nature communications

Article

https://doi.org/10.1038/s41467-026-72396-7

Whole-protein screening and multi-modal profiling of antigen-specific CD4$^+$ T cells at single-cell resolution

Received: 30 September 2025

Accepted: 15 April 2026

Published online: 12 May 2026

Check for updates

Rongyu Zhang[1,2], Jingqi Qi[1,2], Michaela McKasson[1], Jongchan Choi[1], Vanessa Gutierrez[1], Conor Brennan[1], Sunga Hong[1], William Chour[1,3], Rachel H. Ng[1,2], Jingyi Xie[1,4], Dan Yuan[1,2], Andrew Webster[1], Simranjeet K. Sidhu[1,2], Abby Anderson[1,2], Daniel Chen[1], Rick Edmark[1], Kim M. Murray[1], Sarah Li[1], Connor McDonald[1], Lee Rowen[1], Shuo Wang[1,2], Yusuf Rasheed[1], Yapeng Su[1,5], Jamie R. Wagner[6], Jia Ming Chen[7], Karla Nawaly[7], Jie Fu[8], Alexandria Duven[9], Stephen J. Forman[6], Mihae Song[10], Saul J. Priceman[11], Christine E. Brown[6], Antoni Ribas[7], Deborah J. Wong[7], Kelly G. Paulson[12,13], Charles W. Drescher[5,12], Cristina Puig-Saus[7], Jason D. Goldman[9,14,15], Cornelia L. Trimble[8] & James R. Heath[1,2,4] ✉

Systematic whole-protein screening and comprehensive profiling of antigen-specific CD4$^+$ T cells are crucial for advancing vaccine design and cancer immunotherapies, yet remain technically challenging. Here, we present a high-throughput platform that utilizes large-scale class II single-chain trimer libraries to detect antigen-specific CD4$^+$ T cells, while simultaneously profiling their antigen specificity, TCRα/β sequences, MHC restriction, whole transcriptomes, and patient/timepoint origins at single-cell resolution. Upon rigorous platform validation, we screened the full SARS-CoV-2 spike receptor binding domain in a longitudinal cohort of 22 participants, identifying 2,188 antigen-specific CD4$^+$ T cells and showing key metrics defining the immunogenicity of class II-restricted viral antigens. We further extended the platform to whole-protein screening of HPV-16 E6/E7 in a cohort of precancerous patients, indicating HPV-specific CD4 TCRs that, upon extensive characterization, demonstrate strong therapeutic potential. By integrating high-throughput antigen screening with high-dimensional, multi-modal cellular characterization, our approach provides detailed insight into CD4$^+$ T cell immunity, potentially guiding vaccine design and next-generation TCR-based cancer immunotherapies.

CD4$^+$ helper T cells play a critical role in orchestrating and regulating immune responses. Upon exposure to foreign or tumor-derived peptides presented by class II major histocompatibility complex (pMHC), they enhance antigen-presenting cell function, support CD8$^+$ T cell differentiation, and promote B cell activation and antibody maturation[1]. In cancer, CD4$^+$ T cells can help sustain anti-tumor immunity and even exhibit cytotoxic functions[2,3]. A detailed understanding of the dynamic roles of antigen-specific CD4$^+$ T cells is thus crucial for mapping complex immune responses and guiding the development of therapies.

A full list of affiliations appears at the end of the paper. ✉e-mail: jheath@isbscience.org

While large-scale screening methods for antigen-specific $CD8^+$ T cells are well-established[4–10], analogous platforms for $CD4^+$ T cells are more challenging due to the complexity of class II MHC antigen presentation. High-throughput peptide-pool stimulation followed by activation marker-based sorting captures TCR sequence, but disrupts the native phenotypic state of the cells, which is crucial for fundamental understanding of $CD4^+$ T cell responses[11–14]. Alternatively, direct ex vivo capture of $CD4^+$ T cells using conventional class II pMHC multimers preserves phenotype but is constrained by a combination of the small size of feasible pMHC libraries ($n \leq 15$), and the reliance on imperfect computational predictions[15–22].

Comprehensive profiling of $CD4^+$ T cell immunity requires high-throughput approaches capable of capturing and analyzing large, high-dimensional datasets across patient cohorts. Current methods often only capture isolated parameters—such as TCR clonotype, antigen specificity, or MHC restriction—limiting their ability to provide a holistic view of $CD4^+$ T cell responses.

Here we present a high-throughput platform leveraging class II single-chain trimer (SCT) libraries that enables whole-protein screening for antigen-specific $CD4^+$ T cells, while simultaneously profiling multiple key parameters—including antigen specificity, MHC restriction, whole transcriptome, TCR clonotype, and patient/timepoint origin—at single-cell resolution in a single experiment. This platform builds upon our previous report of the APMAT method (antigen-TCR pairing and multiomic analysis of T cells) for the multiomic profiling of antigen-specific $CD8^+$ T cell responses, and so we call it APMAT-CD4. We validated the performance of SCTs against conventional pMHCs and demonstrated the SCT library-based approach for direct ex vivo detection of common pathogen-specific $CD4^+$ T cells in healthy donors. We then applied the platform to screen the entire SARS-CoV-2 spike receptor-binding domain (RBD) in a longitudinal cohort of 22 participants, identifying 2188 antigen-specific $CD4^+$ T cells. This high-dimensional dataset, capturing six orthogonal modalities at single-cell resolution, revealed class II-restricted immunogenic antigens and documented dynamic phenotypic shifts over the course of infection. To further demonstrate its broad applicability, we extended the platform to cancer and performed whole-protein screening of HPV-16 oncogenic proteins E6 and E7 in patients with precancerous lesions for CD4 TCR repertoire profiling, uncovering functionally heterogeneous TCRs with strong therapeutic potential. By integrating high-throughput whole-protein screening with multi-modal single-cell profiling, our platform offers a comprehensive and scalable framework for studying $CD4^+$ T cell immunity across disease contexts, accelerating discovery of key immunological mechanisms and therapeutic targets.

Results

Class II SCT library enables reliable ex vivo identification of antigen-specific $CD4^+$ T cells

We previously reported on the APMAT method for the high-throughput profiling of class I-restricted $CD8^+$ T cell responses to viral and tumor neoantigens[7–10]. An enabling technology of APMAT was the use of class I SCT libraries. SCTs are mammalian cell-expressed proteins engineered to emulate pMHC molecules[10]. APMAT-CD4 builds upon this foundation through the development of a modular platform for the scalable generation of human class II SCT libraries (Fig. 1), extending prior murine constructs[7,23]. For class II, the SCT architecture links HLA α, antigen, and HLA β extracellular domains via flexible glycine-rich linkers (L1 and L2) (Fig. 1a). A partial invariant chain (pIi) enhances SCT stability, while the linker configuration maintains structural integrity[23,24]. An antigen library is genetically encoded into the template for a given MHC allele using two overlapping primers designed via a custom algorithm (Fig. 1b, Methods, and Supplementary Data 1). This approach supports high production efficiency while avoiding peptide synthesis, and achieves a >99% success rate as confirmed by PCR (Supplementary Data 1). Conserved regions for pIi and linker sequences allow rapid antigen exchange and allele substitution. SCT protein expression is achieved by the transient transfection of human-derived Expi293F cells in a multi-well format, followed by biotinylation and purification workflows optimized for large-scale production.

To benchmark performance, we first validated the specificity of class II SCTs against conventional pMHCs. A DRB1*01:01-restricted (DR1) SCT presenting the influenza $HA_{306\text{-}318}$ epitope effectively stained a matched influenza-specific $CD4^+$ T cell line, demonstrating binding performance comparable to standard pMHC tetramers with no evidence of non-specific binding from an HIV-specific SCT negative control (Supplementary Fig. 1a). Sensitivity was evaluated by spiking known frequencies of influenza-specific T cells into healthy donor PBMCs. SCT tetramers could detect target cells as low as 0.05% with a recovery rate similar to conventional pMHC tetramers (Supplementary Fig. 1b).

We then expanded validation to additional prevalent class II HLA alleles[25,26], selecting literature-validated TCR-antigen pairs for one HLA-DP and three HLA-DR alleles[27–31] (Supplementary Data 2). TCRs were cloned into TCR^{KO} Jurkat cells via lentiviral transduction and evaluated for binding against their cognate SCTs. Each SCT specifically recognized its matched TCR, with no evidence of cross-reactivity to unrelated SCTs (Fig. 1c and Supplementary Fig. 1c). Notably, SCTs displayed post-translational modifications consistent with natural antigen processing, in contrast to synthetic peptide-based approaches (Supplementary Fig. 1e). We further investigated whether SCT expression is influenced by peptide motifs using a regression-based 2-mer association analysis. Several 2-mers, including FP, PF, and VR were positively correlated with SCT expression ($\rho = 0.223–0.304$), while others like SS, DL, CR, and RS showed weak to moderate negative correlations ($\rho = -0.389$ to -0.223, FDR <0.05) (Supplementary Fig. 1d). Given that class II peptide-MHC binding is shaped by anchor residues—especially the P1 peptide-binding pocket, which favors aromatic amino acids (F, Y, W)[32,33]—we tested whether these residues are enriched among positively correlated 2-mers. Fisher's exact test confirmed a significant enrichment of F/Y/W in the positively associated group ($P = 0.0004$), and their underrepresentation among negatively correlated 2-mers (Fig. 1d). This pattern aligns with known HLA-DR1 antigen presentation preferences[32,33] and suggests that SCT expression reflects epitope compatibility with MHC class II. These findings underscore class II SCTs as a robust platform for generating pMHC-like reagents capable of highly sensitive and specific $CD4^+$ T cell detection.

