome. J. Autoimmun. *34*, 400–407. https://doi.org/10.1016/j.jaut.2009.10.004.

44. Kapsogeorgou, E.K., Christodoulou, M.I., Panagiotakos, D.B., Paikos, S., Tassidou, A., Tzioufas, A.G., and Moutsopoulos, H.M. (2013). Minor salivary gland inflammatory lesions in Sjögren syndrome: do they evolve? J. Rheumatol. *40*, 1566–1571. https://doi.org/10.3899/jrheum.130256.

45. Skopouli, F.N., Fox, P.C., Galanopoulou, V., Atkinson, J.C., Jaffe, E.S., and Moutsopoulos, H.M. (1991). T cell subpopulations in the labial minor salivary gland histopathologic lesion of Sjögren's syndrome. J. Rheumatol. *18*, 210–214.

46. Mingueneau, M., Boudaoud, S., Haskett, S., Reynolds, T.L., Nocturne, G., Norton, E., Zhang, X., Constant, M., Park, D., Wang, W., et al. (2016). Cytometry by time-of-flight immunophenotyping identifies a blood Sjögren's signature correlating with disease activity and glandular inflammation. J. Allergy Clin. Immunol. *137*, 1809–1821.e12. https://doi.org/10.1016/j.jaci.2016.01.024.

47. Joachims, M.L., Leehan, K.M., Lawrence, C., Pelikan, R.C., Moore, J.S., Pan, Z., Rasmussen, A., Radfar, L., Lewis, D.M., Grundahl, K.M., et al. (2016). Single-cell analysis of glandular T cell receptors in Sjögren's syndrome. JCI Insight *1*, e85609. https://doi.org/10.1172/jci.insight.85609.

48. Moutsopoulos, H.M., Hooks, J.J., Chan, C.C., Dalavanga, Y.A., Skopouli, F.N., and Detrick, B. (1986). HLA-DR expression by labial minor salivary gland tissues in Sjögren's syndrome. Ann. Rheum. Dis. *45*, 677–683. https://doi.org/10.1136/ard.45.8.677.

49. Fox, R.I., Bumol, T., Fantozzi, R., Bone, R., and Schreiber, R. (1986). Expression of histocompatibility antigen HLA-DR by salivary gland epithelial cells in Sjögren's syndrome. Arthritis Rheum. *29*, 1105–1111. https://doi.org/10.1002/art.1780290908.

50. Thrane, P.S., Halstensen, T.S., Haanaes, H.R., and Brandtzaeg, P. (1993). Increased epithelial expression of HLA-DQ and HLA-DP molecules in salivary glands from patients with Sjögren's syndrome compared with obstructive sialadenitis. Clin. Exp. Immunol. *92*, 256–262.

51. Tsunawaki, S., Nakamura, S., Ohyama, Y., Sasaki, M., Ikebe-Hiroki, A., Hiraki, A., Kadena, T., Kawamura, E., Kumamaru, W., Shinohara, M., and Shirasuna, K. (2002). Possible function of salivary gland epithelial cells as nonprofessional antigen-presenting cells in the development of Sjögren's syndrome. J. Rheumatol. *29*, 1884–1896.

52. Molina, C., Alliende, C., Aguilera, S., Kwon, Y.-J., Leyton, L., Martínez, B., Leyton, C., Pérez, P., and González, M.-J. (2006). Basal lamina disorganisation of the acini and ducts of labial salivary glands from patients with Sjögren's syndrome: association with mononuclear cell infiltration. Ann. Rheum. Dis. *65*, 178–183. https://doi.org/10.1136/ard.2004.033837.

53. Manoussakis, M.N., Dimitriou, I.D., Kapsogeorgou, E.K., Xanthou, G., Paikos, S., Polihronis, M., and Moutsopoulos, H.M. (1999). Expression of B7 costimulatory molecules by salivary gland epithelial cells in patients with Sjögren's syndrome. Arthritis Rheum. *42*, 229–239. https://doi.org/10.1002/1529-0131(199902)42:2<229::AID-ANR4>3.0.CO;2-X.

54. Kapsogeorgou, E.K., Moutsopoulos, H.M., and Manoussakis, M.N. (2001). Functional expression of a costimulatory B7.2 (CD86) protein on human salivary gland epithelial cells that interacts with the CD28 receptor, but has reduced binding to CTLA4. J. Immunol. *166*, 3107–3113. https://doi.org/10.4049/jimmunol.166.5.3107.

55. Ping, L., Ogawa, N., and Sugai, S. (2005). Novel role of CD40 in Fas-dependent apoptosis of cultured salivary epithelial cells from patients with Sjögren's syndrome. Arthritis Rheum. *52*, 573–581. https://doi.org/10.1002/art.20789.

56. Gong, Y.-Z., Nititham, J., Taylor, K., Miceli-Richard, C., Sordet, C., Wachsmann, D., Bahram, S., Georgel, P., Criswell, L.A., Sibilia, J., et al. (2014). Differentiation of follicular helper T cells by salivary gland epithelial cells in primary Sjögren's syndrome. J. Autoimmun. *51*, 57–66. https://doi.org/10.1016/j.jaut.2013.11.003.

57. Fye, K.H., Terasaki, P.I., Michalski, J.P., Daniels, T.E., Opelz, G., and Talal, N. (1978). Relationship of Hla-Dw3 and Hla-B8 to Sjögren's syndrome. Arthritis Rheum. *21*, 337–342. https://doi.org/10.1002/art.1780210308.

58. Chused, T.M., Kassan, S.S., Opelz, G., Moutsopoulos, H.M., and Terasaki, P.I. (1977). Sjögren's syndrome associated with HLA-Dw3. N. Engl. J. Med. *296*, 895–897. https://doi.org/10.1056/NEJM197704212961602.

59. Harley, J.B., Reichlin, M., Arnett, F.C., Alexander, E.L., Bias, W.B., and Provost, T.T. (1986). Gene interaction at HLA-DQ enhances autoantibody production in primary Sjögren's syndrome. Science *232*, 1145–1147. https://doi.org/10.1126/science.3458307.

60. Lessard, C.J., Li, H., Adrianto, I., Ice, J.A., Rasmussen, A., Grundahl, K.M., Kelly, J.A., Dozmorov, M.G., Miceli-Richard, C., Bowman, S., et al. (2013). Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren's syndrome. Nat. Genet. *45*, 1284–1292. https://doi.org/10.1038/ng.2792.

61. Khatri, B., Tessneer, K.L., Rasmussen, A., Aghakhanian, F., Reksten, T.R., Adler, A., Alevizos, I., Anaya, J.-M., Aqrawi, L.A., Baecklund, E., et al. (2022). Genome-wide association study identifies Sjögren's risk loci with functional implications in immune and glandular cells. Nat. Commun. *13*, 4287.

62. Elliott, J.I., and Altmann, D.M. (1995). Dual T cell receptor alpha chain T cells in autoimmunity. J. Exp. Med. *182*, 953–959. https://doi.org/10.1084/jem.182.4.953.

63. Hinz, T., Weidmann, E., and Kabelitz, D. (2001). Dual TCR-expressing T lymphocytes in health and disease. Int. Arch. Allergy Immunol. *125*, 16–20. https://doi.org/10.1159/000053792.

64. Schuldt, N.J., and Binstadt, B.A. (2019). Dual TCR T cells: identity crisis or multitaskers? J. Immunol. *202*, 637–644. https://doi.org/10.4049/jimmunol.1800904.

65. Davies, J.M. (1997). Molecular mimicry: can epitope mimicry induce autoimmune disease? Immunol. Cell Biol. *75*, 113–126. https://doi.org/10.1038/icb.1997.16.

66. Cusick, M.F., Libbey, J.E., and Fujinami, R.S. (2012). Molecular mimicry as a mechanism of autoimmune disease. Clin. Rev. Allergy Immunol. *42*, 102–111. https://doi.org/10.1007/s12016-011-8294-7.

67. Leech, S. (1998). Molecular mimicry in autoimmune disease. Arch. Dis. Child. *79*, 448–451. https://doi.org/10.1136/adc.79.5.448.

68. Zhou, S., Cai, Y., Wang, M., Yang, W.-D., and Duan, N. (2018). Oral microbial flora of patients with Sicca syndrome. Mol. Med. Rep. *18*, 4895–4903. https://doi.org/10.3892/mmr.2018.9520.

69. Verstappen, G.M., Kroese, F.G.M., and Bootsma, H. (2021). T cells in primary Sjögren's syndrome: targets for early intervention. Rheumatology (Oxford) *60*, 3088–3098. https://doi.org/10.1093/rheumatology/kez004.

70. Horvath, S., Nazmul-Hossain, A.N.M., Pollard, R.P.E., Kroese, F.G.M., Vissink, A., Kallenberg, C.G.M., Spijkervet, F.K.L., Bootsma, H., Michie, S.A., Gorr, S.U., et al. (2012). Systems analysis of primary Sjögren's syndrome pathogenesis in salivary glands identifies shared pathways in human and a mouse model. Arthritis Res. Ther. *14*, R238. https://doi.org/10.1186/ar4081.

71. Huang, N., Pérez, P., Kato, T., Mikami, Y., Okuda, K., Gilmore, R.C., Conde, C.D., Gasmi, B., Stein, S., Beach, M., et al. (2021). SARS-CoV-2 infection of the oral cavity and saliva. Nat. Med. *27*, 892–903. https://doi.org/10.1038/s41591-021-01296-8.

72. Verstappen, G.M., Pringle, S., Bootsma, H., and Kroese, F.G.M. (2021). Epithelial–immune cell interplay in primary Sjögren syndrome salivary gland pathogenesis. Nat. Rev. Rheumatol. *17*, 333–348. https://doi.org/10.1038/s41584-021-00605-2.

73. Cai, Y., Lv, D., Li, D., Yin, J., Ma, Y., Luo, Y., Fu, L., Ding, N., Li, Y., Pan, Z., et al. (2023). IEAtlas: an atlas of HLA-presented immune epitopes derived from non-coding regions. Nucleic Acids Res. *51*, D409–D417. https://doi.org/10.1093/nar/gkac776.

74. Marcu, A., Bichmann, L., Kuchenbecker, L., Kowalewski, D.J., Freudenmann, L.K., Backert, L., Mühlenbruch, L., Szolek, A., Lübke, M., Wagner, P., et al. (2021). HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. J. Immunother. Cancer *9*, e002071. https://doi.org/10.1136/jitc-2020-002071.

75. Lee, K.H., Wucherpfennig, K.W., and Wiley, D.C. (2001). Structure of a human insulin peptide–HLA-DQ8 complex and susceptibility to type 1 diabetes. Nat. Immunol. *2*, 501–507. https://doi.org/10.1038/88694.

76. Noble, J.A., Valdes, A.M., Cook, M., Klitz, W., Thomson, G., and Erlich, H.A. (1996). The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. Am. J. Hum. Genet. *59*, 1134–1148.

77. Hu, X., Deutsch, A.J., Lenz, T.L., Onengut-Gumuscu, S., Han, B., Chen, W.-M., Howson, J.M.M., Todd, J.A., de Bakker, P.I.W., Rich, S.S., and Raychaudhuri, S. (2015). Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. Nat. Genet. *47*, 898–905. https://doi.org/10.1038/ng.3353.

78. Raychaudhuri, S., Sandor, C., Stahl, E.A., Freudenberg, J., Lee, H.-S., Jia, X., Alfredsson, L., Padyukov, L., Klareskog, L., Worthington, J., et al. (2012). Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat. Genet. *44*, 291–296. https://doi.org/10.1038/ng.1076.

79. Taylor, K.E., Wong, Q., Levine, D.M., McHugh, C., Laurie, C., Doheny, K., Lam, M.Y., Baer, A.N., Challacombe, S., Lanfranchi, H., et al. (2017). Genome-wide association analysis reveals genetic heterogeneity of Sjögren's syndrome according to ancestry. Arthritis Rheumatol. *69*, 1294–1305. https://doi.org/10.1002/art.40040.

80. Li, Y., Zhang, K., Chen, H., Sun, F., Xu, J., Wu, Z., Li, P., Zhang, L., Du, Y., Luan, H., et al. (2013). A genome-wide association study in Han Chinese identifies a susceptibility locus for primary Sjögren's syndrome at 7q11.23. Nat. Genet. *45*, 1361–1365. https://doi.org/10.1038/ng.2779.

