Gene expression in cord blood and tuberculosis in early childhood: A nested case-control study in a South African birth cohort

Carly A. Bobak[1], Maresa Botha[3], Lesley Workman[3], Jane E. Hill[4], Mark Nicol[5,6,] John W. Holloway[7,8], Dan J Stein[9-11], Leonardo Martinez[2*], Heather J Zar[3*]

Affiliations:

1. Biomedical Data Science, Dartmouth College, Hanover, NH, USA
2. Department of Epidemiology, School of Public Health, Boston University, Boston, MA, USA
3. Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital and South African Medical Research Council Unit on Child and Adolescent Health, Cape Town, South Africa
4. University of British Columbia, Vancouver, BC, Canada
5. Division of Infection and Immunity, School of Biomedical Sciences, University of Western Australia, Perth, WA, Australia
6. Division of Medical Microbiology, University of Cape Town, Cape Town, South Africa
7. Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK
8. NIHR Southampton Biomedical Research Center, University Hospital Southampton, Southampton, UK
9. University of Cape Town, Department of Psychiatry and Mental Health, Cape Town, South Africa
10. South African Medical Research Council (SAMRC) Unit on Risk and Resilience in Mental Disorders, Cape Town, South Africa
11. University of Cape Town, Neuroscience Institute, Cape Town, South Africa

* Co-last authors contributing equally.

Author contributions

CAB analyzed the data. CAB and LM wrote the first draft of the manuscript. HJZ is the principal investigator, obtained funding, conceived, and designed the study, and assisted with drafting of the manuscript. MPN is the lead microbiologist. MB provided operational oversight; LW provided data support. DJS obtained funding and oversaw RNA expression aspects. All

authors contributed to the interpretation of results. All authors reviewed, contributed to, and approved the final manuscript.

Conflicts of interest

None

Funding

CAB was supported by the Burroughs Wellcome Fund institutional program grant unifying population and laboratory-based sciences to Dartmouth College (Grant#1014106). JEH reports grants from the US National Institutes of Health, US Cystic Fibrosis Foundation, The Australian Research Council, and the National Sciences and Engineering Research Council of Canada during the completion of this study. HJZ reports grants from the Bill & Melinda Gates Foundation and from Medical Research Council South Africa, the National Research Foundation South Africa, and the National Institutes of Health during completion of the study.

Manuscript Word Count: 3199

Tables: 1

Figures: 6

References: 50

Running Title: Cord blood gene expression and TB in children

Key Words: tuberculosis, pediatrics, transcriptomics.

Summary: Gene expression in umbilical cord blood demonstrates associations with Tuberculosis infection and disease in early life. Differentially expressed genes overlap with known diagnosis and risk signatures. Implicated biological mechanisms include neutrophil activation and defense responses to bacteria. (**37 out of 40 words**)

Correspondence:

Professor Heather J Zar

University of Cape Town

Department of Paediatrics and Child Health

Red Cross War Memorial Children's Hospital and MRC Unit on Child and Adolescent Health

Cape Town 8100, South Africa

Email: heather.zar@uct.ac.za

Abstract

Background

Transcriptomic profiling of adult tuberculosis patients has become increasingly common, predominantly for diagnostic and risk prediction purposes. However few studies have evaluated signatures in children, particularly in identifying those at risk for developing TB disease. We investigated the relationship between gene expression obtained from umbilical cord blood and both tuberculin skin test conversion as well as incident tuberculosis disease through the first 5 years of life.

Methods

We conducted a nested case-control study in the Drakenstein Child Health Study, a longitudinal, population-based birth cohort in South Africa. We applied transcriptome-wide screens to umbilical cord blood samples from neonates born to a subset of selected mothers (n=131). Signatures identifying tuberculin conversion and risk of subsequent tuberculosis disease were identified from genome wide analysis of RNA expression.

Results

Gene expression signatures revealed clear differences predictive of tuberculin conversion (n=26) and tuberculosis disease (n=10); 114 genes were associated with tuberculin conversion and 30 genes were associated with the progression to tuberculosis disease among children with early infection. Co-expression network analysis revealed six modules associated with risk of tuberculosis infection or disease, including a module associated with neutrophil activation in immune response ($p<0.0001$) and defense response to bacterium ($p<0.0001$).

Conclusions

These findings suggest multiple detectable differences in gene expression at birth which were associated with risk of tuberculosis infection or disease throughout early childhood. Such measures may provide novel insights into tuberculosis pathogenesis and susceptibility.

Abstract Word Count: 233 (**limit: 250**)

Introduction

Approximately one million children develop tuberculosis disease (TB) every year substantially contributing to global pediatric morbidity and mortality[1,2]. Most children who develop TB are under five years of age, an age group especially difficult to diagnose[1]. Identifying children who are likely to develop TB based on exposure status and underlying biology is of critical importance to administer targeted preventive therapy and reduce morbidity and mortality[3–5]. RNA transcriptional profiles have been increasingly used for diagnosis and assessment of TB risk among adults and children[6–9]. The relationship between maternal environment in pregnancy and TB risk of offspring is less well understood. Transcriptional analysis of cord blood may potentially provide insight into immune mechanisms that determine risk of TB infection in early childhood.

Whether children differentially express genes compared to adults in relation to TB risk is debated[4]. Gene signatures for diagnosis of pediatric TB have shown specific transcriptomic profiles associated with microbiologically confirmed or clinically diagnosed TB[7]. A small study in India found that children with TB had distinct gene signatures compared to signatures typically used in adults[9]. There are limited data on gene expression profiles predicting TB infection or disease in children, especially from high TB burden countries . However, both inherited genetic variation[10] and maternal environmental exposures (both pre-conceptionally and during pregnancy) are associated with offspring immunity and risk of respiratory infection[11–13]. Identification of gene expression signatures at birth associated with TB risk in childhood may assist our understanding of prenatal factors associated with offspring immunity allowing for strengthened strategies to prevent TB.

We investigated the relationship between gene expression in cord blood among children who did and did not develop TB infection or disease during early childhood from a prospective birth cohort study in Cape Town, South Africa.

Methods

Participants and study design

In a prospective, South African birth cohort, the Drakenstein Child Health Study (DCHS), we followed children from birth through five years of age as described previously[14–17]. Briefly, pregnant women between 20–28 weeks gestation were enrolled at community clinics in the Drakenstein area. Exclusion criteria were those younger than 18 years of age or intention to leave the area within one year. All deliveries occurred at a central hospital, Paarl hospital, where cord blood was collected by trained staff. Infants were given Bacillus Calmette-Guerin (BCG) vaccination at birth (Denmark strain), per national policy. Active surveillance systems for lower respiratory tract illness and TB were established. Children were followed for TB infection and disease until five years of age.

Tuberculin skin testing was done at 6, 12, 24, 36, 48, and 60 months of age, and at the time of lower respiratory tract infection or suspected TB as previously reported[15,16]. Tuberculin skin test conversion was defined as an induration reaction >10mm in children without HIV or >5mm in participants with HIV. Repeat testing was not conducted on children with any tuberculin response (i.e., >0mm induration) to minimize potential for tuberculin boosting. Children with positive tuberculin skin tests were further screened for TB and referred for preventive therapy.

To diagnose TB, children with a positive tuberculin skin test or who were clinically suspected to have TB were investigated using induced sputum done by trained study staff in duplicate for smear, mycobacterial PCR (Xpert MTB/RIF; Cepheid, Sunnyvale, CA, USA) and liquid culture[15,18]. Chest X-rays were taken in all children suspected of TB and were read and

reported by an experienced clinician; TB was diagnosed by experienced healthcare providers in local TB community clinics. We used standardized consensus definitions for diagnostic classification of TB[19]; confirmed TB, unconfirmed TB, and unlikely TB. Children diagnosed with TB in this cohort consist both confirmed and unconfirmed TB.

Cord blood collection, RNA isolation, and gene expression data processing

A subset of the cohort was selected for transcriptional profiling using biobanked umbilical cord blood samples as previously described[20]. Samples were collected after delivery by clamping, cutting, and draining umbilical cords into kidney dishes. Blood was collected, stored at -80°C in PAXgene RNA tubes. An IlluminaHT-12 v4 beadchip array was used to obtain row probe intensity values. Samples were previously randomized within batches, based on demographics (i.e., sex, maternal diagnoses, maternal alcohol and tobacco use, and mode of delivery) to reduce potential for batch effects[20]. We conducted a case-control study of all previously processed samples nested within the cohort. We defined cases as 1) children that converted their tuberculin skin test over follow-up; and 2) children diagnosed with TB. Controls were participants who did not develop TB or convert their tuberculin skin test over follow-up. Children with missing TB outcomes were excluded.

Umbilical cord blood samples, RNA collection, and gene expression array processing were done as previously described[20] and are further detailed in the Supplemental Methods.

