



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/152556/>

Version: Published Version

Article:

Bah, S.Y., Forster, T., Dickinson, P. et al. (2018) Meta-analysis identification of highly robust and differential immune-metabolic signatures of systemic host response to acute and latent tuberculosis in children and adults. *Frontiers in Genetics*, 9. 457.

<https://doi.org/10.3389/fgene.2018.00457>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Meta-Analysis Identification of Highly Robust and Differential Immune-Metabolic Signatures of Systemic Host Response to Acute and Latent Tuberculosis in Children and Adults

Saikou Y. Bah^{1,2,3}, Thorsten Forster¹, Paul Dickinson¹, Beate Kampmann^{3,4*} and Peter Ghazal^{1,5*}

¹ Division of Pathway Medicine and Edinburgh Infectious Diseases, University of Edinburgh Medical School, Edinburgh, United Kingdom, ² West African Centre for Cellular Biology of Infectious Pathogens, University of Ghana, Accra, Ghana, ³ Vaccines and Immunity Theme, Medical Research Council Unit The Gambia at the London School of Tropical Medicine and Hygiene, Banjul, Gambia, ⁴ Centre of International Child Health, Department of Paediatrics, Imperial College London, London, United Kingdom, ⁵ Systems Immunity Research Institute, School of Medicine Laboratory of Immunity and Metabolism, University of Cardiff, Wales, United Kingdom

OPEN ACCESS

Edited by:

Ramana V. Davuluri,
Northwestern University,
United States

Reviewed by:

Sudesh Pundir,
Pondicherry University, India
Yingtao Bi,
AbbVie (United States), United States

*Correspondence:

Beate Kampmann
bkampmann@mrcc.gm
Peter Ghazal
ghazalp@cardiff.ac.uk

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Genetics

Received: 16 July 2018

Accepted: 18 September 2018

Published: 04 October 2018

Citation:

Bah SY, Forster T, Dickinson P, Kampmann B and Ghazal P (2018) Meta-Analysis Identification of Highly Robust and Differential Immune-Metabolic Signatures of Systemic Host Response to Acute and Latent Tuberculosis in Children and Adults. *Front. Genet.* 9:457. doi: 10.3389/fgene.2018.00457

Background: Whole blood expression profiling is a mainstay for delineating differential diagnostic signatures of infection yet is subject to high variability that reduces power and complicates clinical usefulness. To date, confirmatory high confidence expression profiling signatures for clinical use remain uncertain. Here we have sought to evaluate the reproducibility and confirmatory nature of differential expression signatures, comprising molecular and cellular pathways, across multiple international clinical observational studies investigating children and adult whole blood transcriptome responses to tuberculosis (TB).

Methods and findings: A systematic search and quality control assessment of gene expression repositories for human TB using whole blood resulted in 11 datasets with a total of 1073 patients from Africa, Europe, and South America. A non-parametric estimation of percentage of false prediction was used for meta-analysis of high confidence differential expression analysis. Deconvolution analysis was applied to infer changes in immune cell proportions and enrichment tests applied using pathway database resources. Meta-analysis identified high confidence differentially expressed genes, comprising 372 in adult active-TB versus latent-TB (LTBI), 332 in adult active-TB versus controls (CON), five in LTBI versus CON, and 415 in childhood active-TB versus LTBI. Notably, these confirmatory markers have low representation in published signatures for diagnosing TB. Pathway biology analysis of high confidence gene sets revealed dominant metabolic and innate-immune pathway signatures while suppressed signatures were enriched with adaptive signaling pathways and reduced proportions of T and B cells. Childhood TB showed uniquely strong inflammasome antagonist signature (*IL1RN* and *ILR2*), while adult TB patients exhibit a significant preponderance type I and type II IFN markers. Key limitations of the study include the paucity of data on potential confounders.

Conclusion: Meta-analysis identified high confidence confirmatory immune-metabolic and cellular expression signatures across studies regardless of the population resource setting, HIV status and circulating endemic pathogens. Notably, previously identified diagnostic signature markers for TB show limited concordance with the confirmatory meta-analysis. Overall, our results support the use of the confirmatory expression signatures for guiding optimized diagnostic, prognostic, and therapeutic monitoring modalities in TB.

