nature medicine

Article

https://doi.org/10.1038/s41591-022-02110-9

T cell receptor repertoires associated with control and disease progression following *Mycobacterium tuberculosis* infection

Received: 5 October 2021

Accepted: 25 October 2022

Published online: 5 January 2023

Check for updates

Munyaradzi Musvosvi[1,32], Huang Huang[2,32], Chunlin Wang[2], Qiong Xia[2], Virginie Rozot[1], Akshaya Krishnan[3], Peter Acs[3], Abhilasha Cheruku[3], Gerlinde Obermoser[3], Alasdair Leslie[4,5,6], Samuel M. Behar[7], Willem A. Hanekom[1,4,5], Nicole Bilek[1], Michelle Fisher[1], Stefan H. E. Kaufmann[8,9,10], Gerhard Walzl[11], Mark Hatherill[1], Mark M. Davis[2,12,13,32], Thomas J. Scriba[1,32] ✉, Adolescent Cohort Study team* & GC6-74 Consortium*

Antigen-specific, MHC-restricted αβ T cells are necessary for protective immunity against *Mycobacterium tuberculosis*, but the ability to broadly study these responses has been limited. In the present study, we used single-cell and bulk T cell receptor (TCR) sequencing and the GLIPH2 algorithm to analyze *M. tuberculosis*-specific sequences in two longitudinal cohorts, comprising 166 individuals with *M. tuberculosis* infection who progressed to either tuberculosis ($n = 48$) or controlled infection ($n = 118$). We found 24 T cell groups with similar TCR-β sequences, predicted by GLIPH2 to have common TCR specificities, which were associated with control of infection ($n = 17$), and others that were associated with progression to disease ($n = 7$). Using a genome-wide *M. tuberculosis* antigen screen, we identified peptides targeted by T cell similarity groups enriched either in controllers or in progressors. We propose that antigens recognized by T cell similarity groups associated with control of infection can be considered as high-priority targets for future vaccine development.

Antigen-specific CD4 T cells are necessary for protective immunity against *M. tuberculosis*, the etiological agent of tuberculosis (TB)[1,2]. Experimental and clinical evidence shows that the primary T cell mediators of this protection are interferon (IFN)-γ-expressing helper type 1 T cells (T_H1 cells), although recent evidence from nonhuman primates implicates T_H1/T_H17 cells as probable correlates of protection[3–6]. Comprehensive delineation of αβ T cell responses in *M. tuberculosis*-infected humans has been hampered by the complexity and heterogeneity of clinical phenotypes in TB[7,8], the high interindividual diversity of the major histocompatibility complex (MHC), which restricts antigen presentation to T cells, and the marked diversity of TCRs[9,10], even within single hosts.

Recent advances in single-cell and bulk TCR-sequencing technologies enable characterization of the antigen-specific TCR repertoire with unprecedented throughput and efficiency[11,12]. In addition, advances in analytic approaches, particularly GLIPH[12] and GLIPH2 (ref. 13), allow grouping of TCR sequences that share conserved sequences and motifs in the CDR3 region, which is primarily responsible for the recognition of antigenic peptides bound to molecules of the MHC[9,14,15]. This allows rapid clustering of thousands or millions of TCRs into similarity groups, without having to know for what antigens these TCRs are specific. This enables a broad profiling of T cell specificities, despite the complexity of these responses across individuals and groups[13,15,16]. Together, these tools provide the opportunity to analyze the pathogen-specific T cell

A full list of affiliations appears at the end of the paper. ✉e-mail: thomas.scriba@uct.ac.za

response in a holistic and unbiased manner that was not previously possible.

We hypothesized that distinct *M. tuberculosis*-specific T cell clonotype groups in *M. tuberculosis*-infected individuals are associated with either protection against or risk of disease progression. We applied antigen-specific T cell repertoire profiling and analyses to two well-characterized, longitudinal cohorts of *M. tuberculosis*-infected individuals, some of whom successfully controlled infection (controllers) and others who progressed to TB disease (progressors). We identified mycobacteria-reactive T cell groups with similar TCRs (similarity groups, which probably recognize the same epitope) and compared their frequencies in *M. tuberculosis*-infected controllers and progressors, to define putative protective (enriched in controllers) or pathogenic or nonprotective (enriched in progressors) TCR similarity groups. We then identified the *M. tuberculosis* antigenic epitope and restricted MHC for a subset of TCR members of such similarity groups using genome-wide antigen screening. Particularly important in this respect is that we were able to identify a set of controller-associated *M. tuberculosis* antigens that may be excellent candidates for inclusion in a future TB vaccine.

Results

Defining *M. tuberculosis*-specific T cells and their repertoires

We first determined TCR-αβ sequences expressed by mycobacteria-reactive T cells in controllers and progressors selected from adolescents with evidence of *M. tuberculosis* infection who participated in the Adolescent Cohort Study (ACS), a large epidemiological study of TB[17]. Progressors ($n = 44$) developed microbiologically confirmed, intrathoracic TB over 2 years of follow-up. Controllers ($n = 44$) also had evidence of *M. tuberculosis* infection, but did not develop TB during follow-up. Mycobacteria-reactive T cells were identified by stimulating thawed peripheral blood mononuclear cells (PBMCs) from progressors and controllers with *M. tuberculosis* lysate, comprising both protein and nonprotein antigens, and sorting activated CD4 or CD8 T cells (Fig. 1a and Extended Data Fig. 1a). Activated T cells were identified by their elevated expression levels of CD69 together with CD154 or CD137 for single-cell TCR-sequencing (scTCR-seq). We successfully captured the TCR-αβ repertoire of *M. tuberculosis*-lysate-responsive T cells from PBMC samples collected from 35 controllers and 35 progressors using this scTCR-seq approach (Supplementary Table 1). Among 37,674 sorted T cells from progressors and controllers, 22,276 (59.1%) CDR3α and 21,404 (56.8%) CDR3β sequences were detected, of which 15,272 and 16,517 were unique, respectively (Supplementary Table 2). Higher frequencies of activated T cells were observed after stimulation with *M. tuberculosis* lysate compared with phosphate-buffered saline (PBS), but frequencies of activated T cells between controllers and progressors were not different, nor were the numbers of CDR3β sequences detected (Fig. 2a–c). In addition, frequencies of activated T cells were constant over the 2-year follow-up period (Fig. 2d). Clonal expansions (two or more clones) were observed in scTCR data from all but four samples (Fig. 2e). More than 90% of sorted *M. tuberculosis* lysate-reactive T cells were CD4 T cells, 2.2% were CD8 T cells and 6.5% expressed canonical mucosa-associated invariant T (MAIT) cell CDR3α sequences irrespective of CD4 and CD8 expression (Extended Data Fig. 1b). These results are consistent with previous studies which showed that *M. tuberculosis*-reactive T cells are predominately CD4 T cells[13,15]. Cells expressing known canonical MAIT CDR3α sequences expressed markedly higher levels of CD26, a marker associated with MAIT cells[18], compared with CD4 and CD8 T cells (Extended Data Fig. 1c), demonstrating that the phenotype of single-cell sorted cells faithfully aligns with the TCR identity. Expected levels of messenger RNA expression of known functional markers by sorted CD4, CD8 and MAIT cells further validated the experimental TCR-seq pipeline we used. For example, a higher proportion of *M. tuberculosis* lysate-responsive MAIT cells expressed IFN-γ mRNA transcripts compared with CD4 and CD8 T cells, whereas a higher proportion of CD4 T cells expressed tumor necrosis factor (TNF), interleukin (IL)-2, IL-17 and IL-13 mRNA transcripts than CD8 and MAIT cells, and higher proportions of CD8 T cells and MAIT cells expressed eomesodermin and perforin mRNA transcripts than CD4 T cells (Extended Data Fig. 1d).

Comparison of *M. tuberculosis* TCR groups in single-cell repertoires

We then combined the CDR3β sequences obtained from mycobacteria-reactive CD4 T cells from controllers, progressors and previously published TCR datasets[13,15], amounting to 25,256 CDR3β sequences (Supplementary Table 3). To determine whether *M. tuberculosis*-specific T cells are preferentially enriched at the site of recent or ongoing TB disease, we compared bulk TCR data generated from blood and resected lung tissue samples, collected from an independent cohort of TB patients[19]. *M. tuberculosis* lysate-reactive CD4 TCR sequences were significantly enriched in lung tissue compared with corresponding peripheral blood samples (Fig. 3). By contrast, frequencies of cytomegalovirus (CMV), Epstein–Barr virus (EBV) and influenza A-specific CDR3β sequences did not differ between blood and lung resection samples, consistent with an expansion of *M. tuberculosis*-specific TCRs at the site of recent or ongoing disease.

The incredible diversity and private nature of CDR3β sequences have necessitated the development of clustering methods that group CDR3β sequences that probably share epitope specificities[13,15,16,20–22]. Such clustering methods allow interindividual comparisons of CDR3β sequences that probably share antigen specificity. We sought to determine whether such clusters of TCRs were differentially associated with either controllers or progressors. Using GLIPH2 (ref. 13) to cluster TCR-β sequences expressed by mycobacteria-reactive T cells, we identified 3,417 *M. tuberculosis* TCR similarity groups (Supplementary Table 4). Of the TCR similarity groups, 54% contained CDR3β sequences observed in sorting experiments performed in at least two independent studies[13,15] (Extended Data Fig. 2). This observation strongly implies that most of the TCR similarity groups contained TCRs that target antigens in *M. tuberculosis* lysate.

Previously, we reported that applying filters to the GLIPH2 output parameters narrowed down the number of TCR similarity groups and enriched for groups more likely to have been clustered correctly[13]. We selected TCR similarity groups shared by three of more participants, consistent with three or more unique CDR3β sequences, with enriched common V-genes (vb_score <0.05), with a limited CDR3 length distribution (length_score <0.05) and statistically significant motifs from a reference set of CDR3β sequences (Fisher_score <0.05). This filtering resulted in 290 TCR similarity groups. We then investigated whether any of the selected TCR similarity groups were significantly enriched in sorted *M. tuberculosis* lysate-reactive CD4 T cells from controllers or progressors. Most TCR similarity groups were shared between controllers and progressors, suggesting a high degree of overlap in T cell specificities between the groups (Fig. 4a). However, the 'S%QGTGE' and 'REGGTG%SP' TCR groups appeared to be enriched in progressors (Supplementary Table 5). Although no statistically significant enrichment was observed in these single-cell analyses after multiple correction using the Benjamini–Hochberg method ($q < 0.2$), we reasoned that the low depth achieved with scTCR-seq analysis limited statistical power to detect differences in TCR repertoires between the groups.

The degree of TCR sequence diversity may be associated with control of *M. tuberculosis* or, alternatively, with progression. The large size of the *M. tuberculosis* TCR sequence dataset enabled assessment of TCR similarity group diversity within individuals. For each individual we identified the number of unique clusters with a human leukocyte antigen (HLA)-allele association identified by GLIPH2 per 100 unique CDR3β sequences. It is interesting that, among individuals with HLA-DQB1*06 alleles, we observed a trend toward increased diversity in controllers compared with progressors; however, this was not significant when we accounted for multiple testing (Fig. 4b).

https://doi.org/10.1038/s41591-022-02110-9

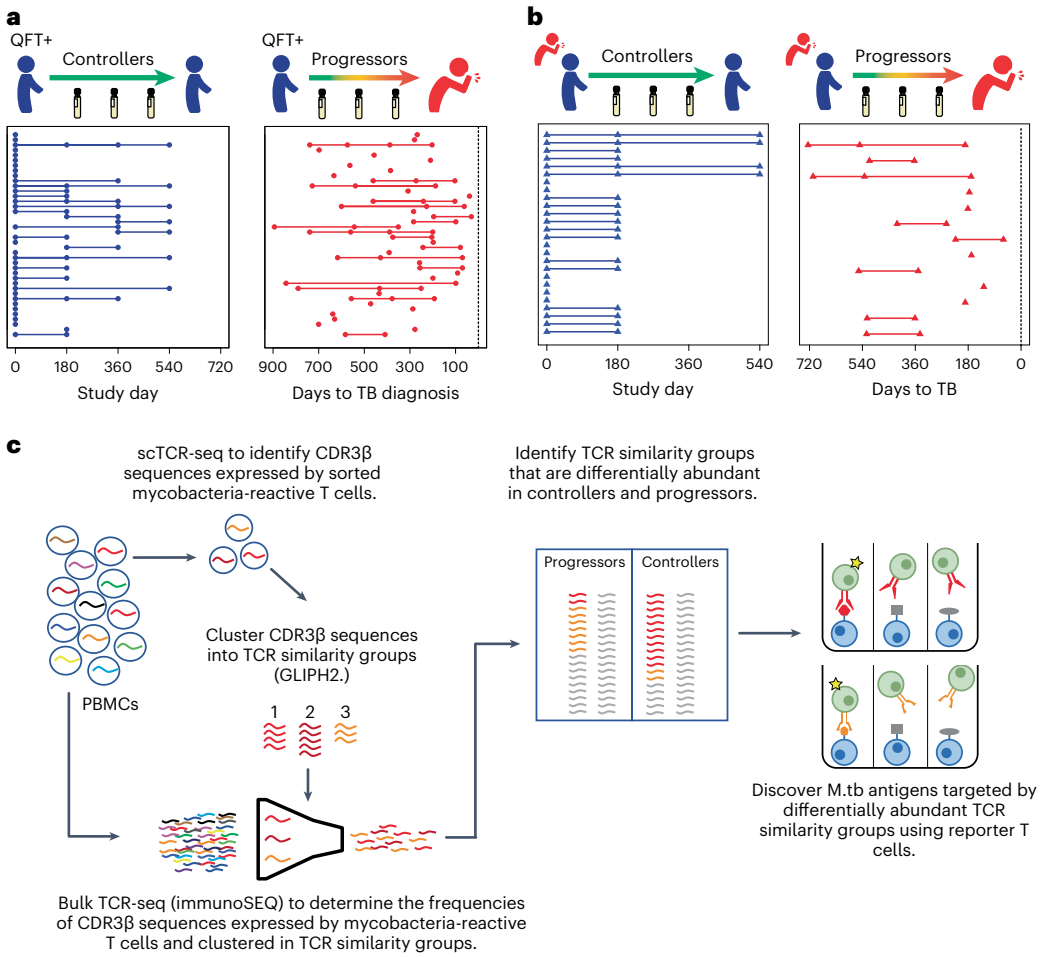


Fig. 1 | Identification of TCR sequences and antigens recognized by *M. tuberculosis* lysate-responsive T cells in controllers and progressors. **a**,**b**, Plots depicting longitudinal study timepoints (dots) at which PBMC samples were analyzed for each individual controller (blue) or progressor (red, synchronized to TB diagnosis) in the ACS (**a**) or the GC6-74 cohort (**b**). Each horizontal line or symbol represents an individual. **c**, Experimental workflow and analysis approach used to identify mycobacteria-reactive CDR3αβ sequences and determine their frequencies. First, scTCR-seq was performed on sorted mycobacteria-reactive T cells expressing the activation markers CD69 and CD154 or CD137 after in vitro *M. tuberculosis* (M.tb) lysate stimulation. GLIPH2 analysis clustered TCR sequences expressed by mycobacteria-reactive T cells into TCR similarity groups. In parallel, bulk TCR-seq was performed on PMBCs (unstimulated) to profile the repertoire and determine the frequencies of CDR3β sequences in each sample. The total frequencies of CDR3β sequences within a GLIPH2 TCR similarity group were determined for each controller and progressor sample using the bulk TCR-seq data. For controllers and progressors with samples collected at multiple study timepoints, the total frequencies of CDR3β sequences within a TCR similarity group were determined for each timepoint. The total frequencies of CDR3β sequences within a TCR similarity group were compared in controllers and progressors. To identify antigens recognized by these antigen-specific T cells, transduced NFAT, reporter stable J76-NFATRE-luc T cell line cells expressing representative TCR-αβ chains from TCR similarity groups found to be differentially abundant in controllers and progressors were coincubated with aAPCs to screen the *M. tuberculosis* proteome. QFT, QuantiFERON-TB Gold.

However, HLA-allele distribution was not associated with controller or progressor status (Fig. 4c and Extended Data Fig. 3), nor was there evidence of more allele subsets for HLA-DQB1*06 than other alleles in this population[23], suggesting that enrichment of *M. tuberculosis* TCR similarity groups in either controllers or progressors did not simply reflect HLA-allele prevalence or allele subset diversity.

M. tuberculosis TCR similarity groups associated with disease outcome

ScTCR-seq of *M. tuberculosis*-specific cells was necessary for identifying TCR similarity groups likely to target *M. tuberculosis* antigens and to identify TCR-α and TCR-β pairs that allow establishment of peptide–MHC specificity. However, scTCR-seq does not allow accurate quantification of clonotypes within the overall TCR repertoire in peripheral blood. To estimate relative frequencies of individual TCR sequences expressed by mycobacteria-reactive T cells, we performed bulk TCR-β repertoire profiling in unstimulated PBMC samples from a subset of ACS study participants ($n = 30$), who had remaining PBMC samples after single-cell sorting, and in a second longitudinal cohort of adult progressors and controllers enrolled into the Grand Challenges 6-74 (GC6-74)[24] (Supplementary Table 1). The GC6-74 cohort comprised South African household contacts of TB patients who either developed microbiologically confirmed pulmonary TB (progressors, $n = 12$) or remained healthy (controllers, $n = 25$) (Fig. 1b). From the combined ACS and GC6-74 bulk TCR-seq data, we selected only CDR3β sequences associated with mycobacteria-reactive T cells (that is, CDR3β expressed in sorted mycobacteria-reactive T cells) (Fig. 1c).

