nature microbiology

Article

https://doi.org/10.1038/s41564-024-01678-x

Age and sex influence antibody profiles associated with tuberculosis progression

Received: 12 October 2023

Accepted: 20 March 2024

Published online: 24 April 2024

Check for updates

Leela R. L. Davies[1,2,17], Chuangqi Wang[3,4,17], Pia Steigler[5,6], Kathryn A. Bowman[1,2], Stephanie Fischinger[1], Mark Hatherill[5], Michelle Fisher[5], Stanley Kimbung Mbandi[5], Miguel Rodo[5], Tom H. M. Ottenhoff[7], Hazel M. Dockrell[8], Jayne S. Sutherland[8], Harriet Mayanja-Kizza[9], W. Henry Boom[10], Gerhard Walzl[11], Stefan H. E. Kaufmann[12,13,14], Elisa Nemes[5], Thomas J. Scriba[5], Douglas Lauffenburger[4], Galit Alter[1,15] ✉ & Sarah M. Fortune[16] ✉

Antibody features vary with tuberculosis (TB) disease state. Whether clinical variables, such as age or sex, influence associations between *Mycobacterium tuberculosis*-specific antibody responses and disease state is not well explored. Here we profiled *Mycobacterium tuberculosis*-specific antibody responses in 140 TB-exposed South African individuals from the Adolescent Cohort Study. We identified distinct response features in individuals progressing to active TB from non-progressing, matched controls. A multivariate antibody score differentially associated with progression (SeroScore) identified progressors up to 2 years before TB diagnosis, earlier than that achieved with the RISK6 transcriptional signature of progression. We validated these antibody response features in the Grand Challenges 6–74 cohort. Both the SeroScore and RISK6 correlated better with risk of TB progression in adolescents compared with adults, and in males compared with females. This suggests that age and sex are important, underappreciated modifiers of antibody responses associated with TB progression.

Up to a quarter of the world's population is estimated to have been infected with *Mycobacterium tuberculosis* (*Mtb*), the aetiologic agent of tuberculosis (TB), and one of the deadliest global pathogens[1]. Only 5–10% of individuals with immune sensitization to *Mtb* consistent with infection ever develop the morbidity, mortality and transmission risks associated with active TB[2]. Identifying individuals at risk of progression to active disease is a cornerstone of the World Health Organization's End TB Strategy[3]. However, the immunology of TB progression remains incompletely understood, and current diagnostics—the tuberculin skin test (TST) and interferon gamma release assay (IGRA)—correlate poorly with progression[4–9]. Better understanding of the range of immune phenotypes of individuals destined to progress to active TB is urgently needed.

Recent proteomic, metabolomic and transcriptomic studies have defined markers of progression to incident TB disease[10–16]. Multi-gene transcriptomic signatures robustly identify individuals at greatest risk of TB progression[12,14,15,17], collectively pointing to an expanded inflammatory response marked by complement activation and type I and II interferon (IFN) signalling that increases approaching the time of TB disease manifestation[18]. IFN has been implicated in *Mtb* pathogenesis both in mouse models and in human patients with TB[19–21], and it has been proposed that the identified inflammatory transcriptomic signatures reflect early or subclinical TB disease[14,18]. However, these signatures are not pathogen specific and may be influenced by other inflammatory states[22,23]. A recent study demonstrated the existence of *Mtb*-specific adaptive immune correlates of progression in distinct antigen-specific T cell receptor (TCR) repertoires that accumulate in progressors versus non-progressors[24].

We and others have demonstrated that *Mtb*-specific antibody signatures vary across TB clinical states[25–33]. These studies have

A full list of affiliations appears at the end of the paper. ✉e-mail: galit.alter@modernatx.com; sfortune@hsph.harvard.edu

 https://doi.org/10.1038/s41564-024-01678-x

revealed unrecognized immune phenotypes, including the presence of *Mtb*-specific humoral responses in the so-called resisters, who are highly exposed to *Mtb* but do not mount the sustained IFN-dominated T cell response that underlies the diagnostic TST and IGRA tests[34,35]. The effects of human immunodeficiency virus (HIV) on humoral immune responses have been assessed[25,30]. However, the impact of other clinical variables such as age and sex on antibody responses in *Mtb*-infected people, or indeed on other TB immune phenotypes, is largely unexplored.

In this Article, we used a systems serology approach to investigate the association of *Mtb*-specific antibody responses with TB disease progression in a well-characterized longitudinal cohort of HIV-negative, IGRA-positive or TST-positive participants, the Adolescent Cohort Study (ACS)[14]. We identified changes in *Mtb*-specific antibody profiles in progressors compared with non-progressor controls, up to 2 years before the diagnosis of active TB, and assessed the reproducibility of these features in a second cohort, Grand Challenges 6–74 (GC6)[12]. These data indicate that progressors have distinct *Mtb*-specific antibody profiles as compared with non-progressors and that age and sex are critical modifiers of the immune phenotypes that define TB progression.

Results

Progressors have distinct *Mtb*-specific antibody responses

Changes in *Mtb*-specific antibody levels, isotype selection and Fc glycosylation are sensitive biomarkers of TB disease states[25–29,34,35]. We therefore sought to determine the extent to which *Mtb*-specific antibody profiles, which can distinguish active TB from latent infection[26,30], pre-date the diagnosis of active TB. We comprehensively profiled the *Mtb*-specific humoral immune response in longitudinal serum samples collected from 36 adolescent progressors before the diagnosis of active TB and 104 matched non-progressors from the ACS cohort[14] (Fig. 1a and Supplementary Table 1). Averaged across all timepoints, overall levels of *Mtb*-specific IgG, the dominant isotype in the blood, trended higher in progressors compared with non-progressors (Fig. 1b). Lipoarabinomannan (LAM)-specific IgG, IgG1, IgG4, IgA2 and IgM were all significantly increased in progressors compared with non-progressors, as were purified protein derivative (PPD)-specific IgG, IgG2 and IgA1, and culture filtrate protein 10 (CFP10)-specific IgG2. *Mtb*-specific Fc receptor binding antibodies were also increased in progressors, including LAM-specific FcγR2A and FcγR2B, PPD-specific FcγR2A, FcγR2B and FcγR3B, heat shock protein X (HspX)-specific FcγR2A, FcγR2B, FcγR3A and FcγR3B, and 1-tuberculosinyl adenosine 1 (TbAd)-specific FcγR3A. LAM- and antigen 85 complex (Ag85)-specific Fc binding of lectins *Sambucus nigra* agglutinin (SNA) and *Ricinus communis* agglutinin I (RCA) were selectively increased in non-progressors, indicating elevated Fc sialylation and galactosylation; decreased Fc sialic acid and galactose have been associated with the inflammatory humoral profile of active TB[27]. We assessed antibody-mediated effector functions, but no differences between progressors and non-progressors were identified in antibody-dependent cellular or neutrophil phagocytosis across several *Mtb*-specific antigens.

We next used a mixed linear model to rank and identify the most differential antibody features between progressors and non-progressors, controlling for the effects of demographic confounders and study timepoints (Fig. 1c). The model identified antibody responses targeting LAM as increased in progressors compared with non-progressors, including LAM-specific total IgG, IgG1, FcγR2A and FcγR2B binding levels. PPD-specific total IgG and IgA1 were also selectively enriched in progressors, albeit to a lesser extent, as were Ag85-specific total IgG, IgG1 and FcγR2A, and TbAd-specific FcγR3A binding. By contrast, no measured *Mtb*-specific antibody features were significantly enriched in non-progressors.

Mtb-specific antibody responses vary longitudinally

Prior analyses of the ACS cohort found that blood transcriptional signatures of progression increased closer to the time of TB diagnosis[12,15,17,18], whereas the frequency of TCR specificities enriched in progressors remained relatively stable across the study period[24]. We therefore examined the longitudinal evolution of antibody features that were significantly increased in progressors as compared with non-progressors (Fig. 1d–g and Extended Data Fig. 1). The temporal and inter-individual variability of each measured feature was high. In progressors, some antibody features, such as LAM-specific IgG and PPD-specific IgA1, increased at timepoints proximal to TB diagnosis. However, others, such as LAM-specific FcγR2A binding and TbAd-specific FcγR3A binding, remained stably elevated in progressors over the duration of follow-up. Thus, while some measured antibody features increased approaching the time of clinical diagnosis, like the transcriptional signatures, others appeared to be longitudinally stable up to 2 years before TB diagnosis, as observed for TCR specificities.

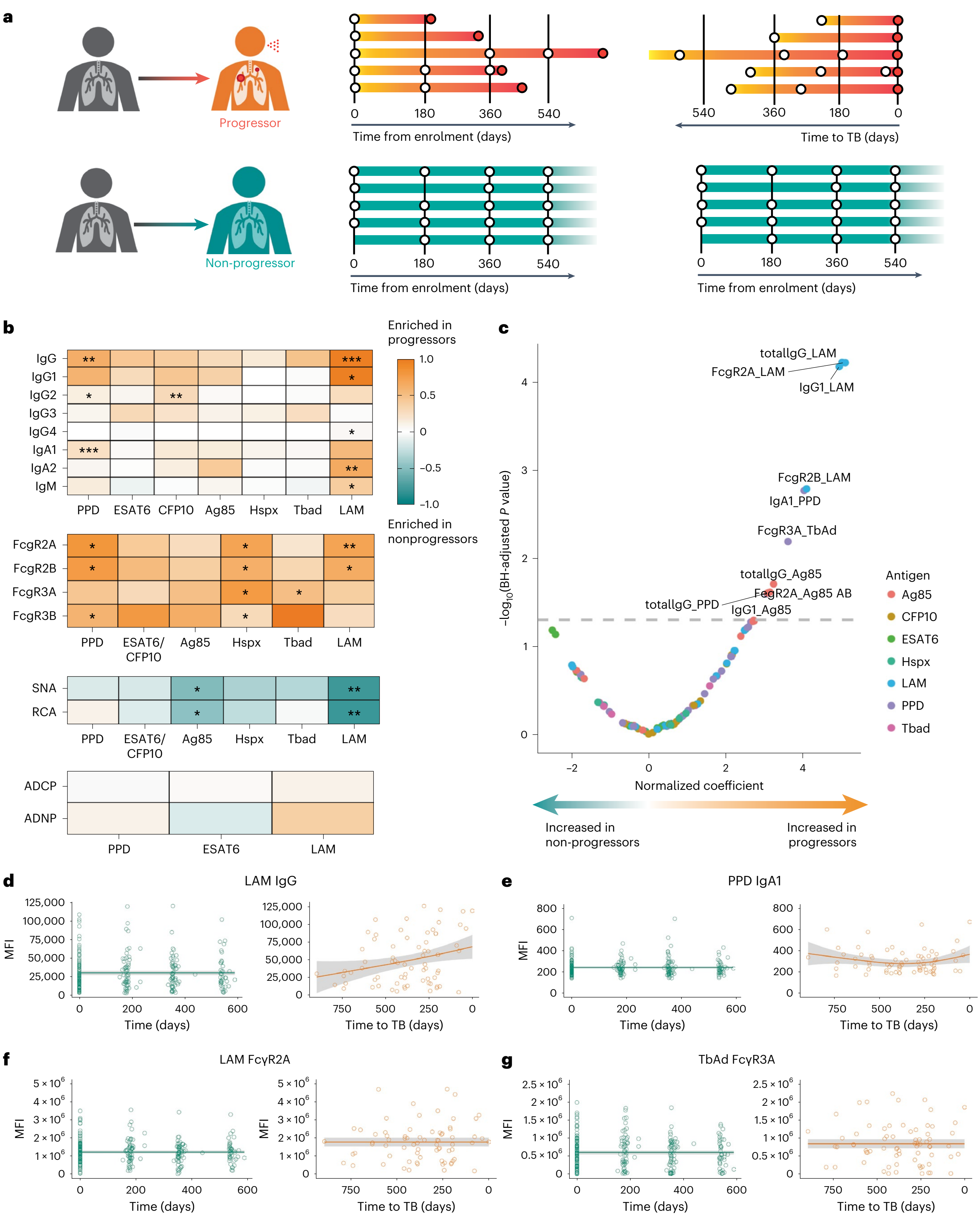
An *Mtb*-specific SeroScore detects TB progression risk

A parsimonious transcriptomic signature was previously developed, RISK6, which differentiates progressors from non-progressors in the ACS cohort[15]. RISK6 signature scores were determined by a pair-ratio approach using three transcripts upregulated in progressors (SERPING1, GBP2 and FCGR1B) and three downregulated in progressors (TRMT2A, SDR39U1 and TUBGCP6). We next asked whether the measured differential antibody features captured the same or distinct immunological processes as those marked by RISK6, by measuring correlations between *Mtb*-specific antibody features and RISK6 scores and transcript expression levels. The two sets of RISK6 transcripts showed the expected correlations with each other, consistent with how they were identified (Fig. 2a). However, the measured antibody features showed minimal correlation with RISK6 and its individual component transcripts (Fig. 2a), suggesting that *Mtb*-specific antibody profiles capture distinct biologic processes.