Keywords: tuberculosis, meta-analysis, immunity, systemic responses, microarray, bioinformatics

INTRODUCTION

Tuberculosis caused by members of the *Mycobacterium tuberculosis* complex is a leading cause of morbidity and mortality due to an infectious agent despite the availability of potent antimycobacterials (WHO, 2015). About a third of the world population is latently infected with *M. tuberculosis* with a life time risk of re-activation to cause active TB and transmission to other susceptible hosts. BCG, the only licensed TB vaccine has variable efficacy ranging from 0–80% with limited efficacy in preventing pulmonary TB in adults who are largely responsible for disease transmission (Colditz et al., 1994; Karonga Prevention Trial Group, 1996). The control of TB is further complicated due to the emergence of antimycobacterial resistance and co-infection with HIV, with TB as the leading cause of death in HIV co-infected patients (Bruchfeld et al., 2015). Furthermore, the host immune response to mycobacterial infection although thought to be primarily T cell mediated is very complex and remains incompletely understood (Cooper, 2009; Dwivedi et al., 2012) and there are currently no validated immune correlates of protection.

To develop the more efficacious vaccines identified as essential for global TB control, it is important to understand the host immune response to *M. tuberculosis*. A promising approach to interrogate the host response at a systemic level in order to identify correlates of protection is systemic analyses of host transcriptional responses, using methods such as microarrays and increasingly RNA sequencing.

Several published studies have used genome-wide transcriptomics to investigate the host systemic response to TB, pathway biology and to identify diagnostic signatures to distinguish active TB from LTBI, uninfected controls, and other diseases (Berry et al., 2010; Maertzdorf et al., 2011a; Bloom et al., 2012; Kaforou et al., 2013; Anderson et al., 2014; Tientcheu et al., 2015). These studies were conducted in different geographical regions, and HIV infection status. Previous studies have reported that a comparison of independently identified differentially expressed genes and diagnostic signatures revealed that less than 80% of differentially regulated genes and only about 60% of interferon (IFN) gamma signature genes were consistently found in all studies (Maertzdorf et al., 2011a). In analyzing whole blood transcriptomics, it is critical to account for cellular

heterogeneity yet an investigation of cellular proportions and pathways by applying cell-specific deconvolution algorithms remains lacking for current TB expression profiling studies. Moreover, independent TB transcriptomics studies conducted for adults and children can be further complicated by the different clinical manifestations, prior history of infection, and for these reasons maybe anticipated to have different systemic gene expression signatures. In this connection, it is worth noting that a comparison between adult and childhood TB expression profiles has not been reported. Moreover, given the wide age range, population, and cellular heterogeneity of these large data-rich studies, it is not too surprising that the clinical usefulness of expression signatures remains uncertain.

Meta-analyses can provide a statistically stringent and powerful approach to integrate and computationally deconvolute large independent datasets and thereby infer with high confidence new and consistently differentially expressed genes and pathways across multiple studies and age groups. Very recently, several studies have conducted a meta-analysis of a subset of publicly available TB transcriptomic studies; however, none of these meta-analyses studies involved a complete and comprehensive range of both adult and childhood TB (Sambarey et al., 2017; Wang et al., 2018) Most notably, the primary data in these studies have yet to be used to identify sentinel immune-metabolic pathways and to deconvolute the proportions of different immune cell types during TB and which are critical in understanding more completely the host systemic responses and for accounting the contribution of specific immune cell type pathways to disease pathogenesis.

In the present investigation, we evaluate the primary hypothesis that a common pathogenic predictive pathway biology response (inclusive of cellular, molecular, and metabolic determinants) underpins TB across different ages and populations. Accordingly, this raises a key question of delineating high confidence qualitative and quantitative pathway biology differences between childhood and adult TB. To address the central hypothesis and question, we used meta-analysis for integrating data from heterogeneous populations, to identify high confidence differentially and similarly expressed marker genes in children and adults, with active and latent TB across multiple international observational case control studies, irrespective of HIV status, geographical location, and circulating endemic pathogens. Furthermore, we performed gene expression deconvolution to determine which cell types constituted the highest proportions in TB patients and are possible contributors to disease pathologies.

Abbreviations: CD, cluster of differentiation; GEO, gene expression omnibus; HIV, human immunodeficiency virus; IL1, interleukin 1; TB, tuberculosis; TLR, Toll-like receptor.