From 290 mycobacteria-reactive TCR similarity groups initially filtered on GLIPH2 output parameters, we further selected TCR similarity groups that had a significant HLA association using Fisher's exact test *P*-value threshold of <0.05 (HLA-alleles defined by two-digit typing). Among 175 TCR similarity group:HLA-allele combinations that met

this criterion, we compared frequencies of TCRs belonging to each similarity group in unstimulated PBMC samples from controllers and progressors bearing the associated HLA-allele (Fig. 5a). A total of 30 TCR similarity group:HLA-allele combinations, comprising 24 unique GLIPH2 TCR similarity groups, were differentially abundant in controllers and progressors at a *P*-value threshold <0.05, after controlling for the false discovery rate (FDR) using the Benjamini–Hochberg method ($q < 0.2$) (Fig. 5a). Twenty TCR similarity group:HLA-allele combinations had higher frequencies in controllers than progressors, whereas ten TCR similarity group:HLA-allele combinations were more abundant in progressors (Fig. 5b,c and Supplementary Table 6).

To investigate the specificity of the disease-outcome-associated TCR similarity groups, we compared frequencies of CMV, EBV and influenza A TCR similarity groups, identified using the GLIPH2-based pipeline (Fig. 5a) in controllers and progressors. Three CMV- (4.4%, 3 of 69), zero EBV- (0%, 0 of 39) and five influenza A-specific (4.1%, 10 of 246) TCR similarity groups were differentially abundant between controllers and progressors (Fig. 5d). To test whether outcome-associated *M. tuberculosis*-reactive TCR similarity group:HLA-allele combinations were nonrandom, we performed permutation analyses using randomized disease outcome labels and determined the number of significantly associated clusters from 1,000 iterations. The 30 *M. tuberculosis*-specific GLIPH2 specificity groups associated with clinical outcome greatly exceeded the numbers obtained from 1,000 iterations with randomized disease outcome; out of the 1,000 iterations, 30 GLIPH2 specificity groups were obtained only 15 times (1.5%) (Fig. 5e). Furthermore, the number of identified CMV-, EBV- or flu-specific TCR groups fell well within the distribution obtained from the analysis with randomized outcome labels. Last, we compared the frequencies of CMV, EBV, influenza A and *M. tuberculosis* TCR similarity groups that are differentially abundant between 274 CMV-infected (CMV^+) and 327 CMV-uninfected (CMV^-) individuals in a bulk TCR-seq dataset published by Emerson et al.[25] (Extended Data Fig. 4). We observed that the frequencies of 14 HLA-associated, CMV-specific TCR clusters (29%,14 of 48) were differentially abundant between CMV^+ and CMV^- individuals. Thirteen clusters were significantly more abundant in CMV^+ individuals and a single cluster was found to be more abundant in CMV^- individuals (Fig. 5f). By contrast, only a single HLA-associated *M. tuberculosis*-specific cluster was differentially abundant between the CMV^+ and CMV^- groups and not a single EBV or influenza A-specific TCR cluster was differentially abundant in CMV^+ and CMV^- individuals (Fig. 5f). Together, these results validate the specificity of our outcome-associated, *M. tuberculosis*-reactive TCR group discovery approach and suggest that the TCR groups identified were nonrandom.

We also used permutation analyses to further assess the robustness of our results. To do so, we first randomly permuted outcome labels 1,000×, calculating *P* values for each cluster using each set of permuted labels. From this, we calculated the distribution of counts of clusters with nominal *P* value <0.05 across the 1,000 iterations (Extended Data Fig. 5). When applying a *P*-value threshold of 0.05 to the (unpermuted, that is, original) progressor versus controller data, 33 TCR similarity group:HLA-allele combinations among the 175 were associated with outcome. Importantly, the total number of significant ($P < 0.05$) clusters exceeded 33 in only 44 of the 1,000 (4.4%) permutations, thus illustrating the presence of signal in the dataset. To identify outcome-associated clusters using a more conservative approach than the Benjamini–Hochberg method, we derived a *P*-value threshold to control the family-wise error rate at 0.05 using the permutations above. Such a threshold was made equal to the 5th percentile of the set of lowest per-permutation *P* values across the 1,000 permutations, yielding a threshold of 0.00001. None of the 175 TCR similarity group:HLA-allele combinations had an outcome-associated *P* value <0.00001 in the unpermuted progressor versus controller data, and so none was significant when controlling the family-wise error rate. We therefore controlled the less conservative FDR in our analyses.

We also sought to investigate the longitudinal kinetics of differentially abundant TCR similarity groups in samples collected at various timepoints before TB diagnosis in progressors, or throughout study follow-up in controllers, modeled by fitting nonlinear splines. Overall, these analyses yielded large 95% confidence intervals (CIs), highlighting the high degree of intersample and interindividual heterogeneity of *M. tuberculosis*-specific TCR data. However, the results suggest that, for many of the clusters identified to be more frequent in controllers, the TCRs were elevated in controllers throughout the study period. Similarly, TCR clusters identified as being more frequent in progressors were also generally elevated in progressors throughout the study period (Extended Data Fig. 6). To determine the influence of each cohort, we compared frequencies of the 30 differentially abundant TCR similarity group:HLA-allele combinations (Fig. 5b,c) in the ACS and GC6-74 cohorts separately. We observed concordant effect sizes between the two cohorts for most clusters, albeit with $P > 0.05$ for a number of clusters (Extended Data Fig. 7).

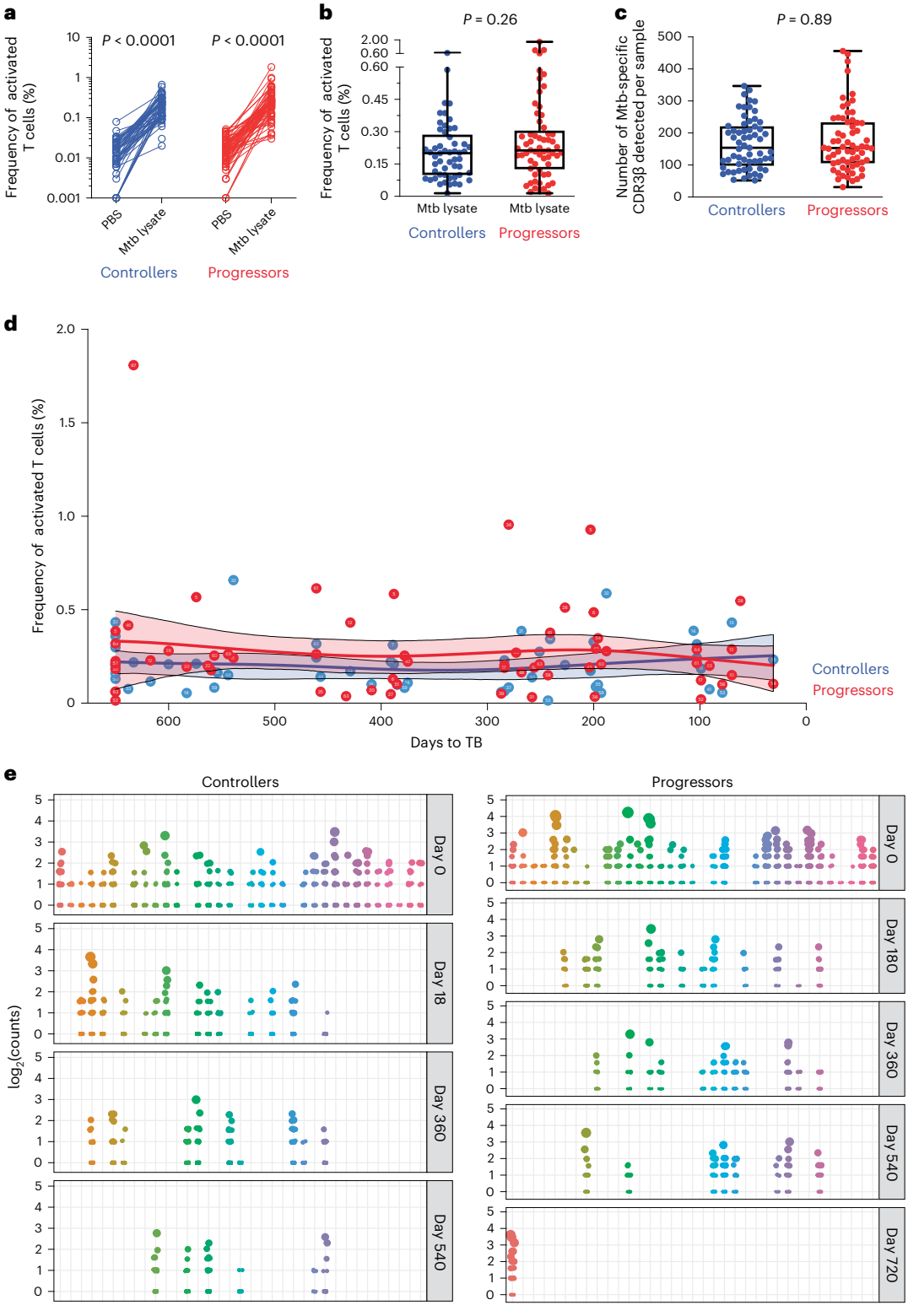
To determine whether our results were robust to the TCR clustering algorithm, we repeated the outcome-associated TCR similarity group discovery analysis using TCRdist3 (Supplementary Table 7), another clustering algorithm[26]. The TCRdist3 pipeline identified 246 unique mycobacteria-reactive metaclone clusters with significant HLA-allele associations. Of these, 46 metaclone cluster:HLA-allele combinations consisting of 33 unique metaclone clusters were differentially abundant in controllers and progressors (Supplementary Table 8). Overall, 67% of GLIPH2-identified clusters associated with clinical outcome were also identified by TCRdist3 (16 of 24), whereas 34.8% of all clinical outcome-associated clusters identified by either GLIPH2 or TCRdist3 were identified by both (Extended Data Figure 8a). For 1,000 randomized permutations, 52 (5.2%) yielded an overlap in TCR clusters between TCRdist3 and GLIPH2 at a proportion >34.8% (Extended Data Figure 8b). Together, these data suggested that our results are largely independent of the TCR clustering algorithm.

Fig. 2 | Similar frequencies and counts of *M. tuberculosis* lysate-reactive T cells in controllers and progressors. a, Plot showing the frequencies of T cells coexpressing CD69 and CD154 or CD69 and CD137 (activated T cells), measured by flow cytometry after PBS (negative control) or *M. tuberculosis* (Mtb) lysate stimulation. Each dot represents an individual sample (controllers, $n = 61$; progressors, $n = 64$). **b**, A plot depicting the background subtracted frequencies of activated T cells. The horizontal lines represent medians, the bounds of the boxes indicate the 25th and 75th percentiles and the whiskers represent the minima and maxima. Each dot represents an individual sample (controllers, $n = 53$; progressors, $n = 61$). The *P* value was calculated using the Mann–Whitney *U*-test (two sided). Note that some samples are from the same participant collected at different study timepoints. **c**, A plot depicting the numbers of detected CDR3β sequences from sorted, *M. tuberculosis*-specific T cells identified by TCR-seq in PBMCs from controllers and progressors in the ACS cohort. Each dot represents an individual sample (controllers, $n = 61$; progressors, $n = 64$). The horizontal lines represent medians, the bounds of the boxes indicate the 25th and 75th percentiles and the whiskers represent the minima and maxima. The *P* value was calculated using the Mann–Whitney *U*-test (two sided). Some samples are from the same participants collected at different study timepoints. **d**, Plot depicting the kinetics (background subtracted) of *M. tuberculosis* lysate-reactive T cells, measured by flow cytometry, after PBMC stimulation with *M. tuberculosis* lysate in controllers and progressors from the ACS cohort. Progressor samples were synchronized according to their time to TB diagnosis and controller samples were synchronized to their matched progressors. The solid lines indicate the modeled nonlinear splines and the shaded bands represent 95% CIs. **e**, Plots depicting clonal expansions of *M. tuberculosis* lysate-reactive T cells in samples from controllers and progressors at different timepoints (in days) after enrollment. Each dot represents a unique CDR3β sequence observed in a sample. The size of the dot is relative to the number of times the sequence was detected. Plots have been aligned by participant on the horizontal axis.

https://doi.org/10.1038/s41591-022-02110-9

Identifying targets of disease-associated TCR groups

Next, we sought to identify antigens and epitopes targeted by TCRs that belong to differentially abundant GLIPH2 TCR similarity groups (that is, similarity groups associated with either controllers or progressors). In an earlier study[15] we observed that TCRs within the SVAL TCR similarity group targeted an Rv1195c (PE13) epitope, restricted by DRB1*15:03, and did not attempt to resolve targets for this TCR similarity group in the present study. Previously, we had also observed that TCRs within the GEAK TCR similarity group recognized an epitope that maps to Rv3874 (CFP-10), restricted by DRB5*01:01 (ref. 15). In the present study, we observed that controllers who possessed DRB1*15 alleles had a higher frequency of the GEAK similarity cluster compared with progressors with DRB1*15. However, we did not observe activation of GEAK TCR-expressing Jurkat T cell clones in the context of DRB1*15:03, but did confirm that Rv3874 was recognized in the context of DRB5*01:01 (Fig. 6a). It is possible that other HLA-alleles in addition to DRB5*01:01 can present the CFP-10 epitope targeted by GEAK TCRs. Therefore, the association of DRB1*15 controllers with



https://doi.org/10.1038/s41591-022-02110-9

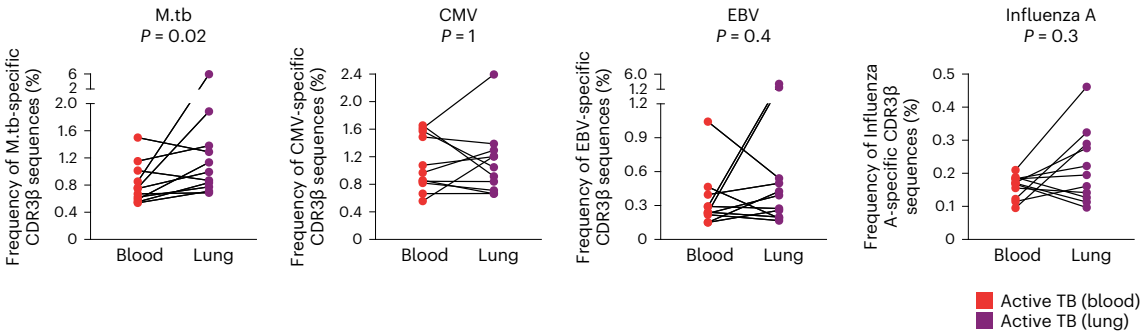


Fig. 3 | CDR3β sequences expressed on *M. tuberculosis* lysate-reactive T cells are enriched in lungs of patients with TB. Frequencies of *M. tuberculosis*- (M.tb-), CMV-, EBV- or influenza A-specific T cell clones, measured as CDR3β sequences in matched blood (PBMCs) and resected lung samples collected from people with severe active TB disease ($n = 11$). The P value was computed using Wilcoxon's signed-rank test (two sided).

higher frequencies of GEAK TCRs may reflect CFP-10 recognition via other HLA-alleles.

The target of the S%EDRGNTE TCR similarity group was resolved to be a Rv3616c (EspA) epitope, restricted by DRB3*01:01 (Fig. 6b–d). We also observed that Jurkat T cell clones expressing TCR sequences in the S%LAAGQET cluster were activated by *M. tuberculosis* lysate in the context of DRB1*04:01 (Fig. 6e), but not in the context of other HLA-alleles tested (Extended Data Fig. 9a–c). We were not able to resolve the antigen/epitope target of the S%LAAGQET cluster after stimulation with the *M. tuberculosis* 300 megapool or *M. tuberculosis* protein screening library (Extended Data Fig. 9d,e). Overall, we were able to determine the antigen targets of TCRs belonging to two controller-associated TCR similarity groups and one similarity group associated with progressors (Fig. 6f).

Last, we compared frequencies of canonical TCR CDR3α sequences of MAIT cells, CD1b-restricted, germline-encoded mycolyl lipid-reactive (GEM) cells and CD1d-restricted invariant natural killer T (iNKT) cells, as well as TCR-δ chains in ACS and GC6-74 controllers and progressors. Similar frequencies of MAIT CDR3α, iNKT and TCR-δ sequences were observed in CDR3α-sequencing data from controllers and progressors. However, progressors had higher frequencies of GEM sequences compared with controllers (Extended Data Fig. 10).

Discussion

In the present study, we broadly surveyed $CD4^+$ T cell responses to *M. tuberculosis* antigens using scTCR-seq to index TCR sequences expressed by mycobacteria-reactive T cells. We combined scTCR-seq with bulk TCR-seq and GLIPH2 analysis to identify controller-associated TCR similarity groups that may be promising targets for TB vaccine development. Traditionally, antigen discovery for vaccine development starts with the most immunogenic antigens from a given pathogen. A number of *M. tuberculosis* antigens used in candidate TB vaccines have been identified in this way[27]. However, *M. tuberculosis* expresses roughly 4,000 gene products[28] and it remains hypothetical that the most immunogenic antigens in natural infection are the most critical immunological targets for disease control, especially as an important resistance strategy for pathogens is to avoid expulsion before transmission. By combining the power of TCR-seq and TCR analysis methods with clinically relevant cohorts, we profiled the αβ TCR response repertoire to *M. tuberculosis* between controllers and progressors without prescribing the antigens involved, and focused on those TCR specificities that associated with clinical outcome.