81. Kamitaki, N., Sekar, A., Handsaker, R.E., de Rivera, H., Tooley, K., Morris, D.L., Taylor, K.E., Whelan, C.W., Tombleson, P., Loohuis, L.M.O., et al. (2020). Complement genes contribute sex-biased vulnerability in diverse disorders. Nature *582*, 577–581. https://doi.org/10.1038/s41586-020-2277-x.

82. Lundtoft, C., Pucholt, P., Martin, M., Bianchi, M., Lundström, E., Eloranta, M.-L., Sandling, J.K., Sjöwall, C., Jönsen, A., Gunnarsson, I., et al. (2022). Complement C4 copy number variation is linked to SSA/Ro and SSB/la autoantibodies in systemic inflammatory autoimmune diseases. Arthritis Rheumatol. *74*, 1440–1450. https://doi.org/10.1002/art.42122.

83. Lundtoft, C., Sjöwall, C., Rantapää-Dahlqvist, S., Bengtsson, A.A., Jönsen, A., Pucholt, P., Wu, Y.L., Lundström, E., Eloranta, M.-L., Gunnarsson, I., et al. (2022). Strong association of combined genetic deficiencies in the classical complement pathway with risk of systemic lupus erythematosus and primary Sjögren's syndrome. Arthritis Rheumatol. *74*, 1842–1850. https://doi.org/10.1002/art.42270.

84. Briggs, D.C., Welsh, K., Pereira, R.S., and Black, C.M. (1986). A strong association between null alleles at the C4A locus in the major histocompatibility complex and systemic sclerosis. Arthritis Rheum. *29*, 1274–1277. https://doi.org/10.1002/art.1780291014.

85. Kumar, A., Kumar, P., and Schur, P.H. (1991). DR3 and nonDR3 associated complement component C4A deficiency in systemic lupus erythematosus. Clin. Immunol. Immunopathol. *60*, 55–64. https://doi.org/10.1016/0090-1229(91)90111-M.

86. Langefeld, C.D., Ainsworth, H.C., Cunninghame Graham, D.S., Kelly, J.A., Comeau, M.E., Marion, M.C., Howard, T.D., Ramos, P.S., Croker, J.A., Morris, D.L., et al. (2017). Transancestral mapping and genetic load in systemic lupus erythematosus. Nat. Commun. *8*, 16021. https://doi.org/10.1038/ncomms16021.

87. Steinsson, K., Jónsdóttir, S., Arason, G.J., Kristjánsdóttir, H., Fossdal, R., Skaftadóttir, I., and Árnason, A. (1998). A study of the association of HLA DR, DQ, and complement C4 alleles with systemic lupus erythematosus in Iceland. Ann. Rheum. Dis. *57*, 503–505. https://doi.org/10.1136/ard.57.8.503.

88. Kristjánsdóttir, H., Bjarnadóttir, K., Hjálmarsdóttir, I.B., Gröndal, G., Arnason, A., and Steinsson, K. (2000). A study of C4AQ0 and MHC haplotypes in Icelandic multicase families with systemic lupus erythematosus. J. Rheumatol. *27*, 2590–2596.

89. Carroll, M.C. (2004). A protective role for innate immunity in systemic lupus erythematosus. Nat. Rev. Immunol. *4*, 825–831. https://doi.org/10.1038/nri1456.

90. White, S., and Rosen, A. (2003). Apoptosis in systemic lupus erythematosus. Curr. Opin. Rheumatol. *15*, 557–562. https://doi.org/10.1097/00002281-200309000-00006.

91. Atkinson, J.P. (1989). Complement deficiency: predisposing factor to autoimmune syndromes. Clin. Exp. Rheumatol. *7* (*Suppl 3*), S95–S101.

92. Simoni, L., Presumey, J., van der Poel, C.E., Castrillon, C., Chang, S.E., Utz, P.J., and Carroll, M.C. (2020). Complement C4A regulates autoreactive B cells in murine lupus. Cell Rep. *33*, 108330. https://doi.org/10.1016/j.celrep.2020.108330.

93. Majumder, P., Gomez, J.A., and Boss, J.M. (2006). The human major histocompatibility complex class II HLA-DRB1 and HLA-DQA1 genes are separated by a CTCF-binding enhancer-blocking element. J. Biol. Chem. *281*, 18435–18443. https://doi.org/10.1074/jbc.M601298200.

94. Raj, P., Rai, E., Song, R., Khan, S., Wakeland, B.E., Viswanathan, K., Arana, C., Liang, C., Zhang, B., Dozmorov, I., et al. (2016). Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity. eLife *5*, e12089. https://doi.org/10.7554/eLife.12089.

95. Vyse, T.J., and Tsao, B.P. (2022). Complement C4, the major histocompatibility complex, and autoimmunity. Arthritis Rheumatol. *74*, 1318–1320. https://doi.org/10.1002/art.42119.

96. Kawasaki, A., Kusumawati, P.A., Kawamura, Y., Kondo, Y., Kusaoi, M., Amano, H., Kusanagi, Y., Itoh, K., Fujimoto, T., Tamura, N., et al. (2023). Genetic dissection of *HLA-DRB1*15:01* and XL9 region variants in Japanese patients with systemic lupus erythematosus: primary role for *HLA-DRB1*15:01*. RMD Open *9*, e003214. https://doi.org/10.1136/rmdopen-2023-003214.

97. Harley, I.T.W., and Sawalha, A.H. (2022). Systemic lupus erythematosus as a genetic disease. Clin. Immunol. *236*, 108953. https://doi.org/10.1016/j.clim.2022.108953.

98. Yin, H., Pranzatelli, T.J.F., French, B.N., Zhang, N., Warner, B.M., and Chiorini, J.A.; NIDCD/NIDCR Genomics and Computational Biology Core (2021). Sclerosing Sialadenitis Is Associated With Salivary Gland Hypofunction and a Unique Gene Expression Profile in Sjögren's Syndrome. Front. Immunol. *12*, 699722. https://doi.org/10.3389/fimmu.2021.699722.

99. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

100. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal *17*, 10–12.

101. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

102. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. *33*, 495–502. https://doi.org/10.1038/nbt.3192.

103. Warner, B.M., Baer, A.N., Lipson, E.J., Allen, C., Hinrichs, C., Rajan, A., Pelayo, E., Beach, M., Gulley, J.L., Madan, R.A., et al. (2019). Sicca syndrome associated with immune checkpoint inhibitor therapy. Oncologist *24*, 1259–1269. https://doi.org/10.1634/theoncologist.2018-0823.

104. Shiboski, C.H., Shiboski, S.C., Seror, R., Criswell, L.A., Labetoulle, M., Lietman, T.M., Rasmussen, A., Scofield, H., Vitali, C., Bowman, S.J., et al. (2017). 2016 American College of Rheumatology/European League Against Rheumatism classification criteria for primary Sjögren's syndrome: A consensus and data-driven methodology involving three international patient cohorts. Arthritis Rheumatol. *69*, 35–45. https://doi.org/10.1002/art.39859.

105. Shiboski, C.H., Shiboski, S.C., Seror, R., Criswell, L.A., Labetoulle, M., Lietman, T.M., Rasmussen, A., Scofield, H., Vitali, C., Bowman, S.J., et al. (2017). 2016 American College of Rheumatology/European League Against Rheumatism classification criteria for primary Sjögren's syndrome: A consensus and data-driven methodology involving three international patient cohorts. Ann. Rheum. Dis. *76*, 9–16. https://doi.org/10.1136/annrheumdis-2016-210571.

106. Huang, L., Holtzinger, A., Jagan, I., BeGora, M., Lohse, I., Ngai, N., Nostro, C., Wang, R., Muthuswamy, L.B., Crawford, H.C., et al. (2015). Ductal pancreatic cancer modeling and drug screening using human pluripotent stem cell- and patient-derived tumor organoids. Nat. Med. *21*, 1364–1371. https://doi.org/10.1038/nm.3973.

107. Huang, H., Wang, C., Rubelt, F., Scriba, T.J., and Davis, M.M. (2020). Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. Nat. Biotechnol. *38*, 1194–1202. https://doi.org/10.1038/s41587-020-0505-4.

108. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420. https://doi.org/10.1038/nbt.4096.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD3 antibody (OKT3)	Thermo Fisher Scientific	Cat: 14-0037-82; RRID:AB_467057
anti-human TCR α/β Antibody	Biolegend	Cat: 306711; RRID:AB_528967
Alexa Fluor 647 anti-human CD8 Antibody	Biolegend	Cat: 344725;RRID:AB_2563452
Brilliant Violet 421™ anti-human CD80 Antibody	Biolegend	Cat: 305222; RRID:AB_10899567
Alexa Fluor® 647 anti-human CD19 Antibody	Biolegend	Cat: 363039; RRID:AB_2750324
Biotin anti-mouse TCR β chain Antibody	Biolegend	Cat: 109204; RRID:AB_313426
APC anti-human CD8 Antibody	Biolegend	Cat: 344722; RRID:AB_2075390
Alexa Fluor® 647 anti-human HLA-DR, DP, DQ Antibody	Biolegend	Cat: 361704; RRID:AB_2563169
FITC anti-human HLA-DQ Antibody	Biolegend	Cat: 361503; RRID:AB_2563223
Alexa Fluor® 647 anti-human CD69 Antibody	Biolegend	Cat: 310918; RRID:AB_528871
APC anti-human CD58 (LFA-3) Antibody	Biolegend	Cat: 330917; RRID:AB_2650886
APC anti-human CD155 (PVR) Antibody	Biolegend	Cat: 337617; RRID:AB_2565815
Alexa Fluor® 647 anti-human CD54 Antibody	Biolegend	Cat: 353113; RRID:AB_2715941
FITC anti-mouse TCR β chain Antibody	Biolegend	Cat: 109206; RRID:AB_313428
Alexa Fluor® 647 anti-human CD3 Antibody	Biolegend	Cat: 317312; RRID:AB_571883
Bacterial and virus strains		
One Shot® Stbl3™ Chemically Competent E. coli	Thermo Fisher Scientific	Cat: C7373-03
ElectroMAX™ DH10B™ Cells	Thermo Fisher Scientific	Cat: 18290015
Chemicals, peptides, and recombinant proteins		
Interleukin-2, human (hIL-2) 25ug	Sigma (Millipore)	Cat: HIL2-RO
Gateway™ LR Clonase™ II Enzyme mix	Thermo Fisher Scientific	Cat:11791100
Gateway™ BP Clonase™ II Enzyme mix	Thermo Fisher Scientific	Cat:11789100
Q5® High-Fidelity 2X Master Mix	New England Biolabs	Cat:M0492L
Lenti-X Concentrator	Takara Bio	631232
Puromycin Dihydrochloride	Thermo Fisher Scientific	Cat:A1113803
Polybrene	Thermo Fisher Scientific	Cat:TR-1003-G
SuperScript™ IV First-Strand Synthesis System with ezDNase™ Enzyme	Thermo Fisher Scientific	Cat:18091150
Ambion™ RNase H, from E. coli, 10 U/μL	Thermo Fisher Scientific	Cat:AM2293
Benzonase Nuclease HC, Purity > 99%	Thermo Fisher Scientific	Cat:71206–3
Nourseothricin Sulfate (Streptothricin Sulfate)	Gold Biotechnology	Cat:N-500-1
NEBuilder® HiFi DNA Assembly Master Mix	New England Biolabs	Cat:E2621L
Dynabeads® Human T-Activator CD3/CD28	Thermo Fisher Scientific	Cat:11132D
jetPRIME	Polyplus	Cat:114-07
Critical commercial assays		
GeneJET Genomic DNA Purification Kit	Thermo Fisher Scientific	Cat:K0721
RNeasy Plus Mini Kit	Qiagen	Cat:74134
Anti-Biotin Microbeads Miltenyi	Miltenyi Biotec	Cat:130-090-485
RosetteSep™ Human CD8+ T Cell Enrichment Cocktail	Stem Cell Technologies	Cat:15023

