

Dynamics of peripheral T cell phenotype and function during tuberculosis disease progression

Virginie Rozot

Virginie.Rozot@uct.ac.za

University of Cape Town <https://orcid.org/0000-0001-9757-1501>

Miguel Rodo

University of Cape Town

Carly Young-Bailie

University of Cape Town

Munyaradzi Musvosvi

University of Cape Town

Constance Schreuder

University of Cape Town

Phu Van

Fred Hutch Cancer Research Center

Greg Finak

Fred Hutch Cancer Research Center <https://orcid.org/0000-0003-4341-9090>

Evan Greene

Fred Hutch Cancer Research Center

Raphael Gottardo

Lausanne University Hospital and University of Lausanne <https://orcid.org/0000-0002-3867-0232>

Holden Maecker

Stanford University School of Medicine <https://orcid.org/0000-0003-0795-9946>

Digby Warner

U Capetown <https://orcid.org/0000-0002-4146-0930>

Valerie Mizrahi

University of Cape Town <https://orcid.org/0000-0003-4824-9115>

Cecilia Lindestam-Arlehamn

Statens Serum Institut

Alessandro Sette

La Jolla Institute <https://orcid.org/0000-0001-7013-2250>

Willem Hanekom

Africa Health Research Institute <https://orcid.org/0000-0003-3070-6154>

Nicole Bilek

University of Cape Town

Michelle Fisher

University of Cape Town

Francesca Little

University of Cape Town

Mark Hatherill

South African Tuberculosis Vaccine Initiative, University of Cape Town <https://orcid.org/0000-0003-3491-1809>

Thomas Scriba

University of Cape Town <https://orcid.org/0000-0002-0641-1359>

Article

Keywords:

Posted Date: August 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10293773/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: **Yes** there is potential Competing Interest. TJS is a co-inventor of a patent of the RISK6 signature. The authors declare no competing interests. R.G. has received consulting income (payments made to the Lausanne University Hospital) from Takeda, Arcellx, Sanofi, Owkin, and declares ownership in Ozette Technologies. R.G. have received research funding from Owkin and 10x Genomics. G.F and E.G. are employees of Ozette technologies and own shares.

Dynamics of peripheral T cell phenotype and function during tuberculosis disease progression

Virginie Rozot^{1#}, Miguel J Rodo^{1,2#}, Carly Young¹, Munyaradzi Musvosvi¹, Constance Schreuder¹, Phu Van³, Greg Finak³, Ethan Greene³, Raphael Gottardo^{3,4}, Holden T. Maecker⁵, Digby Warner⁶, Valerie Mizrahi⁶, Cecilia S. Lindestam Arlehamn^{7,8,9}, Alessandro Sette⁷, Willem A. Hanekom^{1,10,11}, Nicole Bilek¹, Michelle Fisher¹, Francesca Little², Mark Hatherill¹, Thomas J Scriba^{1*}

#contributed equally

*thomas.scriba@uct.ac.za

¹South 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

²Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa

³Department of Biostatistics, Bioinformatics & Epidemiology, Fred Hutch Cancer Research Center, Seattle, WA, USA

⁴Biomedical Data Science Center, Lausanne University Hospital (CHUV), Lausanne, Switzerland_ISREC, EPFL, Lausanne, Switzerland_Swiss Institute of Bioinformatics, University of Lausanne, Switzerland

⁵Institute for Immunity, Transplantation & Infection, Stanford University, Palo Alto, CA, USA

⁶Molecular Mycobacteriology Research Unit, Institute of Infectious Disease and Molecular Medicine, Department of Pathology, University of Cape Town, Cape Town, South Africa

⁷Center for Vaccine Innovation, La Jolla Institute for Immunology, La Jolla, CA, USA

⁸Department of Infectious Disease Immunology, Center for Vaccine Research, Statens Serum Institut, Copenhagen, Denmark

⁹Novo Nordisk Foundation Initiative for Vaccines and Immunity, NIVI Research Center, Department of Immunology and Microbiology, University of Copenhagen, 2200 Copenhagen, Denmark

¹⁰Africa Health Research Institute, Durban, South Africa

¹¹Division of Infection & Immunity, University College London, London, UK

Abstract

Background: Antigen-specific CD4 T cells are essential but not sufficient for controlling *Mycobacterium tuberculosis* (*Mtb*) infection. We characterized changes in *Mtb*-specific CD4, and CD8 and $\gamma\delta$ T cell frequencies and phenotypes during tuberculosis progression.

Methods: *Mtb*-infected adolescents, who remained healthy or progressed to tuberculosis, were followed longitudinally. Bulk and *Mtb*-specific T cells were profiled by intracellular cytokine staining and mass cytometry. Systemic inflammation was profiled by blood transcriptomics and proteomics. Features associated with disease progression or systemic inflammation were identified by mixed-effects modelling.

Results: Among many phenotypes of bulk CD4, CD8 and $\gamma\delta$ T cells, only CD4 T cell activation (HLA-DR+) increased significantly during progression. However, levels of systemic inflammatory biomarkers of tuberculosis were strongly associated with changes in many phenotypic subsets of CD4, CD8 and $\gamma\delta$ T cells. Similarly, among hundreds of *Mtb*-specific functional and phenotypic CD4 and CD8 T cell subsets, only two early-differentiated CD4 subsets (IL-2⁺TNF⁺CD7⁺CD27⁺ and IFN γ ⁺IL-2⁺TNF⁺CD7⁺CD27⁺) significantly decreased, while *Mtb*-specific CD4 T cell activation increased during progression. These changes correlated with systemic inflammation.

Conclusion: Our findings suggest that peripheral blood frequencies and functions of *Mtb*-specific T cell subsets do not associate with tuberculosis progression, but that inflammation drives marked changes in T cell activation and differentiation.

Introduction

The immunological mechanisms that govern control of *Mycobacterium tuberculosis* (*Mtb*), the aetiological agent of tuberculosis (TB), the disease causing the most deaths globally due to a single pathogen, are incompletely understood¹. Despite decades of research, the global TB epidemic persists, highlighting an urgent need for improved vaccines and therapeutic strategies. A central barrier has been the absence of validated immune correlates of protection or correlates of risk in humans. Protective immunity is thought to depend, at least in part, on an antigen-specific T cell response^{2,3}. Important new insights into the role of protective T cell responses have emerged from recent studies in *Mtb*-infected non-human primates, which have profiled formation of granulomas, key sites of the classical host-pathogen interaction which can either restrict or facilitate bacterial persistence⁴⁻⁶. In mice and non-human primates (NHPs), depletion of CD4 T cells accelerates TB pathogenesis, while adoptive transfer confers partial protection. Furthermore, recent studies of intravenous BCG vaccination in NHP highlight a pivotal role for lung-resident Th1/Th17 cells in mycobacterial control^{7,8}, while also implicating cytotoxic lymphocyte functions by CD8 T cells and NK cells as contributors to bacterial containment⁸.

Characterization of protective immune responses against *Mtb* in humans is much more challenging because of limited accessibility to lung or lymphoid tissue and inability to measure *in vivo* bacterial burden. Clinical studies that aim to delineate which immune response attributes are associated with clinical outcome require large cohorts and longitudinal follow-up to measure immune responses before TB disease ensues. We and others have performed longitudinal clinical studies in participants with evidence of prior *Mtb* sensitization or close contact with TB, to identify immune responses that underlie development of disease and identify correlates of risk. Blood transcriptomic

profiling, plasma proteomic or cellular analyses have revealed that inflammatory processes, such as type I/II IFN responses, neutrophil and myeloid inflammation⁹⁻¹⁴, activation of the complement cascade¹⁰ and T cell activation¹⁵ are associated with progression to TB disease.

However, the role of antigen-specific T cells in disease progression remains poorly understood. Cross-sectional studies have reported that mycobacteria-specific T cells from individuals with TB are more activated, and more differentiated, expressing only IFN γ or TNF, or IFN γ ⁺TNF⁺, than those in healthy *Mtb*-infected persons which tend to also co-express IL-2¹⁵⁻¹⁸. Expression of activation markers, HLA-DR, CD38 and Ki67 on *Mtb*-specific T cells are significantly increased in TB patients as compared to *Mtb*-sensitized controls, likely reflecting increased antigen recognition¹⁹. This is further supported by the finding of a comparison of adolescents with recent or remote *Mtb* infection, which showed that recent infection is associated with elevated HLA-DR-expressing *Mtb*-specific T cells^{20,21}.

A longitudinal study in BCG-vaccinated infants showed that frequencies or cytokine co-expression patterns of BCG-reactive Th1 cytokine or IL-17-expressing CD4 or CD8 cells in peripheral blood, measured at 10 weeks of age, were not associated with risk of TB disease²². However, a second study of South African infants from the same population found a weak association between BCG-reactive IFN- γ secreting cells, measured by ELISpot assay, and reduced risk of developing TB disease¹⁵. The precise role of *Mtb*-specific T cells in clinical outcome in humans thus remains incompletely defined. Despite this, mycobacteria-specific T cells expressing the Th1 cytokines, IFN γ , TNF and IL-2, as well as IL-17 are commonly quantified as outcomes of vaccine immunogenicity in preclinical studies and clinical trials of TB vaccine candidates.

Here, we interrogated the role of blood T cells in a subcohort of the adolescent cohort study (ACS) in South Africa^{10,12,14,20,21,23,24}. This unique prospective study followed Mtb-sensitized adolescents longitudinally, collecting blood samples from those who progressed to TB (progressors) and matched controls who remained healthy (controllers). Using intracellular cytokine staining with high-dimensional mass cytometry, automated clustering, and integration with inflammatory markers of TB measured by transcriptomics and proteomics, we mapped the dynamics of bulk and Mtb-specific T cells over-time. We hypothesized that higher frequencies of Mtb-specific Th1 cytokine or IL-17-expressing CD4 or CD8 cells in peripheral blood are associated with non-progression in Mtb-infected individuals. We also hypothesized that T cell activation would steadily increase during TB progression and that unbiased analyses allow identification of controller-associated “protective” T cell responses. We developed and applied a comprehensive intracellular cytokine staining (ICS) assay based on cytometry by time-of-flight (CyTOF, or mass cytometry), which uses elemental metal isotope tags conjugated to monoclonal antibodies to quantify cellular protein expression and customized statistical approaches to comprehensively profile Mtb-specific T cell response magnitude, function and/or phenotype.

Results

Study participants and samples

Among 67 eligible adolescents who developed microbiologically-proven, pulmonary tuberculosis (TB) during study follow-up, 45 were termed progressors and 22 were excluded from analysis because they were living with HIV, pregnant or developed TB within six months of initial sensitisation to *Mycobacterium tuberculosis* (*Mtb*) (Fig. 1; Supplementary Fig. 1; Table 1)⁹. Each progressor was matched to two participants who did not develop TB during the study despite evidence of *Mtb* sensitisation at enrolment. They were matched based on age, sex, ethnicity, school and prior history of TB (Table 1). These 107 individuals are termed controllers.

A mass cytometry assay was performed on peripheral blood mononuclear cell (PBMC) samples from 37 progressors and 36 controllers, selected based on PBMC availability at each time point throughout the longitudinal study, yielding 148 longitudinal samples up to 890 days from TB diagnosis (Table 1; Fig. 1a, 1b). The gating strategy and representative results are in (Supplementary Fig. 2). Whole blood RNA-seq and plasma protein profiling by SomaScan assay (SomaLogic) were previously performed¹⁰ and were available for most of these CyTOF samples (Table 1).

Activated, non-specific CD4 T cell subsets increase during TB disease progression

To characterize phenotypic features of T cells, we built a robust analysis platform, validated against manual gating, using FAUST²⁵ to identify immune cell clusters (Fig. 1b). FAUST-derived CD4, CD8, $\gamma\delta$ T cells, B cells and NK cell frequencies correlated strongly with manual gating (concordance correlation ≥ 0.9) in a subset of manually-gated samples (Supplementary Fig. 3).

Among unstimulated PBMCs, FAUST identified 42 CD4 T cell subsets varying in expression of CD161, perforin, HLA-DR, CD16, CXCR5, CD27, CCR7, CD7 and CD45RA (Fig. 2a). As expected, the most abundant CD4 T cell subset comprised naïve cells co-expressing CCR7, CD27 and CD45RA.

Mixed-effects modelling detected two HLA-DR⁺ CD4 T cell phenotypic subsets that were significantly elevated in TB progressors, relative to controllers (progression effect $p < 0.001$, $q_{\text{fdr}} = 0.071$; Fig. 2a, 2b; Table 2). These increases were observed at proximal (roughly within 1 year of TB diagnosis, $p = 0.012$), rather than distal (more than 1 year of TB diagnosis, $p = 0.5$) timepoints (Fig. 2b; Table 2).

To understand which T cell subsets were associated with systemic inflammation during TB progression, we related their frequencies to a selected set of established markers of progression-associated inflammation: the RISK6¹⁴ six-gene blood transcriptomic signature, plasma MMP-1 (collagenase linked to lung matrix destruction^{10 26}), and complement factor 9 (C9, a complement cascade component and early biomarker of TB progression¹⁰). RISK6 scores, MMP-1 and C9 concentrations were significantly higher in progressors than controllers^{10,12}. Mixed-effects modelling detected four CD4 T cell subsets associated with RISK6, nine subsets associated with C9 and six with MMP-1 (Fig. 2a; Table 3). HLA-DR⁺CD7⁺ CD4 T cell frequencies increased with increasing MMP-1 ($p = 0.009$, $q_{\text{fdr}} = 0.17$; Fig. 2c), consistent with MMP-1 being a marker of lung-tissue destruction²⁶ and although a nominal progressor-specific association was observed ($p = 0.021$), this was not significant after correction for multiple testing ($q_{\text{fdr}} = 0.25$; Fig. 2c; Table 3).