We next defined a minimal set of antibody features associated with progression. We used the least absolute shrinkage and selection

Fig. 1 | ACS progressors exhibit distinct *Mtb*-specific antibody profiles. **a**, Serum collected longitudinally from a cohort of South African adolescents who later progressed to active TB disease ($n = 36$) or who maintained asymptomatic infection ($n = 104$). For analyses in the current study, progressors were aligned by time of diagnosis and non-progressors by time of enrolment. **b**, Systems serologic assays performed against a panel of *Mtb* antigens, including the selection of antibody isotype and subclasses, the binding of Fcγ receptors, the binding to Fc of lectins SNA (recognizes sialic acid) and RCA (recognizes galactose), and the ability to recruit antibody-mediated cellular phagocytosis (ADCP) and neutrophil phagocytosis (ADNP). For each indicated assay, values for each individual were averaged over time. Each heatmap represents $\log_2$(median value in progressors/median value in non-progressors). The statistical significance of the differences between progressors and non-progressors was measured by two-sided Mann–Whitney test followed by Benjamini–Hochberg (BH) correction for multiple comparisons. *$P < 0.05$; **$P < 0.01$. **c**, Mixed-effects linear modelling to evaluate the association between antibody features and progressor status by controlling age, sex, ethnicity, school code and time of sample collection. Likelihood ratio test was used to compare the two paired models, and P values were corrected for multiple comparisons by the BH method. The x axis indicates the effect size as a normalized coefficient of the variable of progression, and the y axis $-\log_{10}$ of the adjusted P values. The dotted line represents the corrected P value of 0.05. **d**, Raw values of LAM-specific IgG measurements for all individuals plotted over time from enrolment (non-progressors, teal) or time to TB (progressors, orange). The solid lines indicate a smooth of median values, using a generalized additive model, and the grey shading indicates one standard deviation. **e**–**g**, Plots for PPD-specific IgA1 (**e**), LAM-specific antibody binding of FcγR2A (**f**) and TbAd-specific antibody binding of FcγR3A (**g**).

https://doi.org/10.1038/s41564-024-01678-x



https://doi.org/10.1038/s41564-024-01678-x

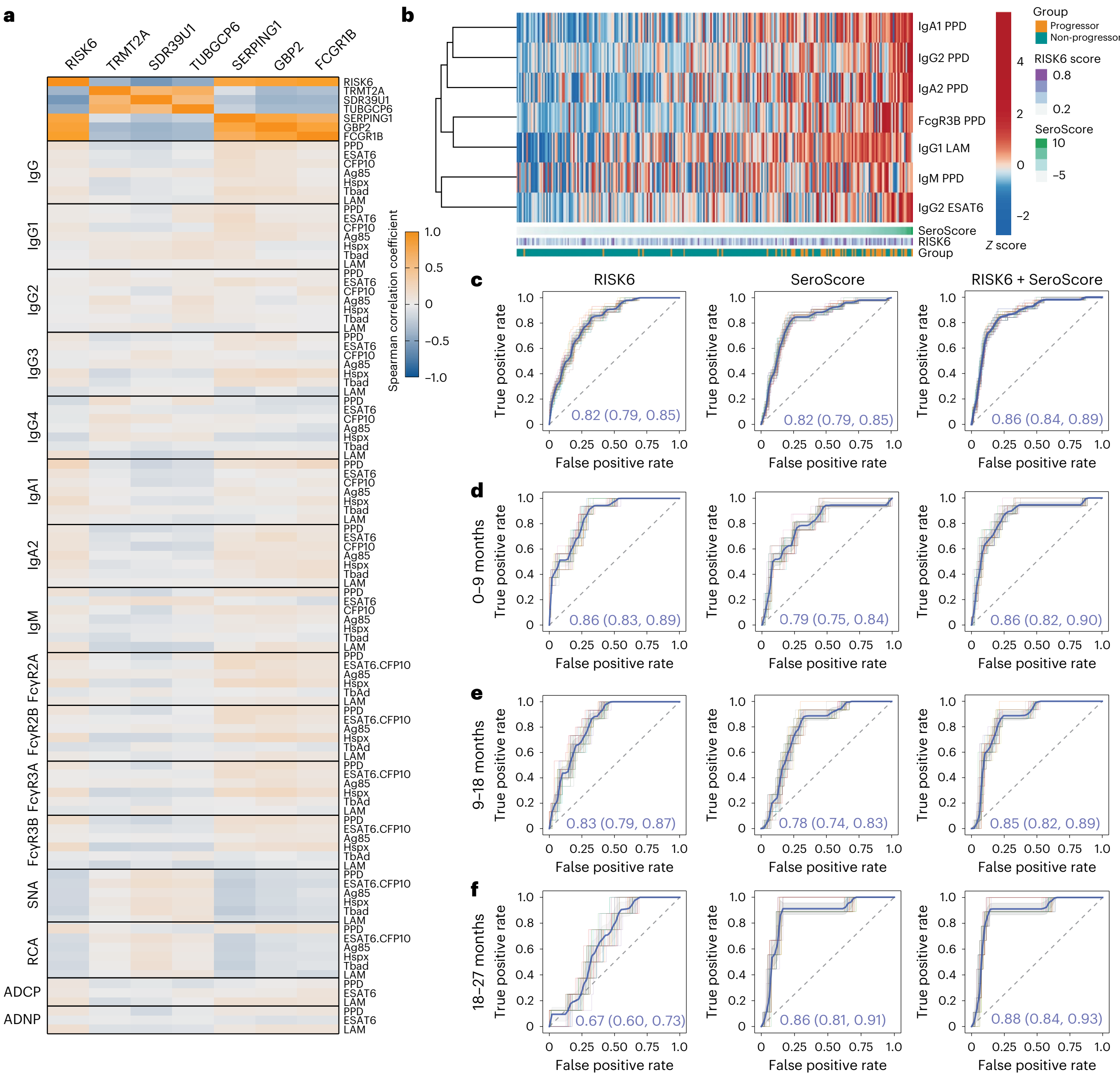


Fig. 2 | An *Mtb*-specific SeroScore differentiates progressors from non-progressors. a, For all individuals at all timepoints, Spearman correlations were calculated between all measured *Mtb*-specific antibody features and RISK6 score ($n = 377$ measurements) or transcript expression levels of each of its six components ($n = 312$ measurements). The heatmap indicates Spearman correlation coefficient for each comparison. **b**, A multivariate SeroScore was developed on the basis of systems serology data in ACS. The heatmap represents *Z*-scored data for the six features included in the SeroScore. Each column represents one individual ($n = 36$ progressors and $n = 104$ non-progressors). Individuals are sorted by overall SeroScore as shown in the track beneath the heatmap. RISK6 score and progressor/non-progressor status of each individual are also indicated in tracks. **c**, ROC curves developed assessing the ability to differentiate progressors ($n = 29$) from non-progressors ($n = 99$) of RISK6 (left), SeroScore (middle) and both RISK6 and SeroScore in combination (right). ROC curves were generated 50 times using randomly selected 80% of samples with group stratification. The mean curve is indicated in blue, with grey shading indicating one standard deviation. The mean AUC with 95% confidence interval is indicated. **d**–**f**, Additional ROC curves developed only including progressors in time windows 0–9 months before diagnosis ($n = 19$ progressors) (**d**), 9–18 months before diagnosis ($n = 18$ progressors) (**e**) and 18–27 months before diagnosis ($n = 10$ progressors) (**f**).

operator (LASSO) technique to identify the most relevant features differentiating progressor and non-progressor groups and identified the combination of features with the highest discriminative ability (Extended Data Fig. 2a). This analysis generated a minimal set of seven features (PPD-specific IgG2, IgA1, IgA2, IgM, FcγR3B, LAM-specific IgG1 and early secretory antigen 6 (ESAT6)-specific IgG2) that enabled resolution of progressors and non-progressors in ACS (Fig. 2b). We used these seven features to define a SeroScore, a multivariate score differentially associated with progression.

To further understand whether the SeroScore and RISK6 captured similar or different biologic processes, we compared their ability to differentiate progressors and non-progressors over time. Across the full study period, RISK6 and SeroScore similarly differentiated progressors, each with an area under the curve (AUC) of 0.82 independently, which improved to 0.86 in combination (Fig. 2c). At 0–9 months, the time window most proximal to diagnosis of active TB, the SeroScore had a median AUC of 0.79, compared with 0.86 for RISK6; combining the two scores in this timeframe did not improve performance beyond that of RISK6 alone (Fig. 2d). At 9–18 months (Fig. 2e) and 18–27 months (Fig. 2f), RISK6 performance declined to AUC of 0.83 and 0.67, respectively, but the SeroScore remained stable with AUC of 0.78 and 0.86. These findings suggest that, particularly at timepoints more remote from diagnosis, the SeroScore captured biologic information that RISK6 did not.

Sex influences scores associated with progression

Sex is a well-established modifier of immune responses[36–38], and the global prevalence of active TB in males exceeds that in females with a ratio of 1.7 (ref. 39). We sought to determine whether sex influenced the association of the SeroScore or the RISK6 signature with progression in the ACS cohort, which is 67.1% female. Across the full study period, both the SeroScore and RISK6 identified male progressors slightly better than females (Fig. 3a,b). We then combined the SeroScore and RISK6 linearly to determine whether they had better ability to identify progressors in combination. The combined score was better able to identify progressors in both male and female subgroups, but the mean AUC among female participants, at 0.84, remained lower than among male participants, where it reached 0.94 (Fig. 3c). When we directly plotted SeroScore by sex and progressor status, male progressors trended towards higher SeroScores and RISK6 scores than female progressors, while both scores were similar in male and female non-progressors (Fig. 3d). Thus, despite being discovered in a predominantly female cohort, both the SeroScore and RISK6 signatures captured progression better among males than females.

The ACS SeroScore detects TB progression among adolescents

We next sought to assess the performance of the ACS SeroScore in a second cohort of 39 progressors and 169 non-progressors from the South African subcohort of GC6, a longitudinal study of TB household contacts[12]. Prior work demonstrated that the ACS-defined RISK6 signature had reduced performance in the GC6 cohort, which was attributed to differences in study design, environmental or temporal exposure to TB, geography of the participants and wider age range (Supplementary Table 2 and Extended Data Fig. 3)[15].

When applied to the GC6 cohort, the ACS-derived SeroScore identified progressors marginally, with an overall mean AUC of 0.60 (Fig. 4a). To more closely align the demographic features of GC6 with ACS, we stratified the GC6 subjects into adolescents, who were 8–20 years of age at enrolment (14 progressors and 63 non-progressors), and adults, who were 21–60 years of age at enrolment (25 progressors and 106 non-progressors). The AUC of the SeroScore differed with age, with a mean value of 0.72 among adolescents and only 0.53 among adults (Fig. 4b). Interestingly, among individuals with an available RISK6 score, the RISK6 test performance also improved among the adolescent GC6 group (Extended Data Fig. 4). As in the ACS cohort, the ability of the ACS-derived SeroScore to identify progressors remained longitudinally stable in GC6 up to 18 months before diagnosis, but the small sample size limited evaluation at timepoints earlier than 18 months (Extended Data Fig. 5).