Following validation of individual SCTs, we evaluated the performance of SCT libraries in direct ex vivo screening for common pathogen-specific $CD4^+$ T cells in healthy donor peripheral blood mononuclear cells (PBMCs). For this task, we assembled a 23-element SCT library presenting DR1-restricted canonical antigens from cytomegalovirus (CMV), Epstein-Barr virus (EBV), influenza, and tetanus (the CEFT library) (Fig. 1e and Supplementary Data 3). A DR1-restricted HIV $Gag_{41\text{-}56}$ SCT was used as the negative control. PBMCs from five HLA-matched healthy donors were enriched for $CD4^+$ T cells and stained with PE- and APC- labeled SCT tetramers (Supplementary Fig. 1f). Double-positive cells were sorted for single-cell sequencing of paired TCRα/β chains, while cells binding to negative control tetramers were excluded (Fig. 1f and Supplementary Fig. 1g). TCRs identified in this way were cloned into TCR^{KO} NFAT-GFP Jurkat cells (Supplementary Data 4) and evaluated for SCT binding (Fig. 1g). We confirmed TCR recognition to cognate antigens, with varying binding intensities suggesting differential recognition capacities among TCRs. To further assess whether SCTs faithfully model antigen presentation on MHCII, we tested these TCRs in a peptide-pulsed activation assay (Fig. 1h) using DRB1*01:01-expressing K562 cells with an empty peptide-binding pocket (DR1-K562)[34–36]. Co-culture with cognate peptides induced NFAT-based activation in all TCR+ Jurkat clones (Fig. 1i)

https://doi.org/10.1038/s41467-026-72396-7

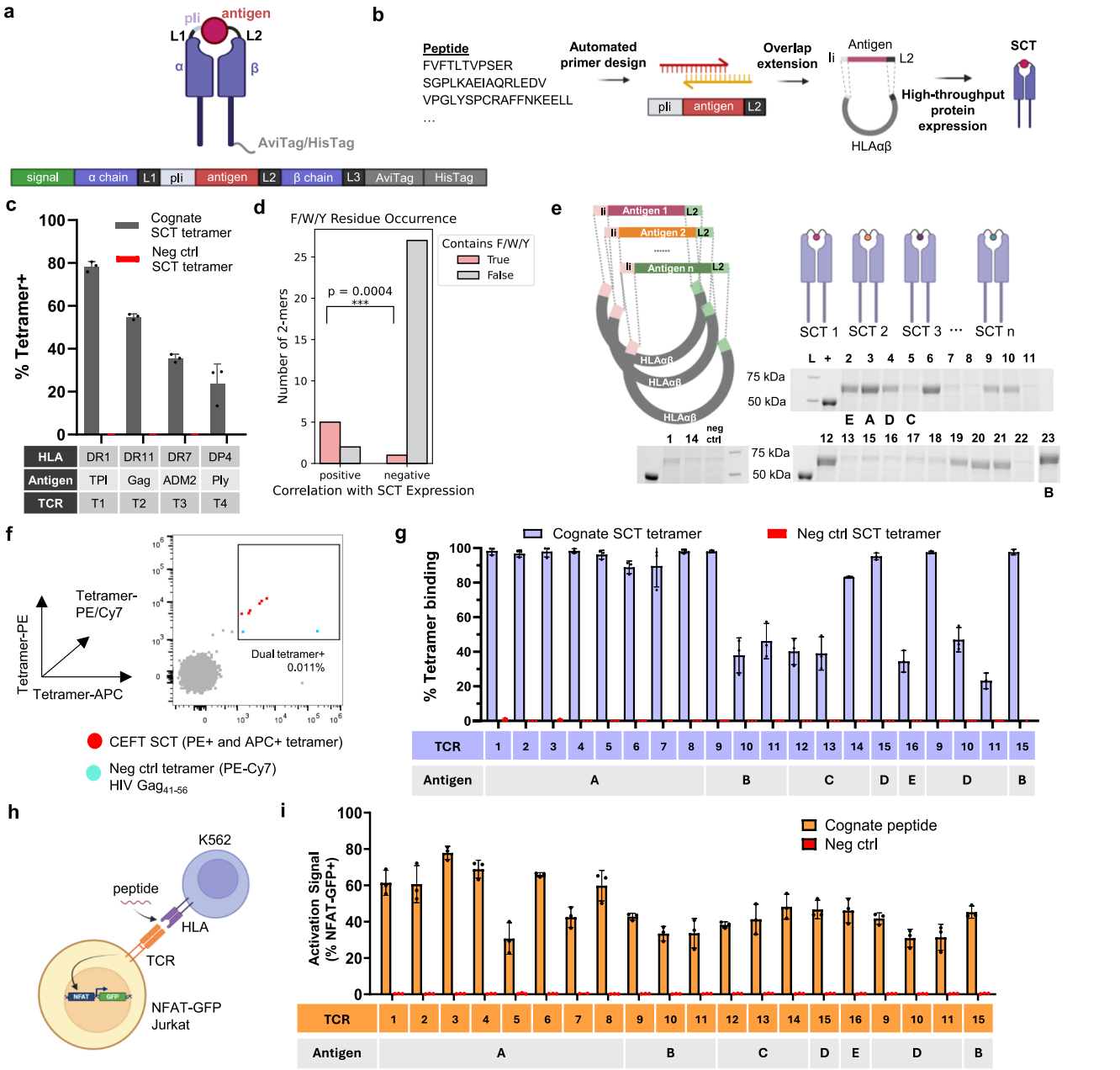

Fig. 1 | Class II SCT library enables the discovery of antigen-specific CD4⁺ T cell in healthy donors. a Genetic design encoding the class II SCT protein. The MHC α and β chains are connected to the antigen through flexible linkers (L1, L2, and a partial invariant chain pIi), followed by purification tags (AviTag/HisTag). A secretion signal is placed upstream of the α chain to facilitate protein export. Created in BioRender[76]. **b** Streamlined workflow for high-throughput SCT plasmid construction and protein expression. High-throughput production of SCT libraries is enabled by automated primer design and an optimized pipeline for plasmid assembly and protein expression/purification. Created in BioRender[76]. **c** Flow cytometric assay to assess binding of SCT tetramers of various common class II HLA alleles to their cognate TCRs ($n = 3$). Previously reported TCR-antigen pairs: DRB1*01:01 (DR1) – mutTPI$_{23-37}$:T28I, DRB1*11:01 (DR11) – HIV Gag$_{299-311}$, DRB1*07:01 (DR7) – ADM2$_{51-65}$, DPB1*04:01 (DP4) – mutPly$_{427-441}$ (Supplementary Data 1). **d** Comparison of the occurrence of aromatic residues F/W/Y in 2-mers that are positively correlated with SCT expression versus those that are negatively associated. The p value was calculated using Fisher's exact test on the full 2 × 2 contingency table comparing F/W/Y presence across groups. **e** A 23-Element DRB1*01:01 CEFT SCT library. SCT expression was evaluated through SDS-PAGE. L, molecular weight ladder; +, protein standard for quantification of expression level. Schematics created in BioRender[76]. **f** Antigen-specific T cells were detected through dual tetramer staining to enhance the true positive rate. An HIV Gag$_{41-56}$ SCT tetramer was used to exclude non-specific binding cells. A representative flow cytometry plot is shown. **g** Tetramer binding validation of 16 identified CD4 TCRs, each reactive to one of the five CEFT antigens (A–E) ($n = 3$). Error bars represent standard deviation. **h** Schematic of peptide-pulsed activation assay. TCR-transduced NFAT-GFP Jurkat cells were co-cultured overnight with peptides and DR1-K562 cells. Created in BioRender[76]. **i** T cell activation measured through the percentage of GFP + Jurkats ($n = 3$). $P < 0.01$ for each group relative to the negative control peptide. Statistical significance was determined by a one-tailed independent t-test assuming equal variances. All replicates are biological replicates. All bar plots ($n = 3$) are presented as mean values ± SD. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-026-72396-7

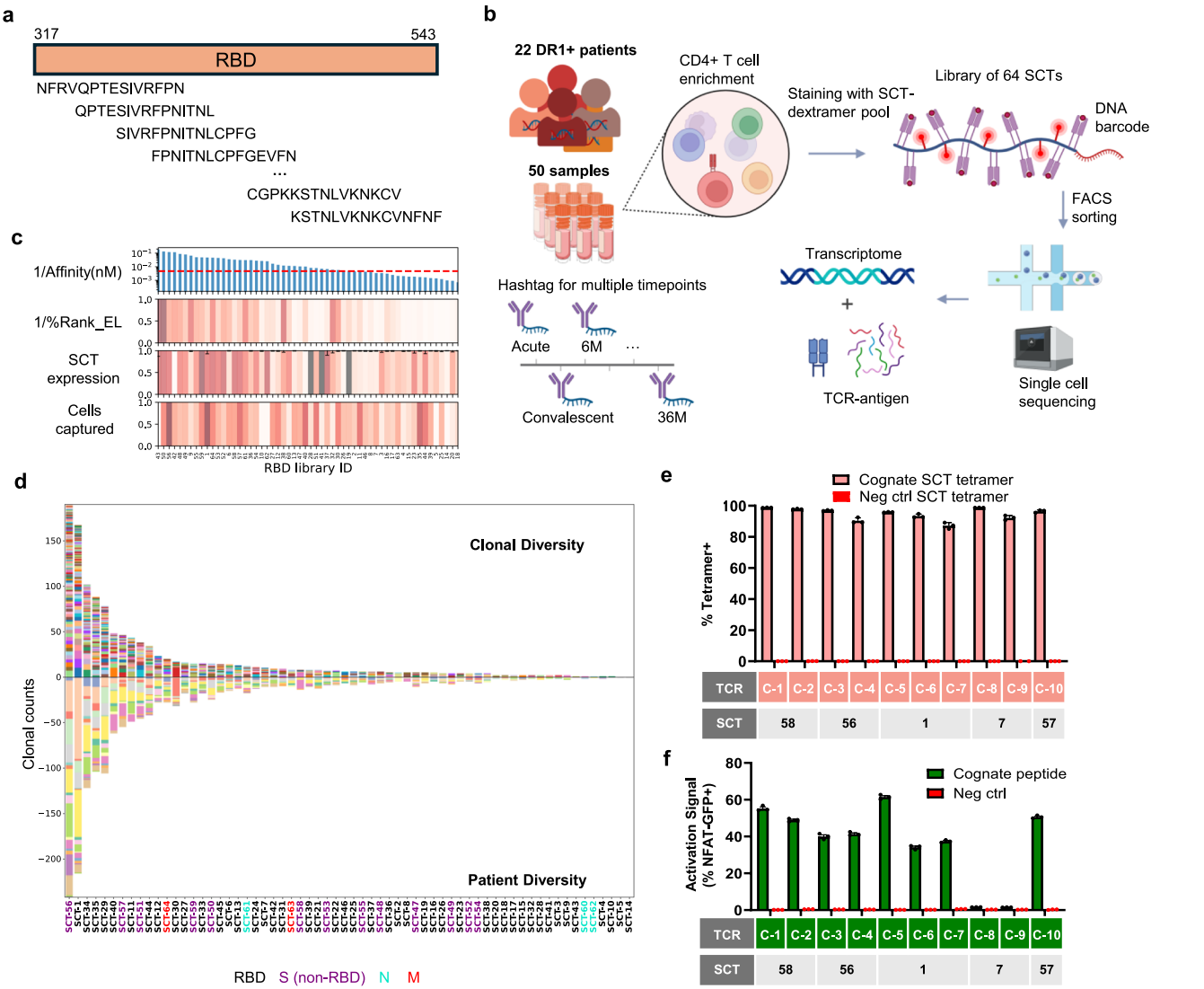

Fig. 2 | High-throughput screening of the entire receptor-binding domain (RBD) of SARS-CoV-2 spike identifies large-scale antigen-specific CD4⁺ T cells. **a** Design of an SCT library with overlapping 15-mer peptides spanning the entire RBD of the SARS-CoV-2 spike protein. **b**. Schematic overview for the experimental workflow of high-throughput screening. 50 PBMC samples from 22 HLA-DR1+ participants were barcoded by timepoint, pooled, enriched for CD4⁺ T cells, stained with the 64-element SCT-dextramer pool, FACS-sorted, and sequenced. Single-cell analysis included RNA expression profiling, TCRαβ sequencing, antigen-MHC pairing, and timepoint identification. Comparison of genetic variants between transcriptome and whole-genome sequencing (WGS) was used to demultiplex patient identity (Methods). Created in BioRender[76]. **c** Correlation between SCT expression levels, quantity of antigen-specific cells captured, and predictions from class II antigen-presentation algorithms. SCT expression quantified prior to purification ($n = 2$–4 biological replicates). Data were shown as mean ± SEM. % Rank_EL, percentile rank of eluted ligands. **d** Clonal and patient diversity in CD4⁺ T cell responses to each SARS-CoV-2 antigen. Antigens are color-coded based on their parent protein. **e** Tetramer binding validation of SARS-CoV-2-specific CD4 TCRs. A representative subset of SARS-CoV-2 TCRs ($n = 10$) from Fig. 2d were selected and validated through SCT tetramer binding assays with biological triplicates. A DR1-restricted influenza A $M1_{17–31}$ SCT was used as the negative control. **f** Peptide-pulsed activation of SARS-CoV-2 CD4 TCRs in NFAT-GFP Jurkats ($n = 3$). **e**, **f** $P < 0.0001$ for each group relative to the negative control peptide, determined by one-tailed independent *t*-test assuming equal variances. Exact *P* values are provided in the Source Data. All replicates are biological replicates. All bar plots ($n = 3$) are presented as mean values ± SD. Source data are provided as a Source Data file.

identified through the SCT library screen. Clones 9, 10, 11, and 15 showed consistent tetramer binding and activation to peptides (B and D) that share the same core sequence (Supplementary Data 5). These results demonstrate that SCT libraries enable direct ex vivo discovery of functionally validated CD4⁺ TCR-antigen pairs with high specificity and sensitivity.