It was notable that most of these subsets were not associated with TB progression. For example, naïve-like CD4 T cells (CD7⁺CD27⁺CD45RA⁺) lacking CCR7 were not associated with TB progression (Fig. 2a), but their frequencies increased with C9 ($p =$

0.001, $q_{\text{fdr}} = 0.057$; Fig. 2d), while CCR7⁺ naïve-like CD4 T cells showed no association with C9 ($p = 0.058$, $q_{\text{fdr}} = 0.31$; Fig. 2e) or RISK6 ($p = 0.37$; Fig. 2f). Cytotoxic CD4 T cells (CD7⁺CD45RA⁺Perforin⁺) increased with RISK6 ($p = 0.008$, $q_{\text{fdr}} = 0.17$; Fig. 2a, 2g). Overall, these data suggest that bulk CD4 T-cell phenotypes appear associated to the inflammatory milieu during TB progression, with activated cells being the only subsets that clearly increase with time to diagnosis.

Non-specific CD8 and $\gamma\delta$ T cells reflect systemic inflammation in TB progression

Among CD8 T cells FAUST identified 28 T cell phenotypic subsets in unstimulated PBMCs (Fig. 3a) based on variability in expression of the same markers as the CD4 phenotypes, except for CXCR5, which was not expressed by CD8 T cells. The naïve-like subset (CD45RA⁺CD27⁺CCR7⁺) was again the most abundant CD8 T cell subset (Fig. 3a).

None of these CD8 phenotypic subsets were significantly associated with TB progression (Fig. 3a), but as for CD4 T cells, many were associated with inflammatory markers: six CD8 phenotypic subsets were associated with RISK6, seven with C9 and four with MMP-1.

In contrast to CD4 T cells, many of these were inversely associated with inflammation. For example, CD7⁺ and CD7⁺CD27⁺CCR7⁺ (naïve) CD8 T cell subsets declined with increasing C9 ($q_{\text{fdr}} = 0.04$ and $q_{\text{fdr}} = 0.01$, Fig. 3b and Fig. 3e, respectively) MMP1 (Fig. 3c, f) or RISK6 (Fig. 3d) in progressors. Conversely, cytotoxic CD8 T cells (e.g. CD7⁺CD45RA⁺Perforin⁺), which are typically highly differentiated, increased with RISK6 ($p = 0.006$, $q_{\text{fdr}} = 0.15$; Fig. 3a, g) in progressors ($p = 0.021$, $q_{\text{fdr}} = 0.25$; Fig. 3g).

Among $\gamma\delta$ T cells FAUST also identified 24 cell subsets expressing CD4 and/or CD8, with relative frequencies consistent with previous reports²⁷ (Fig. 4a). Perforin-expressing CD7⁺CD16⁺ CD45RA⁺ $\gamma\delta$ T cells, comprising more than five percent of $\gamma\delta$ T cells, were associated with progression ($p = 0.001$, $q_{\text{fdr}} = 0.071$; Fig. 4a, b), increasing in frequency at timepoints most proximal to diagnosis ($p < 0.001$; Fig. 4b). Six phenotypes were linked to inflammatory markers but not directly to progression. A representative example is the CD7⁺CD27⁺CCR7⁺CD45RA⁺ $\gamma\delta$ T cell subset; frequencies of this subset associated with C9 and RISK6 ($p = 0.23$, $q_{\text{fdr}} = 0.56$; Fig. 4c and $p = 0.018$, $q_{\text{fdr}} = 0.19$; Fig. 4d).

Overall, these findings suggest that bulk CD8 and $\gamma\delta$ T-cell phenotypes largely reflect systemic inflammation and inflammation signatures during TB progression, with clearer progression-dependent effects than CD4 T-cells.

Mtb-specific Th1/Th17 CD4 T cell responses do not correlate with progression to TB disease

Most cells among the bulk CD4, CD8 and $\gamma\delta$ T cell analyses in unstimulated PBMC are not *Mtb*-specific; therefore stronger associations between phenotypic subsets and TB-associated inflammation can be expected. To explore associations between progression and *Mtb* antigen-specific T cell responses, we characterised *Mtb*-specific CD4, CD8 and $\gamma\delta$ T cell T cells that respond to antigen stimulation (Fig. 1). PBMC samples were stimulated with live *Mtb*, or one of two peptide pools comprising secreted or non-secreted *Mtb* antigens (Supplementary Excel Table). An EBV/CMV peptide pool served as a non-*Mtb* control. Identification of T cell expression of IFN γ , TNF, IL-2, IL-17, IL-22 and/or IL-6 was determined with a novel automated cytokine-gating algorithm, *StimGate* (Supplementary Fig. 4; see “Methods” for details). Strong

correlations between cytokine-expressing cell subset frequencies obtained from *StimGate* and manual gating were observed for CD4 T cells expressing IFN γ , IL-2, TNF, IL-22 and IL-6, for CD8 T cells expressing IFN γ and TNF, $\gamma\delta$ T cells expressing IFN γ and NK cells expressing IFN γ and in some cases TNF and IL-2 (Supplementary Fig. 4; see “Methods” for details).

Antigen-specific CD4 T cells co-expressing combinations of IFN γ , IL-2 and TNF predominated the response to live *Mtb* and secreted *Mtb* proteins, whereas IFN γ -only cells characterised the response to EBV/CMV (Fig. 5a, Supplementary Fig. 5a and 6). A negligible response was induced by non-secreted *Mtb* proteins in most participants (Fig. 5a), and IL-6, IL-17 and IL-22 responses were generally not detected (Supplementary Fig. 5a and 6).

To address our primary hypothesis, we assessed whether *Mtb*-specific CD4 T cells expressing Th1/Th17 cytokines, defined as any co-expression of IFN γ , IL-2, TNF or IL-17, were associated with progression. No significant progression effect was detected for frequencies of Th1/Th17 cytokine expressing CD4 T cell responses to secreted *Mtb* proteins ($p = 0.31$; Fig. 5e), live *Mtb* ($p = 0.47$; Fig. 5f), or EBV/CMV ($p = 0.074$; Fig. 5g).

However, *Mtb*-specific CD4 T cell activation, quantified as the difference in median HLA-DR expression between cytokine-positive and cytokine-negative cells²⁸, was strongly associated with progression, with activated *Mtb*-specific CD4 T cells elevated in progressors ($q_{\text{fdr}} < 0.001$; Fig. 5h); we could not discern a major association between distal or proximal effects for this outcome. EBV/CMV-specific CD4 T cells while scarce (EBV-CMV T-cell responses were predominantly CD8-mediated), showed no significant progression-associated activation ($p = 0.19$; Fig. 5i), consistent with antigen-driven activation unique to *Mtb* exposure during progression. Interestingly,

Mtb-specific CD4 T cell activation showed little evidence of association with RISK6 ($p = 0.11$, $q_{fwr} = 0.64$; Fig. 5j).

To further characterise sub-phenotypes of *Mtb*-specific CD4 T cells, we combined StimGate cytokine gating with FAUST phenotypes and filtering for stimulation-specific responses, which yielded five antigen-specific CD4 subsets (Fig. 1).

Among these, an early-differentiated CD7⁺CD27⁺ CD4 subset co-expressing IL-2 and TNF in response to secreted *Mtb* proteins was significantly associated with TB progression ($p = 0.001$, $q_{fwer} = 0.01$; Fig. 5k), decreasing with time to TB disease (proximal $p = 0.001$; Fig. 5k). A similar pattern was observed for CD7⁺CD27⁺ subsets that co-expressed IFN γ , IL-2 and TNF in response to live *Mtb*; however, this difference was not significant after family-wise correction ($p = 0.047$; $q_{fwer} = 0.42$; Fig. 5i). By contrast, most subsets remained unchanged during progression, as exemplified by the more-differentiated CD7⁺ subset expressing IFN γ only in response to live *Mtb* ($p = 0.7$; Fig. 5m).

We also characterised sub-phenotypes of *Mtb*-specific CD4 T cells using an orthogonal approach based on FlowSOM clustering of cytokine-expressing CD4 T cells, which recapitulated these results, underscoring their robustness (Supplementary Fig. 7). Thus, activation of *Mtb*-specific CD4 T cells, rather than overall cytokine response magnitude, tracks progression to TB disease, and a decline in relative frequencies of circulating early-differentiated IL-2⁺TNF⁺ CD4 T cells may mark advancing disease.

Peripheral *Mtb*-specific CD8 and TCR $\gamma\delta$ T cells are somewhat modulated during progression and systemic inflammation.

We next examined *Mtb*-specific CD8 and $\gamma\delta$ T cells using the same analytical approach applied to CD4 T cells. Frequencies of cytokine-expressing CD8 responses to *Mtb* were low, whether detected after stimulation with live *Mtb* or peptides from secreted *Mtb* proteins (Fig. 6a, b), whereas stimulation with the EBV/CMV peptide pool elicited robust CD8 T cell responses dominated by single-IFN- γ -producing cells, confirming that the assay reliably captured CD8 T cell functionality (Fig. 6c; Supplementary Fig. 5b). By contrast, $\gamma\delta$ T cells responded strongly only to live *Mtb*, predominantly expressing IFN γ and/or TNF (Fig. 6d; Supplementary Fig. 5c).

No significant effect of TB progression was detected for frequencies of cytokine+ *Mtb*- or EBV-specific CD8 T cells ($p = 0.75$ and $p = 0.51$; Fig. 6e, f respectively). However, one EBV/CMV specific- IFN γ +CD7+CD27+ CD8 T-cell subset (negative for the other 7 phenotyping markers, including CCR7 and CD45RA) was lower in progressors than controllers ($p = 0.007$, $q_{\text{fwer}} = 0.052$; Fig. 6g).

No effect of TB progression was observed for frequencies of cytokine+ $\gamma\delta$ T cells responding to live *Mtb* (Fig. 6h), either at the lineage or phenotypic level. By contrast, Th1+ $\gamma\delta$ T cells were significantly associated with levels of MMP-1 ($q_{\text{fwer}} = 0.005$; Fig. 6i) and decreased in progressors ($q_{\text{fwer}} = 0.002$; Fig. 6i). An IFN γ + CD7+CD161+ $\gamma\delta$ phenotype showed a similar trend but was not significant when family-wise correction was applied ($p = 0.006$; $q_{\text{fwer}} = 0.15$; Fig. 6j). Bulk $\gamma\delta$ T cells expressing CD8+CD7+CD27+CCR7+CD45RA+ had a similar nominal association with RISK6 scores, which was not significant after application of family-wise correction ($p = 0.018$, $q_{\text{fdr}} = 0.19$; Fig. 4b).

In summary, neither total *Mtb*-specific CD8 nor $\gamma\delta$ responses demonstrated a strong association with time to TB diagnosis. However, EBV/CMV-specific CD8 responses were the only robust CD8 responses observed, and an IFN γ ⁺ subset was consistently lower in progressors. In addition, *Mtb*-specific $\gamma\delta$ T cells appeared to respond to inflammation in a progression-dependent manner.

Discussion

Progression to TB reflects a dynamic interplay between pathogen burden, tissue pathology, systemic inflammation, and host immune response, including T cell populations. In this longitudinal cohort of Mtb-sensitized adolescents who either remained TB-free or progressed to TB, we comprehensively characterised peripheral CD4, CD8 and $\gamma\delta$ T cell responses using a 45-marker mass cytometry ICS assay. We applied FAUST-based phenotyping and developed StimGate, a new cytokine-gating algorithm, to define bulk and antigen-specific phenotypic subsets. Using longitudinal mixed-effects modelling we identified response measurements associated with disease progression or with transcriptomic and proteomic markers of systemic inflammation.

The primary finding was that the magnitude of canonical Mtb-specific CD4 T cell subsets expressing Th1/Th17 cytokines is not associated with progression to TB. Among many phenotypic CD4 T cell subsets interrogated only two categories were associated with progression: First, activated, HLA-DR-positive Mtb-specific CD4 T cells increased during progression, while second, early differentiated, IL-2+TNF+ CD4 T cells subsets (IL-2+TNF+CD7+CD27+ and IFN γ +IL-2+TNF+CD7+CD27+) decreased during progression.

These results provide several important conceptual insights. Firstly, the lack of association between circulating Th1/Th17 cell frequencies and disease outcome is consistent with results of another study performed in the ACS cohort which showed that frequencies of Mtb-specific activation-induced marker-positive (AIM-positive, CD69+ and CD154+ or CD137+) T cells were not different between TB progressors and controllers. Similarly, frequencies of BCG-reactive Th1 cytokine or IL-17-expressing CD4 or CD8 cells in peripheral blood, measured at 10 weeks of age in

BCG-vaccinated infants, were not associated with risk of TB disease²². Together, these data suggest that the magnitude of T cell responses in peripheral blood do not correlate with protective immunity against TB. It is uncertain whether this result is representative of Th1/Th17 responses in tissue locations relevant to host-pathogen interactions in TB, namely lung and draining lymph nodes²⁹. In multiple non-human primate experiments, immune response features most strongly associated with bacterial control largely localized to tissue compartments, including the bronchoalveolar space (sampled by bronchoalveolar lavage), lung parenchyma, airway mucosa, and lung-draining lymphoid tissues rather than circulation^{7,30-33}. Tissue-resident T cells or granuloma-resident Mtb-specific T cells in lungs exhibit transcriptional programs, TCR-stimulation signatures, spatial positioning, and functional profiles that are not faithfully reflected in circulating blood³⁴⁻³⁶, illustrating that functional attributes most relevant to bacterial control are likely restricted to infected or diseased tissue.

Despite this, we recently identified significant associations between frequencies of Mtb epitope-specific CD4 T cell clonotypes in peripheral blood and risk of TB progression³⁷. Importantly, depending on the Mtb epitope or antigen these T cell clones recognised, some clonotypes were significantly enriched in progressors, while others were significantly enriched in controllers. In line with this, Moguche et al. demonstrated that differential abundance of Ag85B and ESAT-6 produced during early or chronic Mtb infection drives distinct functional and protective T cell programs³⁸. Furthermore, Stetsenko *et. al.* found that not all Mtb-specific T cell clones identified in Mtb-infected individuals can recognize Mtb-infected macrophages *in vitro*, and those that recognized Mtb-infected cells expressed a different set of effector genes to those expressed by T cells specific for viral antigens³⁹. These studies suggest that the

antigen targeted by Mtb-specific T cell responses may be a critical determinant of protective vs pathogenic vs impotent T cell responses, a detail that cannot be resolved when T cell responses are characterised with a complex antigen stimulation such as Mtb lysate, whole organism or peptide pools. We therefore cannot rule out that a more antigen-resolved analysis may have revealed progression or controller-associated T cell features in our study.