Given the increased heterogeneity of the GC6 cohort, we next aimed to independently determine whether distinct *Mtb*-specific antibody features associated with progression existed in GC6. Humoral differences between progressors and non-progressors were more subtle in the GC6 cohort compared with ACS (Fig. 4c). Among adolescents, several *Mtb*-specific antibody responses were selectively enriched among progressors, including markedly increased LAM-specific IgG1, IgG2 and IgG3 levels and binding of FcγR2A, FcγR2B and FcγR3A, though, given the small sample size, none achieved statistical significance after multiple test correction. As observed in ACS, Fc sialylation and galactosylation trended towards an enrichment in non-progressors (Fig. 4d). In a mixed linear model, IgG and IgA1 responses to PPD, LAM and Ag85, and FcγR binding to these antigen-specific antibodies, were significantly enriched in progressors (Fig. 4e). Thus, while the differences were smaller in magnitude, the specific discriminatory *Mtb*-specific antibody profiles observed across progressors and non-progressors in GC6 resembled features identified in the ACS cohort (Fig. 1).

Class-switched *Mtb* antibodies emerge with increasing age

We postulated that the reduced association of antibody features with progression among GC6 adults might be due to chronic *Mtb* exposure increasing background *Mtb*-specific antibody levels among non-progressors. We measured the relationship between antibody features and the age at the time of enrolment in the non-progressors in GC6 (Fig. 4f). We found that class-switched antibody responses, primarily IgG2, IgG3 and IgA1, positively correlated with age at enrolment. Conversely, IgM levels and Fc sialylation and galactosylation, represented by SNA and RCA, demonstrated negative correlations with age. Together, these data demonstrate that, with increasing age, individuals without known active TB disease exhibit broad class-switching of *Mtb*-specific antibody responses and decreased Fc sialylation and galactosylation, antibody features suggestive of increased inflammatory tone[40].

The GC6 SeroScore detects progressors across cohorts

Finally, we hypothesized that profiles of progression identified in a more heterogeneous population would be more likely to yield an epidemiologically concordant score of TB progression. We therefore defined an independent SeroScore in GC6 (Extended Data Fig. 2b) with markers that were disproportionately enriched among progressors (Fig. 5a). The LASSO algorithm is designed to avoid selecting multiple co-correlated variables to limit the risk of model overfitting, and thus may select different features for different datasets, even when the underlying architecture is very similar. Nevertheless, the GC6-derived SeroScore included some of the same features as the ACS-derived SeroScore, including PPD-specific IgA1 and IgG2. Similarly, LAM- and ESAT6-specific total IgG were included in the GC6 SeroScore, whereas subclasses LAM-specific IgG1 and ESAT6-specific IgG2 had been included in the ACS SeroScore.

The GC6-derived SeroScore differentiated progressors from non-progressors in GC6 at all tested time windows, with a longitudinally stable mean AUC of 0.72–0.73 at 0–9 months, 9–18 months and 18–27 months, and over all time windows (Fig. 5b). Similar to the ACS SeroScore, the GC6-derived SeroScore identified progressors better among adolescents than adults, with mean AUC values of 0.82 and 0.69, respectively (Fig. 5c). The G6C SeroScore was able to discriminate progressors in ACS, with an AUC of 0.66 overall (Fig. 5d). It also performed better in males, with an AUC of 0.83 as compared with females with an AUC of 0.63. These findings highlight the presence of similar humoral features differentiating progressors from non-progressors in both the ACS and GC6 cohorts, including enhanced expansion of class-switched *Mtb*-specific IgA and IgG2, and higher levels of LAM-specific antibodies in progressors. We also find that age and sex modify the relationship between these *Mtb*-specific humoral features and TB progression.

Discussion

An explosion of novel profiling tools has begun to define correlates of progression in individuals who go on to develop TB disease[12–18,24]. Here, we investigated the association of *Mtb*-specific humoral profiles with TB progression in a well-characterized cohort of HIV-negative, IGRA-positive South African adolescents[14]. A multivariate *Mtb*-specific SeroScore associated with progression was longitudinally stable for the study duration, up to 2 years before TB diagnosis. The SeroScore

 https://doi.org/10.1038/s41564-024-01678-x

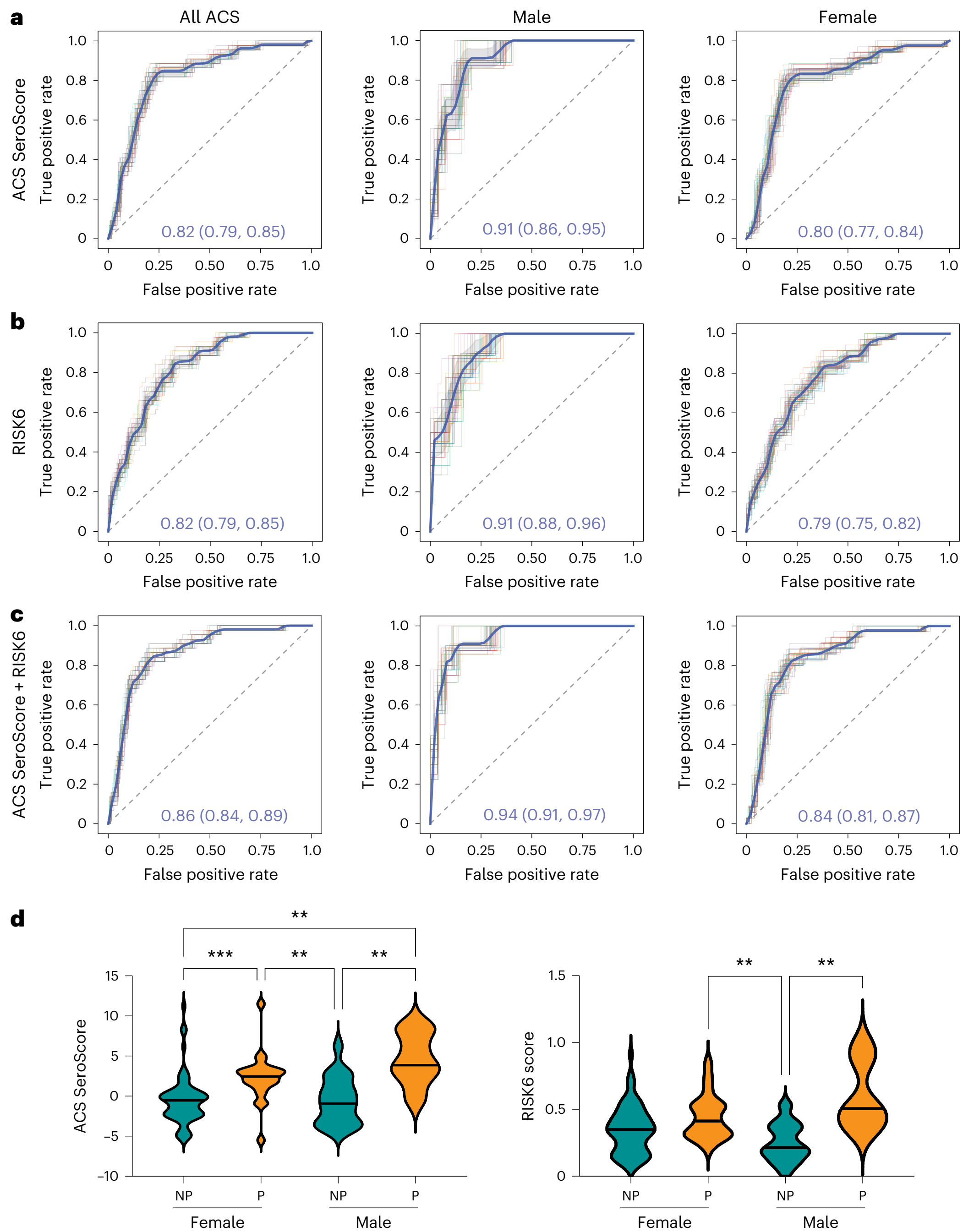


Fig. 3 | Sex modulates the association of SeroScore and RISK6 with TB progression. ROC curves were developed measuring the ability of the ACS-derived SeroScore and RISK6 to differentiate progressors from non-progressors in ACS. **a**, The identification of progressors by the ACS-derived SeroScore among all ACS individuals ($n = 29$ progressors and $n = 99$ non-progressors), males only ($n = 7$ progressors and $n = 37$ non-progressors) and females only ($n = 22$ progressors and $n = 63$ non-progressors). **b**, As in **a**, the performance of RISK6 among all ACS, males only and females only. **c**, The identification of progressors by the ACS-derived SeroScore and RISK6 in combination among all ACS, males only and females only. For **a**–**c**, the mean of 50 curves is shown in blue, with grey shading indicating one standard deviation. The mean AUC with 95% confidence interval is indicated on each plot. **d**, The ACS SeroScore and RISK6 signature score were plotted for female ($n = 22$) and male ($n = 7$) progressors (P, orange) and female ($n = 63$) and male ($n = 37$) non-progressors (NP, teal) from ACS. The groups were compared by Kruskal–Wallis test, with P values <0.05 indicated.

was marked by elevated levels of *Mtb*-specific antibodies including class-switched IgG and IgA isotypes, Fc receptor-binding antibodies, and reduced Fc sialylation and galactosylation. The ACS SeroScore was also associated with risk of progression in the adolescent participants from a more epidemiologically diverse cohort of household contacts, GC6, but it poorly identified progression among adults. In addition, both the ACS SeroScore and a SeroScore derived from GC6 participants better differentiated male as compared with female progressors. Surprisingly, we also found that the performance of the well-studied RISK6 transcriptional signature was modulated by age and sex. Thus, age and sex are critical modifiers of these antibody responses and other immune phenotypes associated with TB progression.

The ability of both SeroScores and RISK6 to detect progressors better among adolescents that adults suggests that adolescent progressors harbour unique immune phenotypes. The observed age-associated increases in class-switched IgA and IgG to multiple