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Singel-Cell-RNA-Seq	Yin et al.[98] Huang et al.[71] This Study	[dbGAP]: [Phs002446.v1.p]
Experimental models: cell lines		
Lenti-X™ 293T Cell Line	Takara Bio	632180
eHEK-293-II	This Study	N/A
HLA-II expressing eHEK-293-II	This Study	N/A
Jurkat, Clone E6-1	ATCC	Cat:TIB-152
SKW-3 Cell line	Gift from Julia Etchin at Dana Farber Cancer Institue	N/A
Oligonucleotides		
See Table S7	This Study	N/A
Recombinant DNA		
pHAGE-CMV:DEST-PGK: $ICAD^{CR}$.	This Study	N/A
pHAGE-CMV:$IFP^{GzB\text{-}Hi}$-PGK: $ICAD^{CR}$	This Study	N/A
pHAGE-TREX-CIITA-PGK-Neomycin	This Study	N/A
pHAGE-EF1L-anti-CD3 scFv-PGK-Puromycin	This Study	N/A
pHAGE-EF1L-anti-H57-597-scFv-PGK-Puromycin	This Study	N/A
pHAGE-TREX-DEST-PGK- Blasticidin	This Study	N/A
pHAGE-TREX-DEST-PGK-Hygromycin	This Study	N/A
pHAGE-TREX-DEST-PGK-Nourseothricin	This Study	N/A
pHAGE-EF1a-DEST-PGK-floxed-mCD19	This Study	N/A
pHAGE-EF1:DEST-PGK: CD4	This Study	N/A
pHAGE-CMV-CD74_V3-DEST-Puromycin	This Study	N/A
Human Peptidome in pHAGE-CMV-CD74_V3-DEST	This Study	N/A
Human Virome in pHAGE-CMV-CD74_V3-DEST	This Study	N/A
Tiling Library in pHAGE-CMV-CD74_V3-DEST	This Study	N/A
HIV Library in pHAGE-CMV-CD74_V3-DEST	This Study	N/A
Mutagenesis Library in pHAGE-CMV-CD74_V3-DEST	This Study	N/A
Software and algorithms		
Bowtie	Langmead et al., 2009[99]	http://bowtie-bio.sourceforge.net/index.shtml
Cutadapt	Martin, 2011[100]	http://cutadapt.readthedocs.io/en/stable/
Cell Ranger v3.0.1	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger
Scanpy	Wolf, 2018[101]	https://github.com/scverse/scanpy
Seurat	Sajita, 2015[102]	https://github.com/satijalab/seurat
Python v3.7	Python Software Foundation	https://www.python.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Stephen Elledge (selledge@genetics.med.harvard.edu).

Materials availability

Information and requests for resources and reagents may be directed to the lead contact.

Data and code availability

Single-cell RNA-seq data have been deposited in the NIH Database of Genotypes and Phenotypes (dbGaP). The accession information is listed in the key resources table. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human Participants

Pancreatic cancer subject recruitment

Peripheral blood and tumor tissue samples were obtained from individuals diagnosed with metastatic pancreatic cancer. Specifically, the TCRs examined in this study and the corresponding tumor organoids were derived from patient 38 (p38) (52 year old male, self-reported Caucasian) diagnosed with stage IV pancreatic ductal adenocarcinoma. The study protocol was approved by the Dana-Farber Cancer Institute Institutional Review Board (IRB), Boston, Massachusetts, and was conducted in accordance with the principles of the Declaration of Helsinki. The samples were collected between 2018-2020.

SjD subject recruitment for single-cell RNAseq

All subjects were seen at the NIH Clinical Center Dental Clinic and provided informed consent to NIH Central IRB approved protocol, 15-D-0051 (NCT02327884, *PI: Warner*) prior to any study procedures. Subjects were evaluated comprehensively following standardized clinical protocols by a multidisciplinary clinical research team, including oral medicine/clinical oral pathology, rheumatology, and ophthalmology, as previously reported.[103] Subjects were classified according to the 2016 SjD classification criteria.[104,105] 6 subjects who did not fulfill SjD 2016 classification criteria(aged 40 to 56, comprising 4 females and 2 male all of White ethnicity) and were otherwise healthy, and 7 SjD (aged 43 to 75, comprising 6 females and 1 male, with the ethnic distribution of 4 Black, 2 Hispanic, and 1 White) subjects who fulfilled American College of Rheumatology Classification Criteria[104,105] were ultimately chosen for scRNA-seq.

SjD subject recruitment for 10X Visium

Labial minor salivary gland biopsies were performed on two individuals (Patient 1: 47 years old, male, white, non-Hispanic, Patient 2 Adult, 56 years old, female, female, white, non-Hispanic.) attending the Oklahoma Sjögren's Research Clinic who met the 2016 SjD classification for SjD.[104,105] This study was approved by the Oklahoma Medical Research Foundation Institutional Review Board, IRB 06-12, IRB 07-12, and IRB 11-18. The socioeconomic status of the participants was not determined for this study.

Cell lines

HEK293T cell lines (Takara Biosciences) were cultured in DMEM GlutaMAX™ (Life Technologies) media.

Primary T cells

Primary T-cells were cultured in either RPMI (Life Technologies) or ImmunoCult-XF (Stemcell Technologies) plus 50U/ml IL-2 (Sigma). DMEM GlutaMAX™ and RPMI media were supplemented with 10% (v/v) FBS (Hyclone) and 1% (v/v) Penicillin/Streptomycin (Life Technologies).

METHOD DETAILS

T-cell extraction and expansion

Apheresis blood samples (Brigham and Women's Hospital) were used for T-cell extraction using the RosetteSep CD8 purification kit (StemCell). The purity of $CD8^+$ extraction was assessed by CD8 (Biolegend) and CD4(Biolegend) antibody staining and FACS. Apheresis blood samples were also used for primary blood mononuclear cells (PBMCs) extraction. PBMCs were purified using a Ficoll (Cytiva Life Sciences) gradient. PBMC cells were subsequently irradiated by 75 Gy ionizing radiation. For expansion, 1E6 centuar T-cells were grown in the presence of irradiated PBMCs(20E6) plus 50U/ml IL-2 (Sigma) and 0.1μg/ml anti-CD3 antibody (Life technologies).

T-cell transduction

1E6 primary $CD8^+$ cells were plated per well in a 24-well plate and stimulated with a 1:1 ratio of Dynabeads Human T-Activator CD3/CD28 (Life Technologies). 24 h following activation, T-cells were transduced with 150ul of concentrated lentivirus expressing centaur TCRs. 48 h following Dynabeads activation cells, beads were removed. 72 h following transduction, transduced cells were sorted using a biotinylated mouse TCR (mTCR) antibody (Clone: H57-597) and MACS® Columns (Miltenyi Biotec). The purification efficiency of transduced T-cells was assessed using the mTCR (Clone: H57-597) and CD4 (Clone: OCT4) antibodies.

Lentiviral production

Lentivirus was produced through co-transfection of lentiviral vectors and packing plasmids (Tat, Rev, Gag-Pol, and VSV-G) into HEK293T-cells (Takara Biosciences). All transfection into HEK293T was done using jetPRIME® (Polyplus). Lentiviral supernatants were collected 48 h post-transfection; for HEK293T transduction, lentivirus was added to cells in the presence of 8 ug/ml Polybrene.

Selection of transduced target cells was performed using Hygromycin (200 μg/ml, Puromycin (2 μg/ml), Blasticidin (10 μg/ml), or Nourseothricin (400 μg/ml). Primary T-cells that were transduced with lentivirus in the absence of Polyberne.

Knockout of endogenous B2M gene

To knock out the endogenous B2M in HEK293T-cells, CRISPR/Cas9 ribonucleoprotein (*RNP*) complexes were used. In brief, three synthetic gRNAs (TrueGuideTM) were mixed individually with Cas9 protein (TrueCutTM Cas9 Protein v2) and delivered via electroporation (NeonTM Transfection) into HEK293T-cells. MHC-negative cells were sorted using the Pan-MHC (Biolegend, Clone W6/32) and B2M (Biolegend, Clone 2M2) antibodies. To achieve a pure B2M population, cells were FACS sorted three times.

B2M-sgRNA-1: AGTCACATGGTTCACACGGC
B2M-sgRNA-2: TCACGTCATCCAGCAGAGAA
B2M-sgRNA-3: TGAAGATGCCGCATTTGGAT

Mutation of endogenous HLA-II genes

To mutate out the endogenous HEK293T HLA-II alleles (DPB1*04:01:01, DQA1*01:02:01, DQB1*06:02:01, DRB1*15:01:01) that are expressed following CIITA transduction CRISPR/Cas9 ribonucleoprotein (*RNP*) complexes were used. In brief, gRNAs targeting either HLA-DRB, HLA-DP, or HLA-DQ (IDT) were mixed individually with Cas9 protein (TrueCutTM Cas9 Protein v2) and delivered sequentially via electroporation (NeonTM Transfection) into CIITA transduced HEK293T-cells. HLA-II negative cells were sorted sequentially using HLA-DR (L243), HLA-DP (B7/21), or HLA-DQ (SPVL3) antibodies.

HLA-DRB-sgRNA-1: TTTGCCATGTAGGAACCTCC AGG
HLA-DQB-sgRNA-1: CAGACACAACTACGGGGTTG TGG
HLA-DQA-sgRNA-1: ATTTGGAGGTTTTGACCCGC AGG
HLA-DPB-sgRNA-1: CACGTGACGGATTTCTACCC AGG
HLA-DPA-sgRNA-1: CGTCACGTGGCTGTGCAACG GGG

Knockout of endogenous TCRβ gene

To knock out the endogenous TCRb in primary $CD8^+$ cells, CRISPR/Cas9 ribonucleoprotein (*RNP*) complexes were used. In brief, primary $CD8^+$ were initially activated using magnetic CD3/CD28 beads (Invitrogen Dynabeads). Two synthetic gRNAs (TrueGuideTM) were mixed individually with Cas9 protein (TrueCutTM Cas9 Protein v2) and delivered via electroporation (NeonTM Transfection) into primary $CD8^+$ cells. Cells null for TCR β (Clone: IP26) were FACS sorted and subsequently transduced with lentivirus to constitutively express centaur TCRs. Cells that stained positive with mTCR (Clone: H57-597) and CD4 (Clone: OCT4) and negative TCR (Clone: IP26) were sorted and used for downstream studies.

TRBC-sgRNA-1: GCAGTATCTGGAGTCATTGA
TRBC-sgRNA-2: GGAGAATGACGAGTGGACCC

Generation of CIITA and $IFP^{GzB\text{-}Hi}$ $ICAD^{CR}$ constructs

The $ICAD^{CR}$ cassette was cloned downstream of the PGK promoter to create pHAGE-CMV:DEST-PGK: $ICAD^{CR}$. Using site-directed mutagenesis, the GzB cleavage sequence (VGPD$^{\downarrow}$FGR) of IFP^{GzB} was mutated to VGPD$^{\downarrow}$SGR to generate the $IFP^{GzB\text{-}Hi}$ reporter. The $IFP^{GzB\text{-}Hi}$ reporter was cloned by Gateway cloning (Thermo Fisher) into pHAGE-CMV:DEST-PGK: $ICAD^{CR}$ to generate pHAGE-CMV:$IFP^{GzB\text{-}Hi}$-PGK: $ICAD^{CR}$. CIITA (Uniprot: P33076) cDNA was cloned into pHAGE-TREX-Dest-PGK-Neomycin expression construct using Gateway cloning.

Generation of membrane bound anti-CD3 scFv

The heavy and light chains of a monoclonal antibody clone: OKT3, were assembled as a single construct and separated by a flexible linker composed of Glycine and Serine (VH-linker-VL). For membrane targeting, we added an N-terminal CD8α signal peptide (Uniprot: P01732.1: 1-21aa), followed by a myc-tag (EQKLISEEDL) to assess membrane surface expression. To tether the scFv to the cellular membrane, we added a C-terminal PDGFR transmembrane peptide (Uniprot: P09619.1 512-561aa). The final construct was synthesized as a gBlocks (IDT) and cloned into the pDONR221 vector (Invitrogen) using Gateway cloning. The membrane-bound anti-CD3 scFv was ultimately cloned into pHAGE-EF1L-DEST-PGK-Puromycin. The construct was introduced in eHEK-293-II cells expressing either the $IFP^{GzB(low)}$ or $IFP^{GzB(Hi)}$ genes via lentiviral transduction.