MATERIALS AND METHODS

Data Acquisition

Gene expression data repositories (GEO and ArrayExpress) were searched for datasets on human response to TB in English for meta-analysis with the following criteria: (I) Search term; {(human TB whole blood) AND “Homo sapiens”[porgn:__txid9606]} was used to retrieve datasets relating to TB gene expression. (II) Each study should be conducted using whole blood not peripheral blood mononuclear cells (PBMCs) as PBMCs when correctly purified have little granulocytes which have been shown to contribute to *M. tuberculosis* pathogenesis (Lowe et al., 2012). (III) The study should include active TB and one of either latently infected individuals (LTBI) or uninfected controls. (IV) Studies conducted using custom arrays or using qPCR were excluded because such array does not contain all genes in the genome and could be biased for specific cell types. Furthermore, following database searches publications associated with selected datasets were read to confirm clinical definitions and whether patients were treated prior to blood collection for microarray. Where datasets were not accessible or raw datasets were not available, authors were contacted for raw datasets, but further data were not made available. All searches were done on publicly available data repositories; no unpublished abstracts and studies were considered. A second assessor role (TF and PD) refereed issues arising from primary data and for workflow accuracy with any discrepancies resolved by discussion and consensus with BK and PG. A detailed workflow of data acquisition, curation, and analysis can be found in **Supplementary Figure S1**. As defined by the original studies, whole blood for transcriptome studies was collected before initiation of treatment and in one study before or within 24 h of initiation of treatment (Kaforou et al., 2013). All searches were done on or before May 2016. While reporting guidelines have yet to be standardized for meta-analysis of gene expression, we adhered as close as possible to guidelines for the meta-analysis of observational studies in epidemiology (MOOSE, **Supplementary Table S2**).

Quality Control, Filtering, and Summarization

The arrayQualityMetrics package in bioconductor (Gentleman et al., 2004; Kauffmann et al., 2009) was used to assess the quality of each dataset in the R statistical environment. After quality control, each dataset was independently normalized before differential expression analysis. Datasets were generated using different array platforms (Illumina, Affymetrix, or Agilent), which have different probe identifiers (IDs) representing different genes on the array chip. Therefore, probes were summarized to Entrez gene levels by selecting probes with the largest median across the samples regardless of clinical phenotype to represent a specific gene. Genes present in at least two datasets were selected for inclusion in meta-analyses to account for as many genes as possible across different array platforms. Prior to extensive statistical analysis, an exploratory analysis of pairwise correlations between fold changes of the datasets was used to

identify outlier datasets that could bias the meta-analysis. Only datasets with good pairwise correlations to other datasets were selected for downstream analysis.

Differential Expression Statistical Analysis

To determine genes differentially expressed in active TB, dataset specific and meta-analysis approaches were performed using the RankProd package in the R statistical environment for estimation of percentage of false prediction (Hong et al., 2006; R Development Core Team, 2011). Meta-analyses for adult and childhood TB were conducted separately. For each individual study, three pairwise contrasts were made where possible; these were active TB versus LTBI, active TB versus uninfected controls, and LTBI versus uninfected controls. Second, the same comparisons were made in a rank product meta-analysis, resulting in an overall estimate of gene expression fold changes and statistical significance across all studies under consideration. The childhood studies lacked uninfected controls in the original studies; therefore, contrasts were performed with LTBI alone and assume that LTBI exhibits a less immune stimulatory status. Differentially expressed genes (adjusted p -value < 0.05, absolute fold change 1.5) obtained from the meta-analysis and from the individual study analysis were compared to identify with high confidence differentially expressed genes for downstream systemic pathway biology analysis. Given the fundamental differences between manifestations of TB in adults and children (Alcaïs et al., 2005), differentially expressed genes obtained from adult and childhood TB were also compared to identify genes commonly differentially expressed in both and those uniquely expressed in either patient population.

Whole Blood Cellular Deconvolution

The cellular components of whole blood can vary in different physiological states and during infections and illness and could influence results obtained from whole blood differential expression analysis. Accordingly, the CellMix package (Gaujoux and Seoighe, 2013) was used to deconvolute the proportions of different immune cell types in active TB patients, latently infected individuals, and uninfected controls. The signatures used for the deconvolution are based on individual immune cell signatures identified based on cell specific gene expression analyses as previously described (Abbas et al., 2005). The normalized and \log_2 transformed dataset was used for the deconvolution using *gedBlood* function in the *CellMix* package (Gaujoux and Seoighe, 2013). Student's t -test was used to perform pairwise statistical analysis of cell proportions between different clinical phenotypes in each dataset. For each cell type, the median of the controls (either LTBI or uninfected controls) was subtracted from the active TB patients for graphical representation per dataset.