Differential gene expression (DGE) analysis was conducted using the *limma* package[21]. DGE was used to identify genes that were significantly differentially expressed for three outcomes: (i) infants who did and did not convert their tuberculin skin test (TST) before three years old (the

former herein referred to as ‘early converters’); (ii) infants who did and did not develop TB before five years old; and (iii) among early converters only, infants who did and did not develop TB by 5 years of age. We considered current maternal smoking status adjusted for HIV as a fourth outcome. We employed an exploratory significance threshold of α=0.005 for DGE[20,22] and used Gene Set Enrichment Analysis (GSEA) to map genes to biological pathways [23]. A pathway Z-score was assigned to each sample as in [24]. Weighted gene co-expression network analysis (WGCNA) [25] was conducted to identify and characterize gene modules for their associations with each of the three TB outcomes. We used gProfiler[26] to identify enriched pathways within each module, and visualized these using an Enrichment Map[27]. CIBERSORTx was used to identify absolute abundance of immune cells using transcriptional data[28]. Identified genes, pathways, and modules were compared to two cohorts of pediatric TB patients collected in Kenya and Malawi[7]. Full details on the analytic pipeline are included in the Supplementary Methods.

All computational code is available at https://github.com/CarlyBobak/TBCordBlood. Raw expression data are publicly available through the Gene expression Omnibus (GEO, accession number GSE114852)[29].

Results

Of 144 biobanked cord blood samples, 131 (91.0%) children had available TST and TB diagnosis data and were included in this analysis. Amongst these, all children were followed for 5 years with no loss to follow-up or death. Maternal HIV occurred in 25.2%, self-reported smoking in 28.2%, while prior maternal TB diagnosis was reported in 3.8% of all participants. The population was predominantly of low household income. In total, 25.2% were HIV-exposed, however no children had HIV. Median weight-for-age z-score and height-for-age z-score at 5 years were -0.07 (IQR: -1.21, 1.64) and -0.07 (-1.21, 1.64), respectively. In total, 10 (7.6%) children received preventive therapy.

Among included children, 26 (19.8%) were early converters while 14 (10.7%) developed TB prior to 5 years of age. Amongst converters, 10 out of 26 (38.5%) subsequently developed TB by 5 years of age.

There were no statistically significant differences between tuberculin converters and non-converters in relation to maternal HIV status, maternal TB during pregnancy, self-reported history of maternal TB, TB in the household one year prior to enrollment, or maternal smoking (Table 1). There were no statistically significant differences between children who did and did not develop TB disease in relation to sex, birthweight, or duration of breastfeeding, while self-reported maternal smoking during pregnancy approached statistical significance, with a higher percentage of smoking mothers represented among children with TB (50% versus 26.6% among TB progressors and non-progressors; p=0.07). (Supplementary Table 1).

Differential gene expression reveals signatures from umbilical cord blood and infant TB outcomes

We sought to identify differentially expressed genes in umbilical cord blood between infants who did and did not experience early TST conversion. A total of 114 genes were significant above the exploratory threshold of $p<0.005$. Of genes that met the significance threshold, the largest absolute $\log_2$ fold changes were for *DEFA3* ($p=0.004$), *DEFA1* ($p=0.002$), *HLA-DQAI* ($p=0.001$), and *IFITM3* ($p=0.004$; Figure 1A; Supplementary Table 2).

Principal component analysis (PCA) of the significant genes shows a visible trend clustering early converters compared to participants who did not convert (Figure 1B,1C)[30]. The differences in PC1 were statistically significant ($p=1.8\text{x}10^{-6}$)[31]. PC1 also separated TB-progressors from non-progressors ($p=0.047$). PC2 demonstrated statistically significant differences between mothers who had and did not have a prior TB diagnosis before pregnancy ($p=0.0052$); this analysis had very small sample size of mothers in the prior TB group (n=5; Figure 1D).

Median centered expression values are displayed using a heatmap with unsupervised hierarchical clustering in Figure 2 [32]. Clustering among early converters was present using this umbilical cord gene expression signature. This finding suggests there are distinct gene expression differences associated with greater susceptibility to TB infection.

When focusing our analysis on incident TB, we identified 60 genes that were statistically significant ($p<0.005$) between TB progressors and non-progressors. The most significant genes

included *SULT1A3* (p=8.05x10^{-5}), *HMBS* (p=1.50x10^{-4}), and *NCOA3* (p=0.002; a full list of these results is shown in Supplementary Table 3 and Supplementary Figure 1).

In an analysis of TB disease in the first 5 years restricted to early TST converters, we found 30 associated genes, where the most significant included *PARP1* (p=9.84x10^{-5}), *WDR4* (p=5.34x10-4), and *KLRD1* (p=0.002; Figure 3A; Supplementary Table 4). In principal component analysis (Figure 3B), there were clear differences along the first principal component (Figure 3C) and these differences were statistically significant (p=7.2x10^{-6}). The first principal component was also associated with maternal smoking status (p=0.012). In an unsupervised hierarchical clustering analysis, we observed a strong clustering of six infants, where 5/6 of those infants had mothers who were current smokers at enrolment (Supplementary Figure 2).

A DGE and subsequent GSEA revealed that pathways related to immune response, response to bacterium, and immune cell activation were overlapping across all TB outcomes and maternal current smoking status (Figures 4A/D). These pathways significantly differentiate between TB outcomes (Figures 4B/C). Further smoking DGE and GSEA results and an enrichment map are available in the Supplement.

Biologically relevant modules reveal meaningful networks of gene expression for TST conversion and diagnosis of TB in young children

We used WGCNA to identify interpretable, biologically relevant co-regulated gene modules. A total of 14 modules were identified (Supplementary Tables 10 and 11) and we evaluated module significance by testing for over-representation of disease-related differential gene

signatures across all TB related outcomes and modules. We found that modules M3, M5, M7, M11, and M14 were significantly associated with the early conversion gene expression signature, with M11 having the most significant association (Figure 5A). No modules were related to the development of TB, but M9 was associated with development of TB among children with early conversion (Supplementary Figure 4).

We found a strong correlation between TST conversion and M11 gene connectivity (kME) (Figure 5B). The greater association between gene expression value and early TST conversion, the greater the connection was within the M11 module (cor=0.57, p=0.00023). The M11 eigengene was tested for associations amongst maternal and child health characteristics. Significant associations were found across several characteristics including maternal prior TB diagnosis (p=0.01), infant birth weight (p=0.044) and early TST conversion (p=0.047) (Figure 5C; Supplementary Figures 5 and 6). Children who developed TB among early converters compared to those who neither converted nor developed disease drove this association (p=0.036; Figure 5D).

We identified functionally enriched biologically interpretable pathways for each module (Supplementary Table 12). The top pathways functionally enriched in the M11 module include neutrophil degranulation, neutrophil activation involved in immune response, and neutrophil mediated immunity (all from Gene Ontology: Biological Processes(GO:BP); adjusted p-values of 1.13×10^{-22}, 1.29×10^{-22}, and 2.07×10^{-22} respectively). Other pathways include the innate immune system (Reactome; 3.6×10^{-14}), antimicrobial humoral response, and killing cells of other organisms (GO:BP; 7.47×10^{-13} and 3.05×10^{-12}). Select pathways are visualized in Figure 6D.

In a full enrichment map, we found overlapping pathways (M5, M9, M11, and M14) across all modules significantly associated with early childhood TST conversion or TB diagnosis (Supplementary Figure 7). We also found three extracted subnetworks from the enrichment map using pathways functionally enriched with M11 (Figure 6). These highly linked subnetworks demonstrate major trends in the biological functionality associated with M11 with 6A representative of the defensive immune response to bacteria, 6B associated with the host interaction with symbiont cells, and 6C focused on the cellular response to molecules of bacterial origin.

Using CIBERSORTx cell-type abundance estimation, we found cell-type differences between (i) TB progressors and early TST converters as well as (ii) TB progressors and children who did not TST convert. These differences were present in γδ T-cells, which were decreased in TB progressers (p=0.0016 and p=0.0114) and mast cells (also decreased; p=0.0292 and p=0.0006). Neutrophils were decreased in TB progressors compared to children who did not TST convert (p=0.045); this was not statistically significant when comparing TB progressors and children who TST converted (p=0.1073). B cells were decreased in TST converters compared to non-converters (p=0.0396). Full CIBERSORTx results are available in Supplementary Table 13.

Differentially Expressed Genes and modules in cord blood are present at the time of diagnosis.

Across two independent pediatric TB cohorts which measured gene expression in whole blood at time of diagnosis [7], we found that 78 of 128 measured cord blood genes (60.94%) were differentially expressed (adjusted p<0.05). Similarly, the M11 module was both preserved and high quality in both the Kenyan and Malawi cohorts (preservation $p=1.37x10^{-28}$ *and* $p=7x10^{-20}$*;*

quality p=3.07X10^{-40} and p=3.01x10^{-53}). Additional validation can be found in Supplemental results and Supplemental Figure 8.

Discussion

In this South African birth cohort study following infants through 5 years old in a high TB prevalence area, we found several novel gene expression profiles from umbilical cord blood that differentiated children at risk of TST conversion and incident TB. Several identified genes have established associations in TB pathogenesis, predominantly in adults [7,8,33–41]. These studies have traditionally assumed gene expression changes due to TB exposure, infection, or disease. In this work, we show that differences in gene expression disease may occur prior to birth, suggesting the possibility of genetic or epigenetic predisposition or possible in utero exposure.