Generation of membrane-bound anti-centaur scFv

The heavy and light chains of a monoclonal antibody clone: H57-597, were assembled as a single construct and separated by a flexible linker composed of Glycine and Serine (VL-linker-VH). For membrane targeting, we added an N-terminal CD8α signal peptide

(Uniprot: P01732.1: 1-21aa), followed by a myc-tag (EQKLISEEDL) to assess membrane surface expression. To tether the scFv to the cellular membrane, we added a C-terminal PDGFR transmembrane peptide (Uniprot: P09619.1 512-561aa). The final construct was synthesized as a gBlocks (IDT) and cloned into the pDONR221 vector (Invitrogen) using Gateway cloning. The membrane-bound scFv was ultimately cloned into pHAGE-EF1L-DEST-PGK-Puromycin. The construct was introduced in eHEK-293-II cells via lentiviral transduction.

HLA allele design and cloning

All protein sequences for HLA-II alleles were retrieved from IPD-IMGT/HLA (EMBL-EBI). Both HLA-II α and β chains were assembled as a single construct, where the HLA-β and HLAα chains were separated by a 2A self-cleaving peptide (β-P2A-α). Sequences were reverse translated, synthesized as gBlocks (IDT), and cloned into the pDONR221 vector (Invitrogen) using Gateway cloning. All HLA alleles were inserted in one of 4 below lentiviral destination vectors via Gateway cloning.

pHAGE-TREX-DEST-PGK-Blasticidin
pHAGE-TREX-DEST-PGK-Hygromycin
pHAGE-TREX-DEST-PGK-Nourseothricin
pHAGE-EF1a-DEST-PGK-floxed-mCD19

Combinations of alleles cloned into expression vectors encoding different selection markers were introduced via lentiviral transduction into eHEK-293-II cell lines. Transduced cells were selected based on the presence of a drug selection marker or cell surface markers.

CD4$^+$ TCR design and cloning

The human CD4 gene (P01730) was cloned downstream of the PGK promoter to create the pHAGE-EF1:DEST-PGK:CD4 lentiviral vector. All T cell receptors (TCRs) were engineered as hybrid TCRs, in which the entire human α and β V(D)J regions were preserved and grafted on the mutant mouse α and β constant regions, respectively. The mouse transmembrane and cytoplasmic parts were mutated to concurrently 1) remove a degron within the TCR alpha to increase TCR abundance, 2) enhance TCR association with CD3 for more efficient T-cell signaling, and 3) improve specific pairing of the exogenous TCRs, so mispairing with endogenous TCRs is reduced. Hybrid α and β TCR sequences were then combined into a single fragment separated by a 2A self-cleaving peptides sequence (TCRβ-P2A-TCRα). The hybrid TCRs were codon optimized and subsequently cloned into a pDONR221 entry clone and then transferred into the pHAGE-EF1:DEST-PGK:CD4 vector using Gateway cloning.

Validation experiments

To validate hits from our genome-wide screens with purified peptides, eHEK-293-II target cells were plated at ~2E5 in a 48-well plate in DMEM GlutaMAX™. 16 h later, cells were pulsed with the 1uM of purified peptide for 30 min in a 37C incubator. Centaur T-cells were added to target cells in RPMI plus IL-2. The co-cultured cells were incubated for 8-12h and analyzed for GzB reporter activation by FACS.

Purified peptides used
MAP3K4 Peptide 1: LRKDLEIAAEFRLSA
MAP3K4 Peptide 2: EIAAEFRLSAPVRDL
MAP3K4 Peptide 3: EIAAEFRLSA
MAP3K4 Peptide 4: LRKDLEIAAEFRLSAPVRDL
DDIAS Peptide1: SLNKFLAVLESEIAV
DDIAS Peptide2: ESLNKFLAVLESEIAV
DDIAS Peptide3: FSESLNKFLAVLESEIAV
DDIAS Peptide4: LNKFLAVLESEIAVTQADVS

For endogenously expressed antigens, peptide fragments of interest were reverse translated and synthesized as DNA fragments (gBlocks Integrated DNA Technologies), and Gateway cloned into the pHAGE-CMV-CD74_V3-DEST-Puromycin lentiviral vector. DNA constructed was packaged into lentivirus and introduced into eHEK-293-II target cells. Transduced target cells were co-cultured with centaur T-cells in RPMI plus IL-2. The co-cultured cells were incubated for 8 h and analyzed for GzB reporter activation by FACS.

Library design, cloning, and transduction

Human virome and peptidome libraries: The human peptidome Version 2 (V2) library of 259,345 fragments tiling across the entire human proteome in 90-aa fragments with 45-aa overlap. The human peptidome Version 3 (V3) library (~350,000) was designed to complement our human peptidome V2 library. This library comprises 586,167 fragments that tile across the entire human proteome

in 90-aa fragments with 67-aa overlaps. The human virome library comprised 93,904 tiles of 56-aa fragments collectively tiled across the human virome (206 viral species and over 1,000 different strains).

Human peptidome libraries (V2 and V3) and virome libraries were cloned into the pHAGE-CMV-CD74_V3-DEST-Puromycin lentiviral vector using Gateway cloning. A 1000x representation was maintained during each cloning step to provide uniform library coverage. For screens with the human peptidome V2 library, 200E6 eHEK-293-II cells were transduced with the lentivirus at an MOI of 5 (5000x library coverage). Each screen was performed using seven to eight replicates (200E6 eHEK-293-II cells, 5000x library representation per replicate) and was co-cultured with centaur T-cells for a minimum of 8 h. The human peptidome V3 library, 300E6 eHEK-293-II cells, were transduced with the lentivirus at an MOI of 5 (2500x library coverage) and selected with puromycin (1ug/ml). Each screen was performed using seven to eight replicates (300E6 eHEK-293-II cells, 2500x library representation per replicate). The human virome library was transduced into 100E6 eHEK-293-II cells at MOI of 3 (3000x library coverage). Each screen was performed using seven replicates (100E6 eHEK-293-II cells, 3000x library representation per replicate). Following the co-culture of centaur cells with target cells, $IFP^{GzB\text{-}Hi}$ positive cells (Alexa-700 channel) were sorted on a SONY MA900 FACS machine.

To generate an HIV library, we included ORFs from 10 different HIV strains as annotated in the Uniprot database. ORFs were tiled as 56 aa fragments with 28 aa overlaps between adjacent tiles. Every fragment in the library was barcoded twice using synonymous aa codons. Twist Biosciences synthesized the 2,493-fragment library.

To generate epitopes mutagenesis libraries, epitopes were encoded in the context of 56 aa fragments. For NY-ESO and HIV mutagenesis library design, each amino acid in the annotated epitope (20 aa) was mutated to the other 19 amino acids. For MAP3K4 and DDIAS, each amino acid in the 50 aa fragment was mutated to all other 19 amino acids. Each peptide was reverse translated with synonymous aa codons. Each library also included four unmutated fragments that served as positive controls. To include negative controls during TScan-II screens, multiple mutagenesis libraries were synthesized as a single oligo pool.

To generate tilling libraries for the identification of minimal epitopes, antigenic fragments of 10, 15, 18, and 20 aa in length with 1-aa overlap that spanned the entire length of antigenic segments of HIV, HPV, MAP3K4, and DDIAS were generated. The smaller fragments were incorporated in the context of 200 bp DNA fragments; immediately following the short epitope sequence, we introduced a stop codon to generate epitopes of the desired length. The libraries were synthesized by Twist Biosciences.

The HIV, mutagenesis, and tiling libraries were PCR amplified using primers with BP recombination site overhangs and cloned into pDONR221, then cloned into the pHAGE-CMV-CD74-V3-DEST-Puromycin lentiviral vector using Gateway cloning. A 1000x representation was maintained during the BP and LR cloning steps to provide uniform library coverage. For screens with these libraries, eHEK-293-II cells were transduced with the lentivirus at an MOI of 0.2. Each screen was performed using seven to six replicates (1000x library representation per replicate) and was co-cultured with centaur T-cells (E: T $\sim$0.5) for a minimum of 6h.

TScan-II NGS sequencing

Sorted $IFP^{GzB\text{-}Hi}$ positive eHEK293-II cells were mixed with 5E5 carrier Jurkat (Clone E6-1, ATCC Cat# TIB-152) before genomic DNA (gDNA) purification. gDNA was purified using the GeneJet gDNA purification kit (Thermo Fisher).

For human virome and peptidome libraries, gDNA of 100x representation of the target cells was used as input for each screen. For smaller libraries (10,000 oligos or less), gDNA from 1000X representation of target cells was used as input.

Multiplexed libraries were generated for Illumina sequencing using a 3-step PCR protocol. All PCRs were performed with 1-4μg of template gDNA using the Q5® High-Fidelity 2X Master Mix (New England Biolabs). PCR1 was performed using TScan-II PCR-1 primers that flanked the antigen cassette (TScan-II_PCR1_Forward CTGTACCAACAACAAGGTCGGC; TScan-II_PCR1_Reverse CTGTTTGACCGGTTTCGCTACC). PCR1 samples were all pooled, and 1μl was used as input for PCR2. PCR2 forward primers were designed to create a stagger sequence between the priming site and the Illumina sequence. The 8 TScan-II_PCR2_forward primers were pooled in equal ratios. PCR2 reaction was performed using the pooled forward primers (TScan-II_PCR2_Forward: tccctacacgacgctcttccgatct[]GTTTGTACAAAAAAGCAGG, represents either t, ct, gct, agct, cagct, acagct or tacagct; TScan-II_PCR2_Reverse: gtgactggagttcagacgtgtgctcttccgatctGCCTTATTCCAAGCGGCTTC). PCR was performed to PCR3 to add on the sequencing adaptors and sample-specific indices. One μl of PCR2 was used as input for PCR3 (TScan-II_PCR3_Forward: aatgatacggcgaccaccgagatctacactcttTCCCTACACGACGCTCTTCCG

; TScan-II_PCR3_Reverse: CAAGCAGAAGACGGCATACGAGAT[]GTGACTGGAGTTCAGACGTGT "" represents a 7bp unique sample index). After PCR3, samples were pooled based on their desired read count, and gel was extracted using the NucleoSpin® Gel and PCR Clean-up kit (Takara Biosciences). Samples were sequenced on the Illumina NextSeq machine.

TScan-II data analysis

The design of libraries was performed in Python using customized codes. NGS Illumina data were analyzed using Bowtie, Samtools, and Cutadapt tools on a Linux server. In brief, PCR adaptors used during the cloning of the libraries were trimmed using Cutadapt. The trimmed reads were subsequently aligned using Bowtie against our customized libraries. DataGraph, Python, and Excel were used for data analysis. The relative enrichment of each fragment in each replica was calculated by dividing the factional abundance of the reads in the sorted samples relative to the fractional abundance in the input samples. The geometrical mean across all 6-8 replicas was calculated for plotting.

Characterization of organoid-reactive T-cells

The generation and culture of the organoids was done according to previously established protocols.[30,106] The culture media was replaced every 4 days, and the organoids used in all experiments were between passages 7 and 12.PBMCs were isolated from the peripheral blood of patient 38 (pt38) using Ficoll-Paque density gradient centrifugation. The isolated PBMCs were cultured for 10 days in human T-cell medium (HTM) consisting of serum-free medium (CellGernix, 20801-0100), 10% human AB serum (Innovative Research, IPLA-SerAB-13458), human IL-2 (1000 IU/mL, Prospec, cyt-209), human IL-15 (10 ng/mL, Prospec, cyt-230), human IL-21 (10 ng/mL, Prospec, cyt-408), 1% penicillin-streptomycin (Gibco, 15140-122), 1% amphotericin B solution (Sigma, A2942), and ciprofloxacin (Fisher Scientific, 50255729).After 10 days of culture, 100,000 PBMCs were co-cultured with autologous tumor organoids in a 96-well flat bottom plate (Falcon, 353072) using HTM. Each well contained 100,000 cells of autologous tumor organoids. The co-cultures were maintained for 7 days. On day 7, the cells were transferred to a new 24-well plate (Falcon, 353047), and the PBMCs were stimulated again with autologous organoids at a ratio of 1:1 for an additional 7 days. This step resulted in the generation of organoid-primed T-cells (opT). The phenotype of opT-cells was determined using flow cytometry and CytoF analysis(pt38, JITC, 2021).