High-throughput screening of the entire SARS-CoV-2 receptor-binding domain identifies large-scale antigen-specific CD4⁺ T cells

We next explored the full APMAT-CD4 technique by using class II SCT libraries to systematically screen for antigen-specific CD4⁺ T cell responses against the SARS-CoV-2 spike protein receptor-binding domain (RBD) in a longitudinal patient cohort. This domain, which is a dominant target of protective antibodies[37–41], is encoded by most SARS-CoV-2 vaccines, and CD4⁺ T cells specific to RBD presumably play a role in the development of B cell immunity. Thus, we constructed a 54-element SCT library spanning the entire RBD in four amino acid increments (Fig. 2a) to systematically profile CD4⁺ T cell responses. Of these, 46 SCTs were prepared in usable quantities (Supplementary Fig. 2a and Supplementary Data 6). An additional 18 SCTs representing reported epitopes from spike (S), membrane (M), and nucleocapsid (N) proteins were also included[14,42–44] (Supplementary Data 6 and Supplementary Fig. 2a).

The complete 64-element SCT library was multimerized using fluorophore-labeled, DNA-barcoded dextramers and pooled to stain CD4⁺ T cells from 50 PBMC samples across DRB1*01:01-positive participants ($n = 22$) (Fig. 2b). Samples spanned multiple time points, ranging from acute infection (AC, <1 week of infection) through convalescence (CV, 2–3 months post-acute) to, for some donors, long-term follow-up (6 to 36 months) (Supplementary Data 7). Hashtag antibodies were included to facilitate time point deconvolution. Dextramer-positive CD4⁺ T cells were isolated by FACS and subjected to single-cell sequencing, yielding paired whole transcriptome, TCR sequence, epitope specificity, HLA restriction, and patient/timepoint origin (Fig. 2b). After data processing (Methods), we identified a total of 2188 antigen-specific CD4⁺ T cells, with cells detected across all donors.

Analysis of SCT expression revealed limited correlations with in silico MHC-binding predictions (Fig. 2c and Supplementary Fig. 2b), consistent with known challenges in prediction algorithms. Cell capture was independent of SCT expression level, predicted binding affinity, or percentile rank, in contrast to class I SCT platforms[45]. These results underscore the enabling use of class II SCT libraries for systematic empirical screening.

The broad diversity of T cell clonotypes captured by each SCT, as well as the wide distributions of antigen-specific CD4⁺ T cell across participants, suggested that the observed responses were not skewed by donor, clonotype, or antigen bias (Fig. 2d). Clonal expansion was observed in responses to 42 of 64 antigens and in 49–58% of cells across time points (Supplementary Fig. 2c). We further validated epitope-specific TCRs through tetramer binding and peptide-pulsed activation on selected clonotypes (Fig. 2e, f and Supplementary Data 8), confirming the fidelity of the methodology.

Multi-modal and high-dimensional profiling empowers deep characterization of CD4⁺ T cells at single-cell resolution

The full APMAT-CD4 dataset involved the integration of multiple dimensions of information (Fig. 3a, Methods). Each T cell is profiled through three parallel sequencing libraries: (1) whole transcriptome, (2) TCR αβ sequences, and (3) a surface protein library which decodes MHC-restricted antigen specificity and timepoint origin via DNA-barcoded dextramers and hashtag antibodies. Patient identity is resolved by matching transcriptome-derived genetic variants to corresponding whole-genome sequencing profiles, and a total of 593,905 single-nucleotide polymorphisms (SNPs) were used for complete demultiplexing of patients (Supplementary Fig. 3a).

To dissect CD4⁺ T cell states, all antigen-specific CD4⁺ T cells were clustered by gene expression similarity and projected onto a uniform manifold approximation and projection (UMAP) (Fig. 3b, Methods). Phenotypic annotation based on canonical markers identified major subsets including naïve-like/central memory T_{CM} (SELL, CCR7, TCF7, and LEF1)[46–49], effector memory T_{EM} (S100A4, AHNAK, IL32, and CLIC1)[47,50,51], Th1 (TBX21, IFNG, STAT4, RUNX3, and CCL5)[1,47], Treg (FOXP3)[1,46,47], and Th17 (RORC, KLRB1, and CCR6)[47] (Supplementary Fig. 3b). A subset of Th1 cells upregulates cytotoxic markers (NKG7, CST7, GNLY, PRF1, TNF, and granzymes)[47,48,52], while the exhausted-like phenotype is characterized by high expression of exhaustion markers (PDCD1, TIGIT, LAG3, and CTLA4)[48]. Additionally, a small cluster of cells displays high expression of proliferation markers, including MKI67, MYBL2, BUB1, PLK1, and CCNE1[53]. These clusters reveal distinct phenotypic variations among SARS-CoV-2-specific CD4⁺ T cells.

T cell responses to viral infection are expected to evolve over the course of disease, and by integrating epitope, timepoint, and transcriptional information, this dataset provided an opportunity to map such dynamics for viral-specific CD4⁺ T cells in an HLA-matched cohort (Fig. 3c). Antigen-induced responses were phenotypically polarized at early timepoints, with distinct biases toward naïve-like/T_{CM}, exhausted-like, and T_{EM} (Fig. 3c, Methods). Longitudinal tracking revealed a convergence toward T_{EM} phenotypes, consistent with previous reports on the evolution of CD4⁺ T cell responses following antigen stimulation[54], although very distinct from CD8⁺ T cell dynamics[10].

T cell antigens can exhibit widely varying degrees of immunogenicity in terms of response generated within and across different patients, as well as through downstream factors such as how they influence B cell maturation. We established a quantitative scoring framework incorporating: (1) the number of unique TCR clonotypes per antigen, (2) the fraction of responding donors, (3) longitudinal persistence, and (4) clonal expansion (Methods). These metrics were $\log_2$-transformed, aggregated, and used to rank antigens into high, medium, and low immunogenicity categories (Fig. 3d and Supplementary Fig. 3c), revealing an antigen-dependent T cell response that varies by several orders of magnitude. We note that sample distribution varied across timepoints, with long-term sampling available from four patients who experienced prolonged COVID symptoms. Low to moderate inter-parameter correlations ($\rho = 0.23–0.64$) indicate that each parameter retains a degree of independent information (Supplementary Fig. 3d). To quantitate whether this metric associated the promotion of B cell maturation, we calculated immunogenicity scores for each donor against the RBD antigens at acute infection (see Methods), and then plotted those scores against RBD-specific antibody titers in those same patients at convalescence (Fig. 3e). In fact, the analysis revealed a statistically significant ($P = 0.02$) association, suggesting that CD4⁺ T cell responses against DRB1-restricted RBD antigens may play a role in B cell maturation against closely related epitopes.

We next folded in transcriptomic information to query how antigen immunogenicity influenced phenotype kinetics from acute disease to recovery. The most immunogenic antigens exhibited an exhausted-like CD4⁺ T cell phenotype during acute disease that transitioned towards an effector memory phenotype over time. In contrast, low-immunogenicity antigens induced naïve-biased responses with limited evolution (Fig. 3f). While clonal expansion exhibited little phenotype bias, clonal persistence was largely biased towards Tem and exhausted-like phenotypes (Supplementary Fig. 3e). We also identified multiple public TCRs, which are TCRs shared across individuals (Supplementary Fig. 3f).

Finally, we carried out a case study of a donor (INCOV042) who was suffering from long Covid, and for whom we had serial blood draws extending out to 3 years post initial infection. In fact, we documented continued SARS-CoV-2-specific CD4⁺ T cell functional activity (Supplementary Fig. 2d) that only receded at the 3-year time point. Whether such ongoing CD4⁺ T cell activity is consistent with reports of antigen or viral persistence in some long Covid patients[55–57], or whether it associates with serial vaccinations, will require a larger cohort study. However, this intriguing result suggests the value of such a study for identifying biomarkers of long Covid.

Collectively, this high-throughput, multi-modal approach enables systematic identification and deep characterization of the immunogenicity, at multiple levels, of both class II-restricted viral antigens, and the responding antigen-specific CD4⁺ T cells. It also allows for documentation of how such immunogenic responses to infections vary across patients.

Whole-protein screening of HPV-16 E6/E7 uncovers CD4 TCRs with strong therapeutic potential for cancer immunotherapy

We next explored the application of APMAT-CD4 to cancer, by performing a comprehensive screening of the oncogenic HPV-16 proteins E6 and E7 for CD4 TCR repertoire profiling. While over 200 HPV genotypes have been identified, HPV-16 infection poses the highest risk for cervical, oropharyngeal, anal, vaginal and other HPV+ cancers. The HPV-16 viral oncoproteins E6 and E7 are constitutively expressed in a functionally obligate manner in HPV-16-associated cancers and precursor lesions, such as cervical intraepithelial neoplasia 2/3 (CIN2/3),