The most important differentially expressed genes in each of our three signatures have been previously associated with TB disease in children[7,8] as well as other TB immune responses in adults, cell culture, and murine models[34–40]. Genes that best predicted TST conversion included *DEFA1* and *DEFA3*, both proposed biomarkers for detecting TB from LTBI in children [7,8]; *HLA-DQAI,* which was associated with protection against pulmonary TB in a prior meta-analysis of TB infection in adults[34]; and *IFITM3*, which is implicated in the restriction of mycobacterial growth[35]. Our top differentially expressed genes for TB progression include *SULT1A3*, which is associated with treatment response for TB in adults[36]; and *NCOA3*, which was differentially expressed in miRNA in adults with TB vs adults hospitalized without a TB diagnosis[37]. Among early TST converters, the top differentially expressed genes between those who did and did not progress to TB include *PARP1* and *KLRD1*. *PARP1* has been implicated in mouse experiments, plays a fundamental role in the host response to TB, and is

hypothesized to contribute the sex differences in response to TB[38]. *KLRD1* has been demonstrated to be a potential T-cell linked biomarker in the progression to TB in mice and macaques[39,40]. Furthermore, it has previously been associated with natural killer cell function in both influenza and TB [33,41].

Of note, both *PARP1* and *KLDR1* have known associations with exposure to cigarette smoke, with PARP1 being implicated in both cellular senescence and lung DNA damage[42,43]. Active and passive smoking have long been associated with TB[44,45], and our previous work with this cohort found an association between maternal smoking and subsequent TB risk[16]. This is the first study to show that umbilical cord gene expression changes are associated both with maternal smoking status and development of TB among children with early TST conversion. The overlap in genes linked to smoking and TB, due to our temporal sampling method for exposure and diagnosis, reflects a biological mechanism that partly clarifies the link between maternal smoking and childhood TB outcomes. While an association between maternal smoking and TB progression did not reach statistical significance (p=0.07), given the small sample size of this study the observed differences may provide insight on the factors that increase TB risk in children. Further studies are needed to characterize this relationship.

Sampling gene expression at birth provides a unique opportunity to study TB pathogenesis preceding exposure. We found a collection of gene signature modules which are associated with TB infection and disease. The most significant of these modules was M11, where pathway analysis indicated genes which primarily implicate neutrophil activation. Previous diagnostic signatures have highlighted neutrophil-driven transcriptional changes as critical in adults with TB[46]. Neutrophils are a critical part of innate immunity and are the primary attackers of bacterial infections, and thus may be important for protection against TB[47]. High neutrophil counts in peripheral blood were highly protective of TB among household contacts [48]. Given

that the M11 gene module was negatively associated with TST conversion and TB within early converters, this adds further evidence that neutrophil activation is important in TB protection. This result was further supported using cell type abundance estimates from CIBERSORTx, indicating that circulating neutrophils were lower among TB progressors compared to children who never TST converted or developed TB.

Clusters of pathways implicated by the M11 gene module include those which are representative of the defensive immune response to bacteria and cellular response to molecules of bacterial origin. Similar pathways are often observed in gene expression studies in patients with TB[49] suggesting that changes in these pathways are already present at birth and may influence TB infection and disease in early childhood.

Limitations of this study include the small sample size. The DGE results specifically are underpowered and should be interpreted cautiously. Modular support with WGCNA provides additional evidence that meaningful and interpretable biology is occurring during or prior to birth that influences early childhood TB outcomes. Future large-scale work with additional clinical and biological data from both mother and infant is essential for further elucidating these mechanisms, particularly in addressing possible confounding from unmeasured characteristics. Additionally, *Mycobacterium* TB infection is likely a heterogenous state and it's possible some of our converters may have been false positives or in early stages of disease. BCG boosting might lead to false positive conversion results; to address this issue, we used a conservative conversion cutoff. Furthermore, any child with a positive skin test reaction did not have a repeat skin test[50]. Furthermore, CIBERSORTx estimates of cell type abundances have not been validated on umbilical cord data and should be interpreted carefully. We are also unable to distinguish whether cord blood gene expression being on the causal pathway to postnatal TB infection or being a biomarker of other exposures that directly alter offspring infection risk

(e.g., maternal smoking[11], HIV[12], and stress[13]). However, these are mitigated by key strengths which include intensive participant follow-up and surveillance for TB infection and disease, as well as excellent phenotyping and cord blood RNA expression measurements. The cohort is representative of many populations with low-income economies, where TB continues to be a major cause of child illness and death.

References

1. Yerramsetti S, Cohen T, Atun R, Menzies NA. Global estimates of paediatric tuberculosis incidence in 2013–19: a mathematical modelling analysis. Lancet Glob Heal **2022**; 10:e207–e215. Available at: http://www.thelancet.com/article/S2214109X21004629/fulltext. Accessed 22 March 2022.
2. The World Health Organization. Global Tuberculosis Report 2020. Geneva: 2020. Available at: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2020. Accessed 22 March 2022.
3. Martinez L, Cords O, Horsburgh CR, et al. The risk of tuberculosis in children after close exposure: a systematic review and individual-participant meta-analysis. Lancet **2020**; 395:973–984. Available at: http://www.thelancet.com/article/S0140673620301665/fulltext. Accessed 22 March 2022.
4. Basu Roy R, Whittaker E, Seddon JA, Kampmann B. Tuberculosis susceptibility and protection in children. Lancet Infect Dis **2019**; 19:e96–e108. Available at: https://pubmed.ncbi.nlm.nih.gov/30322790/. Accessed 22 March 2022.
5. Mandalakas AM, Hesseling AC, Kay A, et al. Tuberculosis prevention in children: a prospective community-based study in South Africa. Eur Respir J **2021**; 57. Available at: /pmc/articles/PMC8060782/. Accessed 22 March 2022.
6. Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. PLoS Med **2019**; 16.
7. Anderson ST, Kaforou M, Brent AJ, et al. Diagnosis of Childhood Tuberculosis and Host RNA Expression in Africa. N Engl J Med **2014**; 370:1712–1723. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24785206. Accessed 30 July 2018.
8. Verhagen LM, Zomer A, Maes M, et al. A predictive signature gene set for discriminating active from latent tuberculosis in Warao Amerindian children. BMC Genomics **2013**; 14.
9. Tornheim JA, Madugundu AK, Paradkar M, et al. Transcriptomic Profiles of Confirmed Pediatric Tuberculosis Patients and Household Contacts Identifies Active Tuberculosis, Infection, and Treatment Response Among Indian Children. J Infect Dis **2020**; 221:1647–1658. Available at: https://pubmed.ncbi.nlm.nih.gov/31796955/. Accessed 22 March 2022.
10. Möller M, Kinnear CJ. Human global and population-specific genetic susceptibility to Mycobacterium tuberculosis infection and disease. Curr Opin Pulm Med **2020**; 26:302–310. Available at: https://pubmed.ncbi.nlm.nih.gov/32101905/. Accessed 3 July 2022.
11. Kessous R, Sheiner E, Wainstock T. 727: Maternal smoking during pregnancy and the risk for childhood infectious diseases in the offspring. Am J Obstet Gynecol **2019**; 220:S478–S479. Available at: http://www.ajog.org/article/S0002937818317721/fulltext. Accessed 3 July 2022.
12. Le Roux DM, Nicol MP, Myer L, et al. Lower Respiratory Tract Infections in Children in a Well-vaccinated South African Birth Cohort: Spectrum of Disease and Risk Factors. Clin Infect Dis **2019**; 69:1588–1596. Available at: https://pubmed.ncbi.nlm.nih.gov/30925191/. Accessed 3 July 2022.
13. Robinson M, Carter KW, Pennell CE, et al. Maternal prenatal stress exposure and sex-specific risk of severe infection in offspring. PLoS One **2021**; 16. Available at: /pmc/articles/PMC7845992/. Accessed 3 July 2022.
14. Zar HJ, Barnett W, Myer L, Stein DJ, Nicol MP. Investigating the early-life determinants of illness in Africa: the Drakenstein Child Health Study. Thorax **2015**; 70:592–594.