Interestingly, in the study by Stetsenko *et. al.* all T cell clones that recognised Mtb-infected macrophages and for which the cognate antigen was identified were specific for secreted Mtb antigens that are substrates for the type VII secretion system (T7SS)³⁹. A recent immunopeptidomics study also found that T7SS substrate proteins were highly enriched among antigens presented via MHC class II on Mtb-infected human antigen presenting cells⁴⁰. An older, genome-wide T cell epitope screening study in Mtb-infected individuals reported that T cell responses occurred in three antigenic islands that primarily comprise proteins secreted via the T7SS or proteins that form the T7SS apparatus⁴¹. Our finding that Mtb-specific T cell responses were readily observed after stimulation with a peptide pool comprising secreted antigens, but only very low or no responses detected after stimulation with a smaller non-secreted peptide pool, supports that secreted proteins represent the most immunogenic Mtb antigens.

Another major finding was that activated T cells, defined as HLA-DR+, rather than shifts in cytokine co-expression patterns, were the most reliable distinguishing immunological feature between progressors and controllers. While previous longitudinal studies of immune responses during TB progression have described marked changes in inflammatory and disease-associated immune

outcomes^{10,11,13,15,16,42}, which were largely restricted to innate markers or inflammation mediators, we show that activated HLA-DR⁺ *Mtb*-specific CD4 T cells increased progressively as progressors approached diagnosis, whereas early-differentiated IL-2⁺TNF⁺ CD4 T cells declined. Previous studies have shown that HLA-DR-expression on *Mtb*-specific CD4 T cells is a marker of short-lived effector T cells, while longer-lived memory CD4 T cells are typically HLA-DR-negative^{43,44}, and suggest that the highly activated phenotype of *Mtb*-specific CD4 T cells is indicative of current or very recent antigen exposure. The finding that HLA-DR expression on CMV/EBV-specific T cells was not associated with progression strengthens our conclusion that these effects are *Mtb*-specific, rather than a secondary response to non-specific TB inflammation. Similar HLA-DR⁺ expression on *Mtb*-specific CD4 T cells activation phenotypes have been reported in individuals with recent *Mtb* infection defined by IGRA conversion, reinforcing the interpretation that HLA-DR upregulation marks ongoing or recent antigen encounter in vivo^{20,21}.

These results are consistent with multiple reports of significantly increased T cell activation^{19,45} and decreased *Mtb*-specific CD4 T cells with early differentiation phenotypes and functional profiles in TB patients, compared to healthy, *Mtb*-sensitized individuals^{16,17,46}. Interestingly, protein-subunit vaccine-induced IL-2⁺TNF⁺ CD4 T cells were associated with protection against *Mtb* challenge in murine models^{38,47}, supporting the hypothesis that maintenance of such early differentiated CD4 T cell phenotypes may mediate immune control even in *Mtb*-infected humans. The parallel loss of early-differentiated CD4 T cells, previously associated with vaccine-induced protection in animal models, suggests that persistent antigen exposure drives their differentiation toward more terminally differentiated, less proliferative states, or that these cells are preferentially sequestered from the

periphery. Notably, these differentiation changes were also not observed in EBV/CMV-specific T cells, again suggesting that they are Mtb-specific and antigen-dependent and not a generalized feature associated with TB inflammation.

A third major finding is that systemic inflammation, ascertained through levels of RISK6, C9, and MMP-1, blood biomarkers that represent different aspects of TB-associated immuno-physiology^{10,14,48}, is strongly associated with changes in T-cell phenotypes across bulk CD4, CD8, and $\gamma\delta$ lineages. For instance, HLA-DR expression on bulk CD4 T cells significantly increased during TB progression, a result that is consistent with our previous report of increased activation of bulk CD4 T cells in infants associated with subsequent TB disease risk¹⁵. In our current study strong associations were also detected between T cell differentiation states and inflammatory markers. Changes in 16 of 42 CD4 phenotypic subsets such as CD7+HLA-DR+ and naïve-like CD7+CD27+CD45RA+ CD4 T cells, 12 of 28 CD8 phenotypic subsets including CD7+CD45RA+Perforin+ CD8 T cells and 6 of 25 $\gamma\delta$ phenotypic subsets such as CD8+CD7+CD27+CCR7+CD45RA+ $\gamma\delta$ T cells were associated with levels of at least one of these three inflammatory biomarkers, but only in progressors. By contrast, few were significantly associated with time to TB diagnosis, suggesting that inflammatory processes associated with advancing TB disease are associated with marked shifts in T-cell composition and phenotype. This finding fits with described responsiveness of T cell populations to stimulation with inflammatory cytokines⁴⁹. The finding that these inflammatory effects, rather than antigen specificity, explains much of the variation in bulk T-cell phenotypes highlights a critical challenge for blood-based correlates studies: distinguishing antigen-driven changes from inflammation-driven changes is inherently difficult. This reinforces why cross-sectional comparisons between TB patients and Mtb-sensitized controls may yield misleading results, since systemic

inflammation itself can reshape the T-cell compartment independently of protective or pathogenic mechanisms.

Our study has several limitations. First, it relies exclusively on peripheral blood, whereas Mtb infection and disease are typically localized to lung and lymph node tissues. Protective or pathogenic responses occurring within these tissues are therefore not captured. Second, the timing of Mtb infection or reinfection and onset of disease progression, was unknown for most adolescents, limiting our ability to study the earliest immunological events that precede sensitization or systemic inflammation. Third, although this is the largest progressor cohort studied with high-dimensional cytometry to date, the sample size remains modest for detecting rare antigen-specific subsets. Fourth, analyses of low-frequency cell populations are inherently sensitive to technical variability, although we mitigated this through rigorous algorithmic gating (FAUST, StimGate). Finally, many progressor samples were collected more than a year prior to TB diagnosis and several studies suggest that immunological effects of TB progression only become robustly detectable in peripheral blood within 12-9 months of TB diagnosis^{13,50-52}; immune alterations during very early progression may therefore be underestimated. Finally, we assessed T cell responses using complex antigen preparations (live Mtb or large peptide pools), which capture aggregated Mtb-specific responses across many antigens and may have missed progression-associated T cell responses that recognise individual antigens.

A key strength of this study lies in the use of complementary, automated analytical approaches that enabled an unbiased and sensitive characterization of the cellular immune landscape. By combining mixed-effects modelling with rigorous robustness checks, we were able to assess longitudinal changes associated with time to TB

diagnosis and TB-related inflammation while accounting for repeated sampling and potential confounding. The development of a novel cytokine-gating algorithm, together with FAUST-based phenotyping enabled detection of rare Mtb-specific CD4 T cell subsets with declining frequencies as individuals approached disease. Finally, integration of these cellular readouts with blood transcriptomic and plasma proteomic data provided a systems-level view of antigen-specific immunity within the context of inflammatory sequelae during TB progression. Our insights, together with emerging tissue-focused studies, support a shift toward integrated, tissue-based, multimodal immunology to uncover the true correlates of protection and to guide next-generation TB vaccine and host-directed therapy development.

Methods

Study population

The Adolescent Cohort Study, including selection of progressors and controllers (or non-progressors), was previously described⁹ (Supplementary Fig. 1). Briefly, 6,363 adolescents attending high schools in the Worcester region of the Western Cape, South Africa, were enrolled and followed for twenty-four months. Among those with evidence of *Mycobacterium tuberculosis* (*Mtb*) infection, either by a positive tuberculin skin test (TST) or QuantiFERON TB-Gold In-Tube assay (Qiagen) (QFT), progressors developed microbiologically confirmed (positive by sputum smear microscopy and/or MGIT liquid culture), intrathoracic disease during two years of follow-up^{9,10}. Controllers also had evidence of *Mtb* infection, but did not develop tuberculosis (TB) disease during follow-up, and were matched to progressors for age, sex, ethnicity, school of attendance and prior history of TB disease (Table 1). Participants were excluded from the progressor group if they developed TB within six months of enrolment (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. 44 and 106 participants met the criteria for progressor and controller, respectively (Table 1). Longitudinally collected PBMC samples were available from most participants at six-monthly intervals (Fig. 1a). For each assay (CyTOF, SOMAScan, and RNA-seq [RISK6]), all available progressor samples were included. For SOMAScan and RNA-seq analyses, two matched controller samples were selected per progressor, whereas for CyTOF, one matched controller sample per participant was used due to cost, processing time, and limited remaining PBMC availability (Table 1). The final numbers of participants and samples per assay reflect both the initial selection and subsequent

exclusion of any samples that were not processed because of assay or quality-control failure (Table 1).

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.

PBMC ICS assay

Frozen PBMCs samples were thawed, rested at 37 °C, and either left unstimulated or stimulated in the presence of anti-CD28 (BD, Cat #340975) and anti-CD49 (BD, Cat #340976) with peptide pools (comprising 87 peptides at 1 µg mL⁻¹ each, that represented secreted, immunodominant mycobacterial proteins, or 24 peptides at 1 µg mL⁻¹ each that represented non-secreted proteins (Supplementary Table 1), or immunodominant EBV and CMV peptides at 1 µg mL⁻¹ each) or live *Mtb* (*M. tuberculosis* triple-auxotroph strain derived from H37Rv, through deletion of leucine, pantothenate and arginine synthesis pathways (kindly donated by Bill Jacobs, Albert Einstein University and propagated by Digby Warner, UCT)) for 5 hours at 37 °C. Brefeldin A (Golgi Plug, BD Cat# 555029, at 10 µg mL⁻¹) was added and cells were stimulated for an 7 additional hours at 37 °C.

Experimental procedure for mass cytometry

Sample selection

37 progressors and 37 matched controllers were selected for processing by mass cytometry, which together had 154 samples across all timepoints.

Panel and experimental design.

We designed a mass cytometry panel to detect all major known leukocyte subsets present in normal human peripheral blood (Supplementary Table 2) with a particular focus on T cell subsets. The panel was optimized with respect to i) the choice and combination of markers in the panel, ii) the pairing of metal isotope with a particular antibody, considering the target antigen abundance, iii) the availability and performance of commercial conjugates and in-house conjugated antibodies and related costs, and iv) minimizing consequences of spill-over.

To reduce technical variation, we barcoded and then pooled 5 stimulation tubes for each of 2 participant samples (total of 10 sub-samples) using a combination of 5 distinct metal-conjugated anti-CD45 antibodies (In113, In115, Pd108, Pt198, Bi209) such that each sub-sample is stained with two conjugates (Supplementary Table 3). We aimed to achieve the best signal definition without saturation of highly sensitive channels, which we kept for rare markers. Iridium191/193 were used for DNA staining and the Rh103 DNA intercalator for live-dead staining. The antibody panel was designed and optimized through sequential steps and once finalized, samples from a cohort of 25 healthy, *Mtb*-sensitized donors and 25 TB patients were stained to confirm successful panel performance (data not shown).

Antibody labelling

In-house labelling of antibodies was performed using the MAXPAR labelling kit according to the manufacturer's instructions (Fluidigm), using X8 polymers. Concentrations of unlabelled antibodies prior to conjugation and of conjugates were determined at an optical density of 280 nm and in a 3 μ L cuvette using the Nanodrop

photometer. 100-200 µg antibodies were used per conjugation reaction. After quantification of antibody conjugates, stocks were diluted in PBS-based Ab stabilizer (Candor Bioscience, Wangen, Germany) to adjust concentrations to 0.25 or 0.5 mg mL⁻¹.

Staining procedure

Stimulated PBMC were washed with 1 mL of PBS (Rockland; made in MilliQ water) with 2% EDTA per tube and incubated at room temperature for 10 minutes. After a 1 mL wash in PBS, cells washed again twice with Cell Staining Buffer (CSM, Fluidigm), before performing CD45 cell surface barcoding (Supplementary Table 3) staining for 30 minutes at room temperature (see Supplementary Table 2a). After washing in PBS, the 10 barcoded sub-samples were pooled, and washed again in PBS. Live-dead staining was then performed using Rh103 intercalator following the manufacturers' instructions (Fluidigm) for 30 minutes at room temperature. Cells were washed twice with CSM were performed. The extracellular antibody cocktail (Supplementary Table 2b) was filtered through a 0.1 µm filter column and cells stained for 45 minutes at 4 °C, washed twice with CSM and fixed with CytoPERM/CytoFIX (BD) for 20 minutes at room temperature. Centrifugation of fixed cells was at 2,000 rpm for 10 minutes at 4 °C for all steps. Cells were washed twice with Perm/Wash Buffer (BD, 1/10 dilution in de-ionized water). The intracellular antibody cocktail (Supplementary Table 2c) was prepared and filtered before cells were stained for 45 minutes at 4 °C. Cells were washed twice with Perm/Wash buffer. Cells were incubated overnight in 1 mL of iridium intercalator (Fluidigm) diluted 1/100 in Fix/Perm buffer (Fluidigm) at 4 °C.

Sample preparation for acquisition

Following over-night incubation with iridium, cells were washed 3 times with de-ionized water and were kept pelleted at 4 °C until acquisition. At the time of acquisition cells were resuspended in a volume (calculated based on the cell number) of 1/10 dilution EQ(TM) Four Element beads solution in de-ionized water.

Data acquisition

Data acquisition was performed using Fluidigm CyTOF® Software v6.7. A template was set up with all parameters included in the experiment including channels containing potential environmental contaminants (i.e. Iodine, Barium) to monitor parameters which could impact the instrument performance. Samples were loaded and Flow Cytometry Standard (FCS) files were collected. Acquisition was monitored by assessing the rainplot to ensure a constant sample delivery and identify any potential cell clumps.

Quality control and pre-processing

Mass cytometry FCS files were exported using default settings (arcsinh transformation) and analysed using FlowJo (version 10) (BD Biosciences) or R version 4.2.3⁵³.

Three batches corresponding to six participant-timepoint combinations with five stimulations each were removed because CD3 expression was unimodal. This reduced the number of samples from 154 to 148 and excluded all samples from one controller (leaving 36 controllers).