 https://doi.org/10.1038/s41564-024-01678-x

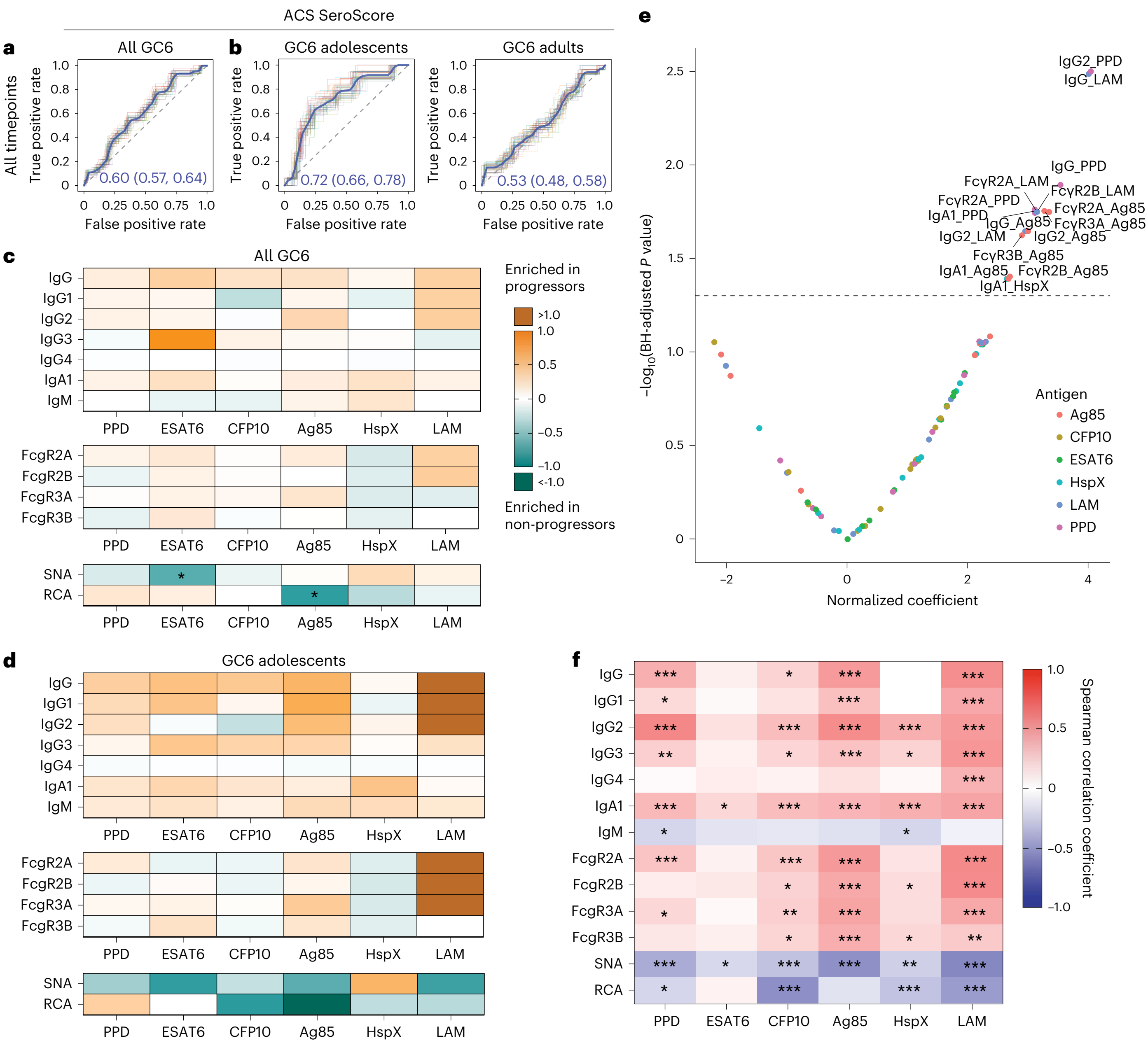


Fig. 4 | *Mtb*-specific antibody profiles correlate with progression in GC6 adolescents. a, ROC curves developed to evaluate the ability of the ACS-derived SeroScore to differentiate progressors ($n = 39$) and non-progressors ($n = 169$) among individuals from the GC6 cohort. The grey shading indicates one standard deviation. The mean AUC is indicated. **b**, ROC curves evaluating the ability of the ACS-derived SeroScore to differentiate progressors from non-progressors among GC6 adolescents (age 8–20 years at enrolment, $n = 14$ progressors and $n = 63$ non-progressors) and adults (age >20 years at enrolment, $n = 25$ progressors and $n = 106$ non-progressors). The mean of 50 curves is shown in blue, with grey shading indicating one standard deviation. The mean AUC with 95% confidence interval is indicated. **c**, *Mtb*-specific systems serology used to profile serum collected longitudinally from the GC6 cohort. For each indicated assay, values for each individual were averaged over time. In the heatmap, each cell represents $\log_2$(median value in progressors/median value in non-progressors). The statistical significance of the differences between progressors and non-progressors was measured by two-sided Mann–Whitney test followed by Benjamini–Hochberg (BH) correction for multiple comparisons. $^*P < 0.05$ after correction. **d**, A heatmap representing averaged values over time for adolescent individuals only. Each cell represents $\log_2$(median value in progressors/median value in non-progressors). The statistical significance of the differences between progressors and non-progressors were measured by two-sided Mann–Whitney test followed by BH correction for multiple comparisons. $^*P < 0.05$ after correction. **e**, Among all GC6 individuals, mixed-effects linear modelling was used to evaluate the association between antibody features and progressor status by controlling age, sex and time of sample collection. Likelihood ratio test was used to compare the two paired models, and P values were corrected for multiple comparisons by the BH method. The x axis indicates effect size as normalized coefficient of the variable of progression, and the y axis $-\log_{10}$ of the adjusted P values. The dotted line represents the corrected P value of 0.05. **f**, A heatmap representing Spearman correlation coefficients between each antibody feature and age among GC6 non-progressors only ($n = 169$). P values for each correlation were adjusted for multiple comparisons by the BH method, and adjusted P values are indicated: $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

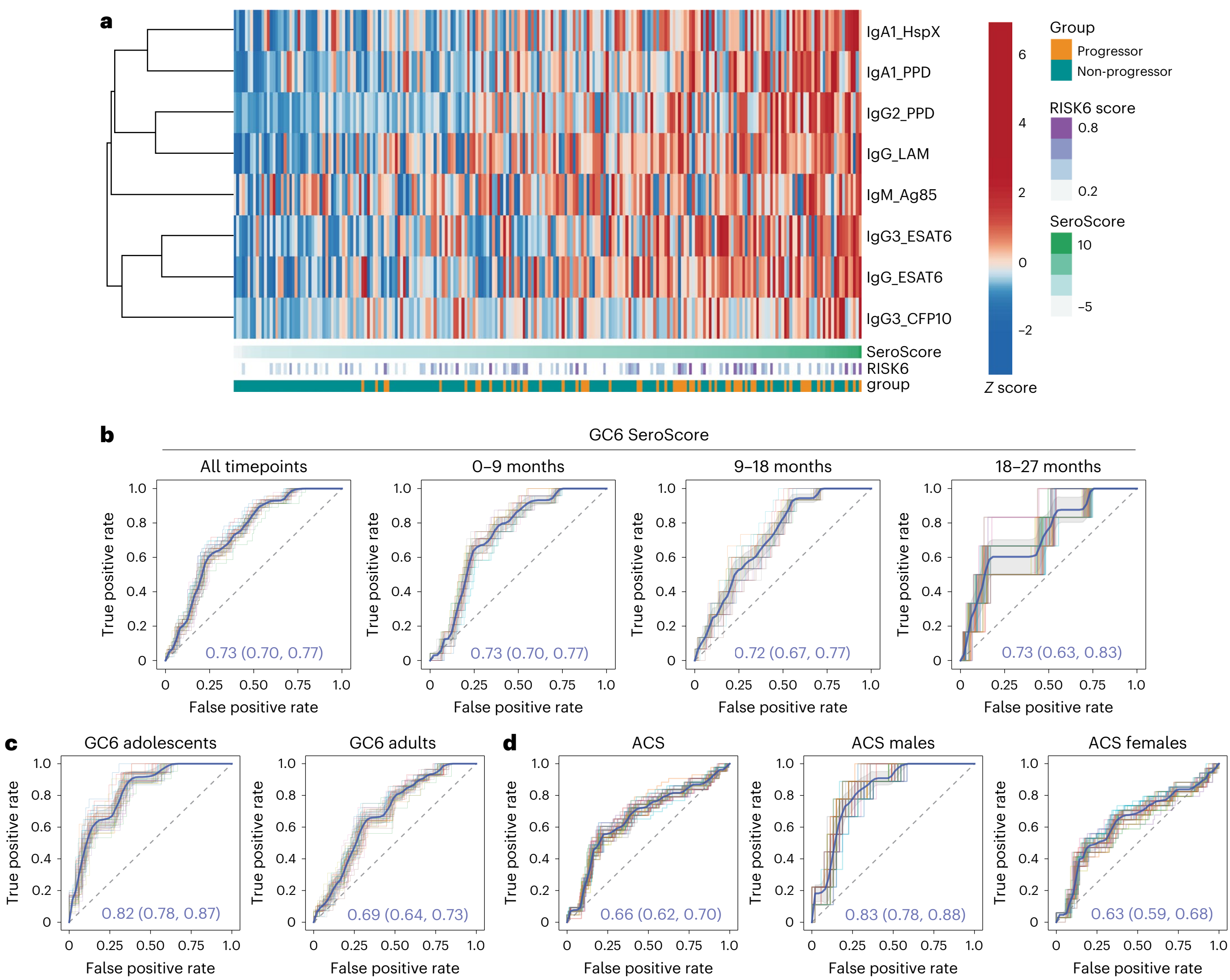


Fig. 5 | A GC6-derived SeroScore detects humoral correlates of progression. **a**, A multivariate SeroScore was developed in GC6. The heatmap represents *Z*-scored data for the six features included in the SeroScore. Each column represents one individual (n = 39 progressors and n = 169 non-progressors). The individuals are sorted by overall SeroScore as shown in the track beneath the heatmap. The RISK6 score and progressor/non-progressor status of each individual are also indicated in tracks. **b**, To evaluate the ability of the GC6-derived SeroScore to identify progressors in the same cohort, ROC curves were developed over the total study period (n = 39 progressors, n = 169 non-progressors) and for progressors in time windows 0–9 months (n = 30 progressors), 9–18 months (n = 19 progressors) and 18–27 months (n = 8 progressors) before the diagnosis of active TB. The mean of 50 curves is shown in blue, with grey shading indicating one standard deviation. The mean AUC with 95% confidence interval is indicated. **c**, ROC curves measure the ability of the GC6-derived SeroScore to identify GC6 progressors in an age-stratified analysis of adolescents (n = 14 progressors and n = 63 non-progressors) and adults (n = 25 progressors and n = 106 non-progressors). **d**, The ROC curves measure the ability of the GC6-derived SeroScore to identify progressors in the full ACS cohort (n = 29 progressors and n = 99 non-progressors), males only (n = 7 progressors and n = 37 non-progressors) and females only (n = 22 progressors and n = 63 non-progressors).

antigens, including the relatively *Mtb*-specific antigen CFP10, suggest that adults have experienced prior exposures to *Mtb*. Alternatively, the observed differences may reflect adolescents' relatively more recent *Mtb* exposure[41,42] or age-specific differences in B and T helper cell responses[43].

Similarly, both transcriptomic and serologic scores were less able to identify progressors among females. Sex has been linked to extensive differences in both innate and adaptive immunity, including altered complement activity, higher IFN levels and increased antibody responses in females[36–38] that may affect transcriptomic and serologic signatures. Additional epidemiologic factors that may affect *Mtb*-specific humoral responses, such as co-morbidities including HIV co-infection, geography and race[44–46], were not evaluated here. While the focus of many clinical correlate studies is the similarity of biomarkers across cohorts, understanding how immune correlates vary across epidemiologically distinct groups may elucidate population-specific determinants of disease risk.

The longitudinal stability of several of the observed serologic responses in progressors is reminiscent of the longitudinal stability of the TCR specificities elevated in the ACS progressors[24]. These longstanding adaptive immune phenotypes suggest that progressor status is determined very early on. It is possible that these responses reflect a higher or more persistent lung mycobacterial burden in progressors before development of symptoms; however, it is also possible that they mark early qualitative differences between a protective and non-protective immune response. Other antibody features, including

LAM-specific IgG2 and PPD-specific IgA1, mirror the increase of inflammatory transcriptomic signatures towards the onset symptomatic TB[13–15,18], suggesting they may be more sensitive markers of B cell surveillance of increasingly inflamed lung tissue. Future longitudinal studies will be essential in detailing the temporal interplay of *Mtb* culture positivity, lung pathology and symptom onset with signatures of progression.

Although antibody correlates of progression have not previously been examined, prior studies of individuals with active TB have identified heterogeneous, broadly increased *Mtb*-specific IgG levels compared with asymptomatic controls, correlating with their bacterial burden[47–50]. Systems serology approaches have shown that active TB is defined by increased levels of IgG across many *Mtb* antigens, higher levels of FcR binding and more inflammatory, less sialylated and galactosylated Fc glycoforms[25–27,30,40]. The antibody features differentially associated with progressors here resemble those differentially enriched in active TB. It is possible that the emergence of these antibody features in progressors long before the diagnosis of active TB reflects smouldering, tissue-level disease. It remains unclear whether the differences in *Mtb*-specific antibodies that distinguish progressors and non-progressors play functional roles in control of infection, and this is an active area of ongoing investigation.

In sum, we have shown here that distinct antibody features are associated with TB disease progression in South African adolescents in an age- and sex-dependent fashion. We find evidence for early *Mtb*-specific humoral immune responses that may mark longstanding subclinical disease, and other antibody features that increase closer to the time of TB diagnosis. The observed alterations in humoral responses among progressors provide critical insights into the immunology of TB progression and emphasize the importance of clinical variables in modulating immune phenotypes of TB disease states.

Methods

Study design

ACS. The ACS was a prospective cohort study that enrolled 6,363 healthy, HIV-negative South African adolescents aged 12–18 years[14]. Among participants with evidence of *Mtb* infection, diagnosed by a positive QuantiFERON TB Gold in-tube assay or positive TST, progressors were those who developed active intrathoracic TB during the follow-up period, defined by either two sputum smears positive for acid-fast bacilli or one positive sputum culture with microbiologically confirmed *Mtb*. For each progressor, at least two QuantiFERON-positive non-progressors were matched by age at enrolment, sex, ethnic origin, school of attendance and presence or absence of previous episodes of TB disease. Adolescents provided written, informed assent, and parents or legal guardians provided written, informed consent. The original clinical study was reviewed and approved by the Human Research Ethics Committee of the University of Cape Town, and the systems serology analysis by Massachusetts General Hospital.