Pathway Analysis

The manually curated innate database (InnateDB; Dunkelberger and Song, 2010) was used for pathway analysis to interrogate biological pathways associated with differentially regulated genes in active TB. Up and downregulated genes were analyzed

separately and for each, three separate analyses were made; genes differentially regulated (adjusted p -value < 0.05 and fold change > 1.5) in both adult and childhood TB, those regulated only in adults, and those regulated only childhood TB. To generate graphical representations of significant pathways, adjusted p -values were $-\log_{10}$ transformed and plotted on the x -axis for each pathway.

RESULTS

Data Curation, Quality Control, and Analysis Workflow

Twelve microarray datasets with total sample size of 1110 patients (consisting of active TB patients, LTBI, and uninfected controls) passed predefined criteria. Of the 12 datasets, nine were from adults (825 patients) and three datasets were obtained from children (285 subjects). The studies analyzed here include both HIV negative and HIV positive patients as defined per the original studies. **Table 1** contains information about datasets included in this analysis and their respective GEO accession numbers. **Table 1** also shows the geographical locations of sample collections sites, year the study was conducted, array platforms used for the microarray, and the age definition of the patients. Before extensive differential expression analysis, an exploratory analysis was done on the twelve datasets, as outlined in the analysis workflow in **Figure 1A** (**Supplementary Figure S1**). In this exploratory analysis, the top 1000 most variable probes in the datasets were used to perform a hierarchical unsupervised clustering to determine if expression of these genes can cluster samples from different clinical phenotypes and identify outlier datasets. The clustering largely grouped active TB patients

from LTBI and uninfected controls but there was no clear distinct clustering of latently infected individuals and uninfected controls as shown in **Figure 1B**. Pairwise correlations of the different datasets were done to identify outlier datasets. Dataset (GSE34608) was found not to be well correlated with the rest of the other datasets (**Table 1** and **Supplementary Figures S2–S4**) and was therefore removed from the differential expression analysis pipeline. Furthermore, after QC for each dataset, poor quality samples were removed. The number of samples removed for each dataset can be found in **Supplementary Figure S1** and the outlier dataset also removed resulting in 11 datasets consisting of 1073 subjects taken forward for meta-analysis.

Host Systemic Perturbation in Active Tuberculosis: Re-analysis of Individual Studies

After normalization and probe summarization across datasets, 18,945 Entrez IDs in adults and 18,629 Entrez IDs in children passed filtering. These gene probes were subjected to standardized re-analysis for differential expression analysis for individual studies and meta-analysis across all studies using the RankProd package. **Tables 2A–C** show a summary of differentially expressed genes obtained in the individual dataset analyses showing that there were more upregulated genes than downregulated genes, which could be due to increase immune modulation in response to the infection.

In the active TB versus LTBI comparison, GSE28623 had the largest number of differentially regulated genes while GSE19439/GSE19444 had the smallest number of differentially expressed genes as shown in **Table 2A**. Comparing active TB with uninfected controls, the same datasets had the smallest and the largest number of differentially regulated genes, respectively,

TABLE 1 | Summary of datasets included in the study.

	GEO accession number	Year	Platform	Country	Age	HC	Latent TB	Active TB	Comments and (ref)
1	GSE19439	2010	Illumina GPL6947	UK	Adults	12	17	13	Batch corrected and merge datasets from same study and site Berry et al., 2010
2	GSE19444	2010	Illumina GPL6947	UK	Adults	12	21	21	
3	GSE19442	2010	Illumina GPL6947	South Africa	Adults	NA	31	20	Berry et al., 2010
4	GSE25534	2010	Agilent GPL1708	South Africa	Adults	9	34	33	Maertzdorf et al., 2011b
5	GSE28623	2011	Agilent GPL4133	Gambia	Adults	37	25	46	
6	GSE34608	2012	Agilent GPL6480	Germany	Adults	18	NA	8	Excluded; low correlation with other datasets Maertzdorf et al., 2012
7	GSE37250	2013	Illumina GPL10558	South Africa/ Malawi	Adults	NA	167	195	Kaforou et al., 2013
8	GSE56153	2012	Illumina GPL6883	Indonesia	Adults	18	NA	18	Ottenhoff et al., 2012
9	GSE73408	2016	Affymetrix GPL11532	USA	Adults	NA	35	35	Walter et al., 2016
10	GSE39939	2014	Illumina GPL10558	Kenya	children	NA	14	79	Anderson et al., 2014
11	GSE39940	2014	Illumina GPL10558	South Africa/ Malawi	children	NA	54	111	Anderson et al., 2014
12	GSE41055	2013	Affymetrix GPLP5175	Venezuela	children	9	9	9	Verhagen et al., 2013