During a mass cytometry run, signal intensity typically decreases over time. This time effect was adjusted for using a method based on reference beads⁵⁴. Bead data was removed from cell data using an automated approach implemented by the *CATALYST* R package⁵⁵. Deconvolution of the barcoded sub-samples in each batch was

performed automatically⁵⁶, whilst live cells and singlets were manually identified using rhodium and iridium expression levels, respectively. Despite initial normalisation, several samples still exhibited within-run time effects; segments with such time effects were removed using Flowcut⁵⁷. Cell doublets and residual beads persisted in several samples, and were manually removed based on their distinctive isotope label levels.

Gating phenotypes

The FAUST algorithm²⁵ clusters cells using marker modality, and was used as it is robust to batch effects, correctly identifies the number of clusters under non-normality and labels clusters semantically. FAUST identified 301 phenotypic clusters based on 16 multimodal markers. To validate FAUST's accuracy, frequencies of 8 cell lineages were compared between FAUST (population definitions in Supplementary Table 4) and expert manual gating (gating strategy in Supplementary Fig. 2). The comparison was performed on 60 samples from 12 participants spanning progression status and visit type, across five stimulation conditions (Fig. 1a). Pearson correlations between FAUST and manual frequencies exceeded 0.9 for all lineages except NKT-like cells ($r = 0.07$), which were excluded (Supplementary Fig. 3). As V α 7.2 was unimodal, FAUST could not identify MAIT cells. This paper focuses on T cells; FAUST identified 42, 28 and 24 CD4, CD8 and $\gamma\delta$ T cell clusters, respectively.

Gating cytokine-positive cells

Description of algorithm

Identifying cytokine-positive cells is challenging because of their rarity. They typically do not form distinct modes or separate well from cytokine-negative cells in marker distributions. To overcome this limitation, we developed a novel algorithm, StimGate, sensitive and robust to batch effects, to identify low-frequency cytokine-expressing cells.

For each stimulated sample and its matched unstimulated sample, the density of the cytokine expression was calculated. The per-cell probabilities were summed to obtain a background-subtracted count of cytokine-expressing cells. A classification threshold was then derived to produce this background-corrected total. When multiple cytokines were measured for a given cell lineage (e.g. CD4 T cells), the classification threshold for each cytokine was adaptively adjusted. Specifically, for cells already positive for any other cytokine, the threshold for an additional cytokine was lowered to the nearest anti-mode of its expression density, thereby making the criterion more permissive.

Within-study validation

To validate StimGate's accuracy, background-subtracted frequencies between StimGate and expert manual gating across 6 cytokines and 5 cell lineages were compared using the same 60 samples as FAUST's validation. Pearson correlations typically exceeded 0.9 (Supplementary Fig. 4b). Weaker correlations were observed for cytokine-lineage combinations with especially low background-subtracted frequencies likely indicative of no genuine response to stimulation (Supplementary Figs. 4a, 6), such as CD4 T cells expressing IL-22 in response to live *Mtb* stimulation.

Statistics

Time to TB model

Model formulation

Ignoring explicit participant and sample indices for clarity, responses Y for a given covariate pattern X were modelled as

$$Y \sim D(\mu, \text{Var}[Y|X]), \quad (1)$$

for some distribution D with mean $\mu = E[Y|X]$. The mean was linked to covariates through a monotonic link function g :

$$g(\mu) = \beta_1 + X_2\beta_2 + X_3\beta_3 + X_4\beta_4 + f_{[d,b_1,b_2,\beta_{5,1},\beta_{5,2},\dots,\beta_{5,d}]}(X_5) + v. \quad (2)$$

Here:

- β_1 is the intercept;
- β_2 is the effect of male sex;
- β_3 represents the effect of each additional (scaled) day since study entry;
- β_4 is effect of the indicator variable for progressor status, modelling the *distal* effect of time to TB;
- $f_{[d,b_1,b_2]}(X_5; \beta_{5,1}, \dots, \beta_{5,d})$ is a natural cubic spline with d degrees of freedom and boundary knots b_1, b_2 , modelling the *proximal* effect of time to TB X_5 ; and
- v is a participant-specific random effect.

For a progressor sample taken collected within 630 days from TB diagnosis, we define

$$X_5 = 630 - x, \quad (3)$$

where x is the number of days to TB diagnosis. Thus $X_5 = 0$ when $x = 630$ (far from diagnosis), $X_5 = 630$ when $x = 0$ (at diagnosis), and $X_5 = 380$ when $x = 250$ (intermediate timepoint). In words, X_5 measures the elapsed time since the fixed reference point 630 days before diagnosis; larger values indicate closer proximity to disease.

For such progressor samples, the spline term $f_{[\cdot]}(X_5; \cdot)$ contributes to $g(\mu)$ and therefore captures changes in the immune response as diagnosis approaches. For progressor samples collected more than 630 days before diagnosis, and for all controller samples, we set $X_5 = 0$. In that reference state the indicator term β_4 (0 for controllers, 1 for progressors) captures any baseline difference between controllers and progressors at or more than 630 days before TB diagnosis.

We selected a 630-day threshold to balance biological plausibility with sample availability. This cut-off captures a meaningful number of progressor samples more than 630 days before TB diagnosis, and assumes that immune responses remain largely stable before the 630-day period preceding TB diagnosis.

The proximal effect was modelled nonlinearly because the relationship between time to diagnosis and immune response was not assumed a priori. Throughout, natural cubic splines were used, with internal knots at 480 and 330 days from TB diagnosis and boundary knots at 600 and 180 days.

A mixed-effects framework accounted for repeated, irregularly spaced samples using the *R* package *glmmTMB*⁵⁸. Sex and days since study entry were included as covariates to adjust for confounding, as immune responses often differ by sex and may be modulated by increasing exposure to *Mtb*.

Confidence intervals were obtained with *emmeans*⁵⁹ as Wald intervals on the linear predictor, marginalising over random effects and back-transforming to the response scale.

Testing

Equation (2) decomposes TB progression into a *distal* component, β_4 , and a *proximal* component, $f_{[.]}(X_5; \cdot)$, represented by the spline coefficients $\beta_{5,1}, \dots, \beta_{5,d}$ (Table 2). A progression effect was defined as at least one of these components being non-zero. The primary hypothesis tested for a progression effect (Table 2); to aid interpretation the distal and proximal components were also tested separately (Table 2).

All tests were performed using two-sided likelihood ratio tests with degrees of freedom equal to the number of coefficients tested. Multiple-testing error rates were controlled using either the Bonferroni method⁶⁰ (family-wise error rate) or the Benjamini-

Hochberg method⁶¹ (false discovery rate). The choice depended on the number of hypotheses jointly tested, with Bonferroni used for ten or fewer and Benjamini-Hochberg otherwise.

Inflammation markers

Model formulation

Ignoring explicit participant and sample indices for clarity, responses Y for a given covariate pattern X were modelled as

$$Y \sim D(\mu, \text{Var}[Y | X]), \quad (4)$$

for some distribution D with mean $\mu = E[Y | X]$. The mean was linked to covariates through a monotonic link function g :

$$\begin{aligned} g(\mu) = & \beta_1 + X_2\beta_2 + X_3\beta_3 + X_4\beta_4 + f_{[d,b_1,b_2]}(X_5; \beta_{5,1}, \dots, \beta_{5,d}) \\ & + X_4 \times f_{[d,b_1,b_2]}(X_5; \beta_{6,1}, \dots, \beta_{6,d}) + v. \end{aligned} \quad (4)$$

Here:

- β_1 is the intercept;
- β_2 is the effect of male sex;
- β_3 is the effect of each additional day since study entry;
- β_4 is effect of the indicator variable for progressor status, modelling a constant difference between controllers and progressors regardless of the inflammatory marker level;
- $f_{[d,b_1,b_2]}(X_5; \beta_{5,\cdot})$ is a natural cubic spline (with d degrees of freedom and boundary knots b_1, b_2) modelling the main association between the inflammation marker X_5 and the response;

- $X_4 \times f_{[d,b_1,b_2]}(X_5; \beta_{6,\cdot})$ is the interaction term, modelling how the inflammation-marker response differs in progressors versus controllers (the $\beta_{6,\cdot}$ are the spline coefficients for this difference); and
- v is a participant-specific normally distributed random effect.

Mixed effects models with sex and days since entry confounders were fitted using the *R* package *glmmTMB*⁵⁸; the spline terms used $d = 2$ degrees of freedom with knot locations set automatically based on percentiles of X_5 .

Confidence intervals were obtained with the *R* package *emmeans*⁵⁹ as Wald intervals on the linear predictor, marginalising over random effects and back-transforming to the response scale.

Testing

The primary hypothesis tested whether the inflammation marker was associated with the response (the “Inflammation marker” effect; Table 3). If an association was detected, a secondary question was whether that association depended on progressor status (Table 3).

Testing followed the same approach as for the time to TB model (two-sided likelihood ratio tests, with degrees of freedom equal to the number of coefficients tested). Multiple-testing correction was applied as described above (Bonferroni or Benjamini–Hochberg as appropriate).

Response distributions

Continuous responses constrained to the interval (0,1) were modelled using the beta distribution with a logit link function⁶². Because the beta density excludes the boundaries 0 and 1, observed values equal to 0 or 1 were moved off the boundary by the smaller of (a) replacement with the nearest observed interior value and (b) an

inward shift equal to the observed data range; the option yielding the smaller absolute change was chosen. COMPASS-derived response probabilities⁶³ and polyfunctionality scores were modelled using this approach.

Continuous positive responses without an upper bound were modelled using the gamma distribution with a log link function. The Δ HLA-DR signature²⁰ was modelled using this approach.

Count responses were modelled using the distribution that minimised the Bayesian Information Criterion, choosing between negative binomial with a log link and the beta-binomial with a logit link. Frequency data were analysed as count outcomes, with the number of cells as the dependent variable and the log of the number of cells in the parent population included as an offset.

Robustness check on p -values

To ensure that apparently significant p -values were not clearly driven by outliers, we implemented the following procedure.

1. Fit the model to the original data and record the resulting p -value, denoted p_{orig} .
2. Replace outlying values of pre-specified variable(s) with values closer to the centre of the variable's distribution (details below).
3. Refit the model to the data with the shrunken variable(s) and obtain the corresponding p -value, denoted p_{shrunk} .
4. Replace p_{orig} with p_{shrunk} if $\log_{10}(p_{shrunk}) - \log_{10}(p_{orig}) > 1.5$.

Some change in p -value is expected after shrinking; the aim here is to identify and correct p -values that are implausibly small due to extreme values.

Definition of the shrunken variable

For a given continuous variable $x = (x_1, \dots, x_n)$ we obtain a shrunk version x^* by replacing observations that fall outside data-defined bounds with the nearest bound. Let $\bar{x}_{\text{mean}}$ and $\bar{x}_{\text{med}}$ denote the sample mean and median of x , and let σ_{sd} and σ_{mad} denote the sample standard deviation and median absolute deviation. For $k \in \mathbb{R}_{>0}$ define

$$b_{\min} = \min(\bar{x}_{\text{mean}} - k \sigma_{\text{sd}}, \bar{x}_{\text{med}} - k \sigma_{\text{mad}}),$$

and

$$b_{\max} = \max(\bar{x}_{\text{mean}} + k \sigma_{\text{sd}}, \bar{x}_{\text{med}} + k \sigma_{\text{mad}}).$$

Each observation x_i with $x_i < b_{\min}$ or $x_i > b_{\max}$ was replaced by the nearest bound (i.e. by $b_{\min}$ or $b_{\max}$, respectively), producing x^* .

Data availability

The processed mass cytometry data (live, debarcoded single cells, as well as CD4, CD8 and TCR $\gamma\delta$ lineages) generated in this study have been deposited in Zenodo under the identifier 10.5281/zenodo.18402521. The processed count matrices used for statistical modeling (DataTidyACSCyTOFFAUST and DataTidyACSCyTOFCytokinesTCells) and the clinical data (DataTidyACSClinical) are available at the SATVILab GitHub organization (<https://github.com/SATVILab>) and are archived on Zenodo under the same identifier, 10.5281/zenodo.18402521. The processed RNA-seq and SomaScan proteomics datasets analyzed in this study are also archived on Zenodo with the identifier 10.5281/zenodo.18402521. Raw RNA-Seq data are available from the GEO database (accession numbers GSE79362 and GSE103147). The raw mass cytometry data are available from the corresponding author upon reasonable request.

Code availability

The custom code for the StimGate algorithm and the analysis pipeline is available at the SATVILab GitHub organization (<https://github.com/SATVILab>), where the repos.list file lists the relevant repositories, including stimgate (cytokine-gating algorithm implementation) and CompACSCyTOFTCells (pipeline setup and dependencies). To facilitate reproducibility, a Docker container capable of regenerating the analysis environment and results is available at <https://github.com/orgs/SATVILab/packages/container/compacscytoftcells/511043745>. All code repositories and the container image are archived on Zenodo with the identifier 10.5281/zenodo.18402521.