https://doi.org/10.1038/s41467-026-72396-7

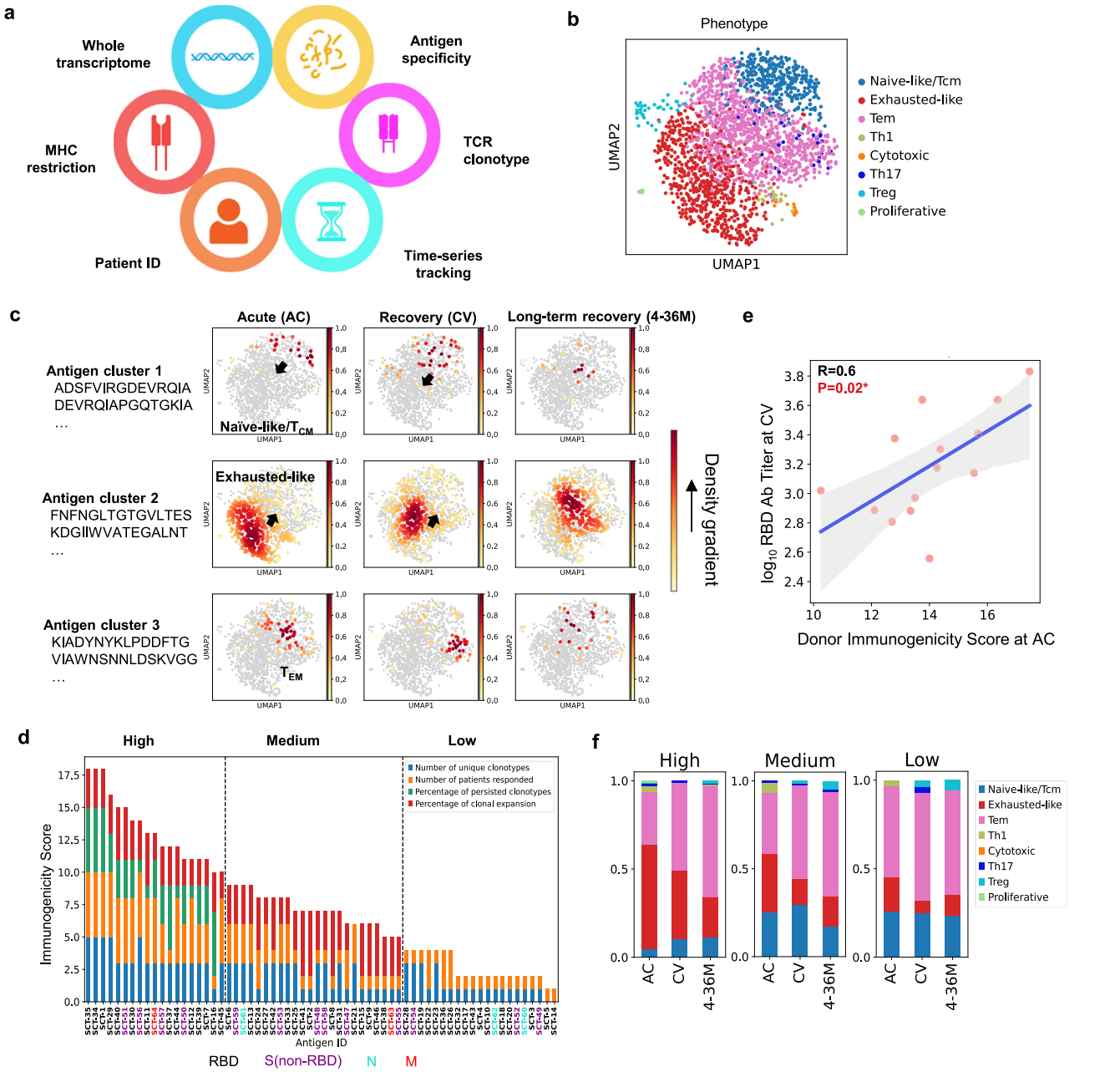

Fig. 3 | Multi-modal, high-dimensional profiling empowers deep characterization of CD4⁺ T cells at single-cell resolution. a Graphical description of the high-dimensional data extracted from each single antigen-specific CD4⁺ T cell. Created in BioRender[76]. **b** UMAP projection of single-cell transcriptomic profiles from SARS-CoV-2-specific CD4⁺ T cells. **c** Longitudinal trajectories of CD4⁺ T cell phenotypes stratified by antigen clusters. Antigens are grouped based on early response patterns at timepoint AC, and the corresponding CD4⁺ T cell phenotypes are tracked over time to reveal the immunological dynamics for each cluster. **d** Immunogenicity ranking of DRB1*01:01-restricted SARS-CoV-2 antigens. Each property contributing to immunogenicity was scored on a $\log_2$ scale (0 to 5) and summed to generate a composite score. **e** Correlation of donor immunogenicity score with antibody level. Scatter plots fitted with linear regression lines and 95% confidence intervals (gray shaded areas) showing the correlation of donor immunogenicity score with RBD antibody titers. Statistical significance was determined using a two-sided Pearson correlation; correlation coefficients (*R*) and *P* values are shown. **f** Phenotypic evolution of CD4⁺ T cells induced by antigens of low, medium, and high immunogenicity. Source data are provided as a Source Data file.

making them ideal targets both for TCR profiling and for providing insights into class I and II antigen recognition. While HPV-specific CD8⁺ TCR-T therapies have demonstrated clinical benefit[58–61], the contribution of HPV-reactive CD4⁺ T cells—despite growing evidence of their critical roles in anti-tumor immunity[3]—remains poorly defined.

Unlike class I MHC, which typically presents peptides of fixed length, class II MHC has an open binding pocket that accommodates variable-length antigens consisting of a 9-mer core flanked at the N- and/or C-termini by peptide flanking regions (PFRs). Recent studies suggest that PFRs may modulate CD4⁺ TCR recognition, yet their functional relevance remains poorly understood[62–65]. To investigate this concept, we designed a comprehensive 92-element SCT library spanning the full E6 and E7 protein, covering all possible lengths (13-25-mer) of putative antigens, with special focus on screening peptide families of varying lengths that share the same core antigen (Fig. 4a, Methods, and Supplementary Data 9). Eighty-five SCTs were prepared in usable quantities after purification for downstream screening. We included an additional 11 common viral antigen-specific SCTs (CEF) as

https://doi.org/10.1038/s41467-026-72396-7

positive controls to complete a 96-element SCT library. We tested whether sensitivity of individual SCTs is maintained in large SCT pools by evaluating the recovery of spiked-in C-3 Jurkat cells into H2 Jurkat cells at 1, 0.1, and 0.01% frequencies (Supplementary Data 10 and Supplementary Fig. 1h). Flow cytometry analysis showed that 70–90% of C-3 cells were successfully recovered at all conditions, demonstrating that sensitivity is preserved even when pooling up to 96 SCTs, and confirming the scalability of the platform.

We applied this library to longitudinal, participant-matched PBMCs from patients with HPV-16+ CIN2/3 enrolled in a neoadjuvant protocol of a therapeutic HPV-16 vaccination study. The vaccine regimen included a heterologous DNA-prime, recombinant vaccinia boost targeting HPV-16/18 E6 and E7, with and without topical application of the TLR7/8 agonist, imiquimod (Aldara®). PBMCs were collected at baseline, post-vaccination, at resection, and four weeks postoperatively ($n = 3$), with one additional donor from an

Fig. 4 | Whole-protein screening of HPV-16 E6 and E7 and TCR repertoire profiling uncovers diverse CD4 TCR functionality in precancerous patients. **a** Antigen coverage across oncogenic proteins E6 and E7 in the SCT library for whole-protein screening. Left panel schematics created in BioRender[76]. **b** Schematic overview of TCR repertoire profiling integrated with other key modalities. Created in BioRender[76]. **c** Antigen specificity validation of HPV-specific CD4 TCRs (H1–H5) through tetramer binding and peptide-pulsed activation assays ($n = 3$). ****$P < 0.0001$, *$P < 0.05$, ns $P > 0.05$ labeled for each group relative to the negative control peptide. **d** Polyfunctionality of TCR-transduced primary CD4$^+$ T cells assessed by cytokine production (IFNγ, TNF, IL2, and GZMB) via ELISA assay following antigen stimulation ($n = 3$). OD_{450} values were normalized to TCR expression levels across each cytokine. ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, and unlabeled tiles indicate $P \geq 0.05$, all relative to negative control peptides. **e** Cytotoxicity of TCR-transduced CD4$^+$ T cells against DR1-K562 cells pulsed with cognate HPV-16 E6 peptides, monitored over 42 h ($n = 2$). The heat map summarizes mean target cell killing at the 42-h time point. Non-cognate peptides served as negative controls for each TCR. **f** Cross-reactivity of TCR H2 and H5 against human self-antigens ($n = 2$). H2- and H5-TCR-transduced T cells were co-cultured with LCLs pulsed with peptides identified from BLAST search and motif scans (Supplementary Data 9). H2 TCR used a single 1 μM dose; H5-TCR was dose titrated at 1, 2, and 10 μM. **g** Alloreactivity screen. H2- and H5-TCR-transduced CD4$^+$ T cells were tested against 41 LCL lines representing 92 distinct class II HLA alleles ($n = 2$). DRB1*01:01-positive LCLs ($n = 4$) pulsed with peptide L-2 ($E6_{91–107}$) or F-2 ($E6_{129–142}$) served as positive controls. Alloreactivity of H5 to HLA-DRB1*13:05-positive LCLs is marked with arrows. **h** Confirmation of H2 TCR reactivity to naturally processed E6 antigen. DRB1*01:01+ LCL (GM17281B) was titrated with full-length E6 protein (0–10 μg/mL), co-cultured with H2 TCR-transduced CD4$^+$ T cells, and evaluated for IFNγ secretion ($n = 2$). A DRB1*01:01- LCL (GM12244A) served as a negative control. Statistical significance was determined by a one-tailed independent *t*-test assuming equal variances. All replicates are biological replicates. All bar plots where $n = 3$ are presented as mean values ± SD. Source data are provided as a Source Data file.

artesunate treatment trial ($n = 1$)[66] (Supplementary Fig. 4a and Supplementary Data 11). Samples were prelabeled with hashtag antibodies for patient and timepoint demultiplexing, stained with DNA-barcoded SCT dextramers and surface protein antibodies, sorted, and subjected to single-cell sequencing to capture TCR repertoire, transcriptome, antigen specificity and surface protein expression (Fig. 4b, Methods). UMAP clustering annotated distinct CD4 + T cell phenotypes (Supplementary Fig. 4b, d). Integration of data across modalities was validated through assessing the accuracy of alignment between transcriptomic and surface protein expression (Supplementary Fig. 4e). We also identified CEF-specific TCRs that matched or clustered with publicly available TCRs through GLIPH analysis[36,67], further benchmarking the accuracy of large SCT pools in capturing cognate TCRs (Supplementary Data 12, 13). Notably, the exhausted-like phenotype, which was prominent (and known[48]) in the SARS-CoV-2-specific CD4$^+$ T cells (Fig. 3b) were absent in this context.

Across the cohort, we identified 112 HPV-16-specific CD4$^+$ T cells. We selected a subset of 25 CD4 TCRs for in-depth functional characterization, with a preference for clonally expanded and persistent clonotypes, and identified five CD4 TCRs (H1–H5) as E6-specific (Supplementary Fig. 4c and Supplementary Data 14). The epitopes for TCRs H1, H3, H4, and H5 all share the same core, while H1, H2, and H4 are each cross-reactive to antigens sharing the same core (Supplementary Data 14), underscoring the role of the core epitope in TCR engagement[68]. We tested these TCRs through an increasingly sophisticated set of binding and functional assays. We first validated their specificity through tetramer binding (Fig. 4c, left). Peptide-pulsed activation of TCR-transduced NFAT-GFP Jurkats revealed that tetramer binding does not always translate into downstream activation, at least in this assay (Fig. 4c, right). This result highlights the critical role of PFRs in modulating functional outcomes for these two TCRs.