The recognition of tumor cells by opT-cells was confirmed using the M30 enzyme-linked immunosorbent assay (ELISA). The assay was performed to evaluate the reactivity of opT-cells against tumor cells. To test if the engineered-T-cells recognize tumor organoids, 0.1 million organoids per well were co-cultured with 0.1 million engineered-T-cells in a HTM in a 96-well flat bottom plate. After 24h and 48h co-culture, the supernatants were collected and tested for IFN-γ secretion by ELISA (Mabtech, 3420-1H-20). For the HLA class I and HLA class II blocking experiment, purified antibodies without azide (final 10ug/ml) were used. The antibodies used include Pan-HLA Class I antibody (clone W6/32, Biolegend, 311428), Pan-HLA Class II antibody (clone Tu39, BD Pharmingen, 555556), HLA-DR (L243), HLA-DP (B7/21), HLA-DQ (SPVL3).

The expanded CD4$^+$ T-cell clone possessed a unique TCRβ chain and two distinct TCRα(α,α') chains. To assess which TCR pairs (TCRβ-TCRα or TCRβ-TCRα') can recognize the tumor organoids, we synthesized the two TCRs (DHM6 & DHM7) using the single TCRβ chain matched with either TCRα or TCRα' chains and introduced each into primary T-cells. TCR-transduced T-cells were co-cultured with the tumor-derived organoids in the presence of HLA-I and class II antibodies. DHM7 was inactive, but the DHM6 TCR exhibited a substantial and distinguishable recognition of the corresponding patient-derived organoids (matched), as evidenced by the secretion of IFN-γ. Importantly, DHM6 did not exhibit reactivity to a pancreatic cancer organoid derived from a different patient's (unmatched) tumor sample. Using a pan HLA-II blocking antibody, we confirmed that the DHM6 recognition of the tumor organoid is HLA-II dependent Furthermore, using a set of HLA-II allele-specific (HLA-DR, DP, or DQ) blocking antibodies, we identified HLA-DP as the alleles mediating DHM6 recognition of the tumor organoid.

To sequence the TCRα and β chains, genomic DNA extracted from opT-cells was used for TCR sequencing using the Adaptive Biotechnologies platform. Deep sequencing was performed for both TCRα and β chains. The obtained sequencing data were analyzed using immunoSEQ Analyzer V.3.0. The top-ranking CDR3β sequence was found to have a frequency of 90.4%. The corresponding matched CDR3α chain was identified based on the sequencing data ratio.

Pancreatic cancer neo-antigen screen

Each oligo was designed to encode 41 amino acid fragments, encompassing the mutation flanked by 20 amino acids on either side. The WT peptide fragments and fragments encoding mutations from non-matching patients serve as negative controls in our patient-specific neo-antigen screen. All peptide fragments within the library were barcoded four times using synonymous amino acid codons. The TScan-II screen of DHM6 TCR was performed in eHEK-293II target cells expressing the patient-specific HLA-DP alleles, and the neo-antigen library.

Characterization of SjD T-cells

We profiled a large cohort of published SG-infiltrating CD4$^+$ lymphocytes from 10 SjD patients that shared the HLA-DRB1*03:01 and HLA-DQB1*02:01 alleles.[47] To allow for a more in-depth analysis, we also complemented this dataset with TCR sequences SG-infiltrating memory CD4$^+$ lymphocytes from 5 additional SjD patients. Twenty TCR pairs across five SjD patients were ultimately chosen for subsequent genome-scale TScan-II screening. Nineteen of these TCRs stratified for antigen discovery were *i)* directly sourced from SjD patient salivary gland samples, *ii*) exhibited a memory-like phenotype (CD45RA$^-$), and *iii*) were clonally expanded within the tissue. These findings are prototypic characteristics of T cell antigen recognition. A direct correlation between the degree of clonal expansion of these specific TCRs and the severity of phenotypes observed among SjD patients further supports their significance.[47] In addition to the clonally expanded T cells, one additional CD4$^+$ TCR was included based on the presence of a shared CDR3β motif identified across patients through GLIPH analysis.[107] This shared motif was found specifically among the CDR3β of CD4$^+$ T cells present in the salivary gland infiltrates of 3 SjD patients but was absent in non-SjD patients. We synthesized each of the 20 TCRs separately as a TCRβ-α pair. In the instances where a single CD4$^+$ cell expressed two TCRα chains (α or α') and a single TCRβ, we synthesized two separate TCR constructs, each with the single TCRβ chain and two different TCRα chains (TCRβ-α or TCRβ-α').

SjD TScan-II screens

To identify potential self-antigens that SjD-associated TCRs may target, we initially generated a set of twenty centaur T-cell lines by transducing each line with a unique TCR construct. Only transductions that yielded high levels of centaur TCR expression were subsequently purified using MACS® Columns for further TScan-II screens. This stringent selection process ensured that all the T-cell

lines used in the screens had an optimal expression of the transduced TCR, thus maximizing the sensitivity and accuracy of antigen discovery using the TScan-II platform. We then expressed HLA-DR3 and HLA-DQ2 alleles shared across all 5 patients into eHEK-293 target cells transduced with our HLA-II targeted human peptidome V2 library. For pooled TScan-II screens, a mixture of 5-6 centaur T-cell lines was pooled in equal ratios a, each expressing a unique TCR from different patients. Target cells expressing the human peptidome were then co-cultured with the mixture of centaur T-cells at an effector-to-target ratio (E: T) of ~1.

Salivary gland single-cell data tissue acquisition

Minor salivary glands (MSG) (N=13; n=7 SjD and n=6 NV) tissues were removed following standard methods and immediately placed in ice-cold RPMI and dissociated using the Miltenyi Multi-tissue Dissociation Kit A. Single-cell suspensions were serially filtered through 70- and 30-μm filters and rinsed with 1× Hanks' buffered salt solution. Cells were centrifuged at 300 g for 10 min at 4 °C and washed once with 1× Hanks' buffered salt solution. Cell counting and viability were determined using a Trypan blue exclusion assay. Additional glands from these patients were submitted for histopathological assessment, including focus scoring.

Salivary gland scRNA-seq

Single-cell suspensions targeting approximately 5,000-10,000 cells were loaded onto a 10x Genomics Chromium Next GEM Chip B in the NIDCR/NIDCD Genomics and Computational Biology Core Facility. After cell capture, single-cell library preparation was performed following the instructions for the 10x Chromium Next GEM Single Cell 3′ kit v3 (10x Genomics). The libraries were pooled and sequenced on four lanes of a NextSeq500 sequencer (Illumina), adopting the read configuration indicated by the manufacturer.

Salivary gland scRNA-seq data processing

Read processing was performed using the 10x Genomics workflow. Cell Ranger v3.0.1 Single-Cell Software Suite was used for demultiplexing, barcode assignment, and UMI quantification (http://software.10xgenomics.com/single-cell/overview/welcome). Sequencing reads were aligned to the hg38 reference genome (Genome Reference Consortium Human Build 38) using a pre-built annotation package from the 10x Genomics website (https://www.10xgenomics.com/). Samples were demultiplexed using the 'cell ranger mkfastq' function, and gene count matrices were generated using the 'cellranger count' function. The single-cell data were analyzed scanpy. Filtering was performed using the standard quality control steps.

Salivary gland scRNA-seq data analysis

Single-cell data sets were integrated from published data[71,98] and unpublished data generated by the Warner Lab (dbGAP: phs002446.v1.p1). Cells from minor SG from n = 6 subjects who did not fulfill ACR 2016 classification criteria and were otherwise healthy and n = 7 SjD subjects who fulfilled American College of Rheumatology Classification Criteria[104,105] were integrated into a single object containing 12 clusters.

10X Visium tissue harvest

Harvested tissue was washed with saline (Henry Schein, #8900026), blotted on dry gauze, and transferred into a cryomold half-filled with OCT compound (VWR, #25608-930), then finished with additional OCT. Cryomolds were initially frozen by placement on a tray filled with 2-methyl butane (VWR, #100504-954) on dry ice. After the initial freeze, the cryomolds were moved and stored at -80°C for long-term storage.

10X Visium slides preparation, and sequencing

Cryosectioning was performed using a CryoStar™ NX70 Cryostat (ThermoFisher #14-071-401) with an Accu-Edge Low profile microtome blade (Sakura #4689) and following 10X Genomics protocol. Optimal permeabilization of 27 minutes was established. Single sections (10μm) from four subjects were placed on gene expression slides (10X Genomics, #1000187). After cryosectioning, the cut side of the cryoblock was placed facedown into the cryoblock pre-filled with a small amount of fresh, partially frozen OCT. Using partially frozen OCT preserved the cut surface and minimized warming of the cut tissue. After tissue placement on each 10X Genomics slide, slides were placed in a slide mailer and stored inside resealable bags at -80°C for up to 4 days. Fixation and H&E staining were performed according to 10X Genomics CG000160 Rev A protocol. Gene expression (GEX) libraries were prepared using the Visium Spatial Gene Expression kit (10X Genomics, #1000187) and Dual Index Kit TT Set A (10X Genomics #1000215) according to the 10X Genomics protocol. Qubit dsDNA HS Assay kit (ThermoFisher #Q32854) and Qubit 2.0 Fluorometer (ThermoFisher #Q32866) were used for sequencing library quantifications. All amplified cDNAs, and sequencing libraries were quality-controlled using Agilent 2200 or 4200 TapeStation (Agilent #G2964AA or #G2991AA). HighSensitivity D5000 ScreenTapes/Reagents (Agilent #5067-5592/5067-5593) were used to quantify amplified cDNAs. HighSensitivity D1000 ScreenTapes/Reagents (Agilent #5067-5584/5067-5585) were used to quantify GEX libraries. Based on the tissue coverage area (range: 3 - 40 %; average: 14.5%), the sequencing depth per sample was established, and libraries pools were made targeting 50,000 read pairs per spot. Sequencing was performed on Illumina NovaSeq S4 PE150.