We successfully studied the αβ TCR repertoire of *M. tuberculosis* lysate-responsive T cells from 70 controllers or progressors of the ACS cohort, combined with single T cell data from 58 individuals previously analyzed using the same methodology, all from the broader ACS cohort. The GC6-74 cohort included 38 individuals. Thus, we analyzed the TCR-β repertoires to *M. tuberculosis* lysate of 166 *M. tuberculosis*-infected individuals and identified over 3,000 *M. tuberculosis* TCR similarity groups, a fraction that was associated with either control of *M. tuberculosis* or progression. The remainder was no different between the groups. Data in the mouse model indicate that certain T cell specificities are more important for mycobacterial control than others[29]. We therefore targeted TCR similarity groups that correlated with controllers for the identification of specific antigens that could be incorporated into a subunit vaccine, using a genome-wide antigen-screening method that we developed previously[13], and report the identities of relevant T cell targets. This approach has applications for clinical studies of specific T cell responses to vaccination, infection and other immunological indications. Moreover, this approach represents a platform for rational antigen selection for candidate subunit vaccines that has utility for other pathogens as well.

We propose that the targets of TCR clonotype clusters associated with controllers can be considered as high-priority antigens for candidate TB subunit vaccines. Controllers possessed higher frequencies of T cells bearing PE13-specific TCRs. It is interesting that PE13 is a virulence factor that is cotranscribed with PPE18 on the same regulon under the control of Rv0485 (ref. 30). The PE and PPE family of proteins (Pro and Glu in the conserved amino-terminal region) has been implicated as key role players in host–pathogen interactions and have been investigated as potentially promising vaccine targets in murine models[31–33]. Importantly, vaccination with a PPE18 (Mtb39A)-containing polyprotein, fused with PepA (Mtb32A), showed 50% protection against TB disease in a recent landmark, phase IIb trial of the $M72/AS01_E$ vaccine[34]. We also observed that controllers had higher frequencies of a TCR similarity cluster that targets a CFP-10 epitope. CFP-10 is an immunodominant antigen specific to *M. tuberculosis* and is routinely used in IFN-γ release assay (IGRA) tests to identify people infected with *M. tuberculosis*. Deleting CFP-10 and ESAT-6 from the MTBVAC vaccine, a live-attenuated TB vaccine, resulted in increased bacterial burden in the murine model[35]. Together these data suggest that further investigation of PE13 and CFP-10 as vaccine targets is warranted.

Progressors had higher frequencies of T cells bearing TCRs within the S%EDRGNTE group, which targets EspA. It is of interest that vaccination with EspA-containing subunit vaccines reduced bacterial control in mice after *M. tuberculosis* challenge[36,37]. The higher frequencies of certain T cell clones in progressors may result from clonal T cell expansion in response to increased bacterial burden during progression, as indicated by increased activation of *M. tuberculosis*-specific CD4 T cells[38] and higher inflammation[39] in ACS progressors than controllers. These data are therefore consistent with in vivo recognition of these antigens by T cells. It remains possible that the progression-associated T cell responses identified in the present study can also contribute to immunopathology[40–43]. This highlights the need for further assessment

https://doi.org/10.1038/s41591-022-02110-9

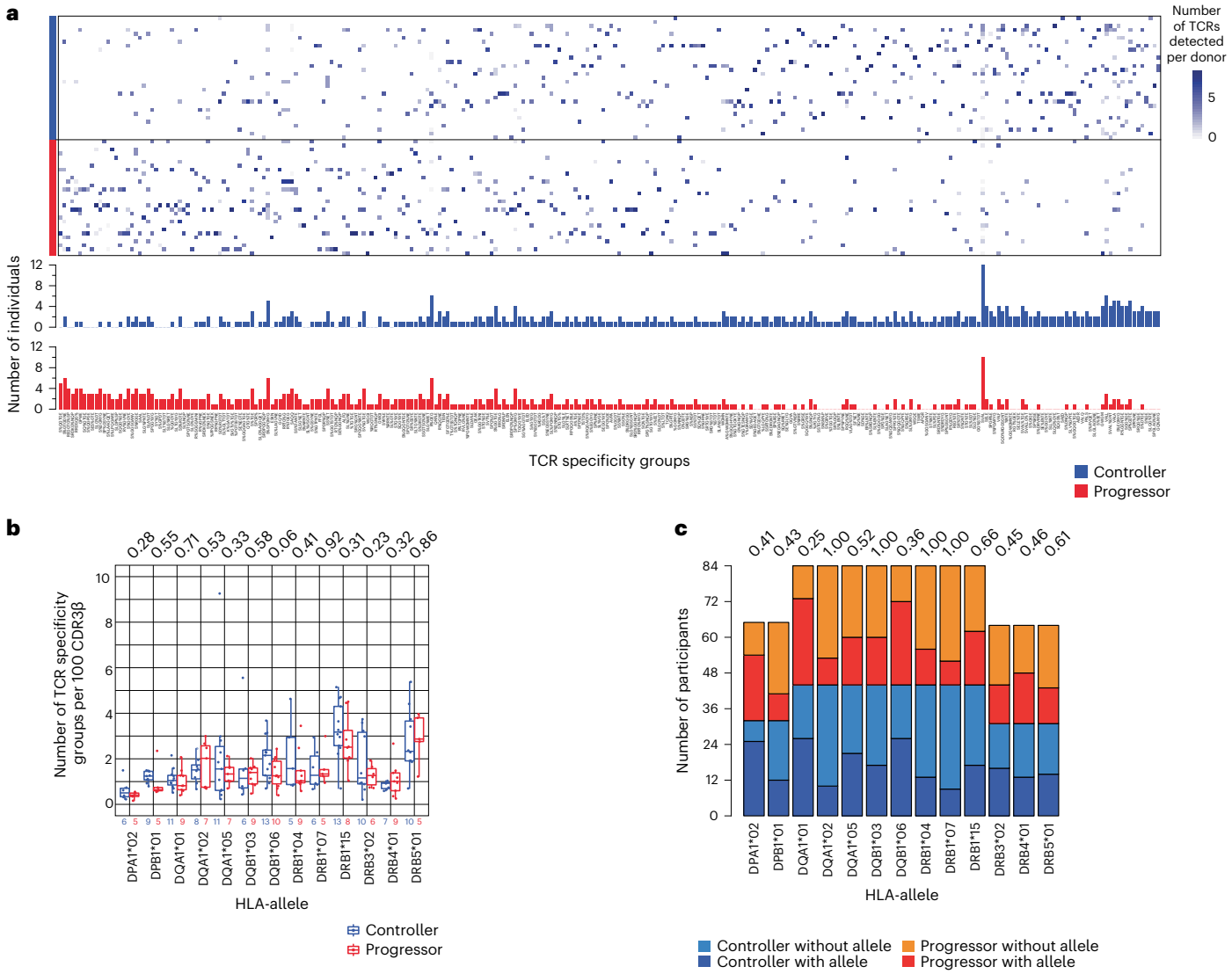


Fig. 4 | Mycobacteria-reactive TCR similarity groups overlap considerably between controllers and progressors. a, Heatmap depicting mycobacteria-reactive GLIPH2 TCR similarity groups (columns) identified in scTCR-seq in controllers and progressors (rows) from the ACS. The color represents the presence (blue) or absence (white) of sequences that belong to each TCR similarity group observed after *M. tuberculosis* lysate stimulation. Similarity groups are ranked according to their detected prevalence in progressors (right) or controllers (left). The barplot below depicts the number of donors possessing CDR3β sequences that belong to the indicated similarity group. The amino acid motif that is shared by TCR sequences clustered together is used to denote the cluster. In some instances GLIPH2 allows for a wildcard (that is, any amino acid) at a specific location within the shared motif; this is indicated by '%'. **b**, Box and whisker plots depicting the number of mycobacteria-reactive TCR similarity groups detected by scTCR-seq per 100 mycobacteria-reactive CDR3β sequences in ACS controllers and progressors. A higher value denotes greater diversity among mycobacteria-reactive CDR3β sequences. Note that not all mycobacteria-reactive CDR3β sequences fall into a similarity group. The midline represents the median, the box the interquartile range and the whiskers the 95% CI. Two-tailed Student's *t*-test: *P* values are shown above the plot. The number of samples from controllers and progressors are indicated below each plot. **c**, Bar plots depicting the number of ACS controllers and progressors with or without the indicated HLA-allele. Fisher's exact test (two sided) *P* values are listed above each bar.

of T cell responses to antigens that we have identified in relevant experimental preclinical and clinical studies.

It is likely that T cell responses associated with TCR similarity groups that we observed in the present study may have been primed by Bacillus Calmette–Guérin (BCG) vaccination and/or nontuberculous mycobacteria exposure before *M. tuberculosis* infection. It is therefore difficult to determine the roles of BCG vaccination and exposure to nontuberculous mycobacteria or *M. tuberculosis* infection in driving controller-associated TCR similarity groups, although the TCR similarity group that recognizes CFP-10 is expected to be *M. tuberculosis* specific. Nevertheless, our results support the possibility that both BCG vaccination and/or *M. tuberculosis* infection may be important in driving the expansion of TCR similarity groups associated with control. For example, PE13 is expressed by both BCG and *M. tuberculosis*. Regardless of the source of priming, we propose that targeting and expanding T cell clones associated with controllers by vaccination will result in better protection from TB progression. Furthermore, we acknowledge that most CDR3β sequences were not clustered into similarity groups. Of the 16,517 unique CDR3β sequences that we observed in the mycobacteria-reactive T cell population, 5,687 unique CDR3β sequences were successfully clustered into 3,417 similarity groups. The modest proportion (34.4%) of CDR3β sequences that could be clustered together probably reflects the diversity and private nature of the TCR repertoire. We were unable to compare the frequencies

https://doi.org/10.1038/s41591-022-02110-9

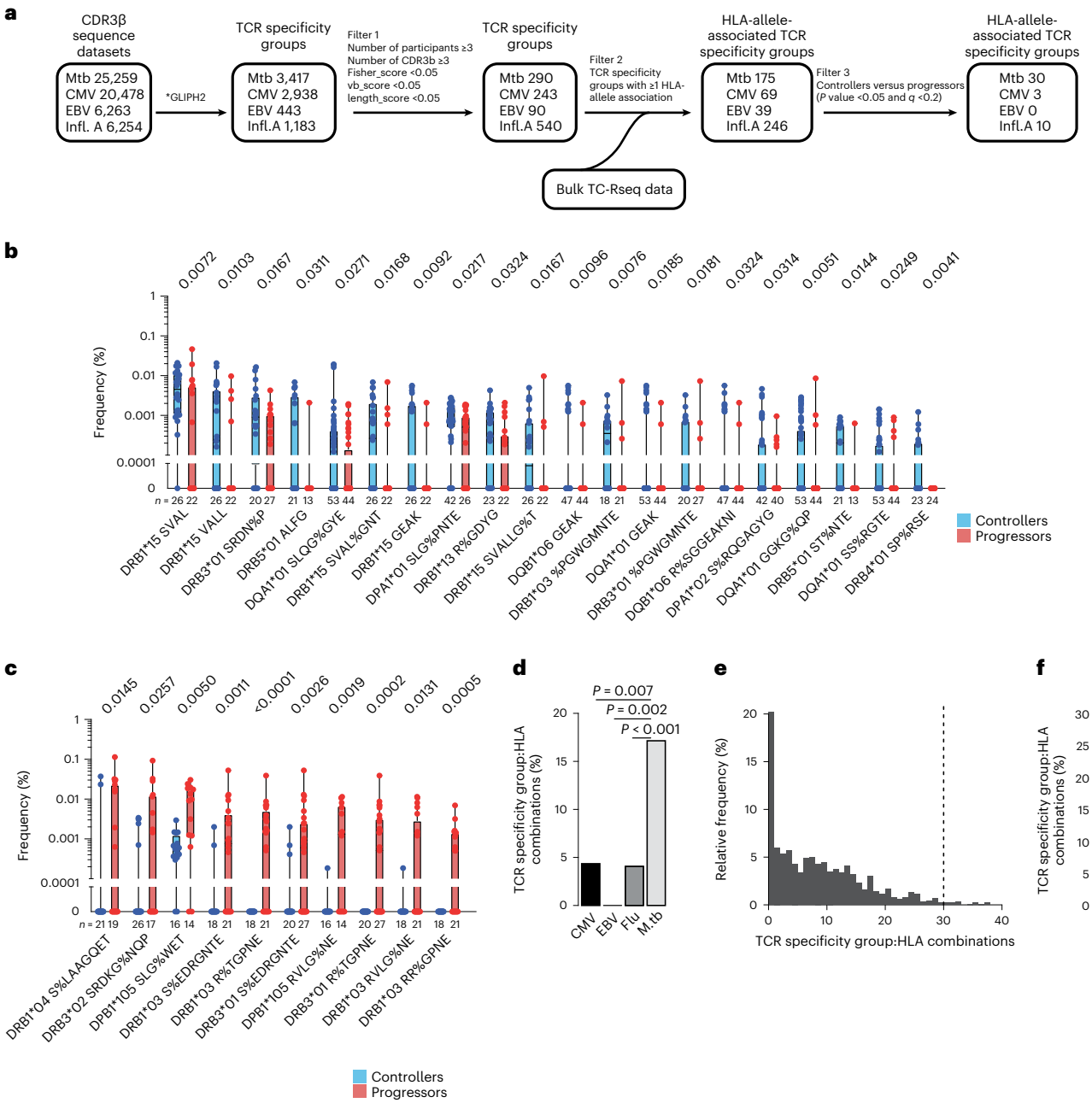
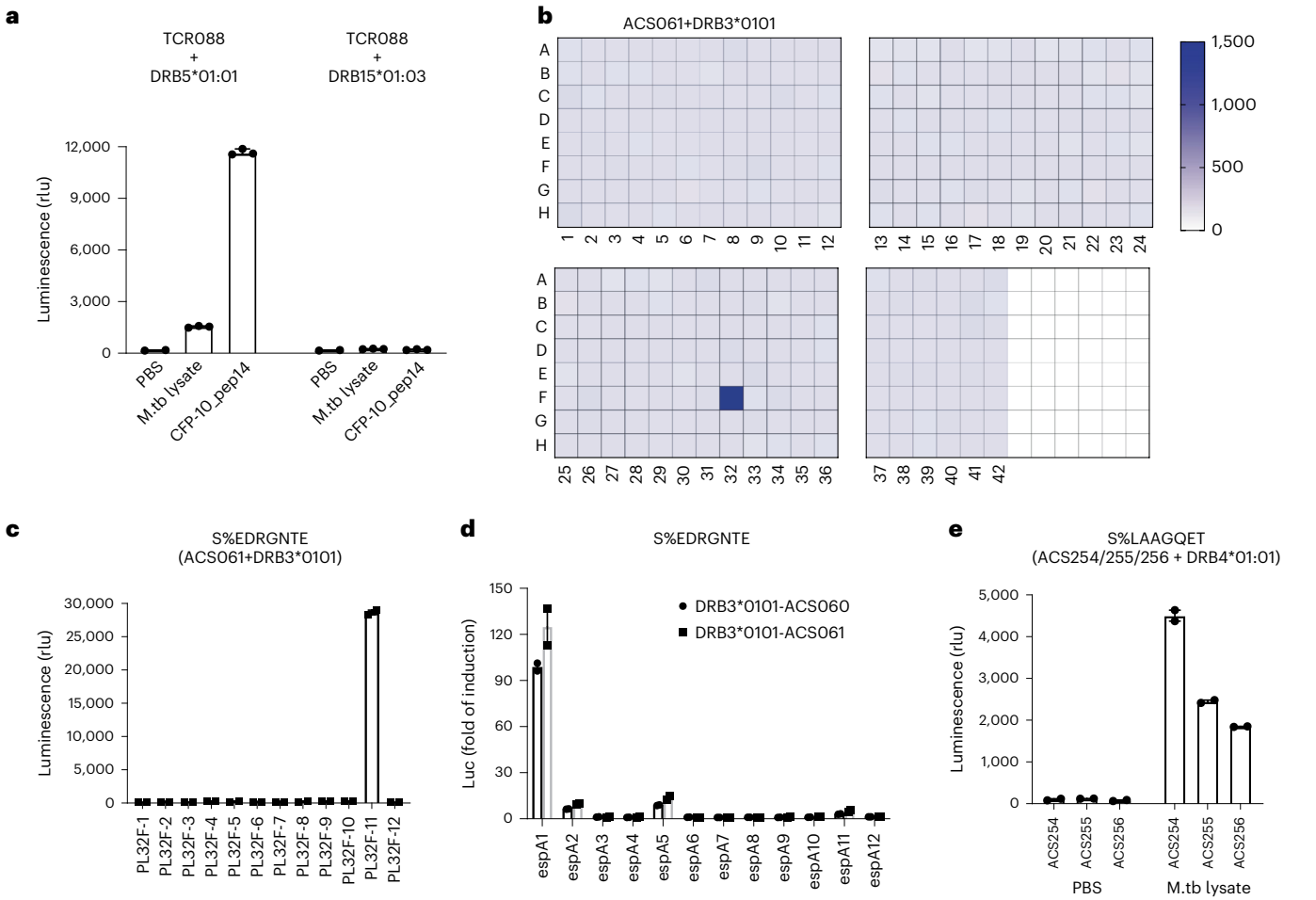