To further investigate their therapeutic potential, we transduced these five TCRs into TCR^{KO} primary CD4$^+$ T cells enriched from healthy donors and confirmed their specificity via tetramer staining (Supplementary Fig. 4f, g). Notably, the knock-in efficiency was markedly lower in CD8$^+$ than CD4$^+$ T cells (Supplementary Fig. 4h). We then evaluated these TCRs through additional functional assays. We first assessed a suite of cytokine secretion (IFNγ, TNF, IL2, and GZMB) in response to peptide-pulsed DR1+ K562 cells (Fig. 4d). TCRs H2, H3, and H5 displayed on-target secretions of all cytokines, while H4 showed minimal secretion for most peptides except F-7, despite strong tetramer binding to the broader peptide family sharing the same core (F-3 to F-7). These results were consistent with cytotoxicity assays in which TCR-transduced CD4$^+$ T cells, with varying expression levels (Supplementary Data 15), were co-cultured with DR1+ K562 cells pulsed with cognate peptides and labeled with Cytolight Rapid Dye (Fig. 4e). TCRs H2 and H5 exhibited the most consistent functional activation across all assays, and so were advanced for additional pre-clinical evaluations.

We next assessed TCRs H2 and H5 for target selectivity. We performed alanine scanning of their 9-mer core epitopes, identifying the key recognition motifs of *xxxxYNKPx* (H2) and *FHNxRGRWx* (H5) (Supplementary Fig. 4i). BLAST-based homology searches identified human self-antigens with similar motifs (Supplementary Data 16). Co-culture assays using HLA-matched, EBV-transformed B-lymphoblastic cell lines (LCLs) pulsed with these peptides showed no evidence of cross-reactivity, indicating strong target specificity (Fig. 4f).

We then screened for alloreactivity to common class II HLAs using a panel of 41 LCLs expressing 92 distinct class II HLA alleles. H5 exhibited reactivity toward two LCL lines sharing the HLA-DRB1*13:05 allele, whereas H2 demonstrated no evidence of alloreactivity (Fig. 4g and Supplementary Data 17). Finally, titration of the full-length E6 protein pulsed to a DRB1+ LCL line showed functional activation of H2-transduced T cells, confirming that the L-1 antigen is naturally processed and presented by DRB1+ 01:01 (Fig. 4h). Collectively, the H2 TCR emerges with strong therapeutic potential for clinical translation.

These results demonstrate the power of class II SCT libraries for large-scale screening of functional CD4 TCR repertoires in cancer, and also highlight the need for a multi-step analytic pipeline for CD4 TCR evaluation. This approach enables precise mapping of peptide families, functional interrogation of PFRs, and identification of therapeutically relevant CD4 TCRs with minimal off-target risk—laying the groundwork for translational CD4 TCR-T immunotherapies.

Discussion

CD4$^+$ T cells are essential orchestrators of adaptive immunity, with critical roles in infectious disease, cancer, and autoimmunity. However, systematic and deep characterization of antigen-specific CD4$^+$ T cell responses at scale remains technically challenging. Conventional approaches either have limited throughput or lack resolution of key biological parameters, including MHC restriction, phenotypic state, and patient/timepoint identity.

We reported here APMAT-CD4 as a scalable platform using class II SCT libraries to enable high-throughput, whole-protein screening for antigen-specific CD4$^+$ T cells while co-profiling for epitope specificity, MHC restriction, TCRα/β genes, whole transcriptome, patient identity, and time-series tracking. We benchmarked the performance of individual SCTs, showing a comparable specificity and sensitivity to conventional pMHCs, and demonstrated the value of SCT libraries for discovering antigen-CD4 TCR pairs in healthy donors. To demonstrate the power of the platform in high-throughput multimodal profiling, we performed full-length screening of the SARS-CoV-2 spike receptor-binding domain in a longitudinal cohort of COVID-19 participants. The resulting high-dimensional dataset permitted comprehensive kinetic analysis of antigen-specific CD4$^+$ T cells, and

informed a multi-component metric for defining class II-restricted immunodominance. Given the importance of CD4$^+$ T cells in B cell maturation, the ability to quantitate immunogenic viral antigens should provide a compelling approach towards enabling rational and precision vaccine design. The five most immunogenic antigens identified in this study are all RBD-derived epitopes—a trend consistent with previous findings highlighting the critical role of anti-RBD IgGs in conferring resistance to SARS-CoV-2[38]. Notably, participants who exhibited the strongest CD4$^+$ T cell responses to RBD antigens also exhibited the highest titers of anti-RBD antibodies. This correlation may suggest that a comprehensive analysis of viral antigen-specific CD4$^+$ T cell response can provide useful guidance for vaccine design.

We further demonstrated the versatility of the platform for CD4 TCR repertoire profiling in cancer settings through a whole-protein screen for HPV-16 E6 and E7 specific TCRs in therapeutically vaccinated patients with precancerous HPV-16+ lesions. These findings highlight that while the SCT platform captures CD4 TCR-antigen binding activities, the biological behaviors of each TCR binder need to be thoroughly evaluated through a series of functional assays (Figs. 1i, 2f, and 4c–h and Supplementary Fig. 4e–h). In particular, the peptide flanking residues of a core antigen can exert a profound influence on T cell functional activation (as reported elsewhere[64]), which is distinct from what is commonly observed for CD8$^+$ T cells specific to class I-restricted epitopes.

A limitation of APMAT-CD4 is our primary focus on the single class II HLA allele, DRB1*01:01, although the results of Fig. 1c and Supplementary Fig. 1c suggest that this is not an intrinsic limitation of the method. Nevertheless, expanding SCT libraries to cover additional class II HLA alleles is an important step towards developing a more comprehensive analysis of antigen-specific T cell responses, but may require template optimization for certain alleles. Such work is currently underway.

Collectively, we demonstrated APMAT-CD4 as a versatile, high-throughput platform that empowers large-scale identification and multi-modal characterization of antigen-specific CD4$^+$ T cells, and permits many HLA-matched PBMC biospecimens to be analyzed in parallel. APMAT-CD4 resolves key biological variables at single-cell resolution and offers a generalizable solution for dissecting antigen-specific CD4$^+$ T cell responses in diverse contexts, including infectious diseases, tumor immunology, autoimmunity, vaccine design, and fundamental T cell immunobiology.

Methods

Ethical statement

The study cohort is a subset of the INCOV and UNCOVR cohorts published previously[10,48,57]. The research presented here complies with all relevant ethical regulations. Procedures for the INCOV study were approved by the Institutional Review Board (IRB) at Providence St. Joseph Health with IRB study number STUDY2020000175 and the Western Institutional Review Board (WIRB) with IRB study number 20170658. This research complies with all relevant ethical regulations. All enrolled participants provided written in-person informed consent, and samples were de-identified prior to analysis. The samples from the HPV vaccine study were collected from a phase 2b trial registered at ClinicalTrials.gov (number NCT01304524) and EudraCT (number 2012-001334-33). The clinical trial protocols were reviewed and approved by the Institutional Review Board at Johns Hopkins Hospital (NA_00002176, NA_00023308). All study participants gave written informed consent before undergoing screening for study eligibility and enrollment. Twenty-two HLA-DRB1*01:01 individuals from the INCOV and UNCOVR cohort (12 females and 10 males) were selected for this study. PBMCs collected at acute timepoint (AC, <1 week of infection), convalescence (CV, 2–3 months post-acute) to, for some donors, long-term follow-up (6 to 36 months), were used for antigen-TCR pairing assay in this study. No covariate-relevant analysis related to this manuscript.

Primer auto-generation algorithm to generate primer sets encoding the antigen fragment of SCTs

The algorithm (https://github.com/rzhang101/classII_sct_primer_auto_generation.git) takes in a user-defined list of antigens in one-letter coded amino acid sequences. An initial gene fragment including the DNA sequences of the antigen generated through reverse translation and codon optimization, the pIi domain, and L2 are constructed. An initial set of forward and reverse primers are designed to allow at least 16 base pairs (bp) of hybridization. The set of primers is then tested through three checkpoints to confirm PCR effectiveness: (1) if the melting temperature of the 20 bp hybridization region is between 50 to 69 °C using nearest neighbor thermodynamics; (2) if at least 16 bp are annealed at the annealing temperature and if the annealed primers are the majority by analyzing the primer structures through NUPACK[69]; (3) if the two primers have at least 16 bp annealed at the extension temperature and if a hairpin structure exists that might hinder the extension process. If any of the checkpoints fail, the primers undergo a random selection of alternative codons until a codon combination passes all checkpoints. The algorithm outputs a list of forward and reverse primers ready for purchase. All primers used to construct the antigen fragments of the SCTs used in the manuscript are listed in Supplementary Data 1.

Class II SCT expression

The plasmid used for the expression of class II SCTs was built based on a pcDNA3.1 vector (Invitrogen, V86020). An SCT plasmid template encoding a class II allele was assembled by a two-fragment Gibson assembly, each encoding for the alpha or beta chain of the allele. The antigen fragment was produced by an overlap extension PCR of the forward and reverse primers designed by the primer auto-generation algorithm (IDT). The PCR was carried out in a 20 uL reaction containing 1.5 pmol of each primer, 7 uL nuclease-free water, and 10 uL KOD Hot Start Master Mix (Millipore Sigma, 71842). The amplified fragment was purified (Qiagen, 28104) and Gibson (New England Biolabs (NEB), E2621S) assembled into the plasmid backbone containing the HLA allele in a 5:1 insert to vector ratio or 0.25 pmol per fragment and 0.05 pmol of vector. The Gibson product was then transformed into Top10 competent cells (Invitrogen, C404003) and the colonies were picked and cultured, followed by extraction of the amplified plasmids (27104). Plasmid sequences were confirmed by Genewiz, MCLAB, or Plasmidsaurus.

The mammalian cell-based Expi293 transfection system (Gibco, A14635) was used for SCT expression. SCT transfection was performed in a 24-well plate format for high-throughput production. Briefly, for each SCT, 1.25 mL of Expi293 cells at 3 M/mL were seeded and transfected with the SCT plasmid packaged based on the manufacturer's protocol. After 18–24 h of incubation, Enhancer 1 and 2 were added, followed by an incubation of 72 hours. The secreted SCT proteins were collected four days after the initial transfection. An aliquot of the supernatant was denatured to quantify the SCT expression level through SDS-PAGE (Bio-Rad, 4568035). The relative intensity of the SCT expression, subtracting the negative control, was recorded. Supernatant of expressed SCTs were concentrated and buffer-exchanged with 20 mM bicine (pH 8.3) (Bufferad, B2239-83), and then biotinylated for 1.5 h at room temperature and then overnight (Avidity, BirA500). The SCTs were his-tag purified on the next day (Biotage, PTR-92-20-03) and desalted into PBS (Thermo Fisher, 89883). Concentration of the SCT protein was measured through nanodrop and stored in 20% glycerol/PBS. The CEFT SCT library was designed based on the antigen list used for ELISpot stimulation on healthy donor PBMC samples from ImmunoSpot. Examples of SCTs on protein gels are included in the Source Data.

N-linked de-glycosylation assay

SCTs were de-glycosylated according to the manufacturer's protocol (NEB, P0704). Specifically, 2 ug of each purified SCT was incubated at 100 °C for 10 min with the 10X denaturing buffer, and chilled on ice. PNGase enzyme, 10% NP-40, GlycoBuffer, and water were mixed well and incubated at 37 °C for 1 h. The PNGase-processed proteins were then reduced and run on an SDS-PAGE protein gel in comparison with non-PNGase-processed SCTs.