Available at: https://pubmed.ncbi.nlm.nih.gov/25228292/. Accessed 22 March 2022.
15. Martinez L, Nicol MP, Wedderburn CJ, et al. Cytomegalovirus acquisition in infancy and the risk of tuberculosis disease in childhood: a longitudinal birth cohort study in Cape Town, South Africa. Lancet Glob Heal **2021**; 9:e1740–e1749. Available at: http://www.thelancet.com/article/S2214109X21004071/fulltext. Accessed 22 March 2022.
16. Martinez L, le Roux DM, Barnett W, Stadler A, Nicol MP, Zar HJ. Tuberculin skin test conversion and primary progressive tuberculosis disease in the first 5 years of life: a birth cohort study from Cape Town, South Africa. Lancet Child Adolesc Heal **2018**; 2:46–55. Available at: http://www.thelancet.com/article/S2352464217301499/fulltext. Accessed 22 March 2022.
17. Martinez L, Gray DM, Botha M, et al. The Long-Term Impact of Early-Life Tuberculosis Disease on Child Health: A Prospective Birth Cohort Study. Am J Respir Crit Care Med **2023**; Available at: https://pubmed.ncbi.nlm.nih.gov/36746196/. Accessed 18 March 2023.
18. Zar HJ, Martinez L, Ncayiyana JR, et al. Vitamin D Concentrations in Infancy and the Risk of Tuberculosis Disease in Childhood: A Prospective Birth Cohort in Cape Town, South Africa. Clin Infect Dis **2022**; 74:2036–2043. Available at: https://academic.oup.com/cid/article/74/11/2036/6358095. Accessed 19 June 2022.
19. Graham SM, Cuevas LE, Jean-Philippe P, et al. Clinical Case Definitions for Classification of Intrathoracic Tuberculosis in Children: An Update. Clin Infect Dis **2015**; 61:S179–S187. Available at: https://academic.oup.com/cid/article-lookup/doi/10.1093/cid/civ581. Accessed 27 August 2019.
20. Breen MS, Wingo AP, Koen N, et al. Gene expression in cord blood links genetic risk for neurodevelopmental disorders with maternal psychological distress and adverse childhood outcomes. Brain Behav Immun **2018**; 73:320–330. Available at: /pmc/articles/PMC6191930/?report=abstract. Accessed 21 December 2020.
21. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res **2015**; 43:e47–e47. Available at: http://academic.oup.com/nar/article/43/7/e47/2414268/limma-powers-differential-expression-analyses-for. Accessed 30 July 2018.
22. Althouse AD. Adjust for Multiple Comparisons? It's Not That Simple. Ann Thorac Surg **2016**; 101:1644–1645. Available at: http://dx.doi.org/10.1016/j.athoracsur.2015.11.024. Accessed 16 December 2020.
23. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A **2005**; 102:15545–15550. Available at: https://dartmouth.sharepoint.com/sites/networkJC/Lists/Papers/Attachments/3/Gene set enrichment analysis- A knowledge-based approach for interpreting genome-wide expression profiles.pdf. Accessed 6 June 2018.
24. Bobak CA, Abhimanyu, Natarajan H, et al. Increased DNA methylation, cellular senescence and premature epigenetic aging in guinea pigs and humans with tuberculosis. Aging (Albany NY) **2022**; 14:2174–2193. Available at: https://pubmed.ncbi.nlm.nih.gov/35256539/. Accessed 24 July 2022.
25. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics **2008**; 9:559. Available at: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559. Accessed 24 December 2020.
26. Raudvere U, Kolberg L, Kuzmin I, et al. G:Profiler: A web server for functional enrichment

analysis and conversions of gene lists (2019 update). Nucleic Acids Res **2019**; 47:W191–W198. Available at: https://academic.oup.com/nar/article/47/W1/W191/5486750. Accessed 30 December 2020.
27. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. PLoS One **2010**; 5:e13984. Available at: http://dx.plos.org/10.1371/journal.pone.0013984. Accessed 30 July 2018.
28. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol **2019**; 37:773–782. Available at: https://doi.org/10.1038/s41587-019-0114-2. Accessed 11 March 2021.
29. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res **2002**; 30:207–210. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/30.1.207. Accessed 26 May 2018.
30. Pearson K. LIII. On lines and planes of closest fit to systems of points in space . London, Edinburgh, Dublin Philos Mag J Sci **1901**; 2:559–572. Available at: https://www.tandfonline.com/doi/abs/10.1080/14786440109462720. Accessed 4 January 2021.
31. Wilcoxon F. Individual Comparisons by Ranking Methods. Biometrics Bull **1945**; 1:80.
32. Lance GN, Williams WT. Computer Programs for Hierarchical Polythetic Classification ('Similarity Analyses'). Comput J **1966**; 9:60–64. Available at: https://academic.oup.com/comjnl/article-lookup/doi/10.1093/comjnl/9.1.60. Accessed 31 January 2021.
33. Bongen E, Vallania F, Utz PJ, Khatri P. KLRD1-expressing natural killer cells predict influenza susceptibility. Genome Med **2018**; 10.
34. Oliveira-Cortez A, Melo AC, Chaves VE, Condino-Neto A, Camargos P. Do HLA class II genes protect against pulmonary tuberculosis? A systematic review and meta-analysis. Eur J Clin Microbiol Infect Dis **2016**; 35:1567–1580. Available at: https://pubmed.ncbi.nlm.nih.gov/27412154/. Accessed 22 March 2022.
35. Ranjbar S, Haridas V, Jasenosky LD, Falvo J V., Goldfeld AE. A Role for IFITM Proteins in Restriction of Mycobacterium tuberculosis Infection. Cell Rep **2015**; 13:874. Available at: /pmc/articles/PMC4916766/. Accessed 22 March 2022.
36. O'Garra A, Bloom C, Barry MPR, et al. Early detection of tuberculosis treatment response. 2015; Available at: https://patents.google.com/patent/US20150133469A1/en. Accessed 22 March 2022.
37. Silva CA, Ribeiro-Dos-santos A, Gonçalves WG, et al. Can miRNA Indicate Risk of Illness after Continuous Exposure to M. tuberculosis? Int J Mol Sci **2021**; 22. Available at: https://pubmed.ncbi.nlm.nih.gov/33916069/. Accessed 22 March 2022.
38. Krug S, Ordonez AA, Klunk M, et al. Host regulator PARP1 contributes to sex differences and immune responses in a mouse model of tuberculosis. bioRxiv **2021**; :2021.04.21.440820. Available at: https://www.biorxiv.org/content/10.1101/2021.04.21.440820v1. Accessed 22 March 2022.
39. Esaulova E, Das S, Singh DK, et al. The immune landscape in tuberculosis reveals populations linked to disease and latency. Cell Host Microbe **2021**; 29:165-178.e8.
40. Moreira-Teixeira L, Tabone O, Graham CM, et al. Mouse transcriptome reveals potential signatures of protection and pathogenesis in human tuberculosis. Nat Immunol 2020 214 **2020**; 21:464–476. Available at: https://www.nature.com/articles/s41590-020-0610-z. Accessed 22 March 2022.

41. Chowdhury RR, Vallania F, Yang Q, et al. A multi-cohort study of the immune factors associated with M. tuberculosis infection outcomes. Nature **2018**; 560.
42. Yao H, Sundar IK, Gorbunova V, Rahman I. P21-PARP-1 pathway is involved in cigarette smoke-induced lung DNA damage and cellular senescence. PLoS One **2013**; 8. Available at: https://pubmed.ncbi.nlm.nih.gov/24244594/. Accessed 22 March 2022.
43. Arimilli S, Madahian B, Chen P, Marano K, Prasad GL. Gene expression profiles associated with cigarette smoking and moist snuff consumption. BMC Genomics **2017**; 18. Available at: /pmc/articles/PMC5307792/. Accessed 22 March 2022.
44. Den Boon S, Verver S, Marais BJ, et al. Association Between Passive Smoking and Infection With Mycobacterium tuberculosis in Children. Pediatrics **2007**; 119:734–739. Available at: /pediatrics/article/119/4/734/70153/Association-Between-Passive-Smoking-and-Infection. Accessed 22 March 2022.
45. Jayes L, Haslam PL, Gratziou CG, et al. SmokeHaz: Systematic Reviews and Meta-analyses of the Effects of Smoking on Respiratory Health. Chest **2016**; 150:164–179.
46. Berry MPR, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature **2010**; 466:973–977. Available at: https://www.nature.com/articles/nature09247. Accessed 24 May 2018.
47. Kroon EE, Coussens AK, Kinnear C, et al. Neutrophils: Innate Effectors of TB Resistance? Front Immunol **2018**; 9:2637. Available at: /pmc/articles/PMC6246713/. Accessed 22 March 2022.
48. Leisching GR. Susceptibility to Tuberculosis is associated with PI3K-dependent increased mobilization of neutrophils. Front Immunol **2018**; 9:1669.
49. Bobak C. Ab2olish Tb: Analysis of Blood and Breath 'Omics to Lend Insights and Strategies for Hindering Tuberculosis - ProQuest. 2021; Available at: https://www.proquest.com/docview/2572581105/abstract/D03C8E1FD3A44F5FPQ/1?accountid=10422. Accessed 22 March 2022.
50. Farhat M, Greenaway C, Pai M, Menzies D. False-positive tuberculin skin tests: What is the absolute effect of BCG and non-tuberculous mycobacteria? Int. J. Tuberc. Lung Dis. 2006; 10.

Acknowledgements

We thank the study staff, the clinical and administrative staff of the Western Cape Government Health Department at Paarl Hospital and at the clinics for support of the study, members of the study International Advisory Board for their advice, our collaborators, and the families and children who participated in the study. This study was funded by the Bill & Melinda Gates Foundation (grant number OPP 1017641), Medical Research Council South Africa, National Research Foundation South Africa, and the Wellcome Trust.