References

- 1 Bagcchi, S. WHO's Global Tuberculosis Report 2022. *Lancet Microbe* **4**, e20 (2023). [https://doi.org/10.1016/S2666-5247\(22\)00359-7](https://doi.org/10.1016/S2666-5247(22)00359-7)
- 2 Scriba, T. J., Coussens, A. K. & Fletcher, H. A. Human Immunology of Tuberculosis. *Microbiol Spectr* **5** (2017). <https://doi.org/10.1128/microbiolspec.TBTB2-0016-2016>
- 3 Lewinsohn, D. M. & Lewinsohn, D. A. The Missing Link in Correlates of Protective Tuberculosis Immunity: Recognizing the Infected Cell. *Front Immunol* **13**, 869057 (2022). <https://doi.org/10.3389/fimmu.2022.869057>
- 4 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). <https://doi.org/10.1371/journal.ppat.1004603>
- 5 Cadena, A. M. *et al.* Concurrent infection with Mycobacterium tuberculosis confers robust protection against secondary infection in macaques. *PLoS Pathog* **14**, e1007305 (2018). <https://doi.org/10.1371/journal.ppat.1007305>
- 6 Dijkman, K. *et al.* Disparate Tuberculosis Disease Development in Macaque Species Is Associated With Innate Immunity. *Front Immunol* **10**, 2479 (2019). <https://doi.org/10.3389/fimmu.2019.02479>
- 7 Darrah, P. A. *et al.* Prevention of tuberculosis in macaques after intravenous BCG immunization. *Nature* **577**, 95-102 (2020). <https://doi.org/10.1038/s41586-019-1817-8>
- 8 Gideon, H. P. *et al.* Multimodal profiling of lung granulomas in macaques reveals cellular correlates of tuberculosis control. *Immunity* **55**, 827-846 e810 (2022). <https://doi.org/10.1016/j.immuni.2022.04.004>
- 9 Zak, D. E. *et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* **387**, 2312-2322 (2016). [https://doi.org/10.1016/S0140-6736\(15\)01316-1](https://doi.org/10.1016/S0140-6736(15)01316-1)
- 10 Scriba, T. J. *et al.* Sequential inflammatory processes define human progression from M. tuberculosis infection to tuberculosis disease. *PLoS Pathog* **13**, e1006687 (2017). <https://doi.org/10.1371/journal.ppat.1006687>
- 11 Suliman, S. *et al.* Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis. *Am J Respir Crit Care Med* **197**, 1198-1208 (2018). <https://doi.org/10.1164/rccm.201711-2340OC>
- 12 Penn-Nicholson, A. *et al.* Discovery and validation of a prognostic proteomic signature for tuberculosis progression: A prospective cohort study. *PLoS Med* **16**, e1002781 (2019). <https://doi.org/10.1371/journal.pmed.1002781>
- 13 Gupta, R. K. *et al.* Concise whole blood transcriptional signatures for incipient tuberculosis: a systematic review and patient-level pooled meta-analysis. *Lancet Respir Med* **8**, 395-406 (2020). [https://doi.org/10.1016/S2213-2600\(19\)30282-6](https://doi.org/10.1016/S2213-2600(19)30282-6)
- 14 Penn-Nicholson, A. *et al.* RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. *Sci Rep* **10**, 8629 (2020). <https://doi.org/10.1038/s41598-020-65043-8>
- 15 Fletcher, H. A. *et al.* T-cell activation is an immune correlate of risk in BCG vaccinated infants. *Nat Commun* **7**, 11290 (2016). <https://doi.org/10.1038/ncomms11290>
- 16 Day, C. L. *et al.* Functional capacity of Mycobacterium tuberculosis-specific T cell responses in humans is associated with mycobacterial load. *J Immunol* **187**, 2222-2232 (2011). <https://doi.org/10.4049/jimmunol.1101122>

- 17 Harari, A. *et al.* Dominant TNF-alpha+ Mycobacterium tuberculosis-specific CD4+ T cell responses discriminate between latent infection and active disease. *Nat Med* **17**, 372-376 (2011). <https://doi.org/10.1038/nm.2299>
- 18 Riou, C. *et al.* A subset of circulating blood mycobacteria-specific CD4 T cells can predict the time to Mycobacterium tuberculosis sputum culture conversion. *PLoS One* **9**, e102178 (2014). <https://doi.org/10.1371/journal.pone.0102178>
- 19 Adekambi, T. *et al.* Biomarkers on patient T cells diagnose active tuberculosis and monitor treatment response. *J Clin Invest* **125**, 3723 (2015). <https://doi.org/10.1172/JCI83279>
- 20 Mpande, C. A. M. *et al.* Antigen-Specific T-Cell Activation Distinguishes between Recent and Remote Tuberculosis Infection. *Am J Respir Crit Care Med* **203**, 1556-1565 (2021). <https://doi.org/10.1164/rccm.202007-2686OC>
- 21 Mpande, C. A. M. *et al.* Immune profiling of Mycobacterium tuberculosis-specific T cells in recent and remote infection. *EBioMedicine* **64**, 103233 (2021). <https://doi.org/10.1016/j.ebiom.2021.103233>
- 22 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). <https://doi.org/10.1164/rccm.201003-0334OC>
- 23 Mahomed, H. *et al.* TB incidence in an adolescent cohort in South Africa. *PLoS One* **8**, e59652 (2013). <https://doi.org/10.1371/journal.pone.0059652>
- 24 Penn-Nicholson, A., Scriba, T. J., Hatherill, M., White, R. G. & Sumner, T. A novel blood test for tuberculosis prevention and treatment. *S Afr Med J* **107**, 4-5 (2016). <https://doi.org/10.7196/SAMJ.2016.v107.i1.12230>
- 25 Greene, E. *et al.* New interpretable machine-learning method for single-cell data reveals correlates of clinical response to cancer immunotherapy. *Patterns (N Y)* **2**, 100372 (2021). <https://doi.org/10.1016/j.patter.2021.100372>
- 26 Elkington, P. *et al.* MMP-1 drives immunopathology in human tuberculosis and transgenic mice. *J Clin Invest* **121**, 1827-1833 (2011). <https://doi.org/10.1172/JCI45666>
- 27 Garcillan, B. *et al.* gammadelta T Lymphocytes in the Diagnosis of Human T Cell Receptor Immunodeficiencies. *Front Immunol* **6**, 20 (2015). <https://doi.org/10.3389/fimmu.2015.00020>
- 28 Mpande, C. A. M. *et al.* Mycobacterium tuberculosis-Specific T Cell Functional, Memory, and Activation Profiles in QuantiFERON-Reverters Are Consistent With Controlled Infection. *Front Immunol* **12**, 712480 (2021). <https://doi.org/10.3389/fimmu.2021.712480>
- 29 Scriba, T. J., Maseeme, M., Young, C., Taylor, L. & Leslie, A. J. Immunopathology in human tuberculosis. *Science Immunology* **10**, eado5951 (2025). <https://doi.org/10.1126/sciimmunol.ado5951>
- 30 Dijkman, K. *et al.* Prevention of tuberculosis infection and disease by local BCG in repeatedly exposed rhesus macaques. *Nature medicine* **25**, 255-262 (2019). <https://doi.org/10.1038/s41591-018-0319-9>
- 31 Peters, J. M. *et al.* High-dose intravenous BCG vaccination induces enhanced immune signaling in the airways. *Science advances* **11**, eadq8229 (2025). <https://doi.org/10.1126/sciadv.adq8229>

- 32 Simonson, A. W. *et al.* Intravenous BCG-mediated protection against tuberculosis requires CD4⁺ T cells and CD8 α ⁺ lymphocytes. *The Journal of experimental medicine* **222** (2025). <https://doi.org/10.1084/jem.20241571>
- 33 Ganchua, S. K. C., White, A. G., Klein, E. C. & Flynn, J. L. Lymph nodes-The neglected battlefield in tuberculosis. *PLoS pathogens* **16**, e1008632 (2020). <https://doi.org/10.1371/journal.ppat.1008632>
- 34 McCaffrey, E. F. *et al.* The immunoregulatory landscape of human tuberculosis granulomas. *Nature Immunology* **23**, 318-329 (2022). <https://doi.org/10.1038/s41590-021-01121-x>
- 35 Sawyer, A. J. *et al.* Spatial mapping reveals granuloma diversity and histopathological superstructure in human tuberculosis. *The Journal of experimental medicine* **220** (2023). <https://doi.org/10.1084/jem.20221392>
- 36 Nelson, C. E. *et al.* Mycobacterium tuberculosis peptide-specific T cells in pulmonary granulomas display broad effector functions. *iScience* **28**, 114034 (2025). <https://doi.org/10.1016/j.isci.2025.114034>
- 37 Musvosvi, M. *et al.* T cell receptor repertoires associated with control and disease progression following Mycobacterium tuberculosis infection. *Nat Med* **29**, 258-269 (2023). <https://doi.org/10.1038/s41591-022-02110-9>
- 38 Moguche, A. O. *et al.* Antigen Availability Shapes T Cell Differentiation and Function during Tuberculosis. *Cell Host Microbe* **21**, 695-706 e695 (2017). <https://doi.org/10.1016/j.chom.2017.05.012>
- 39 Stetsenko, V. *et al.* Human CD4⁺ T cells recognize Mycobacterium tuberculosis-infected macrophages amid broader responses. *The Journal of experimental medicine* **222** (2025). <https://doi.org/10.1084/jem.20250460>
- 40 Leddy, O. *et al.* Immuno-peptidomics can inform the design of mRNA vaccines for the delivery of Mycobacterium tuberculosis MHC class II antigens. *Science Translational Medicine* **17**, eadw9184 <https://doi.org/10.1126/scitranslmed.adw9184>
- 41 Lindestam Arlehamn, C. S. *et al.* Memory T Cells in Latent Mycobacterium tuberculosis Infection Are Directed against Three Antigenic Islands and Largely Contained in a CXCR3⁺CCR6⁺ Th1 Subset. *PLOS Pathogens* **9**, e1003130 (2013). <https://doi.org/10.1371/journal.ppat.1003130>
- 42 Scriba, T. J., Netea, M. G. & Ginsberg, A. M. Key recent advances in TB vaccine development and understanding of protective immune responses against Mycobacterium tuberculosis. *Semin Immunol* **50**, 101431 (2020). <https://doi.org/10.1016/j.smim.2020.101431>
- 43 Tippalagama, R. *et al.* HLA-DR Marks Recently Divided Antigen-Specific Effector CD4 T Cells in Active Tuberculosis Patients. *The Journal of Immunology* **207**, 523-533 (2021). <https://doi.org/10.4049/jimmunol.2100011>
- 44 Tippalagama, R. *et al.* Peptide-driven identification of TCRs reveals dynamics and phenotypes of CD4 T cells in tuberculosis. *The Journal of Immunology*, vkaf287 (2025). <https://doi.org/10.1093/jimmun/vkaf287>
- 45 Musvosvi, M. *et al.* T-cell biomarkers for diagnosis of tuberculosis: candidate evaluation by a simple whole blood assay for clinical translation. *Eur Respir J* **51** (2018). <https://doi.org/10.1183/13993003.00153-2018>
- 46 Riou, C., Berkowitz, N., Goliath, R., Burgers, W. A. & Wilkinson, R. J. Analysis of the Phenotype of Mycobacterium tuberculosis-Specific CD4⁺ T Cells to Discriminate Latent

- from Active Tuberculosis in HIV-Uninfected and HIV-Infected Individuals. *Front Immunol* **8**, 968 (2017). <https://doi.org/10.3389/fimmu.2017.00968>
- 47 Lindenstrom, T. *et al.* Tuberculosis subunit vaccination provides long-term protective immunity characterized by multifunctional CD4 memory T cells. *J Immunol* **182**, 8047-8055 (2009). <https://doi.org/10.4049/jimmunol.0801592>
- 48 Mendelsohn, S. C., Mbandi, S. K., Hatherill, M. & Scriba, T. J. Blood transcriptional signatures for tuberculosis testing. *Lancet Respir Med* **8**, 330-331 (2020). [https://doi.org/10.1016/S2213-2600\(20\)30045-X](https://doi.org/10.1016/S2213-2600(20)30045-X)
- 49 Gutierrez-Arcelus, M. *et al.* Lymphocyte innateness defined by transcriptional states reflects a balance between proliferation and effector functions. *Nature Communications* **10**, 687 (2019). <https://doi.org/10.1038/s41467-019-08604-4>
- 50 Scriba, T. J. *et al.* Biomarker-guided tuberculosis preventive therapy (CORTIS): a randomised controlled trial. *The Lancet infectious diseases* **21**, 354-365 (2021). [https://doi.org/10.1016/S1473-3099\(20\)30914-2](https://doi.org/10.1016/S1473-3099(20)30914-2)
- 51 Mendelsohn, S. C. *et al.* Transcriptomic Signatures of Progression to Tuberculosis Disease Among Close Contacts in Brazil. *The Journal of Infectious Diseases* **230**, e1355-e1365 (2024). <https://doi.org/10.1093/infdis/jiae237>
- 52 Mendelsohn, S. C. *et al.* Screening for asymptomatic tuberculosis among adults with household exposure to pulmonary tuberculosis: a prospective observational cohort study. *The Lancet global health* **13**, e1869-e1879 (2025). [https://doi.org/10.1016/S2214-109X\(25\)00276-1](https://doi.org/10.1016/S2214-109X(25)00276-1)
- 53 R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2025).
- 54 Finck, R. *et al.* Normalization of mass cytometry data with bead standards. *Cytometry Part A* **83**, 483-494 (2013). <https://doi.org/10.1002/cyto.a.22271>
- 55 Chevrier, S. *et al.* Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. *Cell Systems* **6**, 612-620.e615 (2018). <https://doi.org/10.1016/j.cels.2018.02.010>
- 56 Zunder, E. R. *et al.* Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. *Nature Protocols* **10**, 316-333 (2015). <https://doi.org/10.1038/nprot.2015.020>
- 57 Meskas, J., Yokosawa, D., Wang, S., Segat, G. C. & Brinkman, R. R. flowCut: An R package for automated removal of outlier events and flagging of files based on time versus fluorescence analysis. *Cytometry Part A* **103**, 71-81 (2023). <https://doi.org/10.1002/cyto.a.24670>
- 58 Brooks, M. E. *et al.* glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal* **9**, 378-400 (2017). <https://doi.org/10.32614/RJ-2017-066>
- 59 emmeans: Estimated Marginal Means, aka Least-Squares Means v. 1.10.5 (R Foundation for Statistical Computing, 2024).
- 60 Bonferroni, C. E. *Teoria statistica delle classi e calcolo delle probabilità*. Vol. 8 (Seeber, 1936).
- 61 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289-300 (1995). <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

- 62 Ferrari, S. L. P. & Cribari-Neto, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* **31**, 799-815 (2004). <https://doi.org/10.1080/0266476042000214501>
- 63 Lin, L. *et al.* COMPASS identifies T-cell subsets correlated with clinical outcomes. *Nature Biotechnology* **33**, 610-616 (2015). <https://doi.org/10.1038/nbt.3187>

Acknowledgements

The authors are grateful to the study participants enrolled in the Adolescent Cohort Study team and acknowledge the considerable contributions of study clinicians, nurses, technicians and clinical research workers. Members of the Adolescent Cohort Study team are listed in Supplementary Note 1.

Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team (hpc.uct.ac.za; <https://doi.org/10.5281/zenodo.10021612>). We thank Bill Jacobs for providing the *Mtb* triple auxotroph via Digby Warner.

This work was supported by the Gates Foundation Global Health grants OPP1066265, OPP1023483 and OPP1065330, and Grand Challenges in Global Health (GC6-74 grant 37772). The Adolescent Cohort Study was also supported by Aeras and BMGF GC12 (grant 37885) for QFT testing. V.R. was supported by the Swiss National Foundation and EDCTP. M.J.R. was supported by the South African Statistics Association.

Author Contributions

V.R., C.Y., M.M., and C.S. contributed to data generation. M.J.R. and F.L. developed analytical methods. M.J.R., V.R., C.Y., F.L., P.V., G.F., E.G., and R.G. analysed data. M.J.R., V.R., and T.J.S. interpreted results. H.T.M. provided mass cytometry training. D.W., V.M., C.L.A., and A.S. designed antigen selection and peptide pools. N.B. and M.F. provided administrative support. W.A.H., M.H. T.J.S. designed, directed and acquired funding. V.R., M.J.R., and T.J.S. drafted the manuscript. V.R. and M.J.R. contributed equally. All authors approved the final version.

Competing Interests

TJS is a co-inventor of a patent of the RISK6 signature. The authors declare no competing interests. R.G. has received consulting income (payments made to the Lausanne University Hospital) from Takeda, Arcellx, Sanofi, Owkin, and declares ownership in Ozette Technologies. R.G. have received research funding from Owkin and 10x Genomics. G.F and E.G. are employees of Ozette technologies and own shares.

Figure Legends

Fig. 1: Technical and analytical pipelines for assessing T cell correlates of tuberculosis (TB) progression. (a) Schematic of participant sampling and peripheral blood mononuclear cell (PBMC) processing by mass cytometry (CyTOF). All participants were QuantiFERON-positive (QFT⁺) and/or tuberculin skin test-positive (TST⁺) at baseline, demonstrating evidence of current or prior Mtb infection. Participants were classified as *progressors* if they developed TB during follow-up or as *controllers* if they remained TB free. Multiple samples were acquired per participant, yielding 150 samples across 37 progressors and 36 controllers. After cryopreservation, samples were thawed, aliquoted, and stimulated for twelve hours. Cells were barcoded for simultaneous processing in small batches to reduce technical variation, then stained with 29 extracellular and 6 intracellular markers to characterise phenotype and function, particularly of T cells, and acquired by CyTOF. See “Methods” for details. **(b)** Pre-processing, clustering and analysis. Single-cell data were pre-processed to obtain normalised, live, debarcoded, and debeaded (bead removal), minimising artefacts related to processing time. FAUST automatically identified detailed cell phenotypes, which were validated against manual gating on a subset of samples. An in-house cytokine-gating algorithm, StimGate, identified cytokine-positive cells within FAUST-identified phenotypes and was also validated against manual gating. Both bulk and cytokine-positive cells were analysed for associations with TB progression and TB-associated inflammation markers using mixed-effect models. Cytokine-positive CD4 T cells were also phenotyped with FlowSOM to confirm the

robustness of antigen-specific results derived from FAUST-defined cytokine-positive phenotypes. See “Methods” for details.

Fig. 2: CD4 T cell phenotypes associated with TB progression or TB-associated inflammatory markers. (a) Summary of FAUST-identified CD4 T cell phenotypic subsets showing their abundance and statistical associations. Bar height indicates the median frequency of each phenotypic subset among CD4 T cells, and vertical error bars show the interquartile range. Bar colour represents the strongest statistical association with either TB progression or inflammatory markers, after false discovery rate (FDR) correction by the Benjamini–Hochberg method. The middle heat map displays FAUST marker expression (blue, positive; blank, negative). The bottom heat map indicates the statistical significance of the association between abundance of the phenotypic subset with TB progression, or with each of three markers of inflammation separately. See “Methods” for further details. (b–g) Representative model outputs. For TB progression models, the x-axis shows days to TB diagnosis. We tested (i) *Distal*: whether controllers differ from progressors ≥ 630 days before diagnosis; (ii) *Proximal*: whether progressor responses change as diagnosis nears; and (iii) *Progression*: any overall difference between groups. For inflammatory-marker models, the x-axis represents RISK6 score or plasma abundance of MMP-1 or C9. We tested (i) the main effect of the marker (e.g. C9), and (ii) its interaction with progression status (e.g. Prog:C9). P-values from two-sided likelihood ratio tests using mixed-effects models are reported, with Benjamini-Hochberg-adjusted FDR adjustments (q_{fdr}). Shaded bands denote 95% Wald confidence intervals for the linear predictor, back-transformed to the response scale (orange = progressors; blue = controllers). The

central line shows the estimated mean value. For visual clarity, the controller band is shown at a constant height across time to TB. See “Methods” for further details.

Fig. 3: CD8 T cell phenotypes associated with TB progression or TB-associated inflammatory markers. (a) Summary of FAUST-identified CD8 T cell phenotypic subsets showing their abundance and statistical associations. Bar height indicates the median frequency of each phenotypic subset among CD4 T cells, and vertical error bars show the interquartile range. Bar colour represents the strongest statistical association with either TB progression or inflammatory markers, after false discovery rate (FDR) correction by the Benjamini–Hochberg method. The middle heat map displays FAUST marker expression (blue, positive; blank, negative). The bottom heat map indicates the statistical significance of the association between abundance of the phenotypic subset with TB progression, or with each of three markers of inflammation separately. (b–g) Representative models of CD8 T cell phenotypic subsets associated with inflammatory markers. The x-axis represents RISK6 score or plasma abundance of MMP-1 or C9. We tested (i) the main effect of the marker (e.g. C9), and (ii) its interaction with progression status (e.g. Prog:C9). P-values from two-sided likelihood ratio tests using mixed-effects models are reported, with Benjamini-Hochberg-adjusted FDR adjustments (q_{fdr}). Shaded bands denote 95% Wald confidence intervals for the linear predictor, back-transformed to the response scale (orange = progressors; blue = controllers). The central line shows the estimated mean value. For visual clarity, the controller band is shown at a constant height across time to TB.

Fig. 4: TCR $\gamma\delta$ T cell phenotypes associated with TB progression or TB-associated inflammatory markers. (a) Summary of FAUST-identified TCR $\gamma\delta$ T cell

phenotypic subsets showing their abundance and statistical associations. Bar height indicates the median frequency of each phenotypic subset among CD4 T cells, and vertical error bars show the interquartile range. Bar colour represents the strongest statistical association with either TB progression or inflammatory markers, after false discovery rate (FDR) correction by the Benjamini–Hochberg method. The middle heat map displays FAUST marker expression (blue, positive; blank, negative). The bottom heat map indicates the statistical significance of the association between abundance of the phenotypic subset with TB progression, or with each of three markers of inflammation separately. **(b-d)** Examples of TCR $\gamma\delta$ T cell phenotypic subset models. For TB progression, the x-axis shows days to TB diagnosis. We tested (i) *Distal*: whether controllers differ from progressors ≥ 630 days before diagnosis; (ii) *Proximal*: whether progressor responses change as diagnosis nears; and (iii) *Progression*: any overall difference between groups. For inflammatory-marker models, the x-axis represents plasma abundance of C9 or RISK6 score. We tested (i) the main effect of the marker (e.g. C9), and (ii) its interaction with progression status (e.g. Prog:C9). P-values from two-sided likelihood ratio tests using mixed-effects models are reported, with Benjamini-Hochberg-adjusted FDR adjustments (q_{fdr}). Shaded bands denote 95% Wald confidence intervals for the linear predictor, back-transformed to the response scale (orange = progressors; blue = controllers). The central line shows the estimated mean value. For visual clarity, the controller band is shown at a constant height across time to TB.

Fig. 5: Antigen-specific CD4 Th1/Th17 and HLA-DR⁺ Mtb-specific T cell response associated with TB progression or TB-associated inflammatory markers. (a-d)

Response probabilities, computed using COMPASS, for CD4 T cell subsets

expressing the cytokine combinations shown in the heatmaps below the box-and-whisker plots, across four antigen stimulation conditions. Only cytokine combinations with sufficient evidence of a response for at least one stimulation are shown in the heatmaps. The thick horizontal lines represent medians, the boxes represent IQR and the whiskers the range. **(e-g)** Representative model outputs for TB progression models for antigen-specific Th1/Th17⁺ CD4 T cells against TB progression.. We tested (i) *Distal*: whether controllers differ from progressors ≥ 630 days before diagnosis; (ii) *Proximal*: whether progressor responses change as diagnosis nears; and (iii) *Progression*: any overall difference between groups. **(h-j)**. Antigen-specific CD4 T cell activation, computed by Δ HLA-DR expression on cytokine⁺ CD4 T cells (Mpande, 2020), against TB progression or RISK6. **(k-m)** Selected antigen-specific CD4 T cell cytokine-expressing phenotypic subsets, identified by FAUST phenotypes, associated with TB progression. Phenotypes are specified in the y-axis titles; subsets are negative for all unlisted phenotyping markers. The notation q_{fwer} denotes the Bonferroni-adjusted p-value that controls the family-wise type I error rate. Statistical models, testing procedures, q-value notation, confidence-band and colour conventions as in Fig. 2b; see “Methods” for details.

Fig. 6: Antigen-specific CD8 and TCR $\gamma\delta$ Th1 responses associated with TB progression or TB-associated inflammatory markers. CD8 T cell **(a,b,c)** and TCR $\gamma\delta$ cell **(d)** response probabilities, computed by COMPASS, for cell subsets expressing the cytokine combinations shown in the heatmaps below the box-and-whisker plots, across four antigen stimulation conditions. The thick horizontal lines represent medians, the boxes represent IQR and the whiskers the range. Only cytokine combinations with sufficient evidence of a response for at least one

stimulation are shown. No CD8 T cell response was observed for non-secreted *Mtb* proteins. For TCR $\gamma\delta$ cells no response was observed for stimulations other than live *Mtb*. **(e-g)** Antigen-specific CD8 T cell responses associated with TB progression. **(e and f)** refer to total CD8 T cell responses; **(g)** refers to a specific CD8 T cell phenotype. Phenotype is specified in the y-axis title; subset is negative for all unlisted phenotyping markers. Antigen-reactive TCR $\gamma\delta$ T cell responses associated with TB progression **(h)** or MMP-1 **(i,j)**. The left and middle plots refer to total TCR $\gamma\delta$ T cell responses; the right plot refers to a specific TCR $\gamma\delta$ T cell phenotype. Phenotype is specified in the y-axis title; subset is negative for all unlisted phenotyping markers. The notation q_{fwer} denotes the Bonferroni-adjusted p-value that controls the family-wise type I error rate. Statistical models, testing procedures, q-value notation, confidence-band and colour conventions as in Fig. 2b; see “Methods” for details.

Table 1: Progressor and controller demographic breakdown and per-assay

counts of participants and samples. Summary of participant demographics and the number of participants and samples included in each assay, stratified by disease outcome group.

	Progressors (n = 44)	Controllers (n = 106)	Total (n = 150)
Demographic data			
Ethnicity, n (%)			
Colored (Cape mixed ancestry)	40 (90.9%)	97 (91.5%)	137 (91.3%)
Black (African ancestry)	4 (9.1%)	9 (8.5%)	13 (8.7%)
Female, n (%)	31 (70.5%)	69 (65.1%)	100 (66.7%)
Age at baseline, years (mean, range)	15.5 (12-18)	15.6 (13-18)	15.6 (12-18)
Previous diagnosis of tuberculosis, n (%)	5 (11.4%)	16 (15.1%)	21 (14%)
CyTOF data			
Participants, n	37	36	73
Samples, n	76	72	148
SOMAScan data			
Participants, n	37	104	141
Samples, n	82	290	372
RISK6 data			
Participants, n	40	105	145
Samples, n	82	286	368

Table 2: Hypotheses for time to TB model.

Hypothesis	Null hypothesis	Non-null interpretation	Primary hypothesis
Progression	$\beta_4 = \beta_{5_1} = \beta_{5_2} = \dots = \beta_{5_d} = 0$	There is distal and/or proximal TB effect	Yes
Distal TB	$\beta_4 = 0$	Progressors differ from controllers beyond 630 days from <i>Mtb</i> diagnosis	No
Proximal TB	$\beta_{5_1} = \beta_{5_2} = \dots = \beta_{5_d} = 0$	The response by progressors changes from 630 days from <i>Mtb</i> diagnosis to TB diagnosis itself	No

Table 3: Hypotheses for TB-associated inflammation marker model.