Cell line and media

Wild-type Jurkat E6-1 was purchased from ATCC (TIB-152). TCRb followed by TCRa knock-out was performed using CRISPR gRNA (IDT) and nucleofection, confirmed through flow cytometry, and sorted for purity. A plasmid encoding for NFAT-GFP reporter (VectorBuilder) was then lentivirally transduced into the TCR- Jurkats and treated with blasticidin (InvivoGen, ant-bl-05) for selection. DR1+ K562 cells were established from WT K562 (ATCC, CCL-243), which lack functional expression of wild-type class II pMHC[70], by lentivirally transducing a plasmid encoding the DRA/DRB1*01:01 allele with the transmembrane and cytoplasmic tail domains. Expression of DR1 on the surface was confirmed through flow cytometry, and FACS-sorted for purity. HEK293T cells were purchased from ATCC (CRL-3216). Healthy donor PBMC samples were purchased from ImmunoSpot.

TCR cloning

Fragments encoding for the V(D)JC regions of alpha and beta chains were designed, codon-optimized and purchased from Twist Bioscience. The fragments were PCR amplified, and Gibson assembled into a pRRL plasmid (generously provided by Dr. Philip Greenberg, Fred Hutchinson Cancer Center) for lentiviral production. The sequence-verified plasmid DNA was transfected into the HEK293T cell line along with packaging plasmids to produce lentiviral particles (Qiagen, 301425), which were transduced into the target cell line. 48–72 h after removal of the virus soup, the TCR expression was measured through flow cytometry.

Tetramer binding assay

Tetramers were prepared by incubating the SCTs with a fluorophore-labeled streptavidin (Invitrogen, SA10041, SA1005, SA1012) in a 4:1 ratio. Usually, 2.5 pmol of SCT and 0.625 pmol of streptavidin was used to generate tetramers for one staining reaction. After a 30-min incubation on ice, 1 uL of 20 uM biotin was added to the tetramers, followed by an incubation of at least 30 min on ice. Cells were incubated in 50 nM Dasatinib (PKI) (Adooq, A10290) at 37 °C for 20–30 min and then centrifuged. Cells were then stained with tetramer at 4 °C for 15–25 min in 50 nM PKI and washed, followed by live dye (Invitrogen, C34858) and TCR-APC (Biolegend, 306718, clone IP26) staining. Cells were then washed and analyzed through flow cytometry.

K-mer regression association assay

Antigens from SCT libraries used throughout the paper (CEFT library, SARS-CoV-2 library, and HPV-16 E6 and E7 library) were used in the analysis. Antigens were decomposed into 2-mers, and their correlation with SCT expression levels was assessed via Spearman correlation, followed by FDR correction using the Benjamini–Hochberg method.

Antigen-specific CD4 T cells staining

The DRB1*01:01 SCT with antigen derived from the HIV Gag protein (HIV $Gag_{41\text{-}56}$ SALSEGATPQDLNTML) was used as the negative control for SCT tetramer staining. SE buffer for cell and tetramer staining was prepared by adding 0.5% BSA and 2 mM EDTA and filtered. PBMCs were thawed in R10 media and pre-enriched for $CD4^+$ T cells using magnetic-activated cell sorting (MACS) according to the manufacturer's protocol (Miltenyi, 130-096-533). In the final step of CD4 + T cell enrichment, cells were centrifuged and resuspended in 1 mL of media. The enriched CD4 + T cells were then incubated in 50 nM PKI at 37 °C for 20 min, followed by centrifugation at 500×*g* for 5 min. Staining solution for each antigen was prepared by adding 2 μL of PE and APC tetramers, along with 2 μL of HIV PE/Cy7 tetramer, to 100 μL of 50 nM PKI/SE buffer. Each sample was resuspended in 100 μL of the tetramer staining solution, incubated at 4 °C for 20 min, washed and centrifuged. A cell surface staining buffer was prepared by adding 1 uL of CD4-BV421 (Biolegend, 317434, clone OKT4) per 100 μL of buffer per 50k CD4 + T cells, and Calcein Green (Invitrogen, C3099) live stain at a final concentration of 0.2 μM. The cells were incubated for 10 min at 4 °C. After staining, the cells were washed with 100 μL of SE buffer to remove excess reagents, centrifuged, and resuspended in 200 μL of PBS.

Single-cell plate sequencing for TCR sequences

The cells were single-cell sorted into 96-well plates containing 12 uL of 1x lysis buffer (Takara, 635013) with RNAse inhibitor (Promega, N2515). The plates were spun down immediately after sorting and flash frozen on dry ice. When ready for sequencing the TCRs, the lysate plate was thawed on ice for 3 min and centrifuged at 1000×*g* for 1 min. A master mix of one-step RT-PCR reaction was prepared according to the manufacturer's protocol (Qiagen, 210212). Briefly, the following reagents were mixed for one reaction: 3.2 uL 5x buffer, 0.64 uL dNTPs at 10 mM, 0.64 uL enzyme mix, 5.12 uL RNAse-free water, 0.55 uL of a mixture of the alpha or beta variable primers (1.75 uM), and 0.48 uL of alpha or beta constant primer at 10 uM. Each lysate was split into two reactions for amplifying the alpha and beta sequences separately. Following RT-PCR, cleanup was performed using 0.8x SPRI beads (Beckman Coulter, A63881). A second PCR further amplified the TCR fragment and extended the fragment with overlap regions for the third PCR that added in the barcoding regions for multiplexed sequencing through Illumina.

Peptide stimulation assay

Peptides were purchased from GenScript and reconstituted in DMSO. 50k TCR+ NFAT-GFP Jurkat cells were co-cultured with DR1-K562 cells at a 1:1 ratio with 2 μg/mL peptides in 100 μL R10 media and incubated at 37 °C for 16–24 h. Cells were then stained with live-cell dye Calcein UV (Invitrogen, C34858), TCR-APC (Biolegend, 306718, clone IP26), and HLA-DR-PE (BioLegend, 307605, clone L243) and analyzed through flow cytometry.

Prediction for SARS-CoV-2 RBD antigen binding to HLA-DRB1*01:01

Predictions for binding affinities and percentile rank of eluted ligands of the HLA-DRB1*01:01-restricted antigens in the SARS-CoV-2 receptor-binding domain SCT library (Wuhan-Hu-1 strain, GenBank: QHD43416.1) were generated through NetMHCpan4.0[71].

High-throughput screening of antigen-specific $CD4^+$ T cells

SCT-dextramer was prepared by mixing SCT and a DNA-barcoded dextramer-PE (ImmuDex) in a 22:1 molar ratio and incubated on ice for 30 min before adding 10 pmol of biotin for another 30 min incubation on ice. The dextramers were pooled together with negative control tetramer-PE-Cy7s. PBMC samples were stained with hashtag antibodies (Biolegend, custom TotalSeq anti-CD298, clone LNH-94) based on patients and/or timepoints. After two washes, PBMCs were pooled for MACS-based $CD4^+$ T cell enrichment through negative selection (Miltenyi, 130-096-533). Isolated $CD4^+$ T cells were incubated in 50 nM PKI solution at 37 °C for 20–30 min, followed by dextramer pool staining in 50 nM PKI for 20–25 min at 4 °C. After two to three washes, the cells were stained with 1:5000 dilution of live dye (Biolegend, 427402), 1 ul per 200k $CD4^+$ T cells of CD4-BV421 (BioLegend, 317434, clone OKT4), and 1 uL per 200k $CD4^+$ T cells of TCR-APC (BioLegend, 306718, clone IP26) for 10 min at 4 °C. The cells were then FACS-sorted, washed and

loaded onto Chromium Next GEM chips for single-cell sequencing (10x Genomics). Gene expression, TCR, and surface protein libraries were prepared and sequenced on Illumina NextSeq2000.

Single-cell sequencing data analysis

The transcriptome, TCR sequences, surface protein levels, and antigen specificity were simultaneously assessed for each cell. Illumina bcl2fastq2 was used to generate fastq files from bcl files. Raw data were processed using the Cell Ranger Single-Cell Software Suite (v3.1.0, 10X Genomics) with GRCh38 as the reference genome. Gene expression data (GEX) were quality-controlled to eliminate low-quality cells: cells with less than 200 or above 5000 unique genes or a mitochondrial content above 15%. Any genes detected in less than three cells were also removed. Gene expression counts for each cell were normalized by the total expression, scaled by a factor of 10,000, and transformed to a log scale. All cells that passed the quality control were clustered through PCA and KNN graphs based on the highly variable genes and projected onto UMAPs. Phenotypes were defined based on enrichment of signature gene sets and assigned to single cells based on clusters identified using the Leiden algorithm. Single pair and orphan beta TCRs were selected and matched to the GEX data. In the SARS-CoV-2 study, patient identity was deconvoluted through comparison of single-nucleotide polymorphisms (SNPs) in transcriptome data with WGS profiles of all patients (described in detail in the Single-cell demultiplexing using genetic variants section). Any doublet cell was identified based on SNPs assigned to multiple patients and excluded from the analysis. DNA sequences encoded for hashtags were analyzed to determine the time origin of SARS-CoV-2 samples or both the time origin and the patient identity for the HPV samples. Raw reads for each hashtag and dextramer were normalized. The time origin and/or patient identity of a cell was assigned based on the hashtag with the highest expression, provided it exceeded the second-highest hashtag by at least 5%. Cells were assigned a specific antigen only if they had a dextramer UMI representing more than 20% of the cell's total dextramer reads. Cells that could not be definitively assigned to a specific dextramer were discarded.

Single-cell demultiplexing of patient origin using genetic variants

Patient identity was resolved by comparing single-nucleotide polymorphisms (SNPs) in transcriptome data with WGS patient profiles, using previously described bioinformatics pipeline[72]. In short, a multi-donor genotype reference of exon single-nucleotide variants (SNV) was constructed from WGS data using GATK (v4.1.9.0)[73]. The same variants were determined for each single cell using 10x Genomics VarTrix (https://github.com/10XGenomics/vartrix). Finally, vireo (v0.5.0)[74] compares single-cell SNVs with the donor genotype reference to assign each single cell to a specific donor or as a doublet.

Phenotype evolution analysis

Antigens were clustered based on the phenotype distribution of their corresponding CD4 + T cells at the initial infection time point (AC). Antigens with fewer than three corresponding T cells were excluded from the analysis. For each antigen, the dominant phenotype of antigen-specific CD4 + T cells was identified and assigned as its initial phenotype cluster. Phenotype distribution of the T cell responses induced by the antigen clusters were analyzed by calculating the density of the cells using Gaussian kernel density estimation at each timepoint and projected onto a UMAP.