Fig. 5 | Differentially abundant mycobacteria-reactive TCR similarity groups in controllers and progressors. a, Analysis workflow used to measure the frequencies of mycobacteria-reactive- (Mtb-) or CMV-, EBV- or influenza A (Infl.A)-specific GLIPH2 TCR groups in controllers and progressors. GLIPH2 analysis was performed and the resulting GLIPH2 similarity groups were filtered initially using the criteria listed under Filter 1. TCR similarity groups with significant HLA-allele associations in the progressor/controller cohort were then selected (Filter 2). Similarity groups that were differentially abundant in controllers and progressors bearing the associated HLA-allele were identified (Filter 3). **b**,**c**, Box and whisker plots depicting frequencies of mycobacteria-reactive TCRs belonging to the indicated HLA-allele-associated TCR similarity groups that were significantly more abundant in controllers (**b**) or progressors (**c**) bearing the indicated HLA-allele. The horizontal lines represent medians, the boxes the interquartile range and the whiskers the range. The number of samples from controllers and progressors is indicated below each plot. Only clusters with a *P* value <0.05 (Mann–Whitney *U*-test, two sided) and $q < 0.2$ (Benjamini–Hochberg FDR) are shown. **d**, Frequencies of CMV- (3 of 69), EBV- (0 of 39), influenza A- (Flu-) (10 of 246) or *M. tuberculosis* (M.tb)-specific (30 of 175) TCR specificity group:HLA combinations that are associated with clinical outcome (significantly more abundant in either controllers or progressors), expressed as a percentage of all TCR specificity group:HLA combinations for that pathogen. The *P* value was calculated using Fisher's exact test (two sided). **e**, Relative frequency plot of the numbers of TCR specificity group:HLA combinations found to be significantly different between the two groups. We performed permutation analyses with 1,000 iterations using randomized disease outcome labels. The vertical line represents the actual number of *M. tuberculosis*-specific TCR specificity group:HLA combinations found to be significantly different between controllers and progressors (30) with correct disease outcome labels. **f**, Frequencies of CMV- (14 of 48), EBV- (0 of 51), influenza A- (0 of 227) or *M. tuberculosis* (1 of 42)-specific TCR specificity group:HLA combinations that are associated with CMV infection status in a previously published cohort[25], expressed as a percentage of all TCR specificity group:HLA combinations for that pathogen. The *P* value was calculated using Fisher's exact test (two sided).

https://doi.org/10.1038/s41591-022-02110-9



f

Cluster	Clone ID	CDR3α	TRAV	TRAJ	CDR3β	TRBV	TRBJ	HLA-a	HLA-b	Epitope	Protein	Association
SVAL	ACS024	CAGTNTGNQFYF	27	49	CASSVALQGVHTQYF	9	2-3	DRA*01:01	DRB1*15:03	MHVSFVMAYPEMLAA	Rv1195 (PE13)	Controller
	ACS025	CAGPTGGSYIPTF	25	6	CASSVALATGEQYF	9	2-7	DRA*01:01	DRB1*15:03	MHVSFVMAYPEMLAA	Rv1195 (PE13)	Controller
GEAK	ACS088	CAVRDPGNTDKLIF	1-2	34	CSARASGGEAKNIQYF	20-1	2-4	DRA*01:01	DRB5*01:01	AAVVRFQEAANKQKQ	Rv3874 (CFP10)	Controller
S%EDRGNTE	ACS060	CAVPNSGYSTLTF	21	11	CAISQEDRGNTEAFF	10-3	1-1	DRA*01:01	DRB3*01:01	MSRAFIIDPTISAID	Rv3616c (EspA)	Progressor
	ACS061	CAAPNSGYSTLTF	21	11	CAISGEDRGNTEAFF	10-3	1-1	DRA*01:01	DRB3*01:01	MSRAFIIDPTISAID	Rv3616c (EspA)	Progressor
S%LAAGQET	ACS254	CATPNQAGTALIF	1-2	5	CASSILAAGQETQYF	19	2-5	DRA*01:01	DRB1*04:01	-	Mtb lysate	Progressor
	ACS255	CAVMNQAGTALIF	1-2	5	CASSLLAAGQETQYF	11-1	2-5	DRA*01:01	DRB1*04:01	-	Mtb lysate	Progressor
	ACS256	CAAPNQAGTALIF	1-2	5	CAWSVLAAGQETQYF	30	2-5	DRA*01:01	DRB1*04:01	-	Mtb lysate	Progressor

Fig. 6 | Antigen discovery for mycobacteria-reactive TCR similarity groups. **a**, Barplot showing the median relative luminescence signal after an 8-h PBS, *M. tuberculosis* (M.tb) lysate or AAVVRFQEAANKQKQ (CFP-10_p14) stimulation of TCR-ACS088 (TCR008) in the context of DRB5*01:01 or DRB1*15:03. The mean ± s.d. ($n = 3$ biological replicates) is shown. **b**, Antigen recognition screening of the whole *M. tuberculosis* proteome (321 subpools displayed in 3.5 plates) by TCR-transfected clone TCR-ACS061, bearing a TCR in similarity group S%EDRGNTE. The color scale indicates the relative luminescence signal after an 8-h stimulation of the clone. **c**, Barplot showing the deconvolution of recognition of the individual proteins from the positive subpool (PL32F), expressed separately and screened against clone TCR-ACS061. The clone was activated by PL32-F11 (Rv3616c), indicating TCR-mediated recognition. The mean ($n = 2$ biological replicates) is shown. rlu, relative luminescence units. **d**, Barplot showing resolution of the Rv3616c epitope using overlapping peptides spanning Rv3616c to identify the epitope recognized by TCR-ACS061. The mean ($n = 2$ biological replicates) is shown. **e**, Barplot depicting *M. tuberculosis* lysate recognition by clones TCR-ACS0254/255/256. The bar represents the median and each closed circle represents a replicate. The mean ($n = 2$ biological replicates) is shown. **f**, Table listing mycobacteria-reactive TCR similarity groups associated with controller or progressor status and their epitope targets.

of CDR3β sequences that were not clustered and therefore may have missed TCRs associated with control or progression.

We restricted analyses of associations with clinical outcome to TCR similarity groups with a significant HLA-allele association. As it is well known that HLA class II peptide binding is highly promiscuous, we expect that a nontrivial proportion of individual HLA-allele-associated T cell response differences between controllers and progressors may have been masked by this peptide–HLA promiscuity, rendering them unidentifiable with our analytical pipeline. This concept underscores the remarkable complexity and vast scope of human T cell recognition of *M. tuberculosis* proteins, as reported previously[44], and supports future studies and orthogonal approaches to such analyses.

Our result is consistent with previous findings of no association between frequencies of T cell responses in BCG-vaccinated or MVA85A-vaccinated infants and clinical outcome[45,46]. Other T cell functions or features may be more relevant to protection. Recent data from intravenous BCG vaccination and experimental *M. tuberculosis* infection of nonhuman primates suggest that T_H1/T_H17 cells were

associated with successful control or even sterilizing immunity[3–6,47]. Future studies that compare the differentiation state, lung homing capacity and phenotypes of antigen-specific T cells expressing controller-associated and progressor-associated TCR similarity groups may shed more light on the roles of these T cell characteristics.

Most TCR-specificity groups apparently had no association with either control or progression. We speculate that this is consistent with the hypothesis that *M. tuberculosis* allows immune recognition of considerable numbers of 'decoy' proteins to distract the T cell response, probably to facilitate persistence. This decoy strategy has been observed in murine studies, which showed that TB10.4 acts as a decoy antigen by inducing a TB10.4-specific CD8 T cell response that poorly recognizes infected macrophages[48,49]. This immunodominant T cell response suppressed subdominant responses and thereby subverted immune control[48,49]. A similar decoy phenomenon, involving an immunodominant epitope for CD4 T cells within the ESAT-6 protein, that subverts subdominant epitopes with greater protective capacity has also been described[50]. However, further exploration is required to adequately test this hypothesis in humans.

We acknowledge that our study has several other limitations. Our comparisons of controllers and progressors are limited to peripheral blood rather than the more relevant lung compartment[51]. It is possible that distinct T cell responses and specificities are present at sites of disease. Our study utilized samples collected exclusively from South Africans. It will be important to determine whether similar TCR similarity groups are associated with controllers from populations with different TB epidemiology, age, environmental conditions and HLA background. We also note that due to the limited sample size we restricted association analyses to two-digit HLA typing and not four-digit typing and were unable to definitively address the role of genetic variation, especially in the MHC locus, on TCR and clinical outcome between progressors and controllers. Similar larger studies using samples collected from other countries with a high TB burden will need to be performed to determine the generalizability of our results. We also note that the use of H37Rv lysate to stimulate PBMCs may have resulted in underrepresentation of TCR sequences induced by the infecting *M. tuberculosis* strain in controllers and progressors. It is not possible to identify the infecting *M. tuberculosis* strains in controllers, although the identities of *M. tuberculosis* strains in progressors were not determined. Despite these limitations, we demonstrated the utility of TCR profiling for the purpose of identifying *M. tuberculosis*-specific T cell clonotypes associated with control of *M. tuberculosis* infection and their target antigens. We note that the antigenic targets for many *M. tuberculosis* TCR similarity groups identified in the present study remain to be resolved. Regardless, the present study has provided an initial list of TCR specificities and a large TCR sequence database that can be used as a valuable tool in the search for candidate TB vaccine antigens.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-022-02110-9.

References

1. O'Garra, A. et al. The immune response in tuberculosis. *Annu. Rev. Immunol.* **31**, 475–527 (2013).
2. Scriba, T. J., Coussens, A. K. & Fletcher, H. A. Human immunology of tuberculosis. *Microbiol. Spectr.* https://doi.org/10.1128/microbiolspec.TBTB2-0016-2016 (2017).
3. Cadena, A. M. et al. Concurrent infection with *Mycobacterium tuberculosis* confers robust protection against secondary infection in macaques. *PLoS Pathog.* **14**, e1007305 (2018).
4. Darrah, P. A. et al. Prevention of tuberculosis in macaques after intravenous BCG immunization. *Nature* **577**, 95–102 (2020).
5. Gideon, H. P. et al. Variability in tuberculosis granuloma T cell responses exists, but a balance of pro- and anti-inflammatory cytokines is associated with sterilization. *PLoS Pathog.* **11**, e1004603 (2015).
6. Gideon, H. P. et al. Multimodal profiling of lung granulomas in macaques reveals cellular correlates of tuberculosis control. *Immunity*. **55**, 827–846 (2022).
7. Pai, M. et al. Tuberculosis. *Nat. Rev. Dis. Prim.* **2**, 16076 (2016).
8. Barry, C. E. 3rd et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat. Rev. Microbiol.* **7**, 845–855 (2009).
9. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 (1988).
10. Sethna, Z. et al. Population variability in the generation and selection of T-cell repertoires. *PLoS Comput. Biol.* **16**, e1008394 (2020).
11. Carlson, C. S. et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat. Commun.* https://doi.org/10.1038/ncomms3680 (2013).
12. Han, A., Glanville, J., Hansmann, L. & Davis, M. M. Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat. Biotechnol.* **32**, 684–692 (2014).
13. 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.* https://doi.org/10.1038/s41587-020-0505-4 (2020).
14. Jorgensen, J. L., Esser, U., Fazekas de St. Groth, B., Reay, P. A. & Davis, M. M. Mapping T-cell receptor–peptide contacts by variant peptide immunization of single-chain transgenics. *Nature* **355**, 224–230 (1992).
15. Glanville, J. et al. Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98 (2017).
16. Dash, P. et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nat. Publ. Gr.* **547**, 89–93 (2017).
17. Zak, D. E. et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* **387**, 2312–2322 (2016).
18. Sharma, P. K. et al. High expression of CD26 accurately identifies human bacteria-reactive MR1-restricted MAIT cells. *Immunology* **145**, 443–453 (2015).
19. Ogongo, P. et al. Differential skewing of donor-unrestricted and γδ T cell repertoires in tuberculosis-infected human lungs. *J. Clin. Invest.* **130**, 214–230 (2020).
20. Chiou, S.-H. et al. Global analysis of shared T cell specificities in human non-small cell lung cancer enables HLA inference and antigen discovery. *Immunity* **54**, 586–602.e8 (2021).
21. Pogorelyy, M. V. et al. Detecting T cell receptors involved in immune responses from single repertoire snapshots. *PLoS Biol.* **17**, e3000314 (2019).
22. Zhang, H. et al. Investigation of antigen-specific T-cell receptor clusters in human cancers. *Clin. Cancer Res.* **26**, 1359–1371 (2020).
23. Thorstenson, Y. R. et al. Allelic resolution NGS HLA typing of class I and class II loci and haplotypes in Cape Town, South Africa. *Hum. Immunol.* **79**, 839–847 (2018).
24. Suliman, S. et al. Four-gene pan-African blood signature predicts progression to tuberculosis. *Am. J. Respir. Crit. Care Med.* **197**, 1198–1208 (2018).
25. Emerson, R. O. et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat. Genet.* https://doi.org/10.1038/ng.3822 (2017).
26. Mayer-Blackwell, K. et al. TCR meta-clonotypes for biomarker discovery with tcrdist3 enabled identification of public, HLA-restricted clusters of SARS-CoV-2 TCRs. *eLife* **10**, e68605 (2021).

27. Andersen, P. & Scriba, T. J. Moving tuberculosis vaccines from theory to practice. *Nat. Rev. Immunol.* **19**, 550–562 (2019).
28. Cole, S. T. et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
29. Bertholet, S. et al. Identification of human T cell antigens for the development of vaccines against *Mycobacterium tuberculosis*. *J. Immunol.* **181**, 7948–7957 (2008).
30. Goldstone, R. M., Goonesekera, S. D., Bloom, B. R. & Sampson, S. L. The transcriptional regulator Rv0485 modulates the expression of a PE and PPE gene pair and is required for *Mycobacterium tuberculosis*. *Virulence.* **77**, 4654–4667 (2009).
31. Brennan, M. J. The enigmatic PE/PPE multigene family of mycobacteria and tuberculosis vaccination. *Infect. Immun.* **85**, e00969–16 (2021).
32. Sampson, S. L. Mycobacterial PE/PPE proteins at the host-pathogen interface. *Clin. Dev. Immunol.* **2011**, 497203 (2011).
33. Qian, J., Chen, R., Wang, H. & Zhang, X. Role of the PE/PPE family in host–pathogen interactions and prospects for anti-tuberculosis vaccine and diagnostic tool design. *Front. Cell Infect. Microbiol.* **10**, 743 (2020).
34. Tait, D. R. et al. Final analysis of a trial of M72/AS01E vaccine to prevent tuberculosis. *N. Engl. J. Med.* **381**, 2429–2439 (2019).
35. Aguilo, N. et al. Reactogenicity to major tuberculosis antigens absent in BCG is linked to improved protection against *Mycobacterium tuberculosis*. *Nat. Commun.* **8**, 16085 (2017).
36. Aagaard, C. et al. Immunization with *Mycobacterium tuberculosis*-specific antigens bypasses T cell differentiation from prior Bacillus Calmette–Guérin vaccination and improves protection in mice. *J. Immunol.* **205**, 2146–2155 (2020).
37. Woodworth, J. S. et al. A Mycobacterium tuberculosis-specific subunit vaccine that provides synergistic immunity upon co-administration with Bacillus Calmette-Guérin. *Nat Commun.* **12**, 6658 (2021).
38. Mpande, C. A. M. et al. Antigen-specific T cell activation distinguishes between recent and remote tuberculosis infection. *Am. J. Respir. Crit. Care Med*. https://doi.org/10.1164/rccm.202007-2686OC (2021).
39. Scriba, T. J. et al. Sequential inflammatory processes define human progression from *M. tuberculosis* infection to tuberculosis disease. *PLoS Pathog.* **13**, e1006687 (2017).
40. Coscolla, M. et al. *M. tuberculosis* T cell epitope analysis reveals paucity of antigenic variation and identifies rare variable TB antigens. *Cell Host Microbe* **18**, 538–548 (2015).
41. Kwan, C. K. & Ernst, J. D. HIV and tuberculosis: a deadly human syndemic. *Clin. Microbiol. Rev.* **24**, 351–376 (2011).
42. Elkington, P. T., Bateman, A. C., Thomas, G. J. & Ottensmeier, C. H. Implications of tuberculosis reactivation after immune checkpoint inhibition. *Am. J. Respir. Crit. Care Med.* **198**, 1451–1453 (2018).
43. Comas, I. et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat. Genet.* **42**, 498–503 (2010).
44. Lindestam Arlehamn, C. S. et al. A quantitative analysis of complexity of human pathogen-specific CD4 T cell responses in healthy *M. tuberculosis* infected South Africans. *PLoS Pathog.* **12**, e1005760 (2016).
45. Kagina, B. M. et al. Specific T cell frequency and cytokine expression profile do not correlate with protection against tuberculosis after bacillus Calmette–Guerin vaccination of newborns. *Am. J. Respir. Crit. Care Med.* **182**, 1073–1079 (2010).
46. Tameris, M. D. et al. Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. *Lancet* **381**, 1021–1028 (2013).
47. Dijkman, K. et al. Prevention of tuberculosis infection and disease by local BCG in repeatedly exposed rhesus macaques. *Nat. Med.* **25**, 255–262 (2019).
48. Yang, J. D. et al. *Mycobacterium tuberculosis*-specific CD4+ and CD8+ T cells differ in their capacity to recognize infected macrophages. *PLoS Pathog.* **14**, e1007060 (2018).
49. Sutiwisesak, R. et al. A natural polymorphism of *Mycobacterium tuberculosis* in the esxH gene disrupts immunodomination by the TB10.4-specific CD8 T cell response. *PLoS Pathog.* **16**, e1009000 (2020).
50. Woodworth, J. S. et al. Protective CD4 T cells targeting cryptic epitopes of *Mycobacterium tuberculosis* resist infection-driven terminal differentiation. *J. Immunol.* **192**, 3247 LP–3243258 (2014).
51. Ogongo, P. et al. Tissue-resident-like CD4+ T cells secreting IL-17 control *Mycobacterium tuberculosis* in the human lung. *J. Clin. Invest.* **131**, e142014 (2021).