10X Visium data analysis

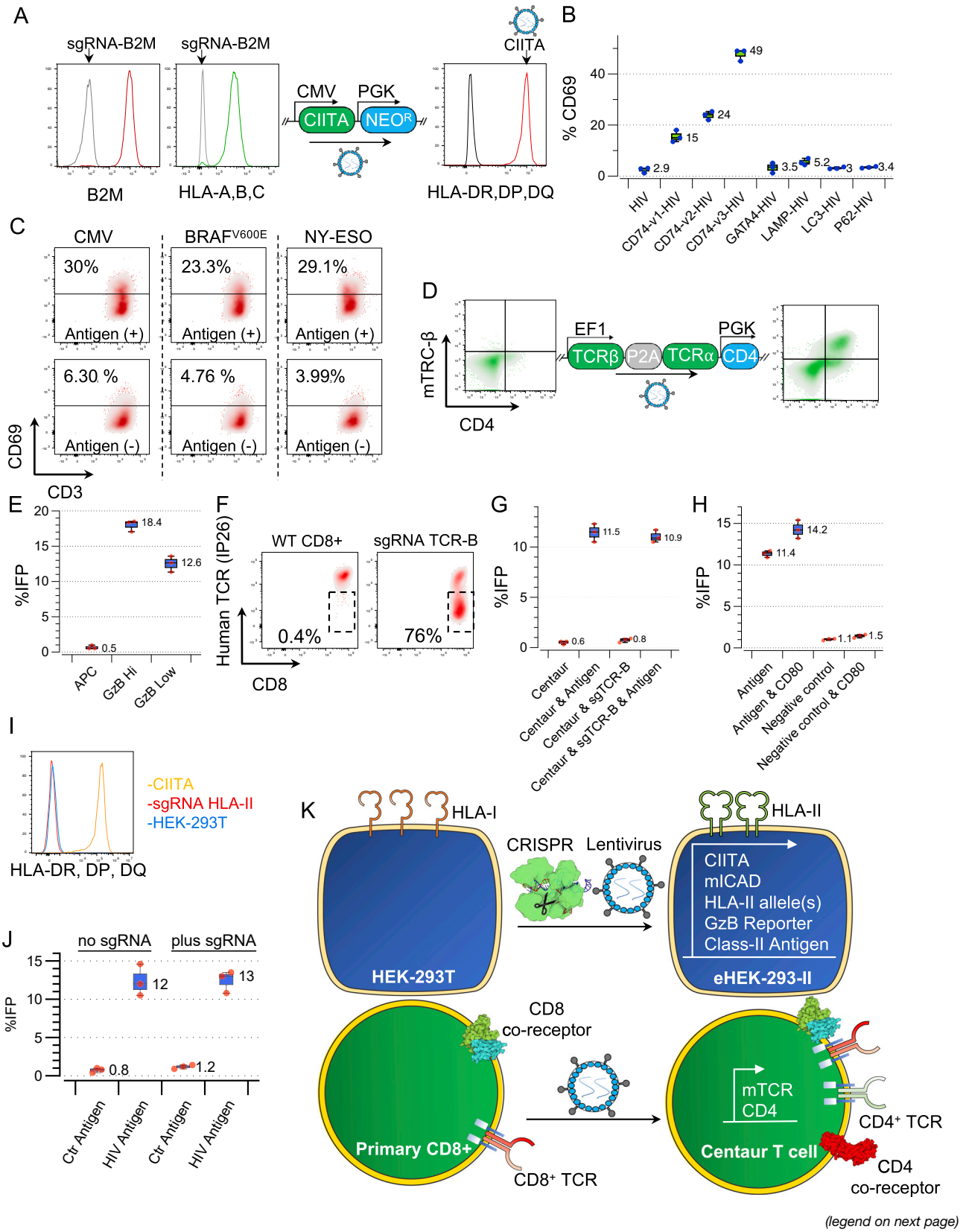
Visium raw data from minor salivary gland samples were converted into fastq files using *spaceranger mkfastq* function of 10x Genomics software (v1.3.1). Reads were subsequently mapped to the human reference genome (GRCH38) using the STAR Methods, and the spatial feature counts were generated using the *spaceranger count* function. *Spaceranger count* data were loaded as a Seurat object using the Load10X_Spatial function in Seurat (v4.0.6).[108] The *SpatialFeaturePlot* function of Seurat was used to visualize the expression of marker genes.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical details for all experiments can be found in the figure legends. For TScan-II screens, the calculation of p-values was done by permutation testing of the geometric mean across enrichment values. Specifically, low input count peptides were first excluded due to high variance in estimates of enrichment. To the remaining peptides, pseudocounts of 1 were added to each library (sort and input libraries) before normalizing: first to library size, then to input normalized counts. These enrichment values were then permuted (n=1,000,000 replicates) within each sort condition to generate a background distribution for the geometric mean of each peptide. False discovery rate correction was done by the Benjamini-Hochberg method. For screens with two barcodes corresponding to the same peptide, where both barcodes had >50 counts in the input library, combined p-values were computed as above after merging enrichment values between barcodes.

CellPress
OPEN ACCESS

Supplemental figures



(legend on next page)

Figure S1. Optimization of antigen processing and TCR signaling for the TScan-II platform, related to Figure 1

(A) FACS histograms depicting the sequential engineering of eHEK-293-II cells. CRISPR-Cas9 ribonucleoprotein (RNP) was used to mutate the B2M gene. Cells that did not express B2M (clone: 2M2) or HLA-A, B, or C (clone: W6/32) on their cell surface were sorted and subsequently transduced with lentivirus to express the MHC-II transactivator (CIITA constitutively). Cells that stably expressed CIITA and had high levels of HLA-DR/DP/DQ (clone: Tü39) were sorted and used for downstream studies.

(B) Boxplots depicting activation levels of F24 TCR-transduced SKW-3 cells following co-culture with HLA-DR11-expressing eHEK-293-II cells in the presence of endogenously expressed and processed HIV antigens. SKW-3 cells were identified using CD3, and their activation was measured via the early activation marker CD69 following co-culture with target cells. The line in the boxplots indicates the median. The box is drawn around the inner quartile range, and the whiskers show minimum to maximum values for all plots shown here and below.

(C) FACS plots depicting activation levels of TCR-transduced SKW-3 cells following co-culture with eHEK-293-II cells in the absence (lower panel) or presence (upper panel) of endogenously expressed (fused to CD74-v3) and processed antigens. TCRs tested include an HLA-DRB1*0701 restricted cytomegalovirus-specific (left), HLA-DQB1*03:02, DQA1*03:01 restricted $BRAF^{V600E}$ (middle), or HLA-DRB1*04 restricted NY-ESO (CTAG1B) antigen.

(D) FACS histograms depicting the generation of centaur primary T cells. Primary $CD8^+$ T cells are transduced with lentivirus to express the modified $CD4^+$TCRs and the glycoprotein CD4. Cells stained positive with mTCR (clone: H57-597) and CD4 (clone: OCT4) were sorted and used for downstream studies.

(E) The frequency of cells that activate the GzB reporter in eHEK-293-II cells expressing either IFP^{GzB} (VGPD$\downarrow$FGR) or $IFP^{GzB\text{-}Hi}$ (VGPD$\downarrow$SGR) upon co-culture with primary $CD8^+$ T cells stimulated with a membrane-bound anti-CD3 antibody (OK3T).

(F) Fluorescence-activated cell sorting plots depicting mutation of endogenous TCR beta chain in primary T cells. CRISPR-Cas9 ribonucleoprotein (RNP) was used to mutate the TCR beta gene. Cells null for TCR beta (clone: IP26) were sorted and subsequently transduced with lentivirus to constitutively express centaur TCR. Cells that stained positive with mTCR (clone: H57-597) and CD4 (clone: OCT4) and negative human TCR (hTCR) (clone: IP26) were sorted and used for downstream studies.

(G) The frequency of target eHEK-293-II cells that activate the GzB reporter was evaluated after co-culture with primary T cells transduced with centaur (Ob1.A12). The centaur T cells used in this study were either wild-type (WT) or TCR null (sgTCR-B). The target eHEK-293-II cells utilized co-expressed the HLA-DRB1*15:01 allele and either full-length Myelin Basic Protein (MBP) fused to CD74-v3 or a control protein fused to CD74-v3.

(H) The frequency of target eHEK-293-II cells that activate the GzB reporter in the presence or absence of CD80 following co-culture with centaur (Ob1.A12) transduced primary. Target eHEK-293-II cells either co-expressed the HLA-DRB1*15:01 allele along with CD80 and full-length Myelin Basic Protein or a control protein fused to CD74-v3.

(I) FACS histograms depicting the mutation of HLA-II alleles. CRISPR-Cas9 ribonucleoprotein (RNP) was used to mutate the HLA-DRB, HLA-DPB, and HLA-DQB genes. Cells null for HLA-DR/DP/DQ (clone: Tü39) were sorted and used for downstream analysis depicted in (J).

(J) The frequency of target cells that activate the GzB reporter following co-culture with centaur (F24) transduced primary. Target eHEK-293-II cells assessed were either null or expressed the endogenous HLA-II alleles. All target cells expressed the HLA-DRB1*1101 allele, along with either the cognate Gag293 antigen or a control protein fused to CD74-v3.

(K) The schematic representation of genetic engineering performed to generate eHEK293-II target cells and centaur T cells.

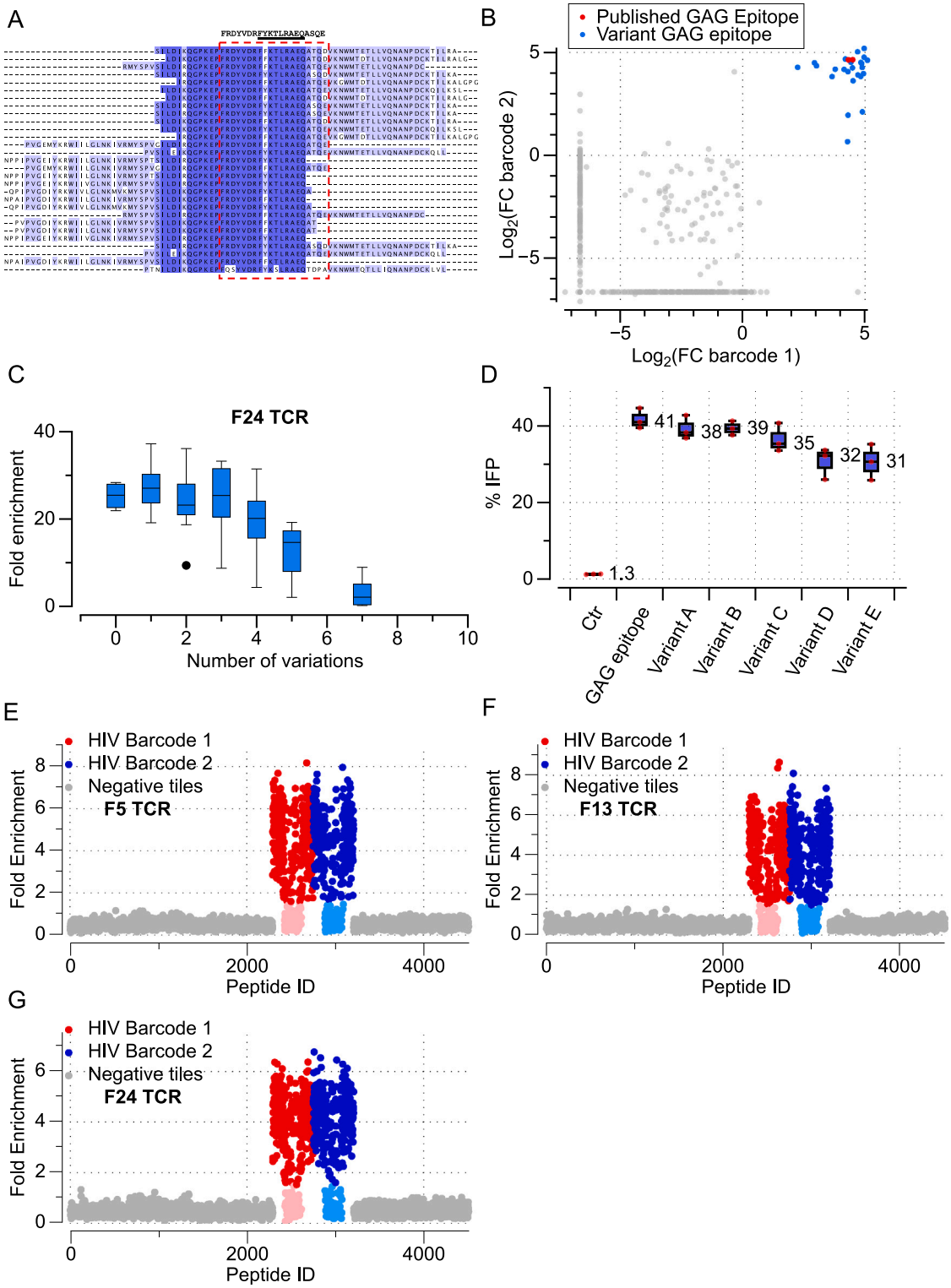


Figure S2. TScan-II screens using the HIV CD4⁺ TCRs, related to Figure 2

(A) Alignment of the enriched fragment from the screen TScan-II HIV genome-wide screen depicted in Figure 2B. The published Gag293 (GAG) epitope sequence is highlighted above the alignment.

(legend continued on next page)

(B) Reproducibility of internal replicates of the genome-wide screen depicted in Figure 2B. Each dot represents one peptide fragment, with the x and y axes plotting the geometric mean of the fold change of each DNA barcode across seven replicates. Red dots represent peptides from an HIV strain with an epitope identical to the published epitope for this TCR. The blue dots represent peptide fragments from variant G293 epitopes.
(C) Fold enrichment of Gag293 variants from the TScan-II screen depicted in Figure 2B. The x axis represents the number of variations, including amino acid substitutions or deletions, relative to the published Gag293 epitope found within the enriched fragments.The line in the boxplots indicates the median. The box is drawn around the inner quartile range, and the whiskers show minimum to maximum values for all plots shown here and below.
(D) Relative activation of the GzB reporter in cells pulsed with HIV peptides co-cultured with F24 T cells. The line in the box plots indicates the median. Variant A: FRDYVDRFFRVLRAEQASQE; variant B: FRDYVDRFYKTLRAEQASQD; variant C: FRDYVDRFFKTLRAEQATQD; variant D: FRDYVDRFFKTLRAEQAT; variant E: FRDYVDRFYKTLRAEQATQE.
(E–G) The TScan-II screens corresponding to the heatmap representation depicted in Figures 2D–2F. eHEK-293-II cells were transduced and selected to express the HLA- DRB1*1101 allele and primary T cells to express the HIV-specific TCRs from an elite controller patient. The mutagenesis library comprises 4,509 fragments, each being 56 aa in length. Each peptide is also barcoded twice using different amino acid codons. Each dot represents one peptide, with the y axis plotting the geometric mean of the fold change of each peptide across six replicates. Fold enrichment is defined as the ratio of the peptide abundance in the sorted population relative to the input library. Red and blue dots represent the first and second replicate (barcodes) of the HIV mutagenesis fragment. The darker and lighter colored dots are HIV mutant peptides that respectively enriched or dropped out.