Immunogenicity score analysis

Four unique properties were assessed for each antigen: (1) the number of donors reactive to the antigen, (2) the number of unique CD4 TCR clonotypes specific to the antigen (normalized for the number of donors), (3) the percentage of unique T cell clonotypes that persisted across at least two time points, and (4) the percentage of clonally expanded TCR clonotypes (adjusted for persistent and cross-donor TCRs). All properties were adjusted to minimize confounding influences. For each property, antigens were ranked from highest to lowest, divided into groups based on a $\log_2$ scale, and assigned scores ranging from 0 to 5 (Supplementary Fig. 3c). The total scores for each antigen were summed, and immunogenicity ranks (high, medium, or low) were assigned based on a $\log_2$ scale of the overall score. To evaluate if there is a correlation between the immunogenicity score and functional outcome, we collected data from our previous study, which measured the SARS-CoV-2 RBD-specific antibodies and neutralizing antibodies of the same patients[57]. To account for differences in antigen breadth across patients, we calculated a weighted average immunogenicity score for each patient (patient score = Σ (immunogenicity_score_antigen_i × frequency_antigen_i)/total_response). Pearson correlation analysis was used to evaluate associations between patient immunogenicity score and RBD antibody titers.

HPV-16 E6 and E7 antigen selection

The class II SCT library encoding the HPV-16 E6 (UniProt ID: P03126) and E7 (UniProt ID: P03129) proteins encompasses 92 antigens varying from 13 to 25 amino acids. The antigens were selected based on the NetMHCIIpan4.0[71] and MixMHC2pred[75] predictions. The top 15% of the antigens predicted to bind the DRB1*01:01 allele were selected. Out of the 92 selected antigens, 85 SCTs were successfully expressed with a concentration applicable for downstream applications. The SCTs were constructed into a pool of dextramers where each DNA barcode corresponds to a unique SCT.

HPV-TCR+ primary CD4 T cell construction

$CD4^+$ T cells were extracted from healthy donor PBMCs via MACS sorting and resuspended at a density of 1 M/mL in Prime-XV media (Fujifilm Irvine Scientific, 91154) with 2% Physiologix serum replacement supplemented with IL-7 (Biolegend, 581904) and IL-15 (BioLegend, 570304), both at a final concentration of 12.5 ng/mL. Each 1 M cells were activated by adding 10 μL of TransAct human CD3/CD28 activator (Miltenyi, 130-111-160,) and incubated for 48 h. After activation, endogenous TCRs were knocked out through CRISPR-Cas9 gene editing and target TCRs were transduced into the primary CD4 + T cells through lentivirus.

ELISA

$DR1^+$ K562 cells were preloaded with peptide for a final concentration of 0.1 mg/mL and incubated at 37 °C for 1–2 h. Excess peptides were washed and removed. Peptide-loaded K562 cells were co-cultured with primary $CD4^+$ T cells in a 1:1 effector-to-target (E:T) ratio for 16–20 h. ELISA assays for IL2, INFγ, TNF, and GZMB were performed according to the manufacturer's manual (R&D Systems, DY202-05, DY285B-05, DY210-05, and DY2906-05).

Cytotoxicity assay

The cytotoxic assay was conducted through IncuCyte. DR1-K562 cells were preloaded with the target peptide in a similar way to the ELISA assays and seeded at 1 M/mL. 10 μL of the 100X working solution of Cytolight Rapid Dye (Sartorius, 4705) was used for every 1 M K562 cells. The cells were incubated for 20 min at 37 °C, and gently mixed every 10 min to allow the dye to fully bind. Any excess dye was washed out, and the supernatant was removed. The cells were then resuspended to a concentration of 500k cells/mL, and 50k cells were seeded and allowed to settle at room temperature for 30 minutes. The apoptosis reagent was prepared by diluting Caspase-3/7 green (Sartorius, 4440) to a final concentration of 20 μM (4x) in culture media and added to cells. The immune effector cells (primary TCR-T cells) were added to the DR1-K562 cells at an E:T ratio of 1:1. The assay plate was

placed in the IncuCyte Live-Cell Analysis System for a 44-h repeat scanning.

Alanine scanning

Peptides were generated for the $E6_{91–107}$ epitope (TCR H2) and $E6_{129–142}$ epitope (TCR H5) with each position replaced with an alanine amino acid (Genscript). Healthy donor T cells, expressing TCR H2, TCR H5, were co-cultured with a DRB1+ K562 cell line that were previously pulsed with a dose titration (0.001–1 μM) of the respective peptides. After ~20 h, supernatants were harvested from the co-cultures and measured for cytokine secretion (R&D Systems, DY285B-05).

Cross-reactivity assessment

BLAST search and motif scan were conducted to identify human self-antigens containing motifs similar to those revealed by alanine scanning. Healthy donor T cells expressing the HPV-TCR of interest were then co-cultured with an HLA-matched EBV-transformed B-lymphoblastic cell lines (LCLs) (Coriell Institute for Medical Research). For the TCR H2, a single dose (1 μM) of each peptide was used. For the TCR H5, 1, 2, 10 μM of each peptide was used. After ~20 h, supernatants were collected, and IFNγ cytokine secretion was determined using the ELISA assay.

Alloreactivity assay

TCR-transduced primary $CD4^+$ T cells were co-cultured overnight with each of the 41 EBV-transformed B cell lines (LCLs) and assessed for IFNγ secretion. As positive controls, four DRB1*01:01+ LCL cell lines were pulsed with 10 uM of $E6_{91–107}$ (TCR H2) or $E6_{129–142}$ (TCR H5) peptides, respectively, and IFNγ secretion was measured.

E6 protein pulsed T cell functional activation assay

A DRB1*01:01+ LCL (GM17281B) was pulsed with E6 protein (Rndsystems, AP-120) at 10, 2, 1, and 0 μg/mL for 3 h, and then co-cultured with H2 TCR-transduced $CD4^+$ T cells overnight. IFNγ cytokine secretion in the supernatant was determined using the ELISA assay. A DRB1*01:01 negative LCL (GM12244A) was used as the negative control.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The single-cell sequencing data generated in this study have been deposited in the GEO repository under accession number: GSE324489, or using the URL: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE324489. Source data in this study are provided in the Supplementary Data/Source Data file. Any additional information required to reanalyze the data reported in this work paper is available from the Corresponding Author upon request. Any additional data in this study are provided in the Supplementary Information. Source data are provided with this paper.

Code availability

The primer auto-generation algorithm for constructing the antigen domain of class II single-chain trimers are deposited on GitHub (https://github.com/rzhang101/classII_sct_primer_auto_generation.git).