[1]South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine, and Division of Immunology, Department of Pathology, University of Cape Town, Cape Town, South Africa. [2]Institute for Immunity, Transplantation and Infection, Stanford University School of Medicine, Stanford, CA, USA. [3]Human Immune Monitoring Center, Stanford University, Stanford, CA, USA. [4]Africa Health Research Institute, Durban, South Africa. [5]School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa. [6]Department of Infection and Immunity, University College London, London, UK. [7]Department of Microbiology and Physiological Systems, University of Massachusetts Chan Medical School, Worcester, MA, USA. [8]Max Planck Institute for Infection Biology, Berlin, Germany. [9]Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany. [10]Hagler Institute for Advanced Study, Texas A&M University, College Station, TX, USA. [11]DST-NRF Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. [12]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. [13]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. [32]These authors contributed equally: Munyaradzi Musvosvi, Huang Huang, Mark M. Davis, Thomas J. Scriba. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: thomas.scriba@uct.ac.za

Adolescent Cohort Study team

Fazlin Kafaar[14], Leslie Workman[14], Humphrey Mulenga[14], Thomas J. Scriba[14], E. Jane Hughes[14], Nicole Bilek[14], Mzwandile Erasmus[14], Onke Nombida[14], Ashley Veldsman[14], Yolundi Cloete[14], Deborah Abrahams[14], Sizulu Moyo[14], Sebastian Gelderbloem[14], Michele Tameris[14], Hennie Geldenhuys[14], Willem Hanekom[14], Gregory Hussey[14], Rodney Ehrlich[15], Suzanne Verver[16] & Larry Geiter[17]

[14]South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine and Division of Immunology, Department of Pathology, University of Cape Town, Cape Town, South Africa. [15]School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa. [16]KNCV Tuberculosis Foundation The Hague Amsterdam Institute of Global Health and Development, Academic Medical Centre, Amsterdam, the Netherlands. [17]Aeras, Rockville, MA, USA.

GC6-74 Consortium

Gerhard Walzl[18], Gillian F. Black[18], Gian van der Spuy[18], Kim Stanley[18], Magdalena Kriel[18], Nelita Du Plessis[18], Nonhlanhla Nene[18], Teri Roberts[18], Leanie Kleynhans[18], Andrea Gutschmidt[18], Bronwyn Smith[18], Andre G. Loxton[18], Novel N. Chegou[18], Gerhardus Tromp[18], David Tabb[18], Tom H. M. Ottenhoff[19], Michel R. Klein[19], Marielle C. Haks[19], Kees L. M. C. Franken[19], Annemieke Geluk[19], Krista E. van Meijgaarden[19], Simone A. Joosten[19], W. Henry Boom[20], Bonnie Thiel[20], Harriet Mayanja-Kizza[21], Moses Joloba[21], Sarah Zalwango[21], Mary Nsereko[21], Brenda Okwera[21], Hussein Kisingo[21], Stefan H. E. Kaufmann[22], (GC6-74 principal investigator)*, Shreemanta K. Parida[22], Robert Golinski[22], Jeroen Maertzdorf[6], January Weiner 3rd[22], Marc Jacobson[22], Hazel M. Dockrell[23], Maeve Lalor[23], Steven Smith[23], Patricia Gorak-Stolinska[23], Yun-Gyoung Hur[23], Ji-Sook Lee[23], Amelia C. Crampin[24], Neil French[24], Bagrey Ngwira[24], Anne Ben-Smith[24], Kate Watkins[24], Lyn Ambrose[24], Felanji Simukonda[24], Hazzie Mvula[24], Femia Chilongo[24], Jacky Saul[24], Keith Branson[24], Sara Suliman[25], Thomas J. Scriba[25], Hassan Mahomed[25], E. Jane Hughes[25], Nicole Bilek[25], Mzwandile Erasmus[25], Onke Nombida[25], Ashley Veldsman[25], Katrina Downing[25], Michelle Fisher[25], Adam Penn-Nicholson[25], Humphrey Mulenga[25], Brian Abel[25], Mark Bowmaker[25], Benjamin Kagina[25], William Kwong Chung[25], Willem A. Hanekom[25], Jerry Sadoff[26], Donata Sizemore[26], S. Ramachandran[26], Lew Barker[26], Michael Brennan[26], Frank Weichold[26], Stefanie Muller[26], Larry Geiter[26], Desta Kassa[26], Almaz Abebe[26], Tsehayenesh Mesele[26], Belete Tegbaru[26], Debbie van Baarle[27], Frank Miedema[27], Rawleigh Howe[28], Adane Mihret[28], Abraham Aseffa[28], Yonas Bekele[28], Rachel Iwnetu[28], Mesfin Tafesse[28], Lawrence Yamuah[28], Martin Ota[29], Jayne Sutherland[29], Philip Hill[29], Richard Adegbola[29], Tumani Corrah[29], Martin Antonio[29], Toyin Togun[29], Ifedayo Adetifa[29], Simon Donkor[29], Peter Andersen[30], Ida Rosenkrands[30], Mark Doherty[30], Karin Weldingh[30], Gary Schoolnik[31], Gregory Dolganov[31] & Tran Van[31]

[18]DST-NRF Centre of Excellence for Biomedical TB Research and MRC Centre for TB Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa. [19]Department of Infectious Diseases, Leiden University Medical Centre, Leiden, the Netherlands. [20]Tuberculosis Research Unit, Department of Medicine, Case Western Reserve University School of Medicine and University Hospitals Case Medical Center, Cleveland, OH, USA. [21]Department of Medicine and Department of Microbiology, College of Health Sciences, Faculty of Medicine, Makerere University, Kampala, Uganda. [22]Department of Immunology, Max Planck Institute for Infection Biology, Berlin, Germany. [23]Department of Immunology and Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. [24]Karonga Prevention Study, Chilumba, Malawi. [25]South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine and Division of Immunology, Department of Pathology, University of Cape Town, Cape Town, South Africa. [26]Ethiopian Health and Nutrition Research Institute, Addis Ababa, Ethiopia. [27]University Medical Centre, Utrecht, the Netherlands. [28]Armauer Hansen Research Institute, Addis Ababa, Ethiopia. [29]Vaccines and Immunity Theme, Medical Research Council Unit, Fajara, The Gambia. [30]Department of Infectious Disease Immunology, Statens Serum Institute, Copenhagen, Denmark. [31]Department of Microbiology and Immunology, Stanford University, Stanford, CA, USA.

Methods

Study populations

The ACS, including a selection of progressors and controllers (also termed nonprogressors), has been previously described[17,24,52]. Briefly, 6,363 adolescents attending high schools in the Worcester region of the Western Cape, South Africa were enrolled and followed for 24 months. Among those with evidence of *M. tuberculosis* infection, by either a positive tuberculin skin test (TST) or a QuantiFERON-TB Gold In-Tube assay (QFT, QIAGEN), 44 progressors developed microbiologically confirmed (positive by sputum smear microscopy and/or MGIT (Mycobacteria Growth Indicator Tube) liquid culture), intrathoracic disease over 2 years of follow-up. Controllers also had evidence of *M. tuberculosis* infection, but did not develop TB disease during follow-up, and were matched to progressors for age, gender, ethnicity, school of attendance and prior history of TB disease. Participants were excluded from the progressor group if they developed TB within 6 months of enrollment (or the first TST- or IGRA-positive sample) to exclude early asymptomatic disease that could have been present at the time of assessment, or if they were HIV infected. Longitudinally collected PBMC samples were available from most participants at 6-monthly intervals (Fig. 1a). A noncash voucher to the value of approximately US$7 per visit was provided to adolescent participants. This voucher could be used at a local shopping mall. The Human Research Ethics Committee of the University of Cape Town approved the study (045/2005) and all participants provided written informed assent, while parents or legal guardians provided written, informed consent. All research was performed in accordance with relevant guidelines/regulations.

The GC6-74 project has also been previously described[17,24]. HIV-uninfected, household contacts of TB cases living in Cape Town, Western Cape, South Africa were enrolled and followed for up to 2 years, with assessments at baseline, 6 months and 18 months. Progressors developed microbiologically confirmed, pulmonary TB during follow-up. Controllers did not develop TB disease during follow-up and were matched at a ratio of 2:1 to progressors. Progressors in whom TB disease developed within 3 months of baseline were excluded to avoid sampling asymptomatic disease at baseline. The TST was performed at enrollment and PBMC samples were collected and stored at enrollment, 6 months and 18 months after enrollment (Fig. 1b). As TB exposure and risk of TB are strongly associated with age, analysis included samples from GC6-74 participants aged <20 years to match the ACS cohort. GC6-74 participants were compensated for transport expenses. The Stellenbosch University Institutional Review Board (N05/11/187) approved the study. Informed consent was obtained from adults, and from minors and their parents or legal guardians.

The adult TB patient cohort has been previously described[19]. The study enrolled patients with microbiologically confirmed active or previous pulmonary TB, who underwent medically indicated lung resections to treat TB or TB sequelae. DNA was extracted from lung tissue and matched blood samples using the appropriate DNAeasy kit (QIAGEN) as per manufacturer's instructions and subjected to bulk TCR-seq (Adaptive Biotechnologies). The Biomedical Research Ethics Committee of the University of KwaZulu-Natal approved the study (BE019/13). Written informed consent was obtained from all participants.

Finally, we utilized published CDR3β sequences generated using immunoSEQ in a CMV infection study[25]. The study used human peripheral blood samples obtained from the Fred Hutchinson Cancer Research Center Research Cell Bank biorepository of 666 healthy bone marrow donors, who underwent CMV serostatus testing. We restricted our analysis to 601 samples (274 CMV^+ and 327 CMV^-) where the sample metadata indicated the participant's HLA-alleles.

ScTCR-seq

Cryopreserved PBMCs from ACS participants were thawed, rested for 6 h and stimulated for 12 h with *M. tuberculosis* lysate (10 μg ml^{-1}, BEI Resources) in the presence of anti-CD49d antibody (1 μg ml^{-1}, BD Biosciences 340976) and anti-CD154-PE antibody (10 μl ml^{-1}, BD Biosciences, catalog no. 555700). PBS and staphylococcal enterotoxin B (1 μg ml^{-1}) were used as negative and positive controls, respectively. Samples from all study timepoints for a participant were processed on the same day. For samples with insufficient cells for a negative and/or control, only the *M. tuberculosis* lysate stimulation was performed. Next, cells were stained with LIVE/DEAD Fixable Aqua Stain (Thermo Fisher Scientific) for 30 min and, thereafter, stained with the following monoclonal antibodies for 60 min in a final volume of 50 μl: anti-CD19 (1 μl, BioLegend, catalog no. 302242), anti-CD14 (1 μl, BioLegend, catalog no. 301842), anti-CD3 (2 μl, BD Biosciences, catalog no. 563800), anti-CD4 (3 μl, BioLegend, catalog no. 300556), anti-CD8 (1 μl, BD Biosciences, catalog no. 561453), anti-TCR-αβ (2 μl, BD Biosciences, catalog no. 306720), anti-CD26 (1 μl, BioLegend, catalog no. 302704), anti-HLA-DR (1 μl, BioLegend, catalog no. 307636), anti-CD69 (0.2 μl, BD Biosciences, catalog no. 340560), anti-CD137 (0.5 μl, BD Biosciences, catalog no. 740798) and BD Horizon Brilliant Stain Buffer (37.3 μl, BD Biosciences, catalog no. 563794). Reverse transcription (RT) and sequence-specific amplification were performed in a series of three nested PCR analyses before sequencing on a MiSeq (Illumina) instrument. Single $\alpha\beta^+$ T cells staining $CD69^+CD137^+$ and/or $CD69^+CD154^+$ were index sorted by FACS (BD FACS Aria-II) into 96-well plates containing 12 μl of One-Step RT-PCR buffer (QIAGEN, 9.6 μl water + 2.4 μl of 5× buffer). Next, 2 μl of phase 1 TCR primers (final concentration 0.06 μM for each C primer and 0.12 μM for each V primer), 1 μl of phase 1 phenotyping primers (final concentration 0.5 μM for each primer), 0.8 μl of enzyme mix, 0.8 μl of dNTP and 0.2 μl of molecular-grade water were added per well. Plates were then placed on a thermocycler and the following thermal profile was used to perform phase 1 PCR: (1) 36 min at 50 °C, (2) 15 min at 95 °C, (3) 30 s at 94 °C, (4) 1 min at 62 °C, (5) 1 min at 72 °C, (6) go to step (3) for 25 cycles, (7) 5 min at 72 °C and (8) hold at 4 °C.

Next, separate 96-well plates were prepared for phase 2 PCR. Note that phase 2 TCR and phenotype PCR occurred in separate plates. Into each well 2 μl of 10× buffer (HotStarTaq DNA Polymerase Kit, QIAGEN), 0.4 μl of dNTP mix, 0.1 μl of HotStarTaq, 2 μl of phase 2 TCR primers (final concentration 0.06 μM each for C primer and 0.12 μM each for V primer), 1 μl of phase 2 phenotyping primers (final concentration 0.5 μM for each primer) and molecular-grade water up to 19 μl were added. Then 1 μl from the phase 1 PCR was used as a template for phase 2 TCR PCR and 1 μl from the phase 1 PCR was used as a template for phase 2 phenotype PCR. Plates were then placed on a thermocycler and the following thermal profile was used to perform phase 2 PCR: (1) 15 min at 95 °C, (2) 30 s at 94 °C, (3) 30 s at 62 °C, (4) 1 min at 72 °C, (5) go to step (3) for 25 cycles, (6) 5 min at 72 °C and (7) hold at 4 °C.

Then, separate plates were prepared for phase 3 PCR (barcoding). Into each well, 2 μl of 10× buffer (HotStar HiFidelity Polymerase Kit, QIAGEN), 0.4 μl of dNTP, 0.1 μl of DNA polymerase, 0.2 μl of paired end mix (50 μM each) and 10.3 μl of molecular-grade water were added. This was followed by 3 μl of 1:300 dilution of 100-μM column BC, 1:75 dilution of 100-μM alpha column to each column, 3 μl of 1:300 dilution of 100-μM row BC to each row and 1 μl of template from phase 2 PCR. Plates were then placed on a thermocycler and the following thermal profile was used to perform phase 3 PCR (barcoding): (1) 15 min at 95 °C, (2) 30 s at 94 °C, (3) 30 s at 62 °C, (94) 1 min at 72 °C, (5) go back to step (2) for 30 cycles, (6) 5 min at 72 °C and (7) hold at 4 °C. Primer sequences can be found in Supplementary Tables 9–14. CDR3β sequences of sorted mycobacteria-reactive T cells from controllers and progressors were compiled using CDR3β sequences from mycobacteria-reactive T cells collected from healthy *M. tuberculosis*-infected adolescents from our previous studies[13,15]. CMV-, EBV- and influenza A-specific CDR3β sequences were obtained from VDJdb[53]. MAIT Match was used for classification of CDR3α as MAIT cell sequences[54]. CDR3α sequences with MAIT Match similarity score ≥0.95 were classified as MAIT cells.

Bulk TCR-seq

Genomic DNA was extracted from unstimulated PBMCs of participants using the QIAGEN QIAamp DNA Blood Mini Kit. The immunoSEQ assay (Adaptive Biotechnologies) was performed to quantify TCR CDR3α and CDR3β sequences[55]. We also accessed a database of published CDR3β sequences obtained using the immunoSEQ assay from adults diagnosed with TB, who underwent clinically indicated lung resections[19].

CDR3β sequences within the bulk TCR dataset that matched amino acid CDR3β sequences of sorted, antigen-specific single T cells, or which had common GLIPH2 CDR3β amino acid motifs, were classified as mycobacteria reactive.

Clustering CDR3β sequences

TCRα and CDR3β sequences generated with single-cell sequencing from T cells stimulated in vitro with TB-specific antigens and sorted based on coexpression of CD69 and CD154 or CD69 and CD137 were included in GLIPH2 (ref. 13) and TCRdist3 (ref. 26) analyses using def ault settings. CDR3β sequences from sorted T cells activated by *M. tuberculosis* lysate stimulation from progressor and control PBMCs at any of the study timepoints and sorted T cells activated by TB-specific antigen stimulation from healthy, *M. tuberculosis*-infected adolescents from two previous studies, also from the larger ACS[13,15], were combined. For GLIPH2 analysis, the pooled CDR3αβ sequence list from progressors, controllers and healthy *M. tuberculosis*-infected adolescents was uploaded to the GLIPH2 server (http://50.255.35.37:8080). We selected GLIPH2 similarity clusters that consisted of three or more unique CDR3 sequences and were present in three or more participants, with a Fisher_score ≤0.05, vb_score ≤0.05 and length_score ≤0.05. Among the identified GLIPH2 similarity clusters, we identified those with significant HLA-allele associations (at the level of two-digit HLA typing), using Fisher's exact test at ≤0.05 (Fig. 5a). Among GLIPH2 similarity cluster:HLA combinations, we then identified those with differentially abundant TCR sequences between controllers and progressors at a *P*-value threshold <0.05 and Benjamini–Hochberg FDR $q < 0.2$.