All samples from progressors with at least 200 ml of available serum were included in the current study (72 samples from 36 individuals). In the absence of prior data on serologic responses in TB disease progression, a formal power calculation was not performed. Sample size was dictated by the number of available progressors in the original cohort and was similar to prior systems serology studies in the setting of HIV or different TB disease states[25,26,30]. At least two non-progressors were matched to each progressor individual by age, sex, school code and ethnicity (264 samples from 104 individuals). Demographics of the included individuals from ACS are presented in Supplementary Table 1.

GC6. Samples were additionally included from participants in the GC6 (refs. 12,16). Briefly, HIV-negative people aged 10–60 years who had household exposure to an adult with sputum smear-positive TB were enrolled to this study and followed for 2 years. In total, 85.7% of South African subjects were TST positive at baseline[12]. Progressors had intrathoracic TB, defined by sputum culture, smear microscopy and clinical signs. For each progressor, four controls were matched according to recruitment region, age category (≤18 years, 19–25 years, 26–35 years or ≥36 years), sex and year of enrolment. Study protocols were approved by the relevant human research ethics committees. Written informed consent was obtained from participants. For adolescents, consent was obtained from parents or legal guardians of adolescents and written informed assent from each adolescent.

Samples from progressors (114 samples from 39 individuals) and matched non-progressors (458 samples from 169 individuals) were included in the current study. For age-stratified analyses, adolescents were defined as individuals aged 8–20 years (14 progressors and 63 non-progressors), and adults as those older than 20 years (25 progressors and 106 non-progressors). While the original study included South Africa, Gambia, Ethiopia and Uganda, to better control for exposure to similar *Mtb* strains as well as environmental non-tuberculous mycobacteria exposure, only South African participants were included in the current study. Demographics of all included individuals from GC6 are presented in Supplementary Table 2.

Measurement of biophysical properties of *Mtb*-specific serum antibodies

A customized, multiplex Luminex assay was used to measure *Mtb* antigen-specific antibody responses across multiple isotypes and subclasses. Antigens included commercially available *Mtb* products: PPD (Staten Serum Institute), LAM (BEI Resources NR-14848), ESAT6 (BEI Resources NR-49424), CFP10 (BEI Resources NR-49425), Ag85 (BEI Resources NR-14855), HspX (BEI Resources NR-49428) and TbAd (a generous gift from the laboratory of Dr Branch Moody). An equal mixture of influenza antigens from HA1(B/Brisbane/60/2008) and HA1(H1N1) (A/New Caledonia/20/99) (Immune Technology Corp.) was used as a positive control, and recombinant HA-tagged ebolavirus glycoprotein minus the transmembrane domain (EBOV GPdTM, Mayflower Bioscience 0501-001) was used as a negative control. All peptide antigens were coupled to carboxylate-modified microspheres (Luminex Corp.) by covalent *N*-hydroxysuccinimide (NHS)-ester linkages by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (Thermo Fisher Scientific) and Sulfo-NHS (Thermo Fisher Scientific) per the manufacturer's instructions. Glycan antigens (LAM and TbAd) were first modified in 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (Sigma-Aldrich) at 9.25 mg ml^{-1} at room temperature for 1 h, desalted with a PD-10 column and then incubated with Luminex beads with rotation overnight at room temperature.

Assays were optimized over a dilution curve, to ensure selection of a dilution within the linear range of the assays. A 1:200 dilution was selected to maximize the dynamic range across control samples and to capture the AUC for the full range of dilutions tested. Diluted serum samples were incubated with pooled microspheres for 2 h at room temperature, and then washed three times with phosphate-buffered saline (PBS) with 0.1% bovine serum albumin and 0.05% Tween to wash away unbound antibodies. Secondary detection reagents included phycoerythrin-conjugated goat anti-human IgG, IgG1, IgG2, IgG3, IgG4, IgM, IgA1 and IgA2 (Southern Biotech) and fluorescein-conjugated SNA and RCA (VectorLabs). For FcR binding, recombinant human FcγR2A, Fcγ2B, Fcγ3A and Fcγ3B (Duke University Protein Production Core) were biotinylated using BirA (Avidity) and conjugated to streptavidin-PE (Phycolink). All secondary incubations were performed over 1 h at room temperature. The median fluorescence intensity for each bead region was measured using an iQue Plus Screener (Intellicyt). All samples were assayed in duplicate, and values were averaged. SNA and RCA measurements were normalized to the corresponding IgG measurements.

Antibody-dependent cellular phagocytosis

Cellular phagocytosis of fluorescent beads coated with PPD, LAM and ESAT6 was performed[51]. The human cell line THP-1 was used to source monocytes from the assay in a reproducible and high-throughput format. Briefly, antigens were biotinylated with 50-fold excess biotin with EZ-link NHS-long chain biotin (Thermo Fisher) following the manufacturer's instructions, and then adsorbed onto 1 µm fluorescent neutravidin beads (Invitrogen) at a 1:1 (µg:µl) ratio of biotinylated polysaccharide to beads. Ten microlitres of a 1:100 suspension of antigen-coupled beads were added to each well of a 96-well plate along with equal volume of serum diluted 1:30, and plates were incubated for 2 h at 37 °C, and then washed with PBS. A total of 25,000 THP-1 cells (human acute monocytic leukaemia cell line, American Type Culture Collection) were added and incubated at 37 °C for 18–20 h. Cells were fixed with 4% paraformaldehyde before data acquisition. Phagocytosis was measured by iQue Plus Screener (Intellicyt). Phagocytic scores were calculated as (per cent bead-positive cells) × (geometric mean fluorescence intensity (MFI))/10,000. Each sample was assayed in two independent technical replicates and averaged.

Antibody-dependent neutrophil phagocytosis

Neutrophil phagocytosis was evaluated[52]. To optimize the signal-to-noise ratios, phagocytosis was performed using total donor leukocytes, and analysis was performed on the neutrophil subset. Briefly, as described for cellular phagocytosis, PPD, LAM and ESAT6 were biotinylated and coupled to 1 µm fluorescent neutravidin beads. Ten microlitres of a 1:160 dilution of coupled beads in PBS were opsonized with 10 µl of serum diluted 1:30 at 37 °C for 2 h. Whole blood was collected from healthy donors, red blood cells were lysed with ACK lysis buffer (Quality Biological), and primary leukocytes were isolated by centrifugation and washed in PBS. A total of 50,000 isolated leukocytes were added per well and incubated for 1 h at 37 °C. The cells were then stained with 10 mg ml^{-1} Pacific Blue anti-human CD66b antibody (BioLegend) and fixed in 4% paraformaldehyde before measurement and analysis on the iQue Plus Screener (Intellicyt). Neutrophils were then gated on $CD66b^+$, and phagocytic scores were calculated as above. Two healthy leukocyte donors were used as biological replicates for each sample and assayed in parallel, and replicates were averaged.

Statistics

Univariate comparisons of individual assayed antibody features were performed in GraphPad Prism 9. Progressor and non-progressor groups were compared using two-sided Mann–Whitney tests, followed by multiple test correction with the Benjamini–Hochberg method. Male and female progressor and non-progressor groups were compared using two-sided Kruskal–Wallis tests. All remaining data visualizations and analyses were performed in Python version 3.9.16 or R version 4.0.2.

Nested mixed linear model

To evaluate the difference of each individual measurement between the progressor and non-progressor groups by controlling the effects of the potential cofounders including demographic features of age, sex and school (ACS only) and the timepoint of sample collection, we used a nested mixed linear model. In detail, we applied two nested mixed linear models (null and full model) without/with progressor group information to assess the significance of the association between measurements and progressor group while controlling for these potential confounding characteristics. We fit two mixed linear models using the maximum likelihood estimation (MLE) and estimated the improvement in model fit by likelihood ratio test (LRT), which follows a chi-square (λ^2) distribution to identify the associated measurements.

$$\text{Null model}: m_{ij} \sim 1 + \text{Sex}_j + \text{Age}_j + \text{District}_j + \text{VisitDate}_j + (1|\text{ID}_j)$$

$$\text{Full model}: m_{ij} \sim 1 + \text{Sex}_j + \text{Age}_j + \text{District}_j + \text{group}_i + \text{VisitDate}_j + (1|\text{ID}_j)$$

$$\text{Likelihood ratio test}: \text{LRT} = -2 \times \frac{\text{MLE in full model}}{\text{MLE in null model}} \sim \lambda^2$$

Here, 'District', identified by the 'SchoolCode', represents the geographical difference, while 'ID' denotes individual participant. The R package lme4 was used to fit the mixed linear model to each measurement and test for differences in measurements, depending on whether each sample belongs to progressor group or non-progressor group. The P value from the likelihood ratio test was adjusted by multiple testing correction using the Benjamini–Hochberg procedure, and the t value (normalized coefficients) associated with the progressor/non-progressor status, 'Group' in the full model, were visualized in a volcano plot using the ggplot function in R package 'ggplot2' (version 3.3.5).

Definition of SeroScores

SeroScores were defined as multivariate antibody signatures differentially associated with progressors. To define SeroScores, measurements from ACS or GC6 were $\log_2$-transformed to correct the skewness of distribution and then Z-scored. Measurements were averaged if more than one sample was collected from the same patient within the given time window.

Next, LASSO regularization[53] was used to select representative features. In detail, 100 sample sets were generated through random sampling. Each sample set included a randomly selected 80% of all samples with group stratification. For each sample set, LASSO feature selection was performed ten times, and features that occurred at least 80% of the time (that is, at least eight among ten times) were selected. The feature selection process was run on 100 generated datasets in parallel. Selected features were ordered by occurrence and the top K (typically 12) features were selected as the final candidates. K was manually chosen in different experiments. The procedure was implemented in the select_lasso function in the systemsseRology (version 1.1) package in R.

All additive feature combinations from the selected set of K candidates were then evaluated exhaustively. The performance of each combination was evaluated by calculating the mean AUC of receiver operating characteristic (ROC) curves generated using a randomly selected 80% of samples for 50 iterations. Feature combinations with the largest mean AUC value were selected to define the SeroScore.

Measurement of receiver operating characteristics

To evaluate the performance of SeroScores and RISK6, respectively and in combination, we generated ROC curves in various conditions. For each condition, we estimated the mean AUC from 50 runs, where for each run 80% of the samples with group stratification were randomly selected. The 95% confidence interval of the AUC value was estimated as ± two standard deviations from the mean, under the assumption of Gaussian distribution. ROC curve visualization and AUC calculation were implemented on the basis of the functions roc_curve, auc and RocCurveDisplay in the Python package sklearn.metrics (version 1.2.1). The 95% confidence interval was programmed using the function norm.interval in the Python package scipy.stats (version 1.10.1).

Reporting summary

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

Data availability

Full systems serology datasets for ACS and GC6 are publically available[54]. RISK6 scores used in this paper were previously calculated for both the ACS cohort[15] and the GC6 cohort (GEO accession number GSE94438). Source data are provided with this paper.

Code availability

All the code used for analysis in this paper is deposited on Zenodo at https://doi.org/10.5281/zenodo.10637145 (ref. 54) and on Github at https://github.com/ChuangqiWang/TB_Progressor (ref. 55).