603 Figures and Tables

604
605 Table 1. Clinical measures for all included study participants (N=131) by TST conversion in
606 infants in the first 36 months of life
607
608
609 Figure 1. Differential gene expression results between infants with TST conversion and infants
610 who do not convert conversion before 36 months of age.
611
612 Figure 2. An unsupervised heatmap with infants who did (blue) and and did not convert their
613 tuberculin skin test prior to 36 months of age (yellow).
614
615 Figure 3: Differential gene expression results between infants who were diagnosed with TB
616 before the age of 5 from those without a TB diagnosis among those who experienced TST
617 conversion before 36 months of age.
618
619 Figure 4: Results from a pathway overlap analysis of all three TB hypotheses and HIV-adjust
620 maternal smoking. Pathways were identified using GSEA[23]. (A) A Venn diagram of all
621 pathways that overlapped between each DEG analysis. (B) The pathway response Z-score of
622 the Defense Response pathway (GO:0006952) across TB outcomes. (C) The pathway response
623 Z-score of the Regulation of Immune Response pathway (GO:0002682) across TB outcomes.
624 (D) A bubble plot illustrating normalized effect size (NES) and p-values of all pathways that
625 were either enriched or depleted in each DEG analysis.
626
627 Figure 5: Weighted Gene Co-expression Network analysis. (A) The 14 discovered gene
628 expression modules and their association with early TST conversion. Error bars represent one
629 standard error on either side of the mean. (B) A regression between the significance of genes in
630 the M11 module with early TST conversion and the gene connectivity within the module. (C) A
631 boxplot of the M11 module eigengene by maternal prior TB diagnosis. (D) A boxplot of the M11
632 eigengene between early TST converters, early converters who developed TB, and children
633 who did not convert their TST in early childhood and who did not develop TB within their first 5
634 years of age.
635
636 Figure 6: Three extracted subnetworks from the WGCNA enrichment results. Nodes represent
637 the biological function, where edges indicate significant overlap of genes associated with the
638 biological functions. Pathways listed are from GO, Reactome, Wikipathways, and KEGG. (A) A
639 subnetwork of biological pathways related to the defense response to bacteria. (B) A
640 subnetwork of responses related to the interaction of the host and symbiont cells. (C) A
641 subnetwork of cellular responses to molecules of bacterial origins. (D) A barplot of select
642 pathways plotted against their gProfiler p-value[26].

643 Table 1. Clinical measures for all included study participants (N=131) by tuberculin skin test conversion in infants
644

	Early TST Converters (N=26)	Children who did not convert their TST (N=105)	Overall (N=131)	P-value
Maternal Characteristics				
Maternal HIV infected	3 (11.5)	30 (28.6)	33 (25.2)	0.12
Tuberculosis Treatment During Pregnancy	1 (3.8)	4 (3.8)	5 (3.8)	1
Prior Maternal Tuberculosis	0 (0)	5 (4.8)	5 (3.8)	0.57
Tuberculosis in Household 1 Year Prior Enrollment	2 (7.7)	14 (13.3)	16 (12.2)	0.64
Maternal Smoking	9 (34.6)	28 (26.7)	37 (28.2)	0.49
Median socioeconomic status score	-0.11 (-1.58, 1.27)	-0.07 (-1.63, 1.27)	-0.11 (-1.58, 1.26)	0.456
Household income, rand per month				0.579
<1000	13 (50.0)	41 (39.1)	54 (41.2)	
1000-5000	10 (38.5)	47 (44.8)	57 (43.5)	
>5000	3 (11.5)	17 (16.2)	20 (15.3)	
Child Characteristics				
Female sex	11 (42.3)	47 (44.8)	58 (44.3)	1.0
Mean Birthweight (kg)	3.07 (0.46)	3.16 (0.53)	3.14 (0.52)	0.38
Median Breastfeeding, months (min, max)	2 (0, 7)	1 (0, 8)	1 (0,8)	0.14
Tuberculosis diagnosis <5 years old	10 (38.5)	6 (3.8)	16 (12.2)	<0.001
Median weight-for-age z-scores at 5 years of age	-0.51 (-1.33, 0.29)	-0.48 (-1.16, 0.43)	-0.07 (-1.21, 1.64)	1.0
Median height-for-age z-scores at 5 years of age	-0.67 (-1.24, -0.15)	-0.58 (-1.27, 0.10)	-0.07 (-1.21, 1.64)	0.647

Abbreviation: TST, tuberculin skin test.

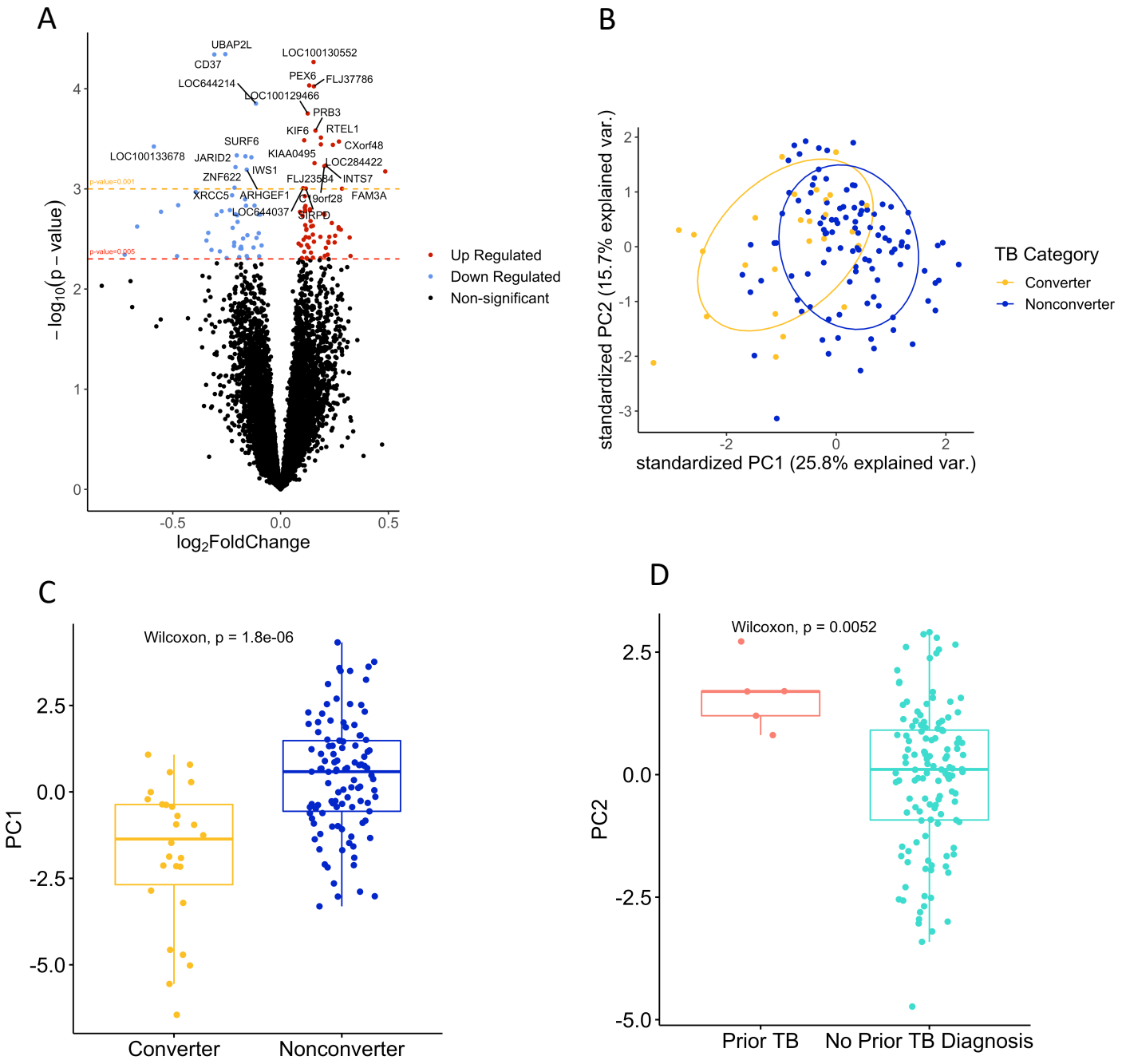


Figure 1: Differential gene expression results between infants with and without early TST conversion. (A) A volcano plot showing differentially up- and down-regulated genes between early converters and non-converting infants. (B) PCA using significantly differentially regulated genes from (A) where early converters are indicated in yellow and those who did not convert are indicated in blue. (C,D) A boxplot demonstrating statistically significant differences in the first principal component in (B) between (C) between infants with and without early TST conversion and (D) between mothers with a known prior TB diagnosis and mothers with no known prior TB diagnosis.