To explore the specificity of the results obtained from the differential abundance analysis of *M. tuberculosis*-reactive TCR clusters, we performed permutation analyses using randomized disease outcome labels and determined the number of significantly associated clusters from 1,000 iterations.

To identify metaclonotypes using TCRdist3, we used a script published by Mayer-Blackwell et al.[26] (https://tcrdist3.readthedocs.io/en/latest/public.html). We applied TCRdist3 analysis to the combined CDR3β sequence from progressors, controllers and healthy *M. tuberculosis*-infected adolescents following the same pipeline as used for GLIPH2 (shown in Fig. 3a).

Nonlinear spline analysis

Nonlinear spline analysis of longitudinal background subtracted frequencies of T cells coexpressing CD69 and CD154 or CD69 and CD137 and differentially abundant *M. tuberculosis*-reative TCR clusters was performed using the smooth.spline function in R with four degrees of freedom; 2,000 iterations were performed to compute the 95% CIs.

Cell culture and cell lines

Candidate TCR-α and -β chains were transduced into the nuclear factor of activated T cells (NFAT) reporter-stable J76-NFATRE-luc T cell line, which is deficient for both TCR-α and TCR-β chains[13]. Candidate HLA-alleles were individually transduced into artificial antigen-presenting cells (aAPCs), which were constructed using lentiviral transduction of CD80 and HLA-DM molecules into K562 cells.

Antigens

M. tuberculosis whole-cell lysate (strain H37Rv) and *M. tuberculosis* gateway clone set (plates 1–42) were kindly provided by BEI Resources. For the whole-proteome production, every 12 open read frame clones from each plate row were pooled together as a subpool and expressed using the Expressway Cell-Free Expression System[13]. Megapool peptides, containing 300 epitopes from 90 *M. tuberculosis* proteins, were kindly provided by A. Sette (La Jolla Institute for Allergy & Immunology). For epitope screening of each identified protein, overlapping peptide libraries were purchased from Elim Biopharm.

Antigen screen

For protein stimulation, 50 µl of aAPCs (10^6 per ml) was preloaded with the Expressway product mixture and individual proteins in a range of 10–10,000 dilutions or protein subpools at 10-fold dilution, for 3 h at 37 °C in the standard cell culture medium. Then, 50 µl of TCR-transduced J76-NFATRE-luc cells (10^6 per ml) were added and cocultured with aAPCs for 8 h. Then cells were harvested and luciferase activity was measured using Nano-Glo Luciferase Assay (Promega). Fold induction of luciferase activity was calculated relative to unstimulated samples. For peptide stimulation, 50 µl of TCR-transduced J76-NFATRE-luc cells (10^6 per ml) was cocultured with 50 µl of HLA-transduced K562 cells (10^6 per ml) in a 96-well plate. A peptide pool or individual peptide was added to the well at 2 µg ml^{-1}. After incubation for 8 h, cells were harvested and luciferase activity was measured.

Statistics and reproducibility

No statistical method was used to predetermine sample size; the sample size was based on the availability of PBMC vials stored from progressors. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. ScTCR-seq plates containing samples from five controllers and six progressors were found to have been contaminated and data from these TCR-seq plates were excluded. Bulk TCR-seq data of one sample from an ACS participant was excluded because the sample did not show strong alignment with corresponding samples from the same participant. Bulk TCR-seq data of ten samples from the GC6-74 were excluded because these samples failed quality control metrics based on sample repertoires. Six failed due to failed material transfer and four did not show strong alignment with corresponding samples from the same participant. Mycobacteria-reactive TCR similarity groups detected by scTCR-seq were compared using a two-tailed Student's *t*-test. We used Fisher's exact test at ≤0.05 to identify GLIPH2 similarity clusters that were HLA associated. GLIPH2 similarity cluster:HLA combinations that were differentially abundant between controllers and progressors were identified using the two-sided Mann–Whitney *U*-test at a *P*-value threshold <0.05 and Benjamini–Hochberg FDR threshold $q < 0.2$.

Reporting summary

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

Data availability

The datasets and scripts to generate the manuscript figures are available at https://github.com/SATVILab/DataTidyMusvosviTCRseq. The raw bulk CDR3α and CDR3β sequence data from the ACS and GC6-74 participants are available at https://doi.org/10.21417/MM2022NM.

References

52. Scriba, T. J. et al. Differential recognition of *Mycobacterium tuberculosis*-specific epitopes as a function of tuberculosis disease history. *Am. J. Respir. Crit. Care Med.* **196**, 772–781 (2017).
53. 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).
54. Nielsen, M. MAIT Match-1.0. https://services.healthtech.dtu.dk/service.php?MAIT_Match-1.0
55. Robins, H. S. et al. Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood* **114**, 4099–4107 (2009).

Acknowledgements
We thank the study participants who enrolled in the ACS, GC6-74 study and the adult TB patient study. We acknowledge the considerable contributions of study clinicians, nurses, technicians and clinical research workers. We thank C. Schreuder, O. Nombida and N. Gupta for assistance with sample shipments and logistics. We thank the Cape Town HIV Vaccine Trials Network Laboratory for use of their FACS facility and the Human Immune Monitoring Center for processing of scTCR-seq assays. We thank V. Mizrahi and S. Gagneux for critical reading and comments on the manuscript. The following reagents were obtained through BEI Resources, National Institute of Allergy and Infectious Diseases, National Institutes of Health: *M. tuberculosis*, strain H37Rv, whole-cell lysate, NR-14822. M.M. was supported by the Carnegie Corporation of New York. This work was supported by the Bill and Melinda Gates Foundation Global Health grants (nos. OPP1066265, OPP1023483 and OPP1065330), the Grand Challenges in Global Health (GC6-74, grant no. 37772) and the Howard Hughes Medical Institute. The Stanford Center for Human Systems Immunology was also supported by Bill and Melinda Gates Foundation grant OPP1113682. The ACS study was also supported by Aeras and BMGF GC12 (grant no. 37885) for QuantiFERON-TB Gold In-Tube testing.

Author contributions
M.M. and H.H. contributed to data generation, analysis, interpretation and drafting of the manuscript. C.W. and P.A. contributed to the analysis and interpretation of the data. Q.X., A.K. and A.C. contributed to the generation of data. V.R., G.O., N.B. and M.F. provided administrative support to ensure sample processing and data generation. A.L. and S.M.B. provided bulk TCR-seq data from match blood and lung tissue samples. W.A.H. designed and directed the ACS study and acquired funds for the collection and storage of samples from the ACS and GC6-74 participants. S.H.E.K. and G.W. designed and acquired funds for the collection and storage of samples from the GC6-74 participants. M.H. developed the study design and acquired funds to complete the study. M.M.D. and T.J.S. developed the study design, acquired funds, interpreted data and drafted the manuscript.

Competing interests
The authors declare no competing interests.

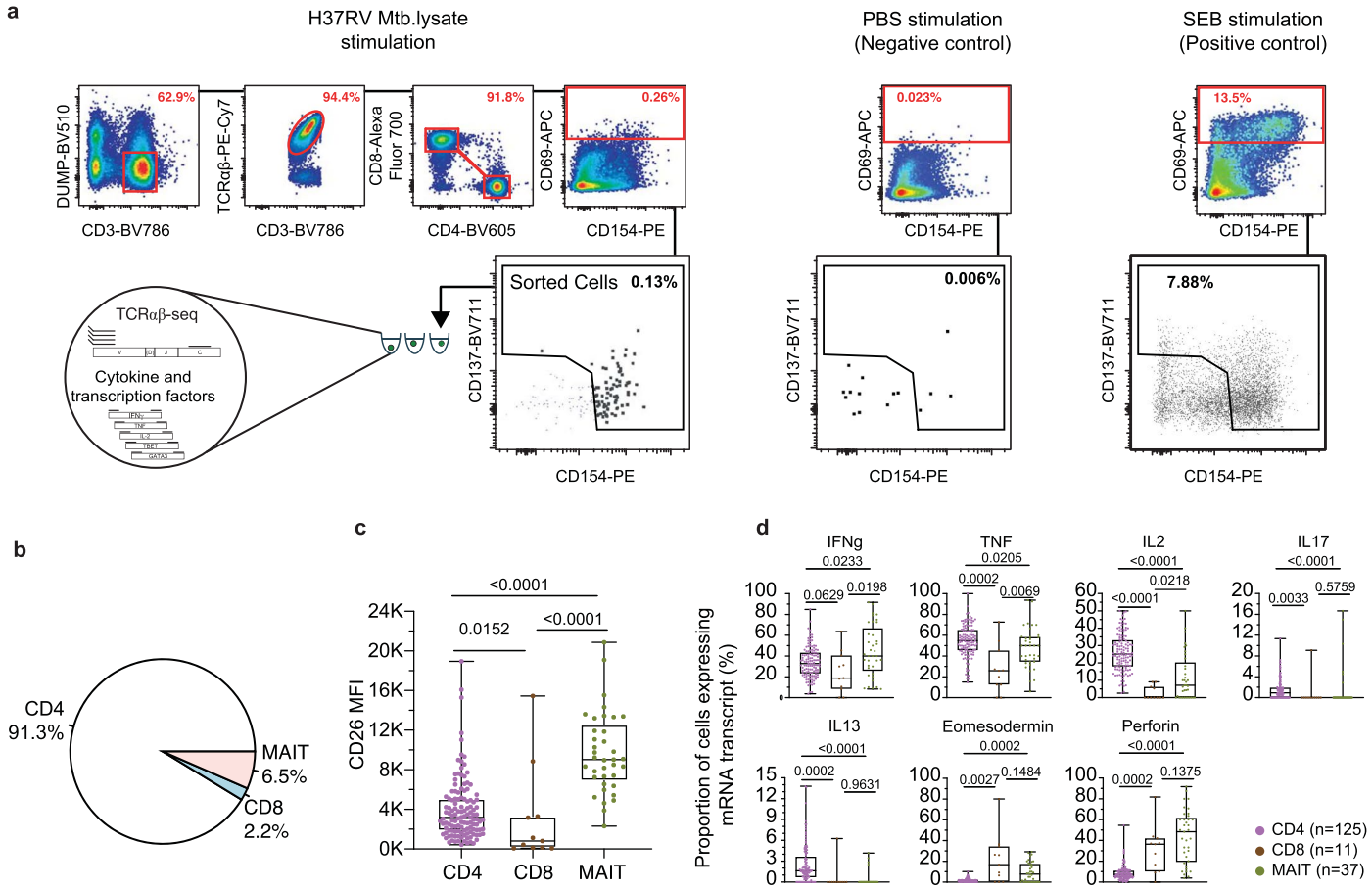
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41591-022-02110-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-022-02110-9.

Correspondence and requests for materials should be addressed to Thomas J. Scriba.

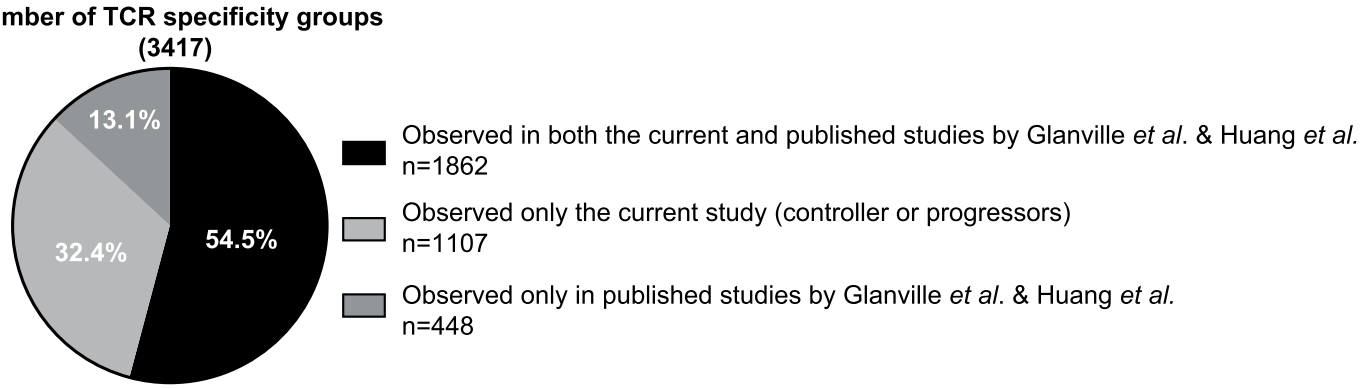
Peer review information *Nature Medicine* thanks Joel Ernst, Paul Ogongo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Saheli Sadanand and Joao Monteiro, in collaboration with the *Nature Medicine* team.

https://doi.org/10.1038/s41591-022-02110-9



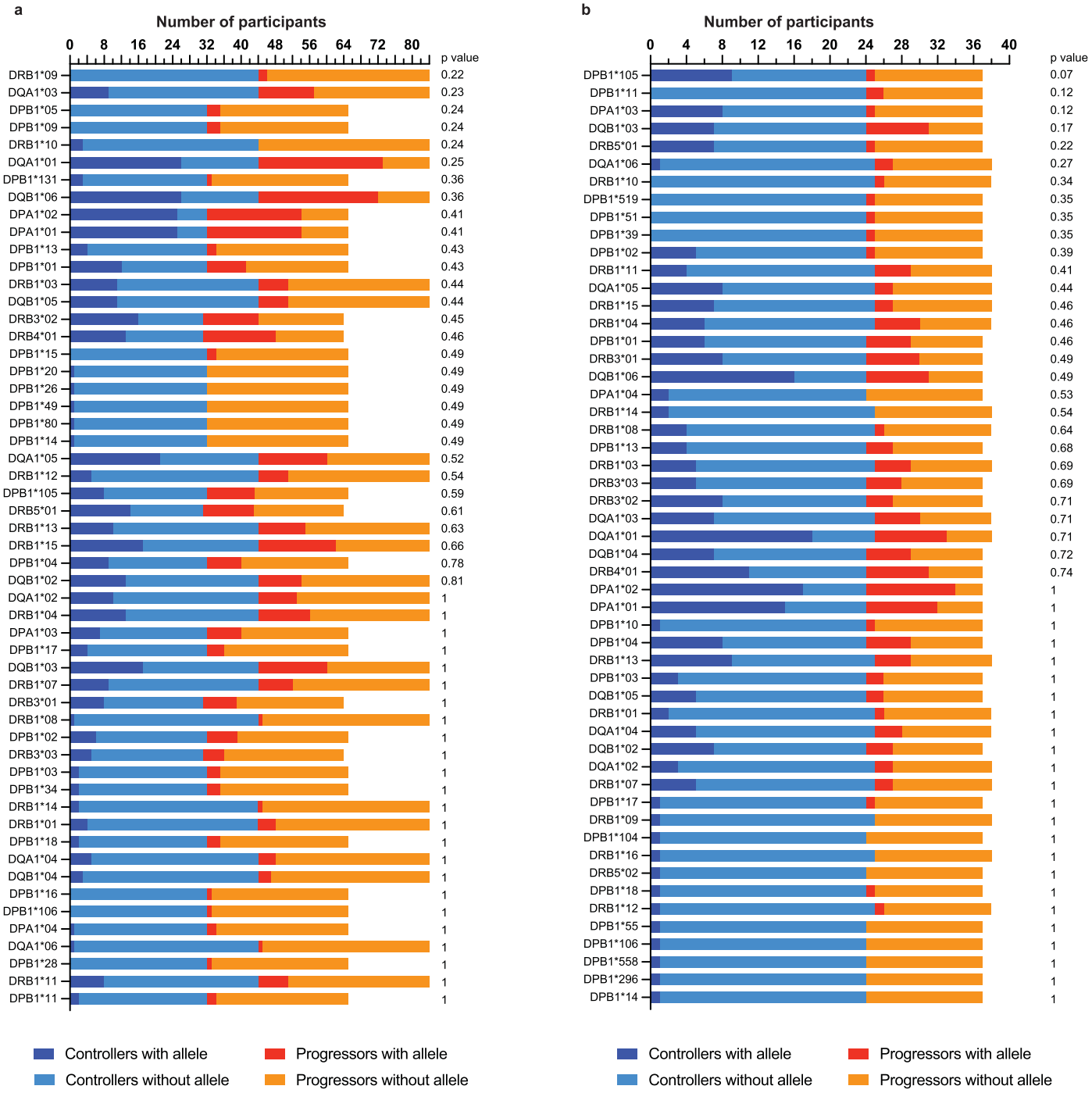
Extended Data Fig. 1 | CD4 T cells are the predominant responding T cell subset to M.tb lysate stimulation. (**a**) Representative flow cytometry plots depicting the gating strategy used to identify and sort activated M.tb lysate-reactive T cells in PBMC stimulated with M.tb lysate. M.tb-specific T cells were defined as live CD3+, TCRαβ+ CD4+ or CD8+ T cells that co-express CD69 and CD154 or CD69 and CD137. Note the value in each plot is the percentage of parent (red) or grandparent (black). Plots depict data from M.tb lysate stimulation, PBS (negative control) and SEB (positive control) stimulations. (**b**) Pie chart depicting the proportions of sorted single T cells that responded to M.tb lysate stimulation by co-expressing CD69 and CD154 or CD69 and CD137, by CD4, CD8 or MAIT cell subset. (MAIT cells were defined as expressing canonical MAIT CDR3α sequences, determined using MAIT Match50). (**c**) Expression levels of CD26, measured as median fluorescent intensity (MFI), on sorted CD4, CD8 or MAIT cells from ACS controllers and progressors. The horizontal lines represent medians, the bounds of the boxes indicate the 25% and 75% percentile and the whiskers represent the minima and maxima. Each dot represents an individual sample and only samples with ≥10 cells were included. (**d**) Proportions of CD4, CD8 or MAIT cells that express the indicated mRNA transcripts in sorted T cells from ACS controllers and progressors. The horizontal lines represent medians, the bounds of the boxes indicate the 25% and 75% percentile and the whiskers represent the minima and maxima. Each dot represents an individual sample and only samples with ≥10 cells were included. A threshold of ≥5 transcript reads per cell was used to classify a cell as positive for the indicated transcript. P-value were calculated with the Mann-Whitney U test (two-sided).