TB, tuberculosis; NA, not available; GEO, Gene Expression omnibus. The year column shows the year the study was done, the platform column is microarray platform used for arraying the samples, country column shows the countries from which the samples were collected, age column is the age category for the patient, and the comment and (ref) column contains the reference and comments where possible if the dataset was remove or not and what other process was done on the datasets.

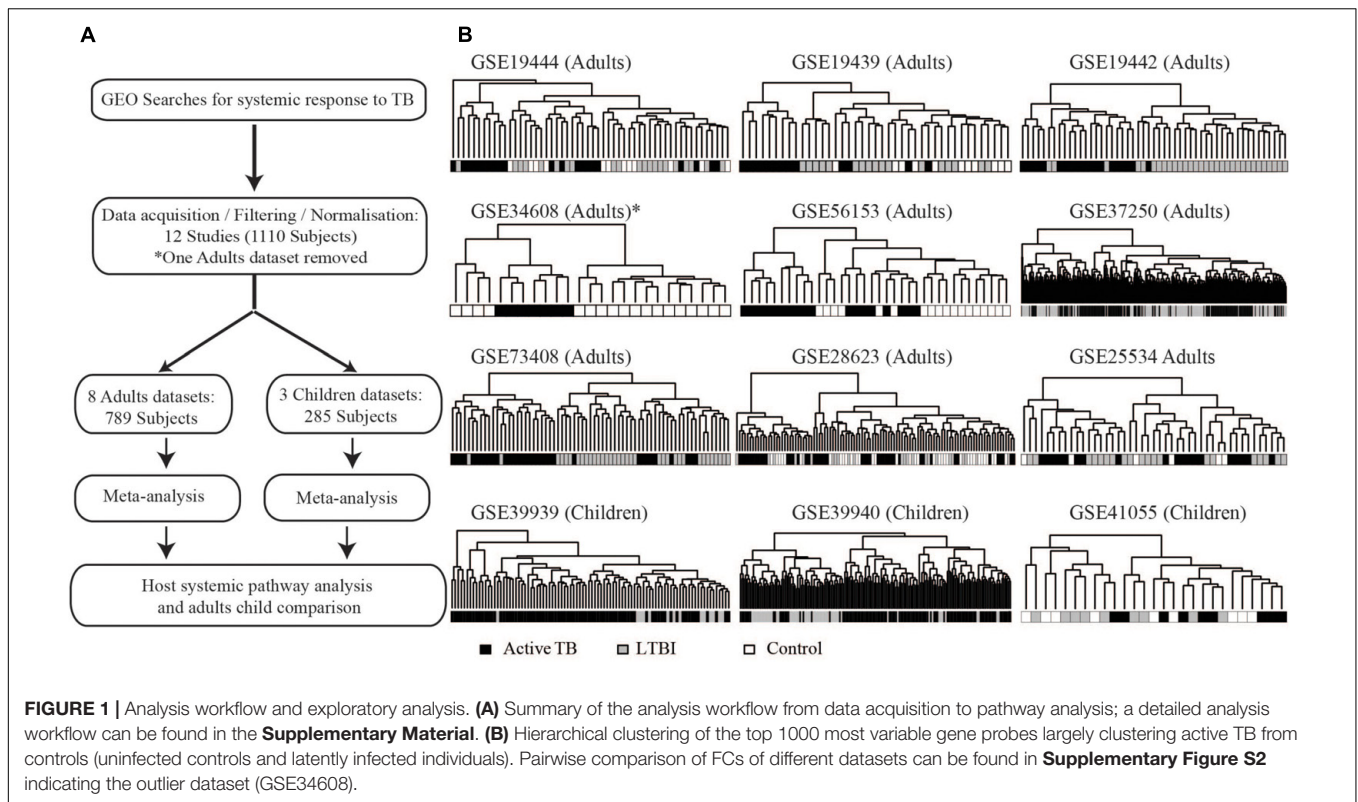


TABLE 2 | Number of differentially expressed genes from individual dataset analysis.