References

1. Luckheeram, R. V., Zhou, R., Verma, A. D. & Xia, B. CD4+T cells: differentiation and functions. *Clin. Dev. Immunol.* **2012**, 925135 (2012).
2. Tay, R. E., Richardson, E. K. & Toh, H. C. Revisiting the role of CD4+ T cells in cancer immunotherapy—new insights into old paradigms. *Cancer Gene Ther.* **28**, 5–17 (2021).
3. Oh, D. Y. & Fong, L. Cytotoxic CD4+ T cells in cancer: Expanding the immune effector toolbox. *Immunity* **54**, 2701–2711 (2021).
4. Gangaev, A. et al. Identification and characterization of a SARS-CoV-2 specific CD8+ T cell response with immunodominant features. *Nat. Commun.* **12**, 2593 (2021).
5. Bentzen, A. K. et al. Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat. Biotechnol.* **34**, 1037–1045 (2016).
6. Ma, K.-Y. et al. High-throughput and high-dimensional single-cell analysis of antigen-specific CD8+ T cells. *Nat. Immunol.* **22**, 1590–1598 (2021).
7. Chour, W. et al. Large libraries of single-chain trimer peptide-MHCs enable antigen-specific CD8+ T cell discovery and analysis. *Commun. Biol.* **6**, 1–13 (2023).
8. Puig-Saus, C. et al. Neoantigen-targeted CD8+ T cell responses with PD-1 blockade therapy. *Nature* **615**, 697–704 (2023).
9. Foy, S. P. et al. Non-viral precision T cell receptor replacement for personalized cell therapy. *Nature* **615**, 687–696 (2023).
10. Xie, J. et al. APMAT analysis reveals the association between CD8 T cell receptors, cognate antigen, and T cell phenotype and persistence. *Nat. Commun.* **16**, 1402 (2025).
11. Moravec, Z. et al. Discovery of tumor-reactive T cell receptors by massively parallel library synthesis and screening. *Nat. Biotechnol.* **43**, 214–222 (2024).
12. Frentsch, M. et al. Direct access to CD4+ T cells specific for defined antigens according to CD154 expression. *Nat. Med.* **11**, 1118–1124 (2005).
13. Poloni, C. et al. T-cell activation–induced marker assays in health and disease. *Immunol. Cell Biol.* **101**, 491–503 (2023).
14. Le Bert, N. et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**, 457–462 (2020).
15. Uchtenhagen, H. et al. Efficient ex vivo analysis of CD4+ T-cell responses using combinatorial HLA class II tetramer staining. *Nat. Commun.* **7**, 12614 (2016).
16. Kozono, H., White, J., Clements, J., Marrack, P. & Kappler, J. Production of soluble MHC class II proteins with covalently bound single peptides. *Nature* **369**, 151–154 (1994).
17. Day, C. L. et al. Ex vivo analysis of human memory CD4 T cells specific for hepatitis C virus using MHC class II tetramers. *J. Clin. Invest.* **112**, 831–842 (2003).
18. Moon, J. J. et al. Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* **27**, 203–213 (2007).
19. Scriba, T. J. et al. Ultrasensitive detection and phenotyping of CD4+ T cells with optimized HLA class II tetramer staining. *J. Immunol.* **175**, 6334–6343 (2005).
20. Kwok, W. W. et al. Frequency of epitope-specific naive CD4(+) T cells correlates with immunodominance in the human memory repertoire. *J. Immunol.* **188**, 2537–2544 (2012).
21. Christophersen, A. et al. Distinct phenotype of CD4+ T cells driving celiac disease identified in multiple autoimmune conditions. *Nat. Med.* **25**, 734–737 (2019).
22. Vyasamneni, R. et al. A universal MHCII technology platform to characterize antigen-specific CD4+ T cells. *Cell Rep. Methods* **3**, 100388 (2023).
23. Thayer, W. P., Dao, C. T., Ignatowicz, L. & Jensen, P. E. A novel single chain I-Ab molecule can stimulate and stain antigen-specific T cells. *Mol. Immunol.* **39**, 861–870 (2003).
24. Chen, X., Zaro, J. L. & Shen, W.-C. Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev.* **65**, 1357–1369 (2013).
25. Gonzalez-Galarza, F. F. et al. 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 (2020).
26. Maiers, M., Gragert, L. & Klitz, W. High-resolution HLA alleles and haplotypes in the United States population. *Hum. Immunol.* **68**, 779–788 (2007).
27. Deng, L., Langley, R. J., Wang, Q., Topalian, S. L. & Mariuzza, R. A. Structural insights into the editing of germ-line–encoded interactions between T-cell receptor and MHC class II by Vα CDR3. *Proc. Natl. Acad. Sci. USA* **109**, 14960–14965 (2012).
28. Galperin, M. et al. CD4+ T cell–mediated HLA class II cross-restriction in HIV controllers. *Sci. Immunol.* **3**, eaat0687 (2018).
29. Benati, D. et al. Public T cell receptors confer high-avidity CD4 responses to HIV controllers. *J. Clin. Invest.* **126**, 2093–2108 (2016).
30. Hu, Z. et al. Personal neoantigen vaccines induce persistent memory T cell responses and epitope spreading in patients with melanoma. *Nat. Med.* **27**, 515–525 (2021).
31. Ciacchi, L. et al. CD4+ T cell-mediated recognition of a conserved cholesterol-dependent cytolysin epitope generates broad antibacterial immunity. *Immunity* **56**, 1082–1097.e6 (2023).
32. Murthy, V. L. & Stern, L. J. The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding. *Structure* **5**, 1385–1396 (1997).
33. Jiang, W. & Boder, E. T. High-throughput engineering and analysis of peptide binding to class II MHC. *Proc. Natl. Acad. Sci. USA* **107**, 13258–13263 (2010).
34. Glanville, J. et al. Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98 (2017).
35. Rosskopf, S. et al. Creation of an engineered APC system to explore and optimize the presentation of immunodominant peptides of major allergens. *Sci. Rep.* **6**, 31580 (2016).
36. Huang, H., Wang, C., Rubelt, F., Scriba, T. J. & Davis, M. M. Analyzing the *Mycobacterium tuberculosis* immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat. Biotechnol.* **38**, 1194–1202 (2020).
37. Ju, B. et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **584**, 115–119 (2020).
38. Starr, T. N. et al. SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* **597**, 97–102 (2021).
39. Andreano, E. et al. Extremely potent human monoclonal antibodies from COVID-19 convalescent patients. *Cell* **184**, 1821–1835.e16 (2021).
40. Brouwer, P. J. M. et al. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **369**, 643–650 (2020).
41. Pinto, D. et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020).
42. Peng, Y. et al. Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nat. Immunol.* **21**, 1336–1345 (2020).
43. Mateus, J. et al. Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science* **370**, 89–94 (2020).
44. Nelde, A. et al. SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. *Nat. Immunol.* **22**, 74–85 (2021).
45. Wells, D. K. et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell* **183**, 818–834.e13 (2020).
46. Szabo, P. A. et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019).
47. Zhang, W. et al. CMV status drives distinct trajectories of CD4+ T cell differentiation. *Front. Immunol.* **12**, 620386 (2021).
48. Su, Y. et al. Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. *Cell* **183**, 1479–1495.e20 (2020).
49. Andreatta, M. et al. Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat. Commun.* **12**, 2965 (2021).
50. Golubovskaya, V. & Wu, L. Different subsets of T cells, memory, effector functions, and CAR-T immunotherapy. *Cancers* **8**, 36 (2016).
51. Soon, M. S., Engel, J. A., Lee, H. J. & Haque, A. Development of circulating CD4+ T-cell memory. *Immunol. Cell Biol.* **97**, 617–624 (2019).
52. Takeuchi, A. & Saito, T. CD4 CTL, a Cytotoxic Subset of CD4+ T Cells, Their Differentiation and Function. *Front. Immunol.* **8**, 194 (2017).
53. Whitfield, M. L., George, L. K., Grant, G. D. & Perou, C. M. Common markers of proliferation. *Nat. Rev. Cancer* **6**, 99–106 (2006).
54. Harrington, L. E., Janowski, K. M., Oliver, J. R., Zajac, A. J. & Weaver, C. T. Memory CD4 T cells emerge from effector T-cell progenitors. *Nature* **452**, 356–360 (2008).
55. Gaebler, C. et al. Evolution of antibody immunity to SARS-CoV-2. *Nature* **591**, 639–644 (2021).
56. Swank, Z. et al. Measurement of circulating viral antigens post-SARS-CoV-2 infection in a multicohort study. *Clin. Microbiol. Infect.* **30**, 1599–1605 (2024).
57. Su, Y. et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell* **185**, 881–895.e20 (2022).
58. Trimble, C. L. & Frazer, I. H. Development of therapeutic HPV vaccines. *Lancet Oncol.* **10**, 975–980 (2009).
59. Ramondetta, L. What is the appropriate approach to treating women with incurable cervical cancer? *J. Natl. Compr. Canc. Netw.* **11**, 348–355 (2013).
60. Doran, S. L. et al. T-cell receptor gene therapy for human papillomavirus–associated epithelial cancers: a first-in-human, phase I/II study. *J. Clin. Oncol.* **37**, 2759–2768 (2019).
61. Nagarsheth, N. B. et al. TCR-engineered T cells targeting E7 for patients with metastatic HPV-associated epithelial cancers. *Nat. Med.* **27**, 419–425 (2021).
62. Zavala-Ruiz, Z., Strug, I., Walker, B. D., Norris, P. J. & Stern, L. J. A hairpin turn in a class II MHC-bound peptide orients residues outside the binding groove for T cell recognition. *Proc. Natl. Acad. Sci. USA* **101**, 13279–13284 (2004).
63. Cole, D. K. et al. Modification of the carboxy-terminal flanking region of a universal influenza epitope alters CD4+ T-cell repertoire selection. *Nat. Commun.* **3**, 665 (2012).
64. Arnold, P. Y. et al. The majority of immunogenic epitopes generate CD4+ T cells that are dependent on MHC class II-bound peptide-flanking residues1. *J. Immunol.* **169**, 739–749 (2002).
65. Sibener, L. V. et al. Isolation of a structural mechanism for uncoupling T cell receptor signaling from peptide-MHC binding. *Cell* **174**, 672–687.e27 (2018).
66. Maldonado, L. et al. Intramuscular therapeutic vaccination targeting HPV16 induces T cell responses that localize in mucosal lesions. *Sci. Transl. Med.* **6**, 221ra13 (2014).
67. Goncharov, M. et al. VDJdb in the pandemic era: a compendium of T cell receptors specific for SARS-CoV-2. *Nat. Methods* **19**, 1017–1019 (2022).
68. Sant'Angelo, D. B. et al. The specificity and orientation of a TCR to its peptide–MHC class II ligands. *Immunity* **4**, 367–376 (1996).
69. Fornace, M. E. et al. NUPACK: analysis and design of nucleic acid structures, devices, and systems. Preprint at *ChemRxiv* https://doi.org/10.26434/chemrxiv-2022-xv98l (2022).
70. Anczurowski, M. et al. Mechanisms underlying the lack of endogenous processing and CLIP-mediated binding of the invariant chain by HLA-DP84. *Gly. Sci. Rep.* **8**, 4804 (2018).
71. Reynisson, B. et al. Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass

spectrometry MHC eluted ligand data. *J. Proteome Res.* **19**, 2304–2315 (2020).
72. Xie, J. & Heath, J. 1539 SARS-CoV-2 genome-wide single-chain-trimer peptide-MHCs reveal CD8+ T cell landscape in COVID-19 patients and distinct TCR repertoire dynamics. *J. Immunother. Cancer* **11**, eabf7550 (2023).
73. Poplin, R. et al. Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at *bioRxiv* https://doi.org/10.1101/201178 (2018).
74. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).
75. Racle, J. et al. Machine learning predictions of MHC-II specificities reveal alternative binding mode of class II epitopes. *Immunity* **56**, 1359–1375.e13 (2023).
76. Zhang, R. https://BioRender.com/v67dds1 (2026).

Acknowledgements

We would like to thank all the participants in the InCoV and UNCOVR studies and the HPV clinical trials for their support and contributions. We thank Y. Su for his advice and guidance on project planning and manuscript revision; P. Troisch for her support in preparing sequencing libraries and sequencing runs; the ISB Biobank team for collecting and processing clinical samples; and the Flow Core and team members at Fred Hutchinson Cancer Research Center and Benaroya Research Institute. We thank M. van Meurs, N. Gu, and K. Wang for their assistance in the projects; F. Gao, V. Adit, and M. Davis for generously sharing plasmid sequences, reagents, and advice on troubleshooting protocols. J.R.H. discloses support for the research of this work from the Andy Hill Cancer Research Endowment (CARE) Fund [BRK_201801-05] and the NIH R01 grant [CA264090-01]. C.L.T. discloses support for publication of this work from the NIH R01 grant [CA263256]. K.G.P. was supported by the Paul G Allen Research Center at Providence-Swedish Cancer Institute. Y.S. was supported by the Damon Runyon Quantitative Biology Fellowship from the Damon Runyon Cancer Research Foundation [DRQ-13-22].

Author contributions

Conceptualization and design, R.Z., W.C., Y.S. and J.R.H.; Acquisition, analysis, or interpretation of data, R.Z., J.Q., M.M., J.C., V.G., C.B., S.H., J.X., D.Y., A.W., S.K.S., A.A., D.C., Y.R., Y.S. and J.R.H.; Creation of new software used in the work, R.Z., W.C., R.H.N., J.X., D.C. and S.W.; Manuscript writing and revision, R.Z., Y.S. and J.R.H.; Clinical sample acquisition, R.E., K.M.M., S.L., C.M., L.R., J.R.W., J.C., K.N., J.F., A.D., S.J.F., M.S., S.P., C.E.B., A.R., D.W., K.G.P., C.W.D., C.P-S., J.D.G., C.L.T. and J.R.H.

Competing interests

C.P.-S. has received honoraria from Merck, Kyowa Kirin, and ImmPACT Bio. J.D.G. has consulted for Gilead, Merck, Ivivyd, obtained grants from Gilead and has contracted research with Gilead and Helix. J.R.H. is a consultant to Regeneron and has received research support from Gilead and Merck. A.R. has received honoraria from consulting with Amgen and Merck, is or has been a member of the scientific advisory board and holds stock in Appia, Apricity, Arcus, Compugen, CytomX, ImaginAb, ImmPact, Inspirna, Kite-Gilead, Larkspur, Lyell, Lutris, MapKure, Merus, Synthekine, and Tango, has received research funding from Agilent and from Bristol Myers Squibb through Stand Up to Cancer (SU2C), and patent royalties from Arsenal Bio. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-72396-7.

Correspondence and requests for materials should be addressed to James R. Heath.

Peer review information *Nature Communications* thanks Ahmed Hegazy and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Institute for Systems Biology, Seattle, WA, USA. [2]Department of Bioengineering, University of Washington, Seattle, WA, USA. [3]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [4]Molecular Engineering & Sciences Institute, University of Washington, Seattle, WA, USA. [5]Fred Hutchinson Cancer Center, Seattle, WA, USA. [6]Department of Hematology & Hematopoietic Cell Transplantation (T Cell Therapeutics Research Laboratories), City of Hope Beckman Research Institute and Medical Center, Duarte, CA, USA. [7]University of California Los Angeles (UCLA), the Jonsson Comprehensive Cancer Center at UCLA, Parker Institute for Cancer Immunotherapy at UCLA, Los Angeles, CA, USA. [8]Johns Hopkins University School of Medicine, Baltimore, MD, USA. [9]Swedish Center for Research and Innovation, Swedish Medical Center, Seattle, WA, USA. [10]Division of Gynecologic Oncology, Department of Surgery, City of Hope, Duarte, CA, USA. [11]KSOM/Norris Center for Cancer Cellular Immunotherapy Research (CCCIR), Department of Medicine, Division of Medical Oncology, Keck School of Medicine of USC, Los Angeles, CA, USA. [12]Paul G. Allen Research Center, Providence-Swedish Cancer Institute, Seattle, WA, USA. [13]Elson S. Floyd College of Medicine, Washington State University, Everett, WA, USA. [14]Providence St. Joseph Health, Renton, WA, USA. [15]Division of Allergy & Infectious Diseases, University of Washington, Seattle, WA, USA. ✉e-mail: jheath@isbscience.org