https://doi.org/10.1038/s41591-022-02110-9



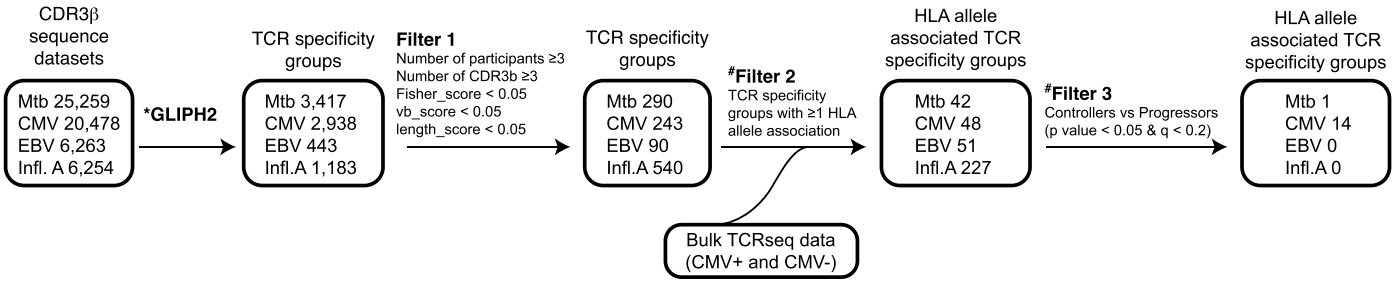
Extended Data Fig. 2 | The majority of M.tb TCR specificity groups contain TCRβ sequences observed in two independent single cell TCR sorting experiments. Pie chart depicting the proportions of 3,417 M.tb TCR specificity groups, which contain TCR sequences identified by single cell TCR sequencing in both the present study and the previously published single cell TCR datasets by Glanville et al.[15] & Huang et al.[13] (black), or those identified only in the present study (dark grey), or reported only in the Glanville et al.[15] & Huang et al.[13] single cell TCR datasets (light grey).

https://doi.org/10.1038/s41591-022-02110-9



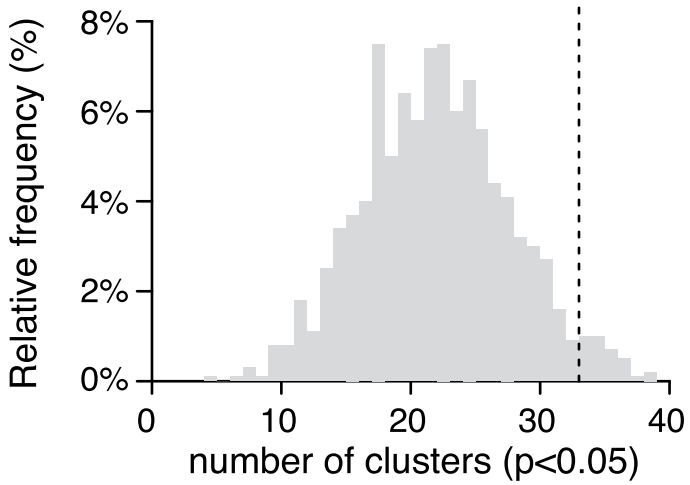
Extended Data Fig. 3 | HLA allele distributions in controllers and progressors are not different. Bar plots depicting the number of controllers and progressors in the (**a**) ACS (n = 88) and (**b**) GC6 (n = 38) cohorts with or without the indicated HLA-allele. The Fisher's exact test (two-sided) was used to compare proportions of HLA alleles between controllers or progressors. Some HLA loci have less participants because a few participants did not have complete HLA typing results at all class I and II loci.

https://doi.org/10.1038/s41591-022-02110-9



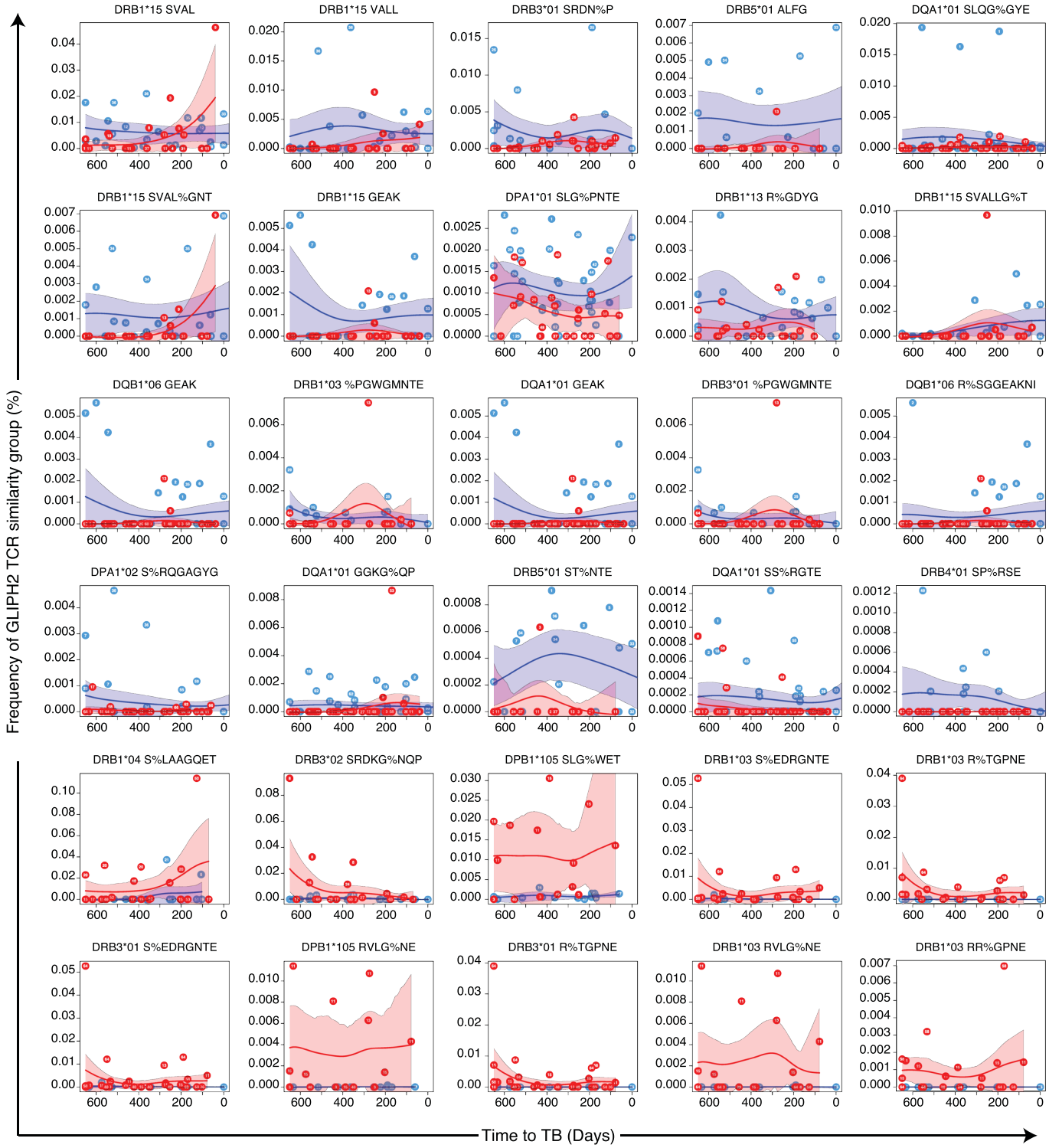
Extended Data Fig. 4 | Differentially abundant CMV-specific TCR similarity groups in CMV+ persons compared to CMV- persons. Analysis workflow used to measure the frequencies of GLIPH2 TCR similarity groupings from mycobacteria-reactive (Mtb) or CMV, EBV or Influenza-A (Infl.A)-specific CDR3β sequences in CMV+ and CMV- persons. GLIPH2 analysis was performed and the resulting GLIPH2 similarity groups were filtered initially using the criteria listed under Filter 1. We then selected TCR similarity groups with significant HLA allele associations in the CMV+/CMV- cohort (Filter 2). Finally, we identified similarity groups that were differentially abundant in CMV+ and CMV- participants bearing the associated HLA allele (Filter 3).

https://doi.org/10.1038/s41591-022-02110-9



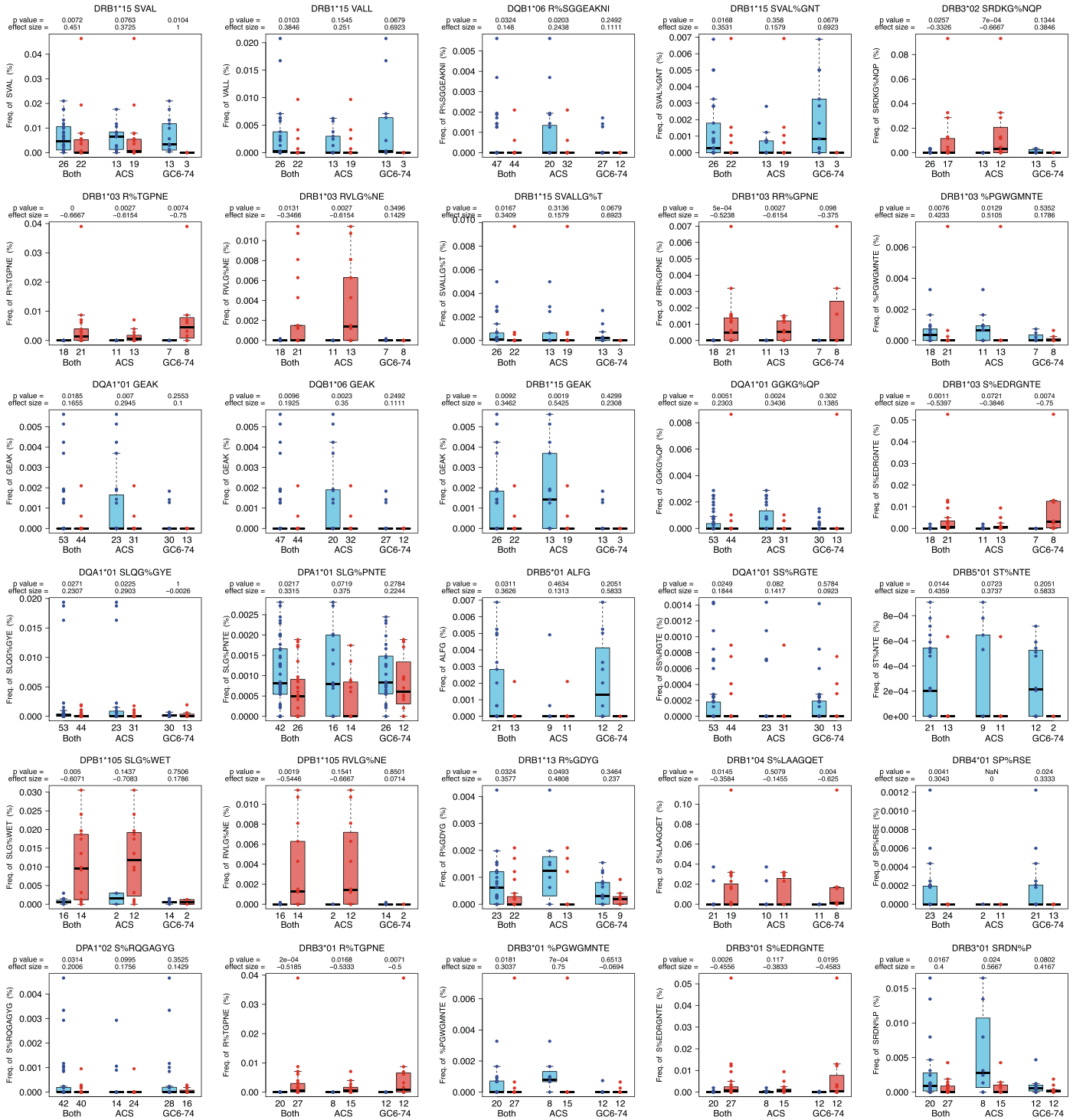
Extended Data Fig. 5 | Number of TCR clusters that pass a nominal p-value threshold in permutation analysis when disease outcome label is randomized. Histogram showing the distribution of the number of TCR clusters identified at a nominal p value < 0.05 when controller and progressors status was randomized in permutation analyses with 1000 iterations. The vertical line on the right indicates the number of TCR clusters observed when the correct controller or progressor classification was used (33). 44 (4.4%) of the 1000 iterations exceeded 33.

https://doi.org/10.1038/s41591-022-02110-9



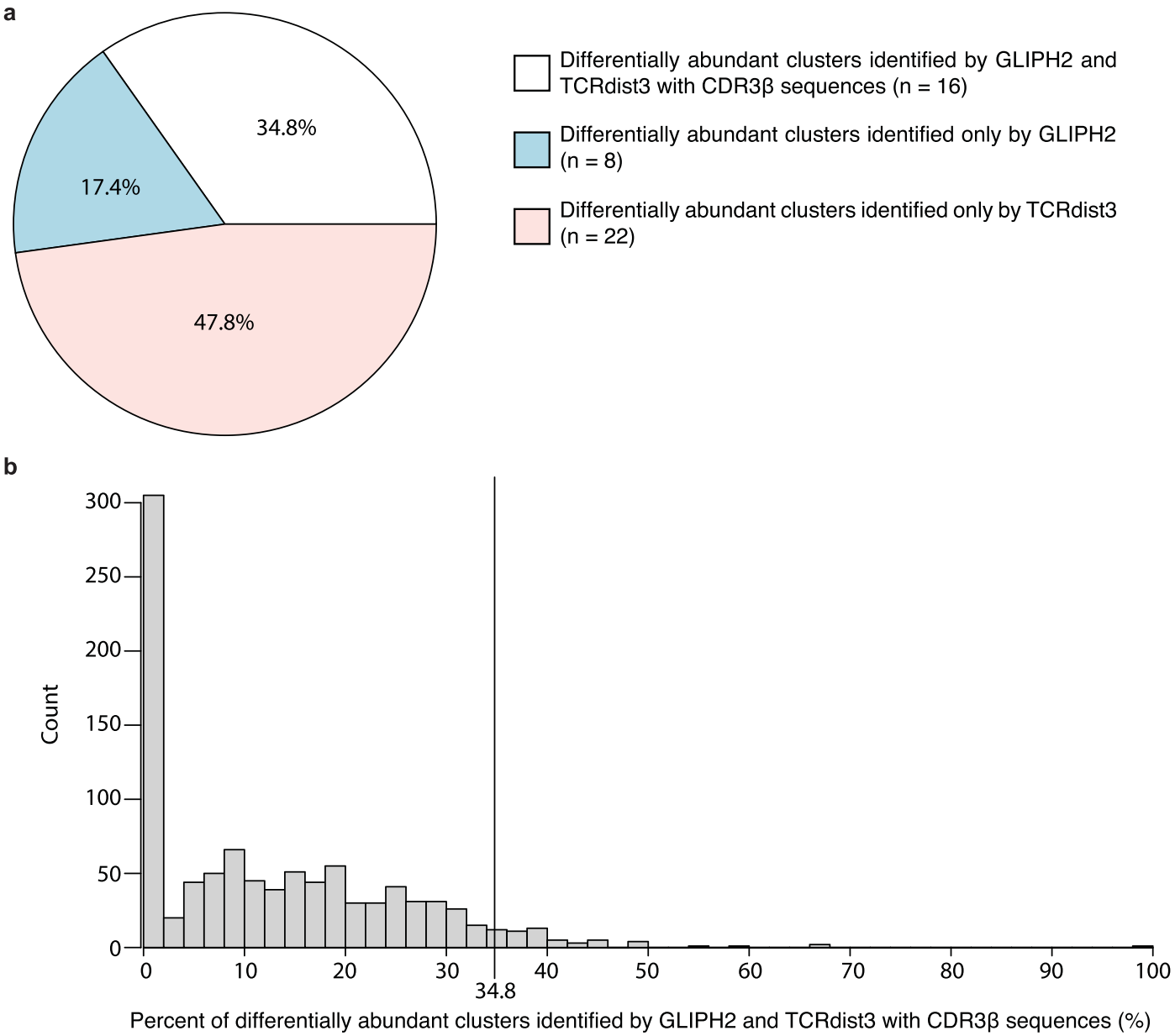
Extended Data Fig. 6 | Longitudinal kinetics of differentially abundant TCR similarity clusters. Non-linear spline plots depicting the longitudinal kinetics of differentially abundant TCR similarity clusters in controllers and progressors expressing the indicated HLA allele. Samples were aligned to time to TB for progressors. For controllers, the 'time to TB' of their respective age-matched progressor was used. The solid lines indicate the modeled non-linear splines and the shaded bands represent 95%CI.