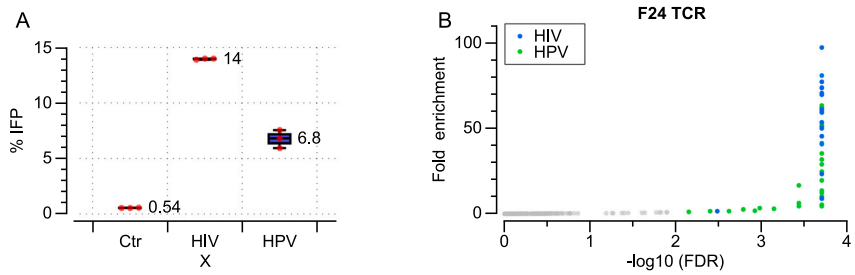


Figure S3. HIV and HPV epitope mapping using the TScan-II platform, related to Figure 3

(A) Validation of the identified antigens in target cells using the GzB reporter following co-culture with HIV (F24) TCR-transduced primary T cells. The boxplots depict the frequency of target cells that activated the GzB reporter a 6-h co-culture experiment. eHEK-293-II cells were transduced and selected to express the HLA-DRB1*1101 allele. Antigens assessed include a control peptide (left), HIV Gag293 protein 289–336 aa (middle), or HPV regulatory E2 protein 29–84 aa (right). All antigens were delivered into target cells via lentiviral transduction. The line in the boxplots indicates the median. The box is drawn around the inner quartile range, and the whiskers show minimum to maximum values.

(B) TScan-II screen of the HIV elite controller F24 TCR using a second library of the top 7,000 enriched fragments from the primary screen (Figure 3B). eHEK-293-II cells were transduced and selected to express the HLA-DRB1*1101 allele and primary T cells to express the HIV-specific TCR (F24). Blue dots represent peptides from the HIV Gag293 protein. Green dots represent peptide fragments from the HPV regulatory E2 protein. p values were calculated by permutation testing, and false discovery rate correction was done by the Benjamini-Hochberg method.

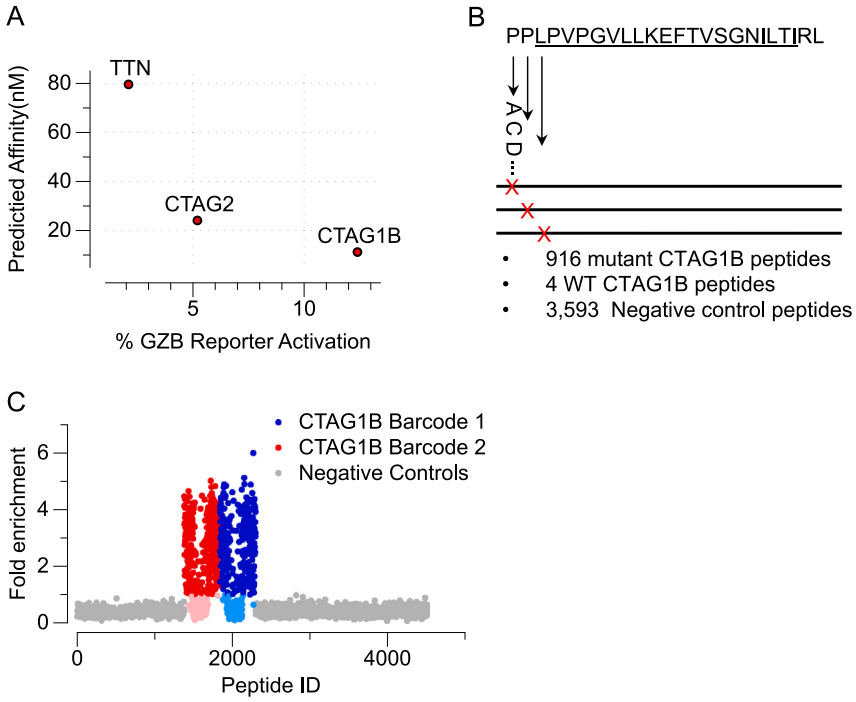


Figure S4. TScan-II mutagenesis screen using NY-ESO CD4$^+$ TCR, related to Figure 4

(A) Predicated affinity of each of the identified epitopes for the HLA-DRB1*04 allele is plotted (y axis) vs. the level of T cell activation as measured by GzB delivery into target cells (x axis). Epitope binding affinity was measured using sliding tiles of 15 amino acids in length on the NetMHCII server.

(B) Design of the CTAG1B epitope comprehensive mutagenesis library. Every position in the CTAG1B epitope (LPVPGVLLKEFTVSGNILTI) is mutated to each of the 19 alternative amino acids. The epitopes are expressed in the context of a 56-aa fragment C-terminal CD74-v3 tag. The two amino acids, N-terminal and C-terminal to the epitope, are also mutated in the 56-aa versions. The library encodes for 4 WT epitopes, 916 mutant CTAG1B epitopes, and 3,593 negative controls.

(C) TScan-II screen of CTAG1B CD4$^+$ TCR (3598-2) against a mutagenesis library. The mutagenesis library comprises 4,509 fragments, each being 56 aa in length. Each peptide is also barcoded twice using different amino acid codons. Each dot represents one peptide, with the y axis plotting the geometric mean of the fold change of each peptide across eight replicates. Fold enrichment is defined as the ratio of the peptide abundance in the sorted population relative to the input library. eHEK-293-II cells were transduced and selected to express the HLA-DRB1*04 allele and primary T cells to express the CTAG1B-specific centaur TCR. Red and blue dots represent the CTAG1B mutagenesis fragment's first and second barcodes (replica). The darker and lighter colored dots are CTAG1B mutant peptides enriched or dropped out, respectively.

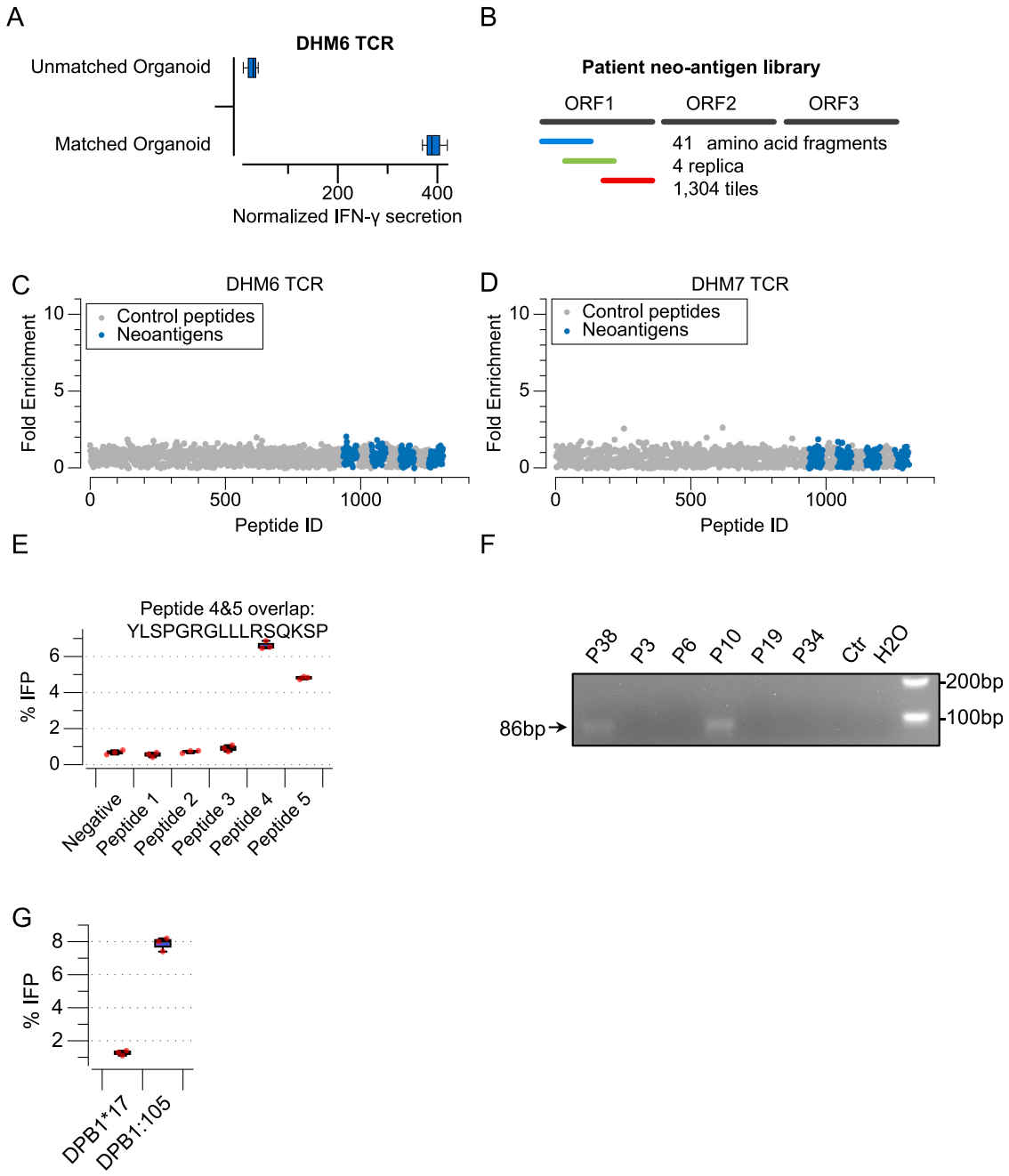


Figure S5. Characterization of pancreatic-cancer-reactive TCRs, related to Figure 5

(A) Characterizing tumor reactivity of tumor-expanded TCRs using patient-derived organoids. TCR-transduced PBMCs co-cultured with either autologous (matched) or non-autologous (unmatched) organoids. IFN-γ secretion was used as a readout of T cell activity.

(B) Schematic of TScan-II human neo-antigen library. The patient neo-antigen library comprises 1,404 fragments that tile across patient mutations with flanking sequences as 41-aa fragments. The WT peptide fragments and fragments encoding mutations from non-matching patients are negative controls in our library. All peptide fragments within the library were barcoded four times using synonymous amino acid codons.

(C and D) TScan-II screen of pancreatic reactive DHM6 and DHM7 TCR against the neoantigen library. Each dot represents one peptide, with the y axis plotting the geometric mean of the fold change of each peptide across eight replicates. Fold enrichment is defined as the ratio of the peptide abundance in the sorted population relative to the input library. eHEK-293-II cells were transduced and selected to express the patient HLA-DP alleles (DPB1*17 and DPB1*105). Blue dots represent four barcoded fragments, with each encoding the patient mutant fragments. Gray dots represent either WT or mutations from non-matching patients and collectively act as negative controls.

(E) Assessment of DHM6 reactivity toward overlapping fragments of LOC100131311. The boxplots depict the frequency of target cells that activated the GzB reporter following a 6-h co-culture experiment. eHEK-293-II cells expressing a control protein were used as negative controls.

(legend continued on next page)

(F) Total RNA was extracted from pancreatic cancer organoids derived from several patients (P3, P6, P10, P19, P34, and P38 [matching patient]). The extracted RNA was utilized for strand-specific cDNA synthesis specific to LOC100131311. The resulting cDNA was then employed to evaluate the expression of LOC100131311 using PCR.