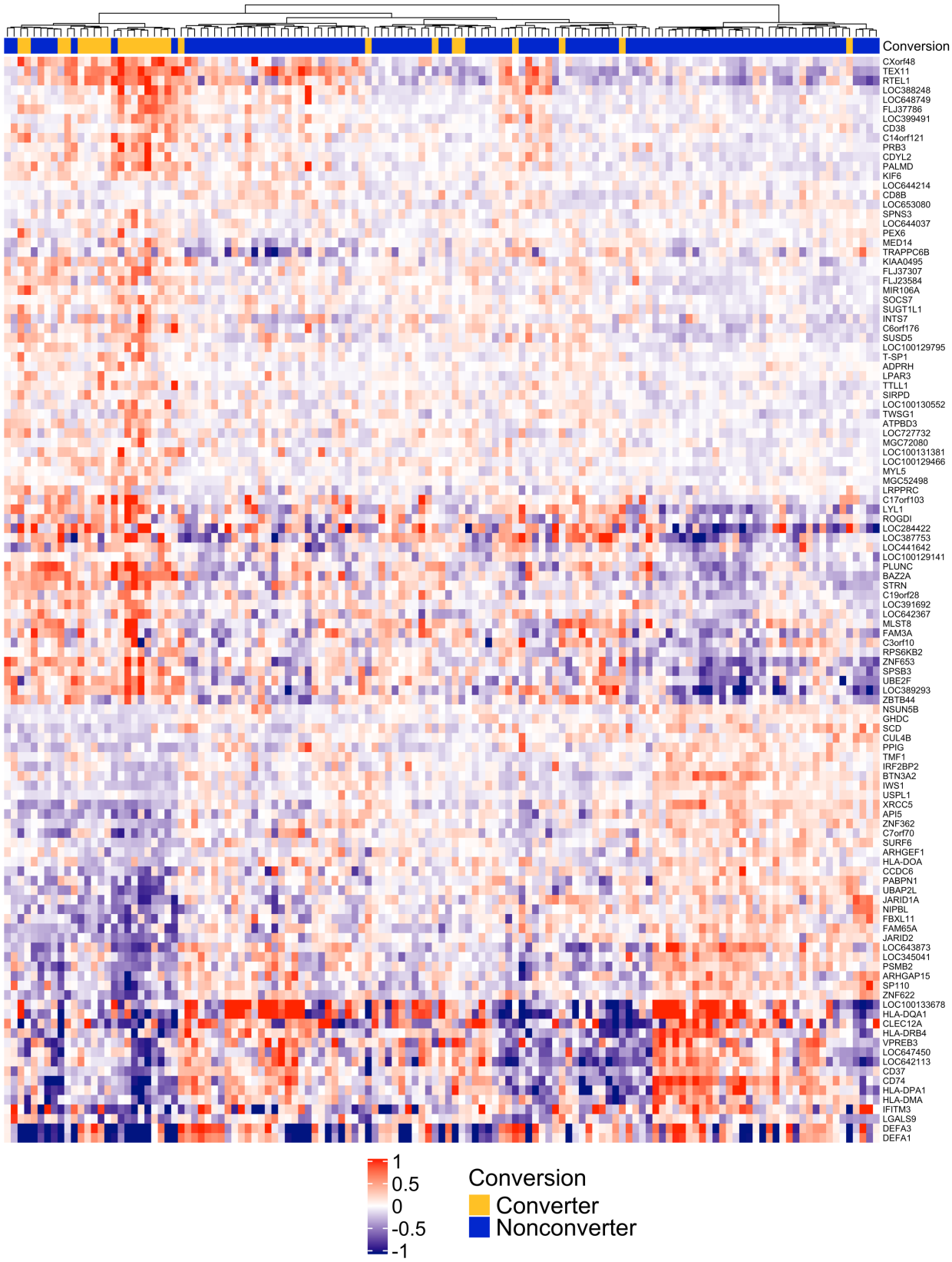

Conversion
CXorf48
TEX11
RTEL1
LOC388248
LOC648749
FLJ37786
LOC399491
CD38
C14orf121
PRB3
CDYL2
PALMD
KIF6
LOC644214
CD8B
LOC653080
SPNS3
LOC644037
PEX6
MED14
TRAPPC6B
KIAA0495
FLJ37307
FLJ23584
MIR106A
SOCS7
SUGT1L1
INTS7
C6orf176
SUSD5
LOC100129795
T-SP1
ADPRH
LPAR3
TTLL1
SIRPD
LOC100130552
TWSG1
ATPBD3
LOC727732
MGC72080
LOC100131381
LOC100129466
MYL5
MGC52498
LRPPRC
C17orf103
LYL1
ROGDI
LOC284422
LOC387753
LOC441642
LOC100129141
PLUNC
BAZ2A
STRN
C19orf28
LOC391692
LOC642367
MLST8
FAM3A
C3orf10
RPS6KB2
ZNF653
SPSB3
UBE2F
LOC389293
ZBTB44
NSUN5B
GHDC
SCD
CUL4B
PPIG
TMF1
IRF2BP2
BTN3A2
IWS1
USPL1
XRCC5
API5
ZNF362
C7orf70
SURF6
ARHGEF1
HLA-DOA
CCDC6
PABPN1
UBAP2L
JARID1A
NIPBL
FBXL11
FAM65A
JARID2
LOC643873
LOC345041
PSMB2
ARHGAP15
SP110
ZNF622
LOC100133678
HLA-DQA1
CLEC12A
HLA-DRB4
VPREB3
LOC647450
LOC642113
CD37
CD74
HLA-DPA1
HLA-DMA
IFITM3
LGALS9
DEFA3
DEFA1
1
0.5
0
-0.5
-1
Conversion
Converter
Nonconverter

Figure 2. An unsupervised heatmap with infants who did (blue) or did not convert their tuberculin skin test in early childhood (yellow). Each row represents a significantly associated gene (p-value < 0.005) and each column is one umbilical cord blood sample. Expression values were median centered. Both columns and rows were clustered using Canberra distance.

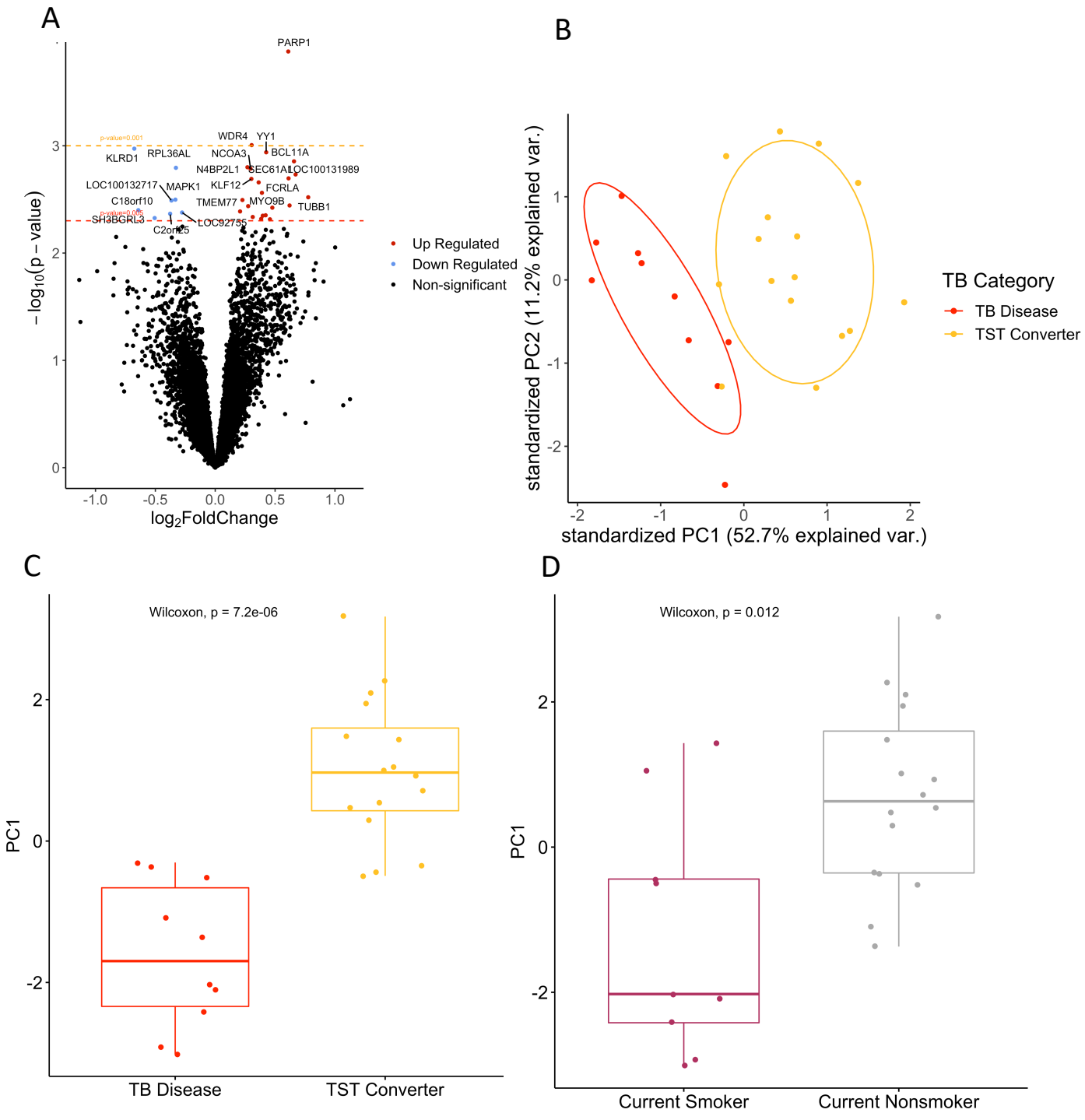


Figure 3: The results from DGE analysis between children who were diagnosed with TB before the age of 5 among all early TST converters. A) The volcano plot showing the effect size by the -$\log_{10}$ p-value. B) The PCA using just the significantly differentially expressed genes C) a boxplot showing statistically significant differences along PC1 between children diagnosed with TB before age 5 and those who were not and D) a boxplot showing statistically significant differences between mothers who were smokers at the time of enrollment and mothers who were not smokers at time of enrollments.