References

1. *Global Tuberculosis Report* (World Health Organization, 2021).
2. Comstock, G. W., Livesay, V. T., Woolpert, S. F. & Woolpert, C. G. The prognosis of a positive tuberculin reaction in childhood and adolescence. *Am. J. Epidemiol.* **99**, 131–138 (1974).
3. *The End TB Strategy* (World Health Organization, 2015).
4. Diel, R., Loddenkemper, R. & Nienhaus, A. Predictive value of interferon-γ release assays and tuberculin skin testing for progression from latent TB infection to disease state: a meta-analysis. *Chest* **142**, 63–75 (2012).
5. Lalvani, A., Berrocal-Almanza, L. C. & Halliday, A. Predicting progression to active tuberculosis: a rate-limiting step on the path to elimination. *PLoS Med.* **16**, e1002814 (2019).
6. Saunders, M. J. et al. A score to predict and stratify risk of tuberculosis in adult contacts of tuberculosis index cases: a prospective derivation and external validation cohort study. *Lancet Infect. Dis.* **17**, 1190–1199 (2017).
7. Johnston, J. et al. How well does TSTin3D predict risk of active tuberculosis in the Canadian immigrant population? An external validation study. *Clin. Infect. Dis.* **73**, e3486–e3495 (2021).
8. Andrews, J. R. et al. The dynamics of QuantiFERON-TB Gold in-tube conversion and reversion in a cohort of South African adolescents. *Am. J. Respir. Crit. Care Med.* **191**, 584–591 (2015).
9. Metcalfe, J. Z. et al. Evaluation of quantitative IFN-gamma response for risk stratification of active tuberculosis suspects. *Am. J. Respir. Crit. Care Med.* **181**, 87–93 (2010).
10. Maertzdorf, J. et al. Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol. Med.* **8**, 86–95 (2016).
11. Sweeney, T. E., Braviak, L., Tato, C. M. & Khatri, P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir. Med.* **4**, 213–224 (2016).
12. Suliman, S. et al. Four-gene pan-African blood signature predicts progression to tuberculosis. *Am. J. Respir. Crit. Care Med.* **197**, 1198–1208 (2018).
13. Penn-Nicholson, A. et al. Discovery and validation of a prognostic proteomic signature for tuberculosis progression: a prospective cohort study. *PLoS Med.* **16**, e1002880 (2019).
14. Zak, D. E. et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* **387**, 2312–2322 (2016).
15. Penn-Nicholson, A. et al. RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. *Sci. Rep.* **10**, 1–21 (2020).
16. Weiner, J. et al. Metabolite changes in blood predict the onset of tuberculosis. *Nat. Commun.* **9**, 5208 (2018).
17. Warsinske, H. C. et al. Assessment of validity of a blood-based 3-gene signature score for progression and diagnosis of tuberculosis, disease severity, and treatment response. *JAMA Netw. Open* **1**, e183779 (2018).
18. Scriba, T. J. et al. Sequential inflammatory processes define human progression from *M. tuberculosis* infection to tuberculosis disease. *PLoS Pathog.* **13**, 1–24 (2017).
19. Manca, C. et al. Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN. *Proc. Natl Acad. Sci. USA* **98**, 5752–5757 (2001).
20. Stanley, S., Johndrow, J., Manzanillo, P. & Cox, J. S. The type I IFN response to infection with *Mycobacterium tuberculosis* requires ESX-1-mediated secretion and contributes to pathogenesis. *J. Immunol.* **178**, 3143–3152 (2007).
21. Berry, M. P. R. et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466**, 973–977 (2010).
22. Mulenga, H. et al. Longitudinal dynamics of a blood transcriptomic signature of tuberculosis. *Am. J. Respir. Crit. Care Med.* **204**, 1463–1472 (2021).
23. Singhania, A. et al. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nat. Commun.* **9**, 2308 (2018).
24. Musvosvi, M. et al. T cell receptor repertoires associated with control and disease progression following *Mycobacterium tuberculosis* infection. *Nat. Med.* **29**, 258–269 (2023).
25. van Woudenbergh, E. et al. HIV is associated with modified humoral immune responses in the setting of HIV/TB coinfection. *mSphere* **5**, e00104-20 (2020).
26. Lu, L. L. et al. A functional role for antibodies in tuberculosis. *Cell* **167**, 433–443.e14 (2016).
27. Lu, L. L. et al. Antibody Fc glycosylation discriminates between latent and active tuberculosis. *J. Infect. Dis.* **222**, 2093–2102 (2020).
28. Fischinger, S. et al. A *Mycobacterium tuberculosis* specific IgG3 signature of recurrent tuberculosis. *Front. Immunol.* **12**, 1–10 (2021).
29. Grace, P. S. et al. Antibody subclass and glycosylation shift following effective TB treatment. *Front. Immunol.* **12**, 1–12 (2021).
30. Nziza, N. et al. Defining discriminatory antibody fingerprints in active and latent tuberculosis. *Front. Immunol.* **13**, 856906 (2022).
31. Irvine, E. B. et al. Robust IgM responses following intravenous vaccination with Bacille Calmette–Guérin associate with prevention of *Mycobacterium tuberculosis* infection in macaques. *Nat. Immunol.* **22**, 1515–1523 (2021).
32. Zimmermann, N. et al. Human isotype-dependent inhibitory antibody responses against *Mycobacterium tuberculosis*. *EMBO Mol. Med.* **8**, 1325–1339 (2016).
33. Li, H. et al. Latently and uninfected healthcare workers exposed to TB make protective antibodies against *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **114**, 5023–5028 (2017).
34. Lu, L. L. et al. IFN-γ-independent immune markers of *Mycobacterium tuberculosis* exposure. *Nat. Med.* **25**, 977–987 (2019).
35. Davies, L. R. L. et al. IFN-γ independent markers of *Mycobacterium tuberculosis* exposure among male South African gold miners. *EBioMedicine* **93**, 104678 (2023).
36. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016).
37. Gaya da Costa, M. et al. Age and sex-associated changes of complement activity and complement levels in a healthy Caucasian population. *Front. Immunol.* **9**, 2664 (2018).
38. Sauerwald, N. et al. Pre-infection antiviral innate immunity contributes to sex differences in SARS-CoV-2 infection. *Cell Syst.* **13**, 924–931.e4 (2022).
39. Gupta, M., Srikrishna, G., Klein, S. L. & Bishai, W. R. Genetic and hormonal mechanisms underlying sex-specific immune responses in tuberculosis. *Trends Immunol.* **43**, 640 (2022).
40. Jennewein, M. F. & Alter, G. The immunoregulatory roles of antibody glycosylation. *Trends Immunol.* **38**, 358–372 (2017).
41. Dodd, P. J. et al. Age- and sex-specific social contact patterns and incidence of *Mycobacterium tuberculosis* infection. *Am. J. Epidemiol.* **183**, 156–166 (2016).
42. Marais, B. J. et al. The natural history of childhood intra-thoracic tuberculosis: a critical review of literature from the pre-chemotherapy era. *Int. J. Tuberc. Lung Dis.* **8**, 392–402 (2004).
43. Siegrist, C. A. & Aspinall, R. B-cell responses to vaccination at the extremes of age. *Nat. Rev. Immunol.* **9**, 185–194 (2009).

44. Haralambieva, I. H. et al. Associations between race, sex and immune response variations to rubella vaccination in two independent cohorts. *Vaccine* **32**, 1946–1953 (2014).
45. Choe, Y. J., Blatt, D. B., Lee, H. J. & Choi, E. H. Associations between geographic region and immune response variations to pneumococcal conjugate vaccines in clinical trials: a systematic review and meta-analysis. *Int. J. Infect. Dis.* **92**, 261–268 (2020).
46. Moir, S. & Fauci, A. S. B cells in HIV infection and disease. *Nat. Rev. Immunol.* **9**, 235–245 (2009).
47. Kunnath-Velayudhan, S. et al. Dynamic antibody responses to the *Mycobacterium tuberculosis* proteome. *Proc. Natl Acad. Sci. USA* **107**, 14703–14708 (2010).
48. Kunnath-Velayudhan, S. et al. Proteome-scale antibody responses and outcome of *Mycobacterium tuberculosis* infection in nonhuman primates and in tuberculosis patients. *J. Infect. Dis.* **206**, 697–705 (2012).
49. Khan, I. H. et al. Plasma antibody profiles as diagnostic biomarkers for tuberculosis. *Clin. Vaccin. Immunol.* **18**, 2148–2153 (2011).
50. Broger, T. et al. Diagnostic performance of tuberculosis-specific IgG antibody profiles in patients with presumptive tuberculosis from two continents. *Clin. Infect. Dis.* **64**, 947–955 (2017).
51. Ackerman, M. E. et al. A robust, high-throughput assay to determine the phagocytic activity of clinical antibody samples. *J. Immunol. Methods* **366**, 8–19 (2011).
52. Karsten, C. B. et al. A versatile high-throughput assay to characterize antibody-mediated neutrophil phagocytosis. *J. Immunol. Methods* **471**, 46–56 (2019).
53. Tibshirani, R. The lasso method for variable selection in the cox model. *Stat. Med* **16**, 385–395 (1997).
54. Wang, C. ChuangqiWang/TB_Progressor: SeroScore.V1. *Zenodo* https://doi.org/10.5281/zenodo.10637145 (2024).
55. TB_Progressor. *GitHub* https://github.com/ChuangqiWang/TB_Progressor TB_Progressor (2024).

Acknowledgements

This work was supported by the National Institutes of Health (K08AI167766 to L.R.L.D. and 75N93019C00071 to S.M.F. and G.A.), the Doris Duke Charitable Foundation (L.R.L.D.) and the Bill and Melinda Gates Foundation (grants OPP1023483, OPP1065330 and Grand Challenges in Global Health (GC6-74 grant 37772)). The ACS study was also supported by BMGF GC12 (grant 37885) for QuantiFERON testing. We gratefully acknowledge the work of the ACS and GC6-74 study groups as well as the participation of all enrollees, without whom this work would not be possible. We further thank B. Moody for generous contribution of the lipid antigen 1-TbAd, BEI Resources for providing *Mtb* antigens and the Duke Protein Production Facility for providing Fc receptors. We are grateful to D. Mugahid for her assistance in reviewing and revising the paper. We thank the Hi-IMPAcTB Data Management team for organizing the (meta)data associated with this project for FAIR sharing, namely, C. Demurjian, E. Koo, S. Levine (MIT BioMicro Center) and D. Mugahid (Harvard School of Public Health).

Author contributions

T.H.M.O., H.M.D., J.S.S., H.M.-K., M.H., W.H.B., G.W. and S.H.E.K. led the original GC6-74 study. L.R.L.D., S.F., T.J.S. and G.A. conceived the current experimental approach. P.S. and S.K.M. contributed to sample and data management. L.R.L.D. and S.F. performed all experiments. C.W. and D.L. designed the analytic approach. C.W., L.R.L.D., K.A.B., P.S. and M.R. performed analysis. Data were interpreted by L.R.L.D., S.F., K.A.B. and C.W. under the supervision of E.N., S.M.F., T.J.S., D.L. and G.A. C.W. and L.R.L.D. created all data visualizations. M.F. contributed to project administration. The paper was drafted by L.R.L.D. and C.W. and was reviewed, edited and approved by all authors.

Competing interests

G.A. is an employee of Moderna Therapeutics and holds equity in Leyden Labs and Systems Seromyx. S.F. is a member of the board of directors of Oxford Nanopore Technologies. L.R.L.D. became an employee of BioNTech US after completion of the work described in this paper. T.J.S. is co-inventor of a patent of the RISK6 signature. S.H.E.K., J.S.S., T.H.M.O., H.M.D., W.H.B., T.J.S. and G.W. are co-inventors on a GC6-derived signature. The remaining authors declare no competing interests.

Ethics and inclusion statement

All clinical research performed in this study was performed in accordance with the Declaration of Helsinki. The clinical samples analysed in this paper were collected as part of the original ACS and GC6-74 observational studies and are described in detail in the original publication[14] and in 'Study design' section above. Samples were collected at all sites only after written informed consent was given by the patients' legal guardian. Subjects in ACS were compensated an amount of R50 (approximately 7 USD) in the form of a non-cash payment such as a voucher at every occasion of a blood draw. Subjects in GC6 were compensated for loss of income and transport costs incurred due to research visits. For the ACS cohort study, protocols were approved by the University of Cape Town Research Ethics Committee, Cape Town, South Africa. For the GC6-74 study, protocols were approved by the institutional review boards of Stellenbosch University, Case Western Reserve University, the Uganda National Council for Science and Technology and the Joint Gambian Government/MRC Ethics Committee. The systems serology analysis was approved by Massachusetts General Hospital. Participant sex and age were self- or guardian-declared.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-024-01678-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-024-01678-x.