(G) Deconvolution of HLA restriction of DHM6 TCR. The boxplots depict the results of GzB reporter activation following a 6 h co-culture experiment. Primary T cells were transduced with DHM6 TCR and were co-cultured independently with eHEK-293-II target cells expressing the antigenic fragment and DPB1*105 or DPB1*17 alleles.

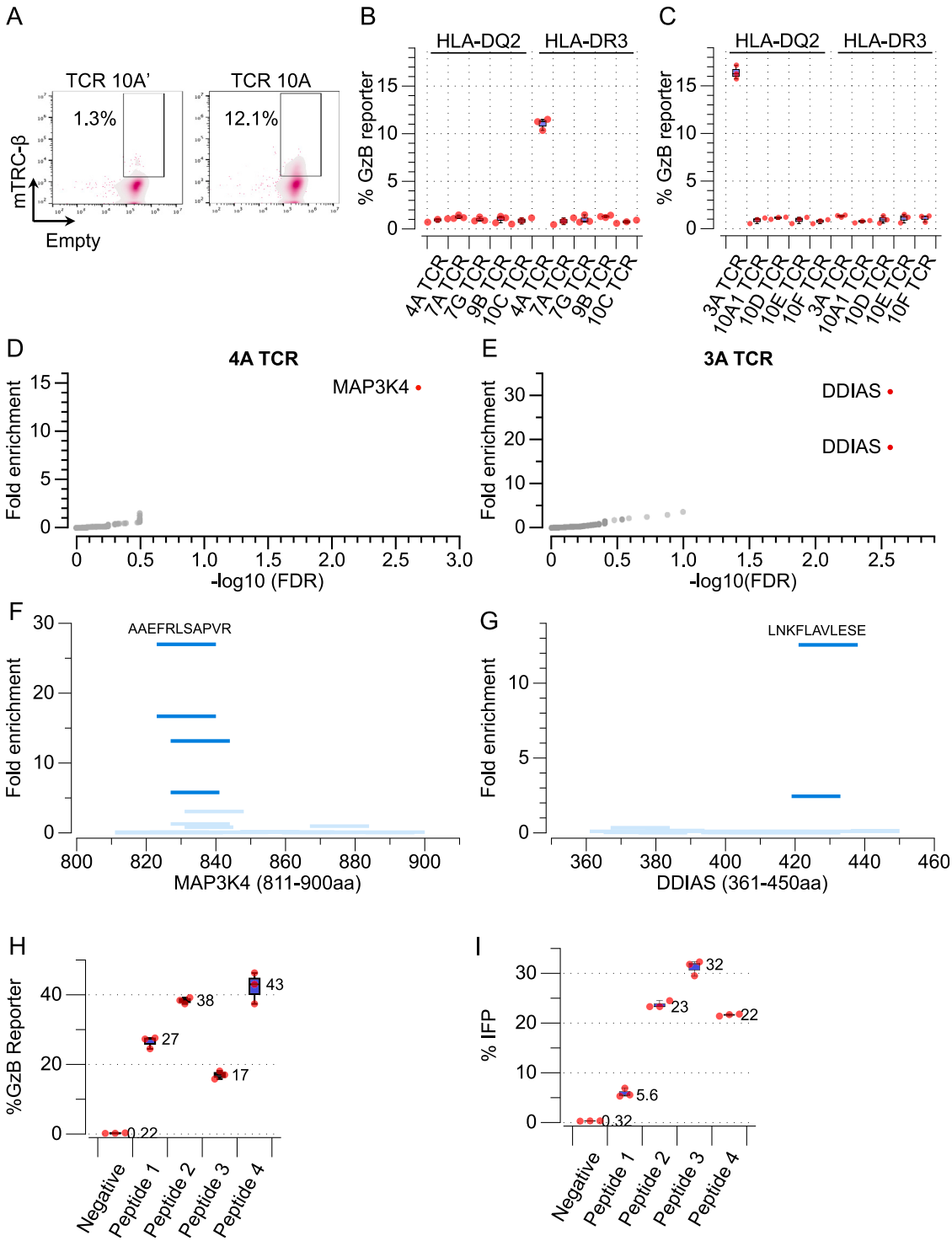


Figure S6. Validation and mapping of DDIAS and MAP3K4 antigens, related to Figure 6

(A) FACS histograms depicting the initial transduction rate of primary T cells using lentivirus encoding two TCRs with identical CDR3β chains matched with two different CDR3α(α or α′) chains from the same clonally expanded $CD4^+$ T cells stained with mTCR (clone: H57-597).

(B) Deconvolution of HLA restriction and TCR reactivity toward MAP3K4 antigen. The boxplots depict the results of GzB reporter activation following a 6-h co-culture experiment. Primary T cells were transduced with 4A, 7A, 7G, 9B, and 10C TCRs and were co-cultured independently with eHEK-293-II target cells expressing MAP3K4 antigenic fragment and HLA-DQB1*02_HLA-DQA1*05 or HLA-DRB1*03:01.

(legend continued on next page)

(C) Deconvolution of HLA restriction and TCR reactivity toward DDIAS antigen. The boxplots depict the results of GzB reporter activation following a 6- co-culture experiment. Primary T cells were transduced with 3A,10A1, 10D, 10E, and 10F TCRs and were co-cultured independently with eHEK-293-II target cells expressing DDIAS antigenic fragment and HLA-DQB1*02_HLA-DQA1*05 or HLA-DRB1*03:01.

(D) TScan-II screen of the 4A TCR using a second library of the top 2,200 enriched fragments from the primary screen shown in Figure 5C. Each dot represents one peptide, with the y axis plotting the geometric mean of the fold change of each peptide across six replicates. eHEK-293-II cells express 5 of patient 4 HLA-II alleles: HLA-DRB1*03:01, HLA-DRB1*13:0, HLA-DQB1*02:01_HLA-DQA1*05:01, HLA-DQB1*06:03_HLA-DQA1*01:03, and HLA-DPB1*04:01_HLA-DPA1*01:03. p values were calculated by permutation testing, and false discovery rate correction was done by the Benjamini-Hochberg method.

(E) TScan-II screen of the 3A TCR using a second library of the top ~5,500 enriched fragments from the primary screen shown in Figure 5D. Each dot represents one peptide, with the y axis plotting the geometric mean of the fold change of each peptide across 6 replicates. eHEK-293-II cells express HLA-DQB1*02_HLA-DQA1*05 and HLA-DRB1*03:01. p values were calculated by permutation testing, and false discovery rate correction was done by the Benjamini-Hochberg method.

(F and G) TScan-II screen of a library of 15- and 18-aa fragments with 1-aa overlap spanned the entire length of 90-aa antigenic segments of MAP3K4 and DDIAS. The length of each line along the x axis indicates the corresponding region of the MAP3K4 covered by each peptide according to a pairwise protein alignment; the y axis corresponds to the fold enrichment of that fragment. Fragments in light blue did not score, whereas darker blue indicates the scoring peptides. The amino acid sequence depicted represents the minimal epitope characterized.

(H) Relative activation of the GzB reporter in cells pulsed with 1 μg of MAP3K4 peptides co-cultured with 4A T cells. Peptide 1: LRKDLEIAAEFRLSA; peptide 2: EIAAEFRLSAPVRDL; peptide 3: EIAAEFRLSA; peptide 4: LRKDLEIAAEFRLSAPVRDL. The line in the boxplots indicates the median. The box is drawn around the inner quartile range, and the whiskers show minimum to maximum values.

(I) Relative activation of the GzB reporter in cells pulsed with DDIAS peptides co-cultured with 3A T cells. The line in the boxplots indicates the median. Peptide 1: SLNKFLAVLESEIAV; peptide 2: ESLNKFLAVLESEIAV; peptide 3: FSESLNKFLAVLESEIAV; peptide 4: LNKFLAVLESEIAVTQADVS. The box is drawn around the inner quartile range, and the whiskers show minimum to maximum values.

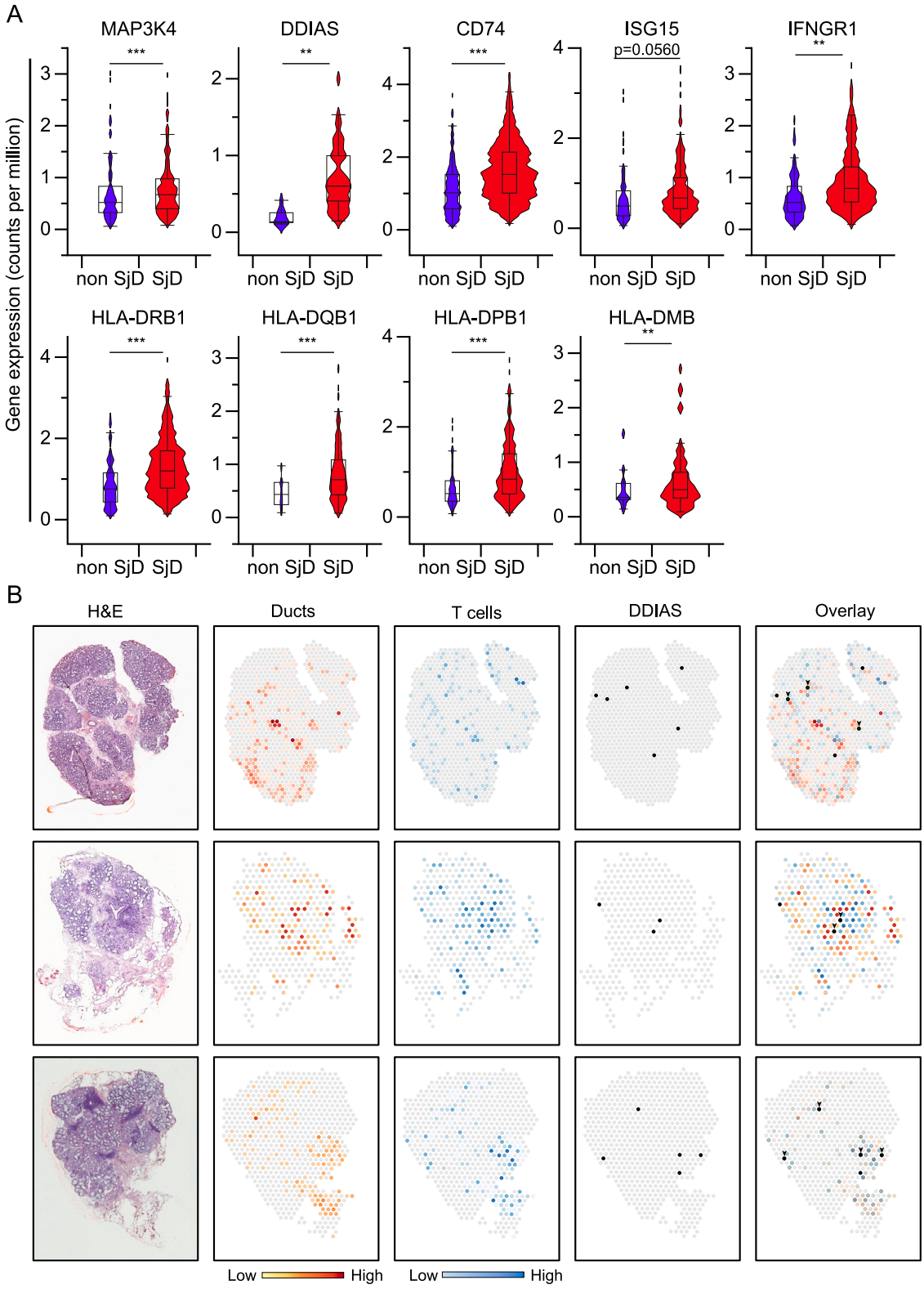


Figure S7. Assessment of antigens and HLA-II transcript abundance within the salivary gland, related to Figure 7

(A) Plot showing the expression of select genes across ductal cells (as defined in Figure 7A) in both healthy individuals and SjD patients. p values were determined by a two-tailed t test and are shown with asterisks: *p < 0.05, **p < 0.01, and ***p < 0.001.

(legend continued on next page)

(B) The spatial expression pattern of DDIAS relative to the location of ductal and T cells in the SG of three SjD subjects was assessed using the 10× Visium transcriptomics. Spotplot representing either ductal cell markers (mean normalized expression of S100A2 and CFTR) or T cell markers (mean normalized expression of CD3D, CD3E, CD3G, and CD45).