Hypothesis	Null hypothesis	Non-null interpretation	Primary for hypothesis
Inflammation marker	$\beta_{5_1} = \beta_{5_2} = \dots$ $= \beta_{5_d} = \beta_{6_1}$ $= \beta_{6_2} = \dots$ $= \beta_{6_d} = 0$	The inflammation marker influences the response	Yes
Progression:inflammation marker	$\beta_{6_1} = \beta_{6_2} = \dots$ $= \beta_{6_d} = 0$	The effect of the inflammation marker depends on progressor status	No

a**Sampling**

37 progressors,
76 cryopreserved
PBMCs



750 500 250 0
Days to TB

36 controllers,
72 cryopreserved
PBMCs



0 200 400 600
Days since entry

CyTOF processing

12 hours stimulation

- Unstimulated
- Live Mtb
- Secreted Mtb proteins peptide pool
- Non-secreted Mtb proteins peptide pool
- EBV and CMV peptide pool

Staining

- CD45 surface barcoding (5)
- Extracellular (29)
- Intracellular (6)

**b****1. Pre-processing**

- I. Normalisation
- II. Bead removal I
- III. Debarcoding
- IV. Singlet gating
- V. Live cell gating
- VI. Time-based gating
- VII. Bead removal II

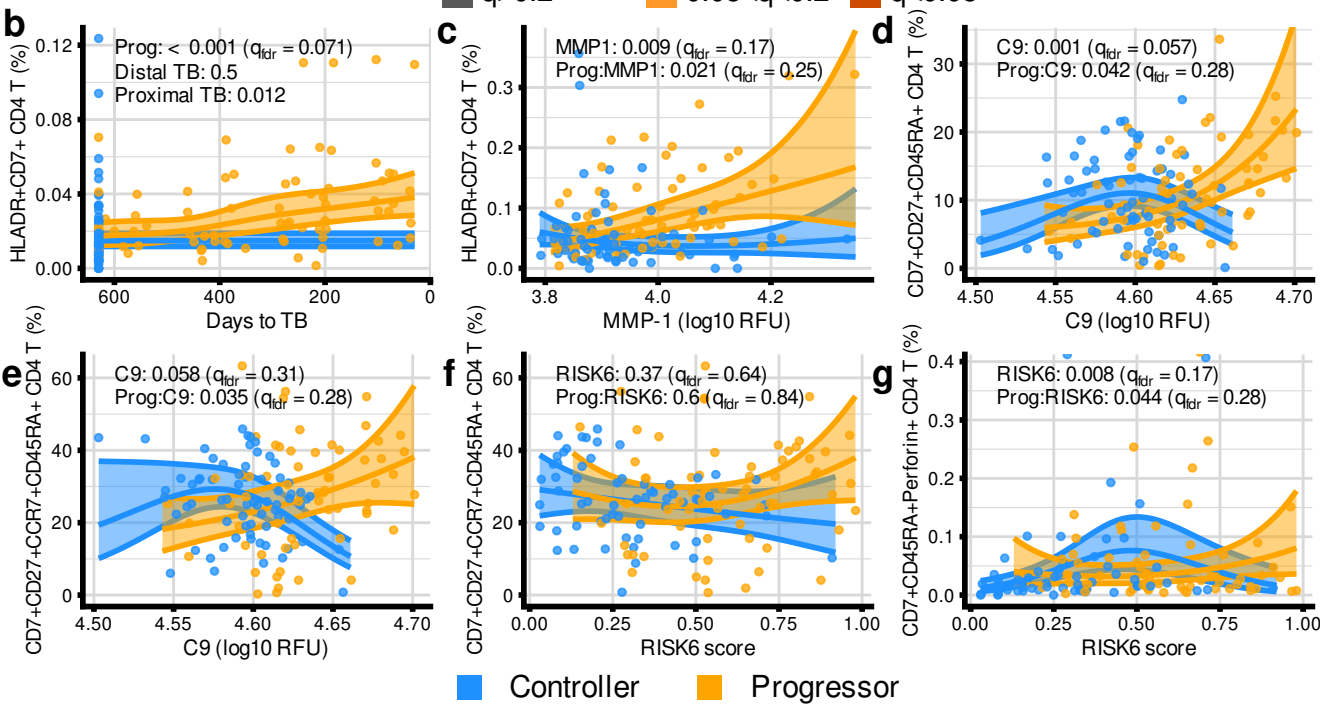
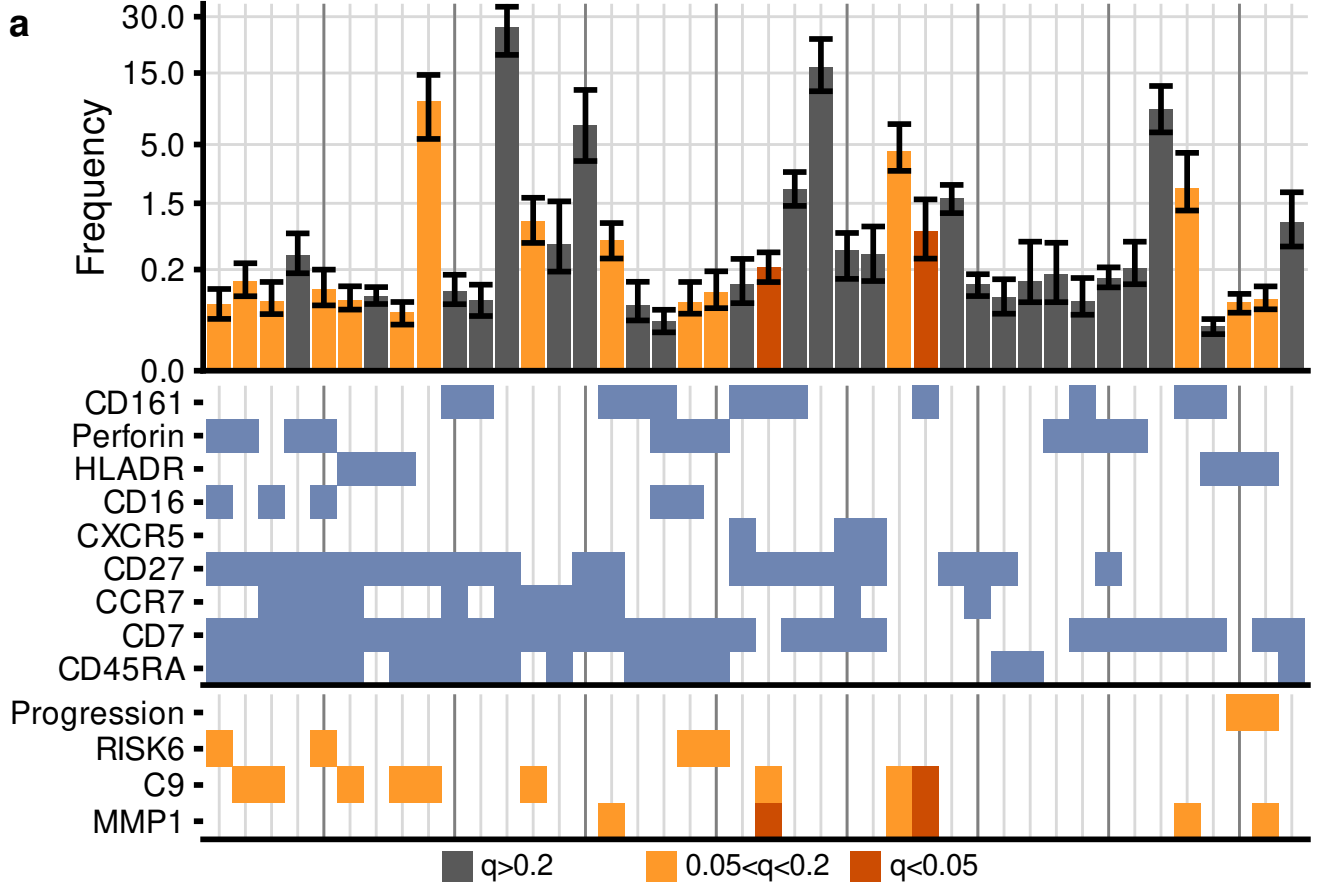
2. Clustering

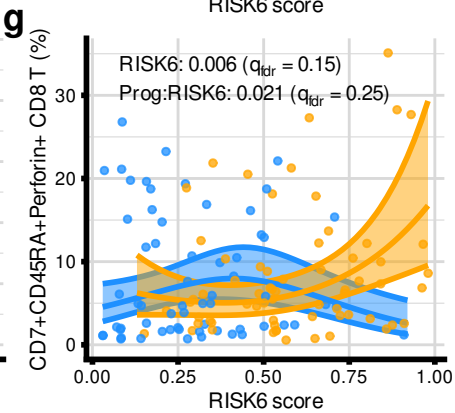
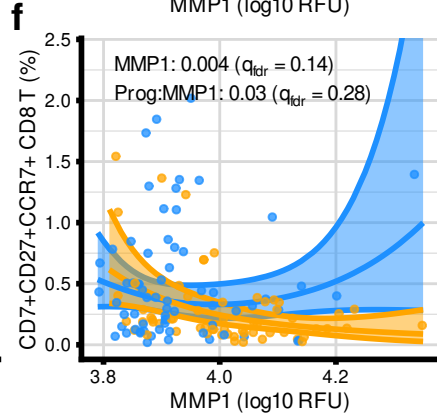
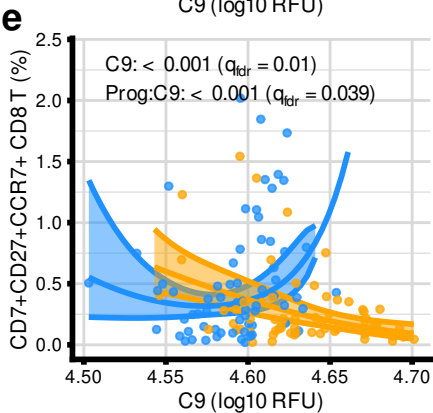
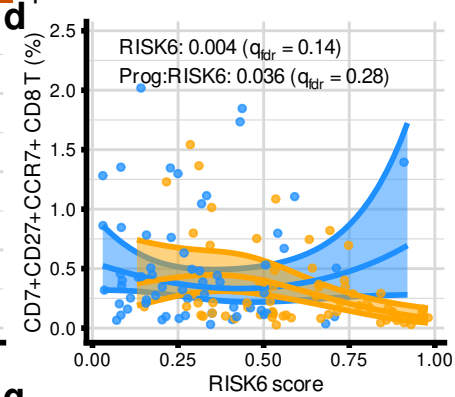
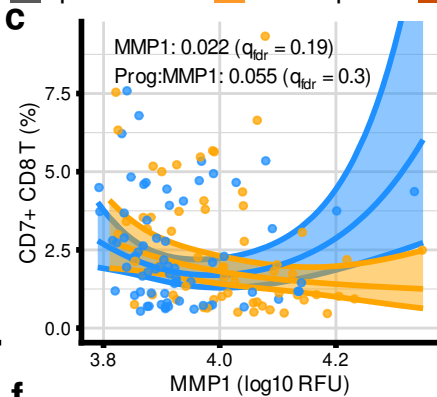
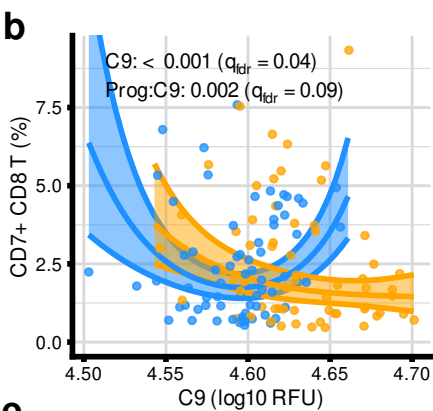
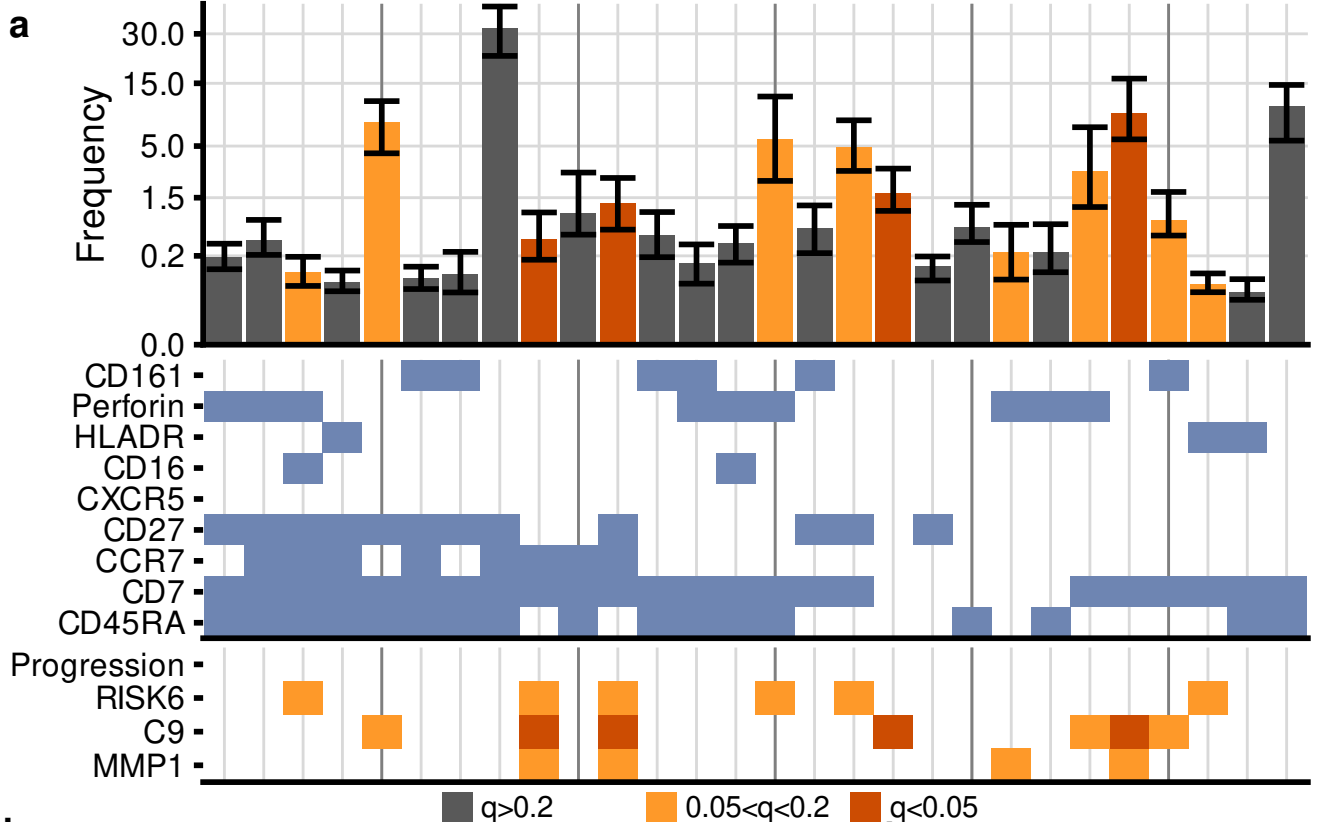
- I. Identify phenotypes
 - Apply *FAUST* to pre-processed cells
- II. Identify cytokine-positive cells
 - Apply *StimGate* to *FAUST*-identified lineages
- III. Cluster all cytokine-positive cells using all markers
 - Apply *FlowSOM* to cytokine-positive cells from step II.