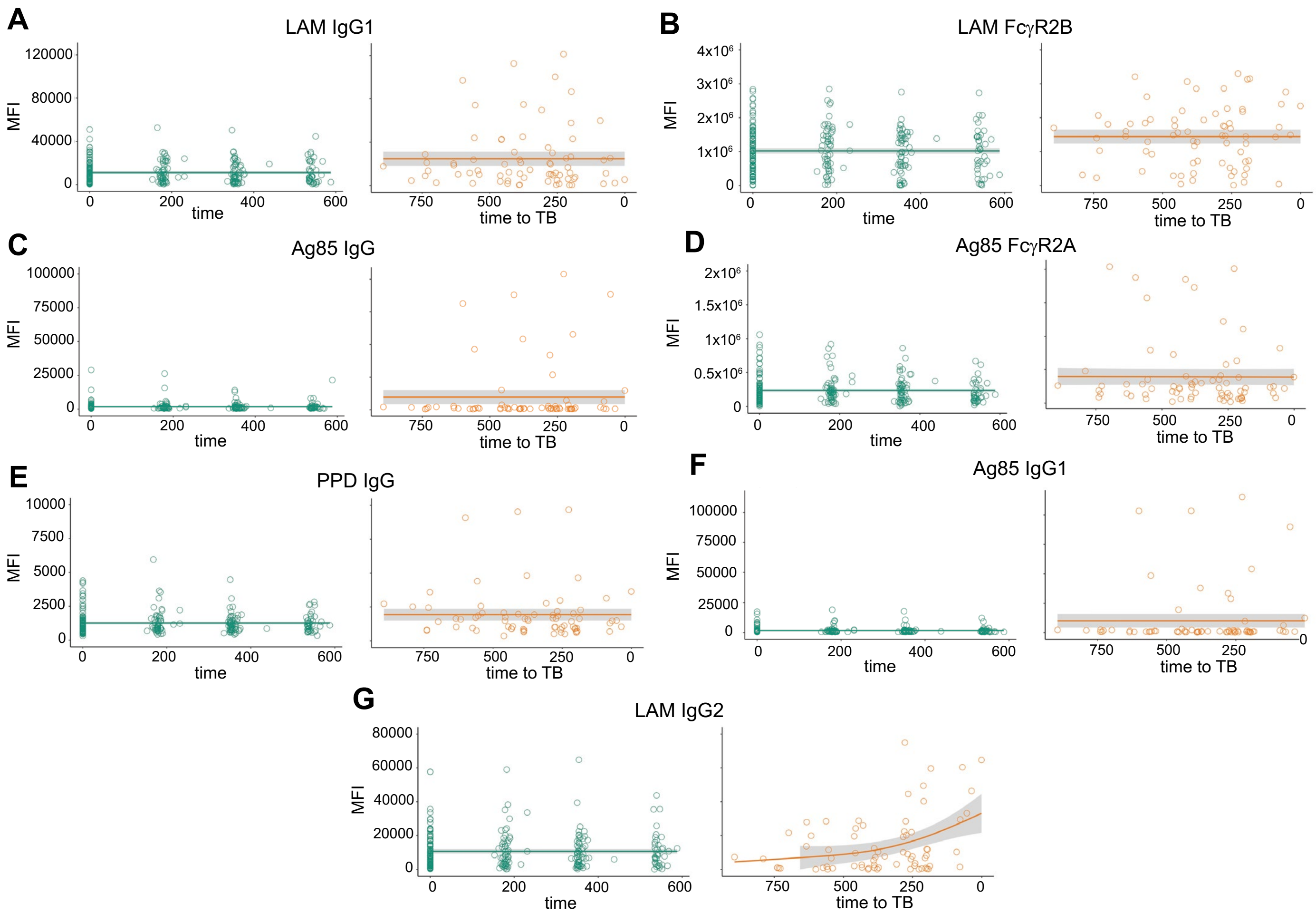
Correspondence and requests for materials should be addressed to Galit Alter or Sarah M. Fortune.

Peer review information *Nature Microbiology* thanks the anonymous reviewers for their contribution to the peer review of this work.

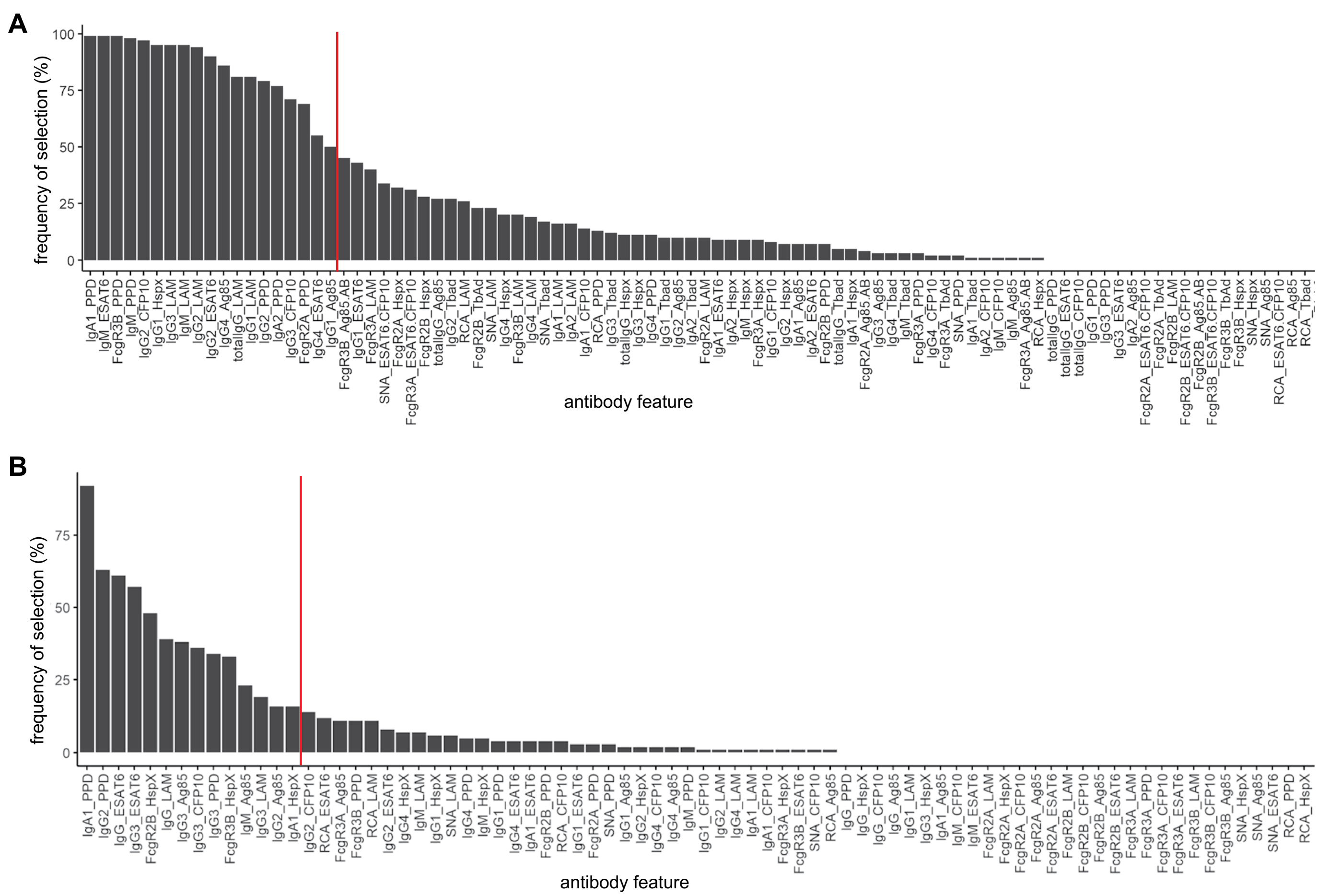
Reprints and permissions information is available at www.nature.com/reprints.

[1]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA. [2]Brigham and Women's Hospital, Boston, MA, USA. [3]Department of Immunology and Microbiology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [4]Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]South African Tuberculosis Vaccine Initiative and Institute of Infectious Disease and Molecular Medicine, Division of Immunology, Department of Pathology, University of Cape Town, Cape Town, South Africa. [6]Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Disease and Molecular Medicine and Division of Immunology, Department of Medicine, University of Cape Town, Cape Town, South Africa. [7]Department of Infectious Diseases, Leiden University Medical Center, Leiden, the Netherlands. [8]Vaccines and Immunity Theme, Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine, Banjul, The Gambia. [9]Department of Medicine and Department of Microbiology, Makerere University, Kampala, Uganda. [10]Tuberculosis Research Unit, Case Western Reserve University, Cleveland, OH, USA. [11]Department of Science and Technology National Research Foundation 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]Max Planck Institute for Infection Biology, Berlin, Germany. [13]Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany. [14]Hagler Institute for Advanced Study, Texas A&M University, College Station, TX, USA. [15]Moderna Therapeutics, Cambridge, MA, USA. [16]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [17]These authors contributed equally: Leela R. L. Davies, Chuangqi Wang. ✉e-mail: galit.alter@modernatx.com; sfortune@hsph.harvard.edu

 https://doi.org/10.1038/s41564-024-01678-x

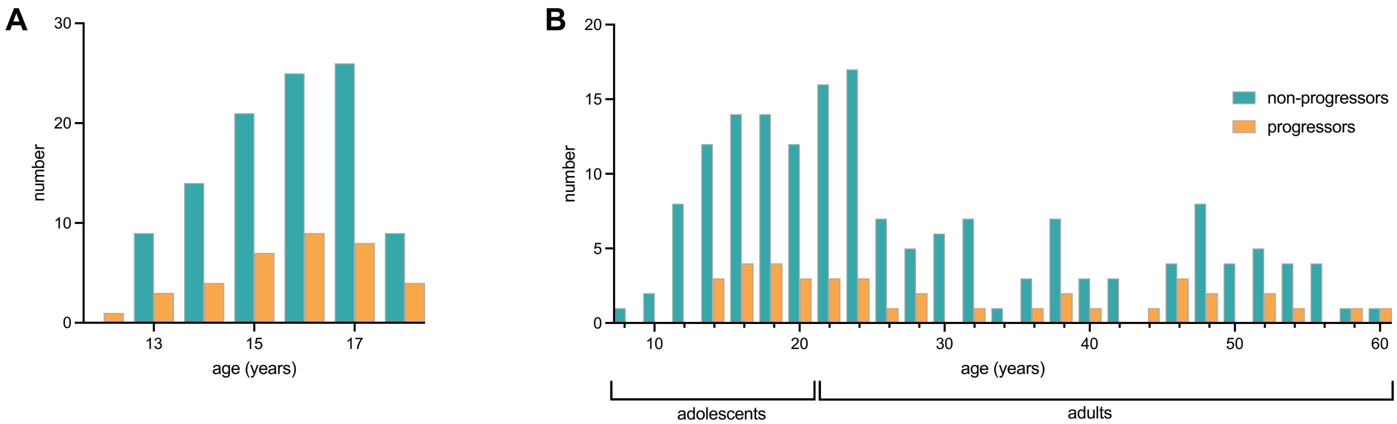


Extended Data Fig. 1 | Temporal trajectories of additional Mtb-specific antibody features significantly enriched in ACS progressors. Raw values of measured antibody features for all indivduals were plotted over time from enrollment (non-progressors, teal) or time to TB (progressors, orange). Solid lines indicate a smooth of median values, using a generalized additive model, and grey shading indicates 95% confidence interval. Data is shown for **a**) LAM IgG1, **b**) LAM FcgR2B, **c**) Ag85 IgG, **d**) Ag85 FcgR2A, **e**) PPD IgG, **f**) Ag85 IgG1, and **g**) LAM IgG2. These include all antibody features found to statistically differ between progressors and non-progressors in mixed effects linear modeling, as well as LAM IgG2, which did not statistically differ but is shown for comparison.