https://doi.org/10.1038/s41591-022-02110-9



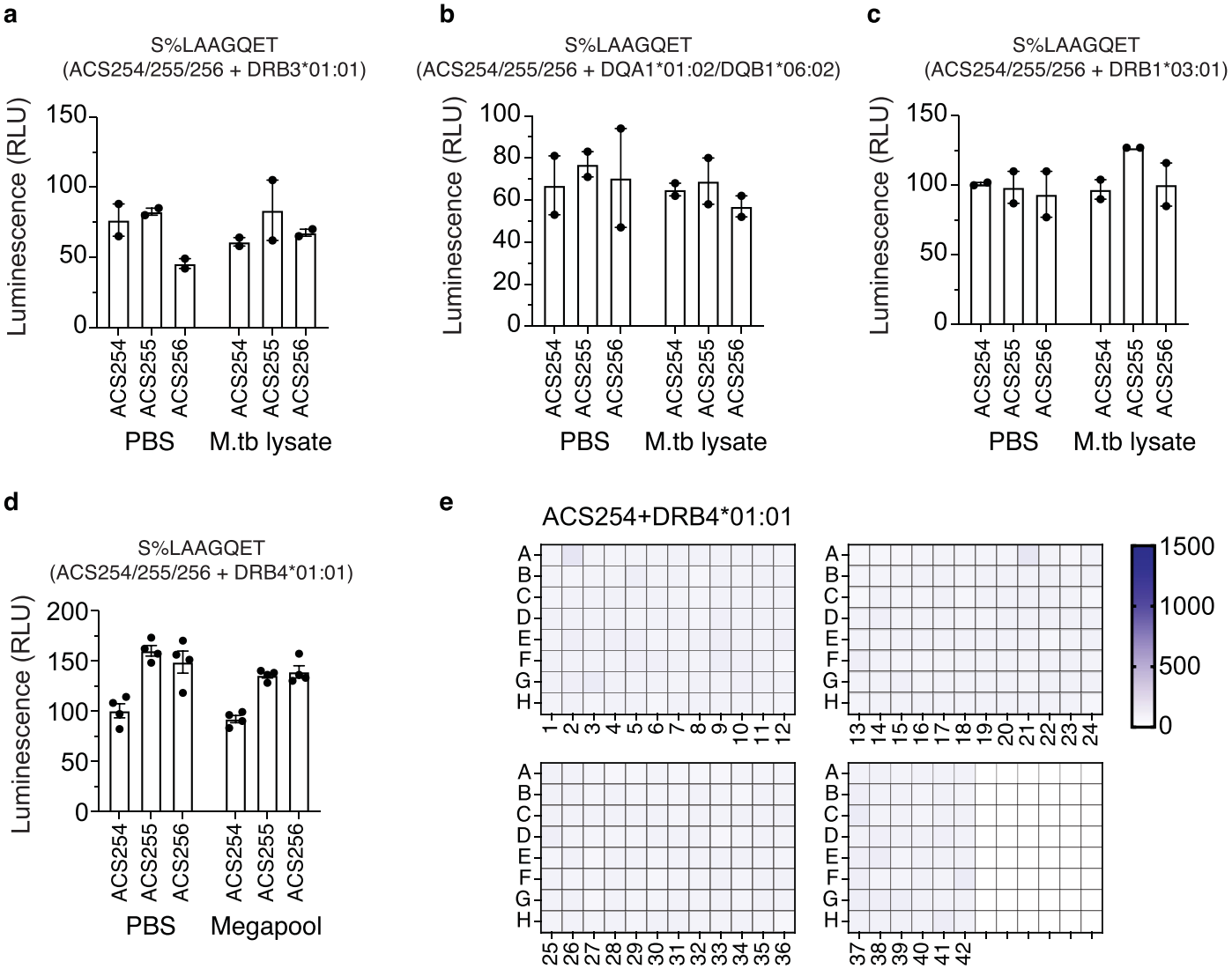
Extended Data Fig. 7 | Frequencies of mycobacteria-reactive HLA-allele-associated TCR similarity groups in each of the the ACS and GC6-74 cohorts. Box and whisker plots depicting frequencies of mycobacteria-reactive TCRs belonging to the indicated HLA-allele-associated TCR similarity groups that were significantly more abundant in controllers or progressors bearing the indicated HLA-allele when data from ACS and GC6-74 participants was combined, or when ACS and GC6-74 samples were assessed separately. The horizontal lines represent medians, the boxes the interquartile range and the whiskers are the range. The number of samples from controllers and progressors are indicated below each plot. The p-value (Mann-Whitney U test, two-sided) and the effect size (Cliff's Delta) are indicated above each plot.

https://doi.org/10.1038/s41591-022-02110-9



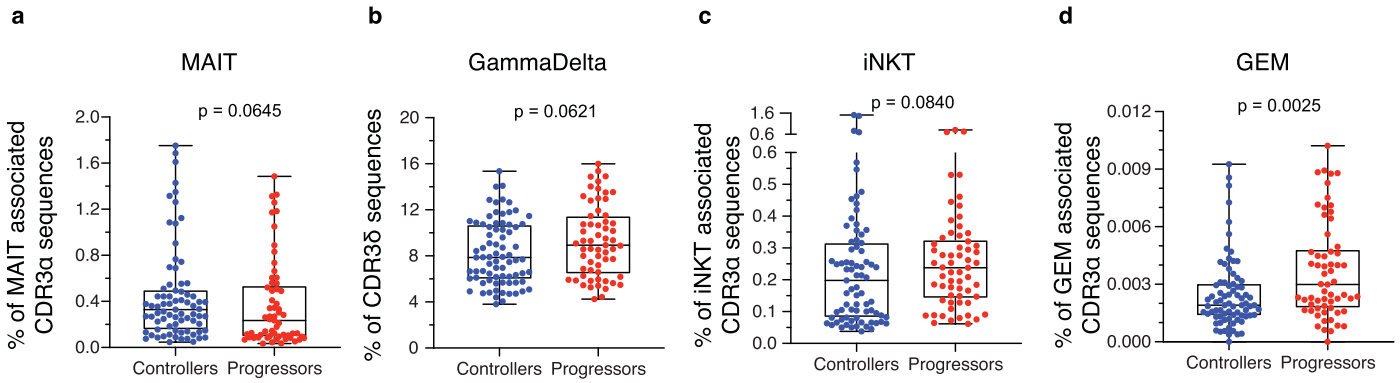
Extended Data Fig. 8 | Considerable overlap between GLIPH2-identified TCR similarity clusters and TCRdist3-identified metaclone clusters found to be differentially abundant between controllers and progressors. (**a**) Pie chart showing the proportions of differentially abundant GLIPH2 similarity clusters that shared at least one CDR3β sequence with a differentially abundant TCRdist3 metaclone cluster. (**b**) Histogram showing the distribution of overlap in TCR clusters with significant association with outcome obtained from GLIPH2 or tcrdist3 when the controller and progressor status was randomized 1000 times. The vertical line on the right indicates the proportion of overlap observed when the correct classification was used (34.8%), which falls at the 94.8th percentile of the 1000 permutations.

https://doi.org/10.1038/s41591-022-02110-9



Extended Data Fig. 9 | S%LAAGQET antigen discovery screen. Barplot showing the relative luminescence signal after an 8-hour PBS or M.tb lysate stimulation of TCR-ACS254, ACS255, or ACS256 in the context of (**a**) DRA*01:01/DRB3*01:01, (**b**) DQA1*01:02/DQB1*06:02 or (**c**) DRA*01:01/DRB1*03:01. The mean and range (n = 2 biological replicates) is shown (**d**) Barplot showing the relative luminescence signal after an 8-hour PBS or Mtb300 megapool stimulation of TCR-ACS254, ACS255, or ACS256 in the context DRA*01:01/DRB4*01:01. The mean and SEM (n = 4 biological replicates) is shown (**e**) Antigen screening of the whole M.tb proteome (321 subpools displayed in 3.5 plates) by TCR-transfected clone TCR-ACS254. Color scale indicates the relative luminescence signal after 8-hour stimulation of the clone.

https://doi.org/10.1038/s41591-022-02110-9



Extended Data Fig. 10 | Frequencies of donor unrestricted T (DURT) cells in controllers and progressors. Box and whisker plot showing frequencies of (a) mucosal associated invariant T (MAIT) cells, (b) γδ T cells, (c), germline-encoded mycolyl lipid-reactive (GEM) T cells, and (d) invariant natural killer T (iNKT) cells, estimated from canonical TCR CDR3α sequences in PBMC samples collected from controllers (n = 77) and progressors (n = 61). Mucosal associated invariant T cells (MAIT), gammadelta, iNKT, and germline-encoded, mycolyl lipid reactive (GEM) T cells were defined as MAIT match score >=0.95, TCRDJ gene, TCRAV10;TCRAJ18 CVVSDRGSTLGRLYF, and TCRAV01-02;TCRAJ09 CAV[RL]. TGGFKTIF, respectively, with '[]' containing the permitted amino acid and '.' denoting any amino acid. The horizontal lines represent medians, the bounds of the boxes indicate the 25% and 75% percentile and the whiskers represent the minima and maxima. Each dot represents an individual sample (controllers, n = 77; progressors, n = 61). The Mann-Whitney U test (two-sided) was used to compare frequencies between groups. P-values have not been corrected for multiple comparisons.

nature portfolio

Corresponding author(s): Thomas J. Scriba

Last updated by author(s): Oct 20, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Single-cell TCR sequences were acquired using Illumina instrumentation and software as described in the method session. The pipeline used to analyze single-cell TCR sequences in this study has been described in the previous published literature: Han, A., Nat Biotechnol 32, 684-692 (2014). Specifically, TCR V, D, and J segments were assigned by VDJFasta-1.0, in combination with blast-2.2.17, HMMER3, and Perl 5.22. Bulk TCR sequencing was performed by Adaptive Biotechnologies' immunoSEQ Technology.
Data analysis	CDR3b sequences from Mtb-reactive CD4 T cells were clustered into GLIPH specificity clusters and metaclone clusters using GLIPH version 2 (http://50.255.35.37:8080/) and tcrdist3-0.2.0 (https://tcrdist3.readthedocs.io/en/0.2.0/). Default parameters were used for GLIPH2 and tcrdist3 analysis. Flow cytometry data was analyzed using FlowJo v10. data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The datasets and scripts to generate the manuscript figures are available at https://github.com/SATVILab/DataTidyMusvosviTCRseq. The raw bulk CDR3 alpha and CDR3 beta sequence data from the ACS and GC6-74 participants is available at http://doi.org/10.21417/MM2022NM.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Participants in the progressor and controller groups were matched for age, gender and ethnicity. Sex and gender-based analysis were not performed due to limited sample sizes. Sex and age disaggregated data is available at https://github.com/SATVILab/DataTidyMusvosviTCRseq and http://doi.org/10.21417/MM2022NM
Population characteristics	The ACS progressor and controller cohort: 44 adolescents, aged 12-18, among 6,363 adolescents, developed microbiologically-confirmed intrathoracic disease during 2 years of follow-up (Progressors). 44 adolescents who had M.tb infection, but did not develop TB disease during follow-up, and were matched to progressors for age, gender, ethnicity, school of attendance and prior history of TB disease (Controllers). The Adolescent Cohort Study was an epidemiological study that enrolled adolescents attending high schools in the Worcester region of the Western Cape, South Africa, and followed up participants for 24 months. The GC6-74 progressor and controller cohort: 12 participants, aged 12-19 years, who developed microbiologically confirmed pulmonary TB for an HIV-uninfected, household contacts cohort were selected. 25 participants matched for age, gender, ethnicity who did not develop TB disease during follow-up were selected from this household contact cohort. The adult TB patient cohort: 11 patients, aged 22-68 with microbiologically confirmed active or previous pulmonary TB, who underwent medically indicated lung resections to treat TB or TB sequelae were enrolled.
Recruitment	The ACS participants were part of larger epidemiological study that enrolled 6,363 adolescents attending high schools in the Worcester region of the Western Cape, South Africa, and followed for 24 months. Healthy household contacts living in Cape Town, South Africa were enrolled into the the GC6-74 study and followed up for 24 months. For both the ACS and GC6 we matched controls by age, gender, ethnicity, but there could be other factors such as force of infection that could be different between progressors and controllers that could influence our results. The adult TB patient who underwent medically indicated lung resections to treat active TB or TB sequelae at King Dinuzulu Hospital and Inkosi Albert Luthuli Central Hospitals in Durban, KwaZulu-Natal cohort were recruited into the adult TB patient cohort. Lung tissue could only be obtained from adult who underwent medically indicated lung resections, therefore our analysis could be bias by studying a participant group with extensive disease.
Ethics oversight	For the Adolescent Cohort Study, the Human Research Ethics Committee of the University of Cape Town approved the study (045/2005) and all participants provided written informed assent, while parents or legal guardians provided written, informed consent. All research was performed in accordance with relevant guidelines/regulations. For the GC6-74 cohort, the Stellenbosch University Institutional Review Board (N05/11/187) approved the study. Informed consent was obtained from adults. Informed assent was obtained from minors and informed consent was obtained from their parents or legal guardians. For the adult TB patient cohort, the Biomedical Research Ethic Biomedical Research Ethics Committee (BREC) of the University of KwaZulu-Natal approved the study (BE019/13). Written informed consent was obtained from all participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size was based on the availability of PBMC vials stored from progressors.
Data exclusions	Single cell TCR sequencing plates containing samples from 5 controllers and 6 progressors was found to have been contaminated and data from these plates was excluded. Data from one sample from the ACS sample was excluded from the bulk TCR sequence database because the sample did not show strong alignment with corresponding samples from the same participant. Ten bulk TCR sequencing samples from the GC6-74 were excluded because these samples failed QC metrics based on sample repertoires. Six failed due to failed material transfer and four samples did not show strong alignment with corresponding samples from the same participant.
Replication	Verification of reproducibility of differentially abundant M.tb TCR similarity groups between controllers and progressors could not be performed because samples from an independent cohort of controllers and progressors were not available
Randomization	Study randomization was not applicable for the study, because the progressors were identified prospectively. To limit bias, we did not randomly select controllers, but selected controllers matched to progressors for age, gender, ethnicity, and prior history of TB disease.
Blinding	Laboratory personnel were not blinded during sample processing or analysis

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-CD49d, BD Biosciences 340976, Clone:L25 anti-CD3-BV786, BD Biosciences, Cat 563800, Clone:SK7 anti-CD4-BV605, Biolegend, Cat 300556, Clone:RPA-T4 anti-CD8-Alexa Fluor 700, BD Biosciences, Cat 561453, Clone:RPA-T8 anti-TCRαβ-PE-Cy7, Biolegend, Cat 306720, Clone:IP26 anti-CD69-APC, BD Biosciences, Cat 340560, Clone:L78 anti-CD137-BV711, BD Biosciences, Cat 740798, Clone:4B4-1 anti-CD154 (CD40L)-PE, BD Biosciences, Cat 555700, Clone:TRAP1 anti-CD26-FITC, Biolegend, Cat 302704, Clone:BA5b anti-HLA-DR-BV421, Biolegend, Cat 307636, Clone:L243 anti-CD14-BV510, Biolegend, Cat 301842, Clone:M5E2 anti-CD19-BV510, Biolegend, Cat 302242, Clone:HIB19
Validation	All monoclonal antibodies were purchased from commercial suppliers (BD Biosciences and Biolegend). The manufacturers state that these antibodies are research use only (ROU) and have been tested for flow cytometry application using human samples. We titrated all antibodies to determine the optimal staining volumes

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	The original Jurkat T cell line, K562 cell line and HEK-293T cell line were obtained from the ATCC. Jurkat 76 T-cell line, was engineered using the Jurkat T cell line from ATCC and kindly provided by Dr. Shao-An Xue (Department of Immunology,

	University College London).
Authentication	Non authenticated
Mycoplasma contamination	Negative
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell line was used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cryopreserved PBMCs from ACS participants were thawed, rested for 6 hours and stimulated for 12 hours with M.tb lysate (10µg/mL, BEI Resources) in the presence of anti-CD49d antibody (1µg/mL) and anti-CD154-PE antibody (10µL/mL). After stimulation, cells were harvested and stained with surface markers. Dead cells were stained using LIVE/DEAD™ Fixable Aqua Dead Cell Stain Kit from Thermo Fisher scientific. Activated CD4+ and CD8+ T cells were single-cell sorted into 96-well plate for single cell TCRa/b sequencing.
Instrument	Sorting was performed on a BD FACS Aria-II (BD Biosciences)
Software	Data were analyzed using R and FlowJo v10.
Cell population abundance	The median frequencies of activated (CD69+ and CD154+ or CD69+ and CD137+) T cells in progressor and controller samples was 0.236% (range, 0.03% -1.832%) and 0.207% (range, 0.02% - 0.667%), respectively. T cells directly in in 96-well plate containing One-Step RT-PCR buffer.To determine the purity of our cell sorts we compared the expression of CD26, a marker associated with MAIT cells, on sorted cells expressing known canonical MAIT CDR3a sequences and compared this to sorted CD4 and CD8 T cells expressing other non MAIT CDR3a sequences. We observed that the expression of CD26 aligned with the T cell subset identities.
Gating strategy	A lymphocyte gate was set on Forward scatter (FSC-A) and side scatter (SSC-A), followed by a singlet gate set on FSC-A and FCS-H. A LIVE/DEAD negative, CD14 negative, and CD19 negative, CD3 positive (live T cells) gate was set, followed by ab TCR positive gate. Next CD4 positive or CD8 positive ab T cells were selected. Gate for activated population (CD69+ and CD154+ or CD69+ and CD137+) was set based on comparing negative (PBS) treatment and M.tb lysate treatment.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.