692

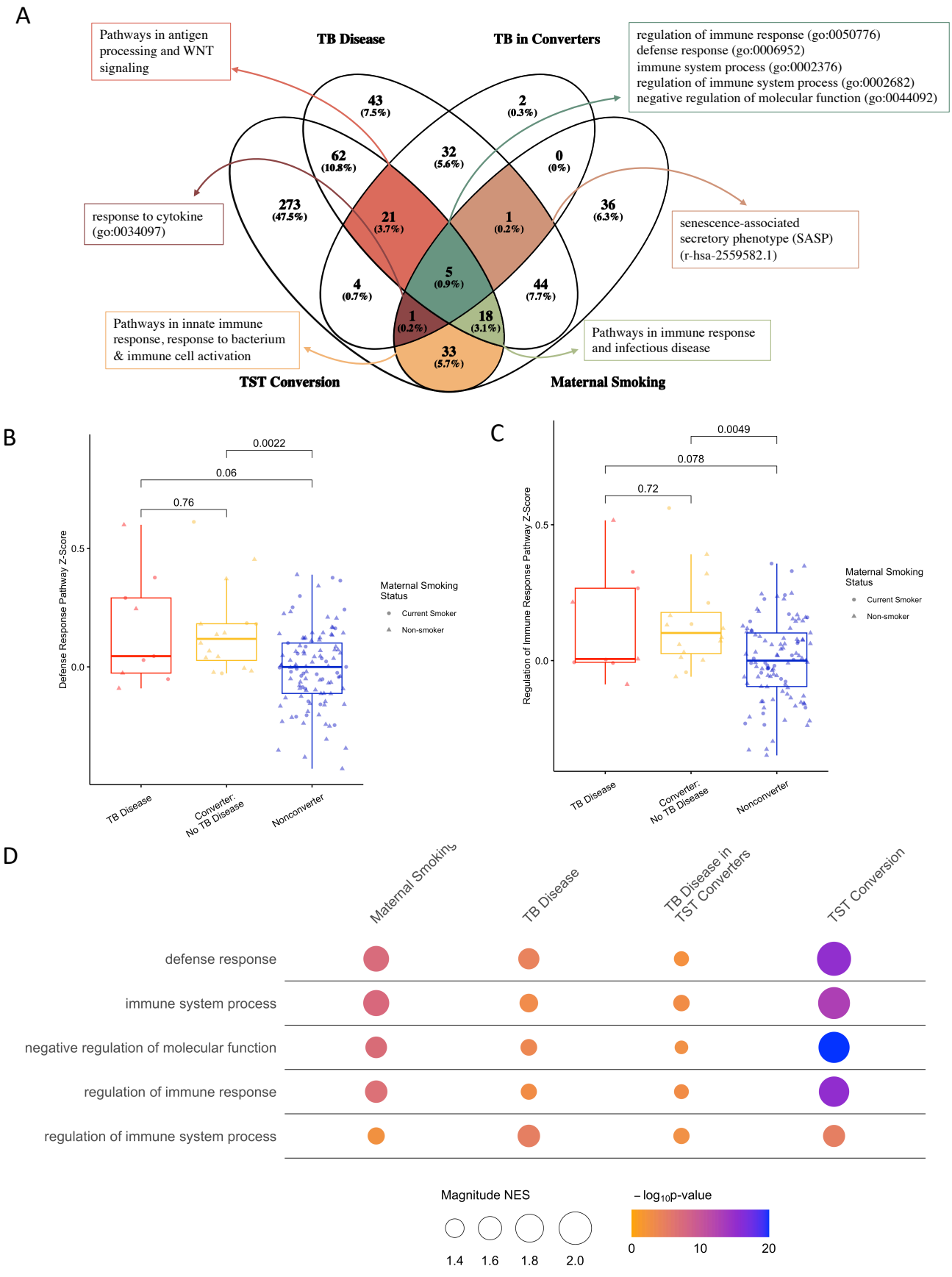
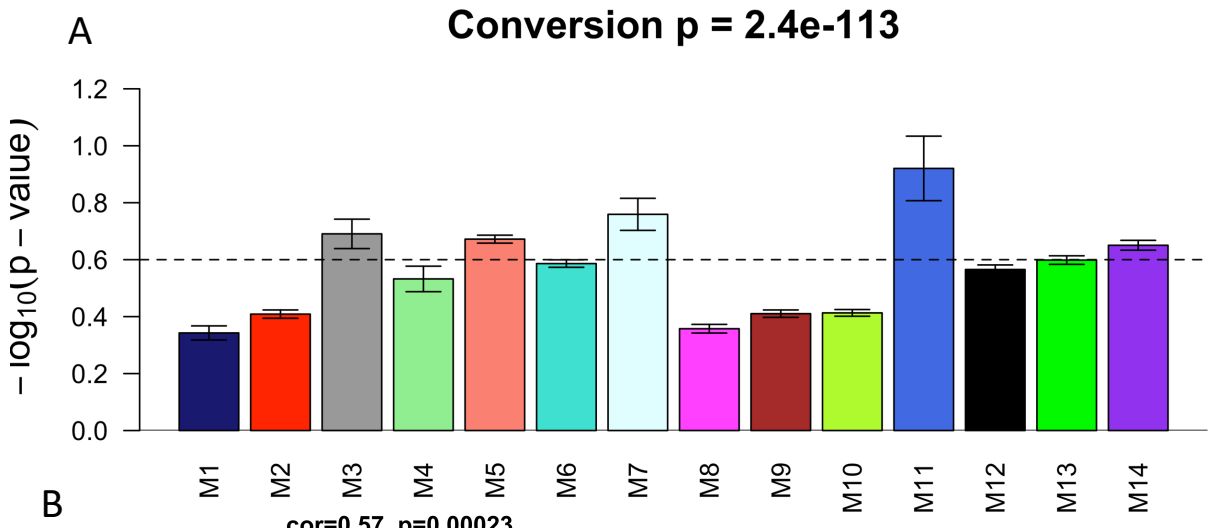
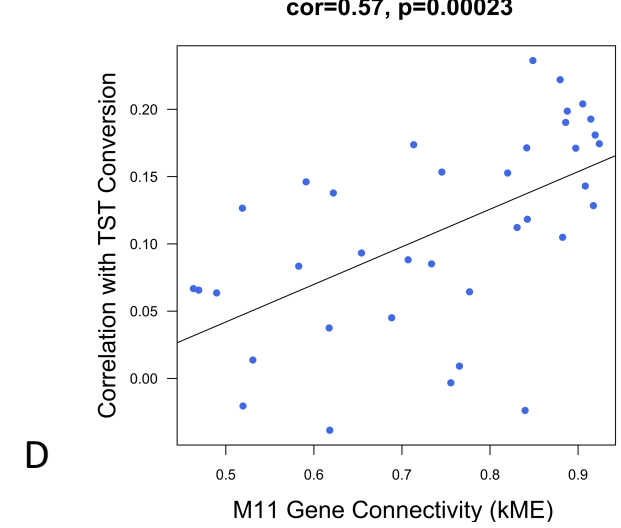
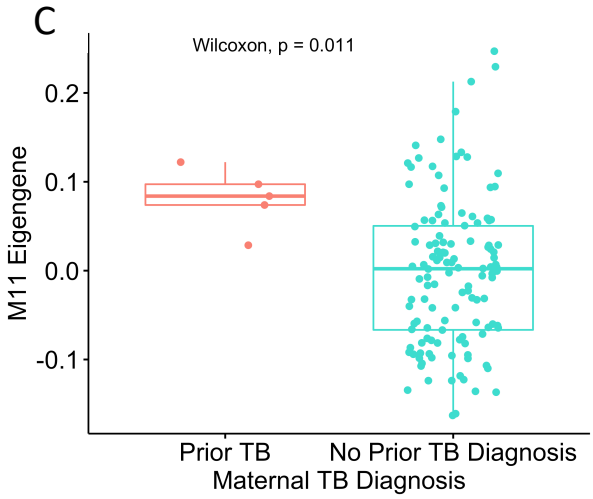

A
Pathways in antigen processing and WNT signaling
TB Disease
TB in Converters
regulation of immune response (go:0050776)
defense response (go:0006952)
immune system process (go:0002376)
regulation of immune system process (go:0002682)
negative regulation of molecular function (go:0044092)
43 (7.5%)
2 (0.3%)
62 (10.8%)
32 (5.6%)
0 (0%)
273 (47.5%)
21 (3.7%)
1 (0.2%)
36 (6.3%)
response to cytokine (go:0034097)
senescence-associated secretory phenotype (SASP) (r-hsa-2559582.1)
5 (0.9%)
4 (0.7%)
44 (7.7%)
1 (0.2%)
18 (3.1%)
33 (5.7%)
Pathways in innate immune response, response to bacterium & immune cell activation
Pathways in immune response and infectious disease
TST Conversion
Maternal Smoking
B
0.0022
0.06
0.76
Defense Response Pathway Z-Score
0.5
0.0
Maternal Smoking Status
Current Smoker
Non-smoker
TB Disease
Converter: No TB Disease
Nonconverter
C
0.0049
0.078
0.72
Regulation of Immune Response Pathway Z-Score
0.5
0.0
Maternal Smoking Status
Current Smoker
Non-smoker
TB Disease
Converter: No TB Disease
Nonconverter
D
Maternal Smoking
TB Disease
TB Disease in TST Converters
TST Conversion
defense response
immune system process
negative regulation of molecular function
regulation of immune response
regulation of immune system process
Magnitude NES
1.4
1.6
1.8
2.0
– log10p-value
0
10
20

Figure 4: Results from a pathway overlap analysis of all three TB hypotheses and HIV-adjust maternal smoking. Pathways were identified using GSEA[23]. (A) A Venn diagram of all pathways that overlapped between each DEG analysis. (B) The pathway response Z-score of the Defense Response pathway (GO:0006952) across TB outcomes. (C) The pathway response Z-score of the Regulation of Immune Response pathway (GO:0002682) across TB outcomes. (D) A bubble plot illustrating normalized effect size (NES) and p-values of all pathways that were either enriched or depleted in each DEG analysis.