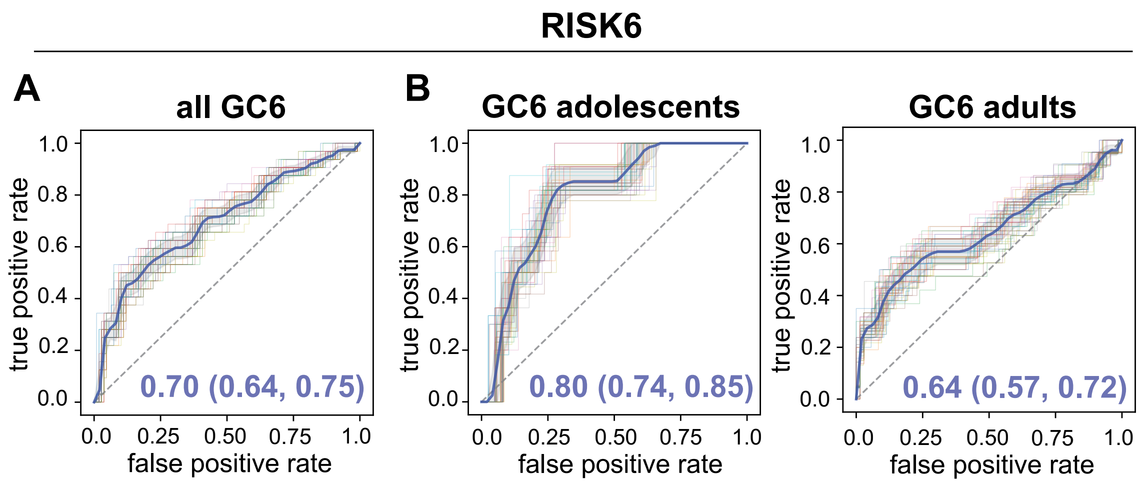
Extended Data Fig. 2 | Selection of antibody features for inclusion in SeroScores. For **a**) ACS, and **b**) GC6, LASSO (Least Absolute Shrinkage and Selection Operator) regularization was applied to 100 randomly selected subsets, each containing 80% of the full dataset, and iterated 10 times. The frequency of selection of each antibody feature is shown. Red lines indicate the threshold defining the features that were evaluated in combination for each SeroScore.

 https://doi.org/10.1038/s41564-024-01678-x



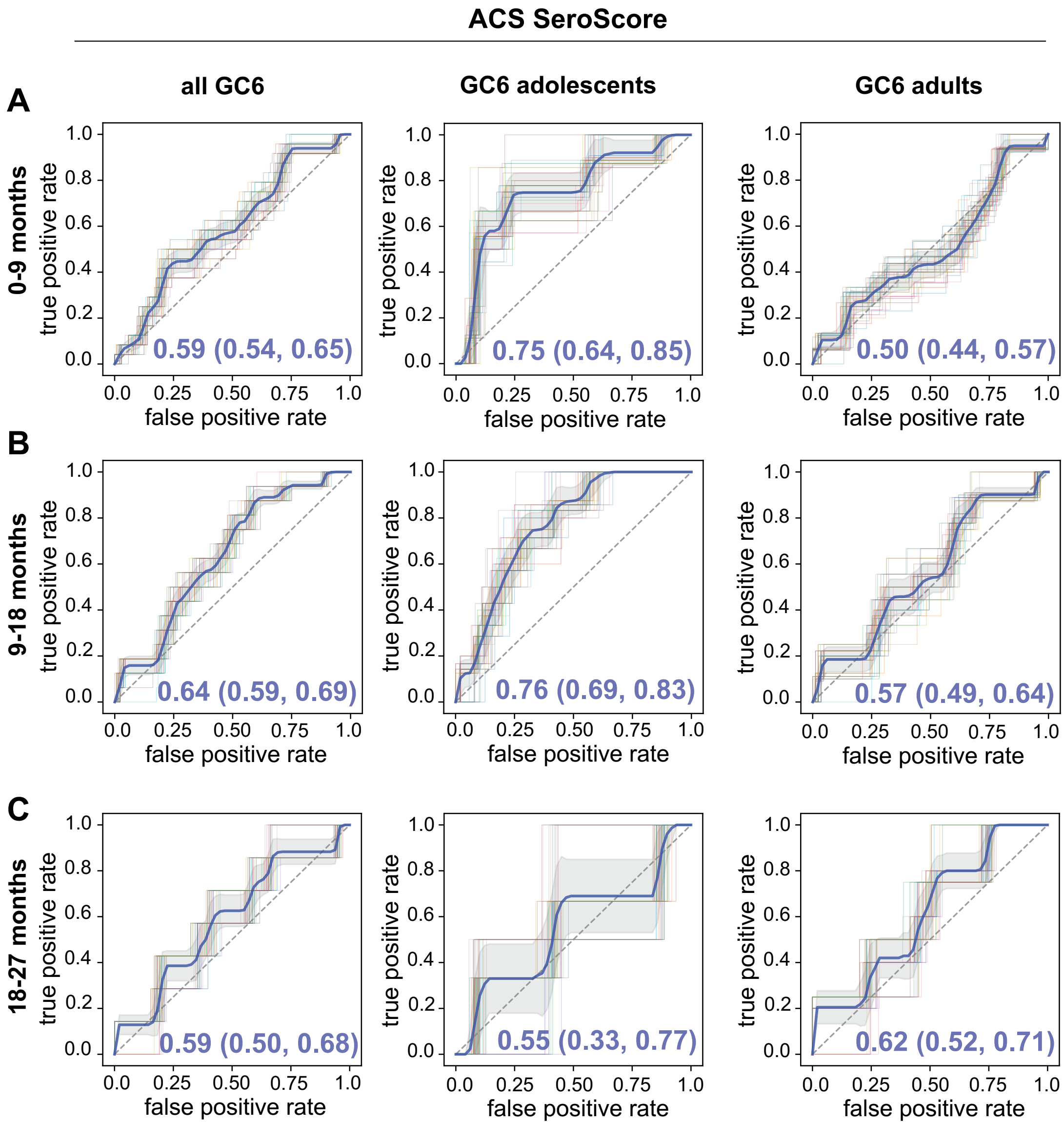
Extended Data Fig. 3 | Age distributions for included subjects from ACS and GC6. a) Histogram of age distributions for ACS progressors and non-progressors. **b**) Histogram of age distributions for GC6 progressors and non-progressors. For GC6 subjects, adolescents were defined as those with age at enrollment between 8 and 20 years, and adults were those with age at enrollment of 21 or more years.

 https://doi.org/10.1038/s41564-024-01678-x



Extended Data Fig. 4 | Ability of RISK6 to identify GC6 adolescent and adult progressors. ROC curves were developed over the full study duration for the subset of individuals in GC6 for whom the RISK6 score was available. A) ROC curve indicates performance of RISK6 among all GC6 individuals (n = 35 progressors and n = 135 non-progressors). B) ROC curve for performance of RISK6 among GC6 adolescents (age 8–20 years, n = 12 progressors and n = 49 non-progressors) and adults (age ≥21 years, n = 23 progressors and n = 86 non-progressors). ROC curves were generated 50 times using randomly selected 80% of samples with group stratification. The mean curve is indicated in blue, with grey shading indicating one standard deviation. Mean AUC with 95% confidence interval is indicated.

 https://doi.org/10.1038/s41564-024-01678-x



Extended Data Fig. 5 | Longitudinal ability of the ACS-derived SeroScore to identify GC6 adolescent and adult progressors. The SeroScore derived in ACS was used to develop ROC curves for GC6 adolescents (age 8–20 years) and adults (≥21 years) over time windows **a**) 0–9 months, **b**) 9–18 months, and **c**) 18–27 months prior to diagnosis of active TB. ROC curves were generated 50 times using randomly selected 80% of samples with group stratification. The mean curve is indicated in blue, with grey shading indicating one standard deviation. Mean AUC with 95% confidence interval is indicated.

nature portfolio

Corresponding author(s): Sarah M Fortune
Galit Alter

Last updated by author(s): 2024 02 20

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	All systems serology data generated on the iQue Plus Screener was collected using Forecyt software.
Data analysis	Univariate comparisons of individual assayed antibody features were performed in GraphPad Prism 9. Mixed linear models were created using the R package "lme4" and visualized using "ggplot2" version 3.3.5 in R version 4.0.2. LASSO features selection using the R package "systemsseRology". ROC curves were generated, and AUC calculated, in Python using package "sklearn.metrics" version 1.2.1, and confidence intervals programmed using "scipy.stats" version 1.10.1. The code used for the analyses in this paper is available on Github at https://github.com/ChuangqiWang/TB_Progressor.

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

RISK6 scores included in analyses within the current manuscript were previously calculated for both the ACS cohort (ref 15) and the GC6 cohort (GEO Accession number: GSE94438). All data generated as part of this manuscript will be available in the main text or the supplementary materials. No custom biological materials were generated during the course of this research. All samples were obtained from the researchers who conducted the original ACS and GC6 observational studies, all of whom can be contacted about sample availability from the respective cohorts.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Both male and female participants, by self-reported sex, were included in both ACS and GC6 studies. Included participants from ACS were 67.1% female, and those from GC6 were 58.7% female. Sex-stratified analyses of the measured Mtb-specific antibody responses are key findings presented in the manuscript, and they are shown in Figures 3 and 5.
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	ACS is a cohort of HIV-negative South African adolescents aged 12-18 years. GC6 is a cohort of HIV-negative people aged 10–60 years who had household exposure to an adult with sputum smear-positive TB, and all included participants here were from South Africa. Available demographic data for each cohort is described in Supplementary Tables 1 and 2.
Recruitment	See https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5392204/#SD2
Ethics oversight	All clinical research performed in this study was performed in accordance with the Declaration of Helsinki. The clinical samples analyzed in this manuscript were collected as part of the original ACS and GC6-74 observational studies and are described in detail in the original publication and in the Experimental Design section. Samples were collected at all sites only after written informed consent was given by the patients' legal guardian. Subjects in ACS were compensated an amount of R50 (approximately 7 US dollars) in the form of a non-cash payment such as a voucher at every occasion of a blood draw. Subjects in GC6 were compensated for loss of income and transport costs incurred due to research visits. For the ACS cohort study, protocols were approved by the University of Cape Town Research Ethics Committee, Cape Town, South Africa. For the GC6-74 study, protocols were approved by the institutional review boards of Stellenbosch University, Case Western Reserve University, the Uganda National Council for Science and Technology, and the Joint Gambian Government/MRC Ethics Committee. The systems serology analysis was approved by Massachusetts General Hospital.

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	Sample size was determined based on sample availability at the clinical sites where the ACS and G6-74 studies were conducted.
Data exclusions	Raw data were excluded if they failed to meet quality control measures, specifically if the Luminex bead count was less than 20 for a given data point. Otherwise, no data were excluded.
Replication	All assays were performed in two technical replicates, or two biological replicates where applicable, and strong correlation between replicates was confirmed as part of quality control of raw data. The univariate and multivariate antibody signatures identified as associated with progression in the ACS cohort were validated in a second cohort, GC6.
Randomization	Participant allocation was not random, but rather included matched cases and controls. ACS and GC6 were both conducted as large

Randomization	prospective cohort studies. A number of individuals within each study progressed to active TB within the study period. All available samples from these progressors were included in the current study, as well as non-progressors matched by available demographic features.
Blinding	Investigators were blinded as to progressor/non-progressor status during experimental data collection. We performed a supervised computational analysis, and thus analysis was not performed in a blinded manner.

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
☒	☐ Plants

Methods

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

Antibodies

Antibodies used	Phycoerythrin (PE)-conjugated goat anti-human antibodies from Southern Biotech: IgG (9040-09), IgG1 (9052-09), IgG2 (9070-09), IgG3 (9210-09), IgG4 (9200-09), IgA1 (9130-09), IgA2 (9140-09), and IgM (9020-09). Anti-human CD66b antibody from BioLegend (305112).
Validation	All antibodies used were validated by the manufacturers. RRID numbers are available for all antibody reagents as follows: AB_2796601 (IgG), AB_2796621 (IgG1), AB_2796639 (IgG2), AB_2796701 (IgG3), AB_2796693 (IgG4), AB_2796656 (IgA1), AB_2796664 (IgA2), AB_2796577 (IgM), AB_2563294 (CD66b).

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	THP-1 cells (human acute monocytic leukemia cell line, American Type Culture Collection);
Authentication	Cell lines were not authenticated after purchase.
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.