3. Mixed-effects modelling

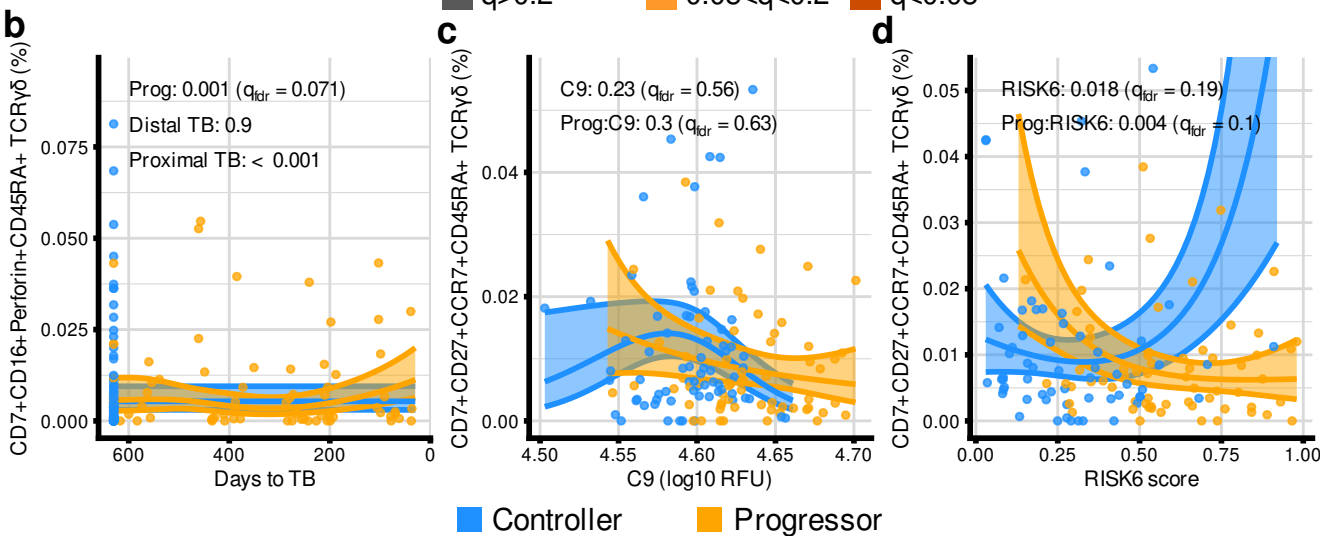
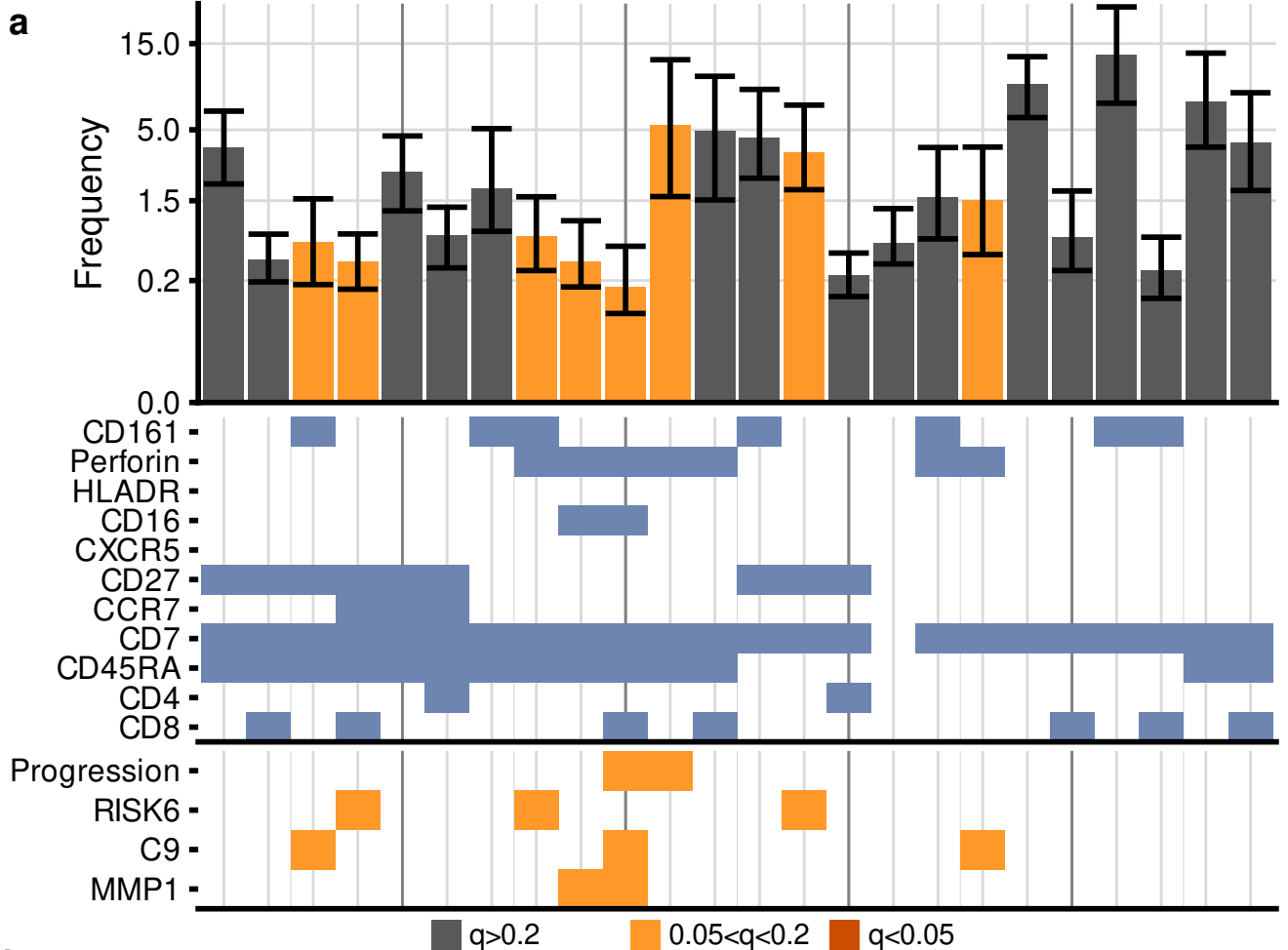
All subsets assessed for associate with time to TB and TB-associated inflammatory markers.

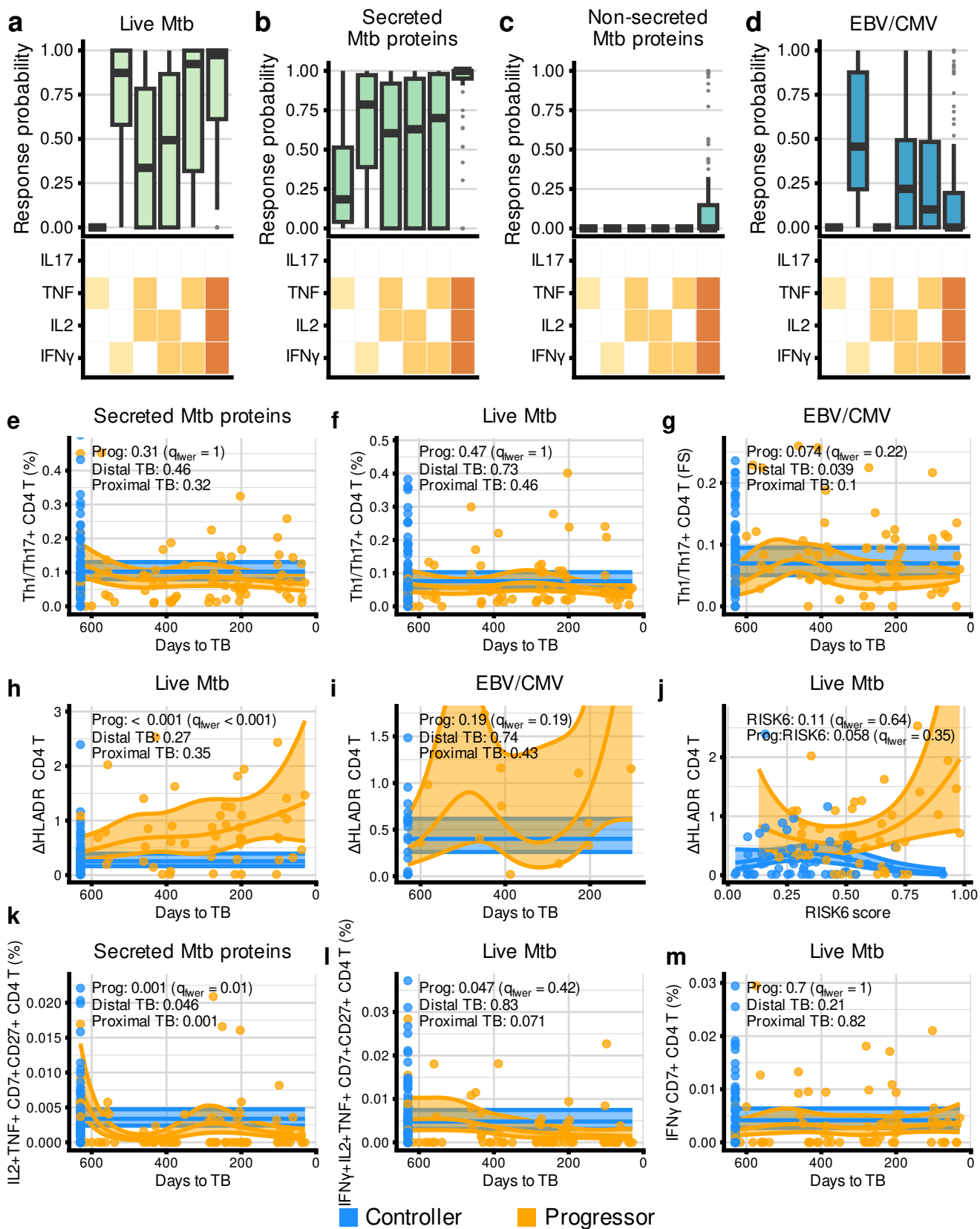
- I. Non-specific
 - *FAUST* phenotypes
- II. Antigen-specific
 - Cytokine-positive lineages
 - Cytokine-positive phenotypes
 - *FlowSOM* clusters

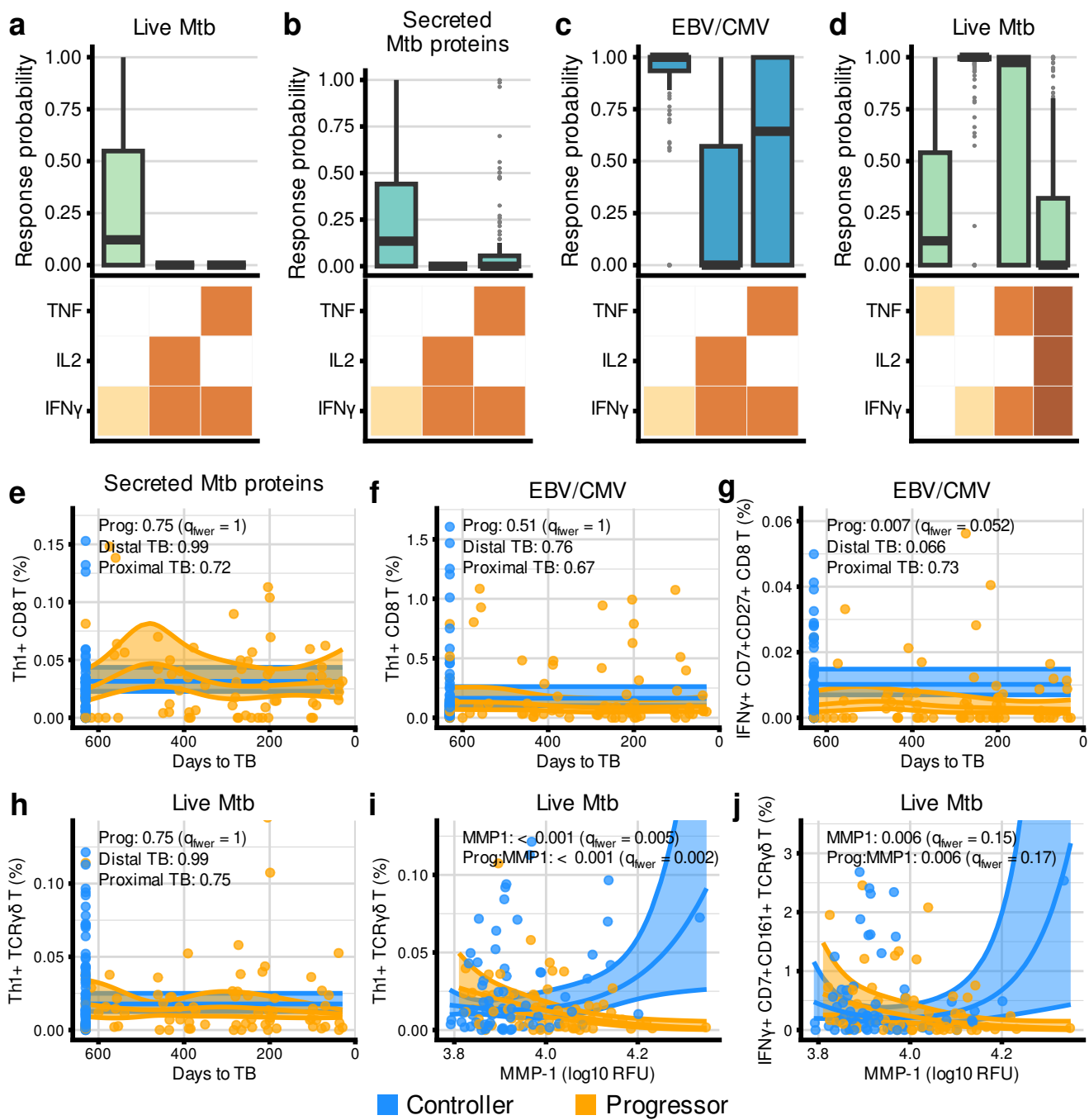




■ Controller ■ Progressor







Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MtbPeptidePools.xlsx](#)
- [RozotRodoetalNatcomformatSupplementaryInformation.docx](#)