A
Conversion p = 2.4e-113
−log10(p − value)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14


B
cor=0.57, p=0.00023
Correlation with TST Conversion
0.20
0.15
0.10
0.05
0.00
0.5
0.6
0.7
0.8
0.9
M11 Gene Connectivity (kME)


C
Wilcoxon, p = 0.011
M11 Eigengene
0.2
0.1
0.0
-0.1
Prior TB
No Prior TB Diagnosis
Maternal TB Diagnosis

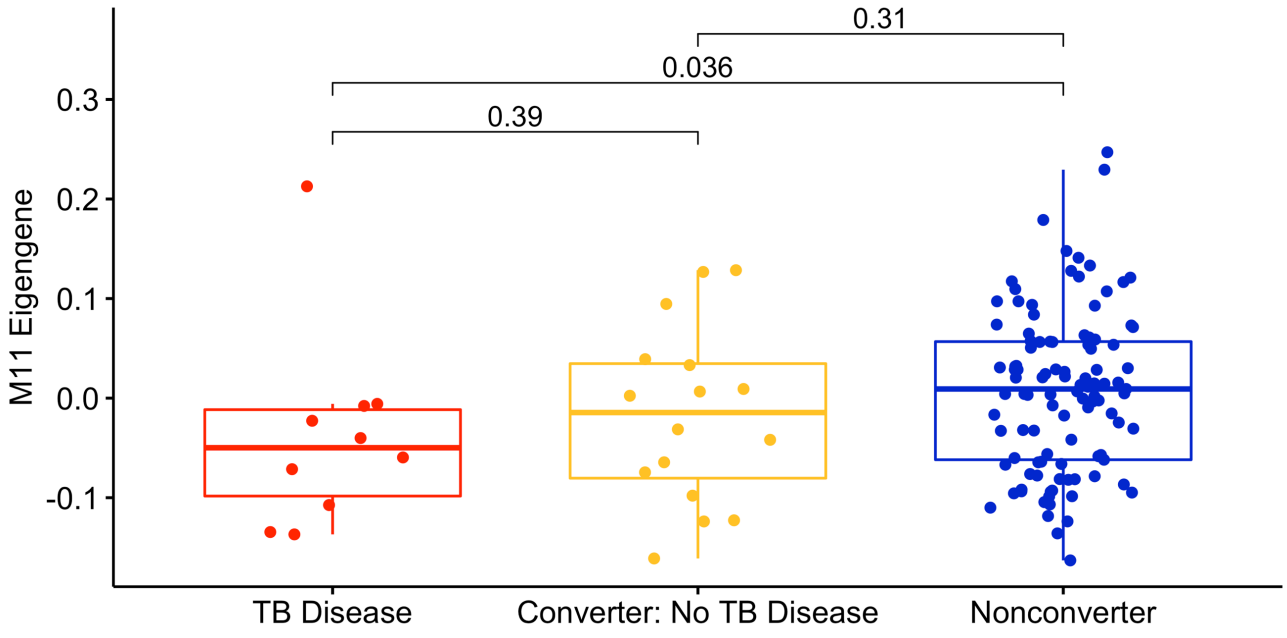

D
0.31
0.036
0.39
M11 Eigengene
0.3
0.2
0.1
0.0
-0.1
TB Disease
Converter: No TB Disease
Nonconverter

Figure 5: Weighted Gene Co-expression Network analysis. (A) The 14 discovered gene expression modules and their association with early TST conversion. Error bars represent one standard error on either side of the mean. (B) A regression between the significance of genes in the M11 module with early TST conversion and the gene connectivity within the module. (C) A boxplot of the M11 module eigengene by maternal prior TB diagnosis. (D) A boxplot of the M11 eigengene between early TST converters, early converters who developed TB, and children who did not convert their TST in early childhood and who did not develop TB within their first 5 years of age.

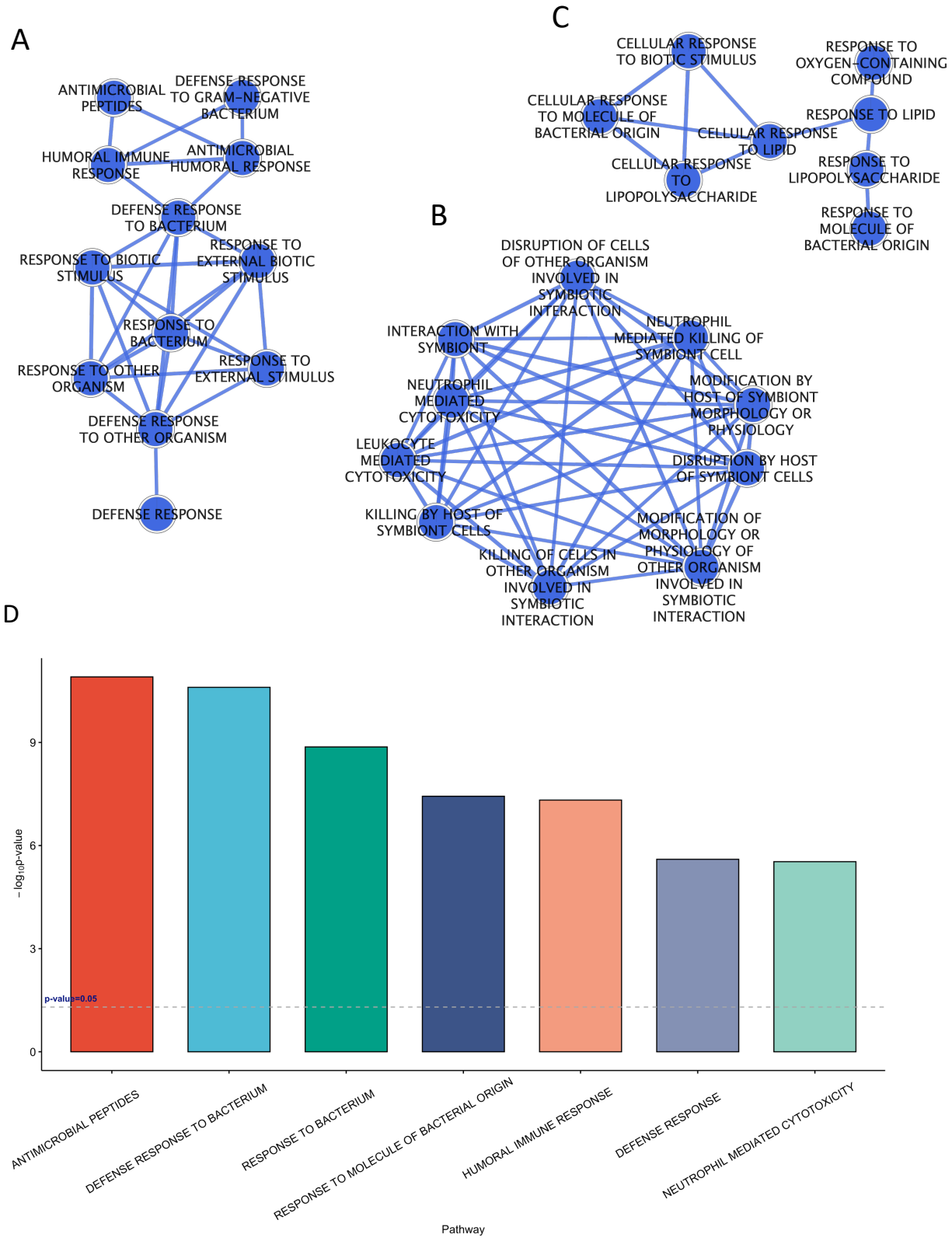


Figure 6: Three extracted subnetworks from the WGCNA enrichment results. Nodes represent the biological function, where edges indicate significant overlap of genes associated with the

biological functions. Pathways listed are from GO, Reactome, Wikipathways, and KEGG. (A) A subnetwork of biological pathways related to the defense response to bacteria. (B) A subnetwork of responses related to the interaction of the host and symbiont cells. (C) A subnetwork of cellular responses to molecules of bacterial origins. (D) A barplot of select pathways plotted against their gProfiler p-value[26].