	GSE19439/ GSE19444		GSE19442		GSE25534		GSE28623		GSE37250		GSE73408	
	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down
(A) Active TB versus latent TB												
1.5 FC	246	122	433	244	292	167	695	466	348	173	256	175
2.0 FC	51	7	98	19	57	5	162	34	73	10	43	4
	GSE19439/GSE19444		GSE25534		GSE28623		GSE56153					
	Up	Down	Up	Down	Up	Down	Up	Down				
(B) Active TB versus uninfected controls												
1.5 FC	241	106	685	800	695	466	813	657				
2.0 FC	69	4	128	94	162	34	192	169				
	GSE39940		GSE39939		GSE41055							
	Up	Down	Up	Down	Up	Down						
(C) Children active TB versus latent TB												
1.5 FC	535	343	451	422	114	16						
2.0 FC	116	28	101	67	9	4						

as shown in **Table 2B**. In active TB versus LTBI, only 81 genes were differentially expressed across all datasets at 1.5-fold change cut-off and 66 genes were differentially regulated across all datasets in active TB versus uninfected controls (**Table 3**). This outcome underscores the magnitude of variability in transcriptional responses in these patient population studies and which could be due to several factors including sample

preparation, array processing, and circulating endemic pathogens in different geographical locations. Therefore, to account for different factors affecting different datasets, a combined meta-analysis based on a rank estimate with permutation was used to identify consistently differentially expressed genes.

Quantitatively children developed more differentially expressed genes compared to adults as shown in **Table 2C**.

TABLE 3 | Genes differentially regulated across all studies in the study specific analyses.

Comparison	Genes significantly expressed across all individual analysis
TB versus LTBI	<i>CARD16, GBP4, GBP5, BATF2, ANKRD22, MAPK14, DUSP3, SAMD9L, FBXO6, GBP1, FER1L3, GZMK, HP, IFIT3, FAM102A, CARD17, FAM26F, NELL2, LAP3, PLSCR1, Septin 4, RAB20, NLR4, RTP4, SERPING1, TLR5, C1QB, AQP10, RSAD2, TIFA, EPSTI1</i> , DHRS9, LHFPL2, KLRG1, CD96, TRIM22, IFITM3, HPSE, IFI44L, IRAK3, SLC26A8, ADM, DEFA4, EBI2, S1PR1, C19orf59, ZNF438, GK, CD274, GYG1, ANXA3, CLEC4D, JAK2, KCNJ15, LIMK2, LMNB1, MMP9, GPR84, LRRN3, SIPA1L2, SORT1, CEACAM1, BLK, NOD2, BLR1, SLC22A4, BMX, SMARCD3, TNFAIP6, WARS, DYSF, HIST2H2BE, PARP9, KREMEN1, FCRLA, SIGLEC5, GPR109B, PSTPIP2, CD2, AIM2, P2RY14
TB versus CON	<i>CARD16, GBP4, GBP5, BATF2, ANKRD22, MAPK14, DUSP3, SAMD9L, FBXO6, GBP1, FER1L3, GZMK, HP, IFIT3, FAM102A, CARD17, FAM26F, NELL2, LAP3, PLSCR1, Septin 4, RAB20, NLR4, RTP4, SERPING1, TLR5, C1QB, AQP10, RSAD2, TIFA, EPSTI1</i> , MYL9, OLFM4, TNFSF13B, VAMP5, MTHFD2, PIK3IP1, ANKRD9, MSRB2, PASK, GBP2, IFI27, IFIT2, IFIT1, IL1B, IL7R, IL15, ITGA2B, NAIP, OSM, BANK1, TMEM140, PRKCH, RPS4Y1, S100A12, IFIH1, STAT1, XK, SLC38A1, CAMP, CASP1, MARCO, TNFRSF25, TNFSF10, IL18RAP, UBE2L6

Red and italicized are regulated in both comparisons.

However, it is noteworthy that the amplitude of the per gene fold change differences between adults and children was comparable. High variability is also notable in the childhood studies with dataset GSE41055 yielding the smallest number of differentially regulated genes with only nine upregulated and four downregulated genes at a twofold change cut-off. The two African datasets had similar numbers of differentially expressed genes. Overall, there were more upregulated genes than downregulated genes (Table 2) in both adults and children.

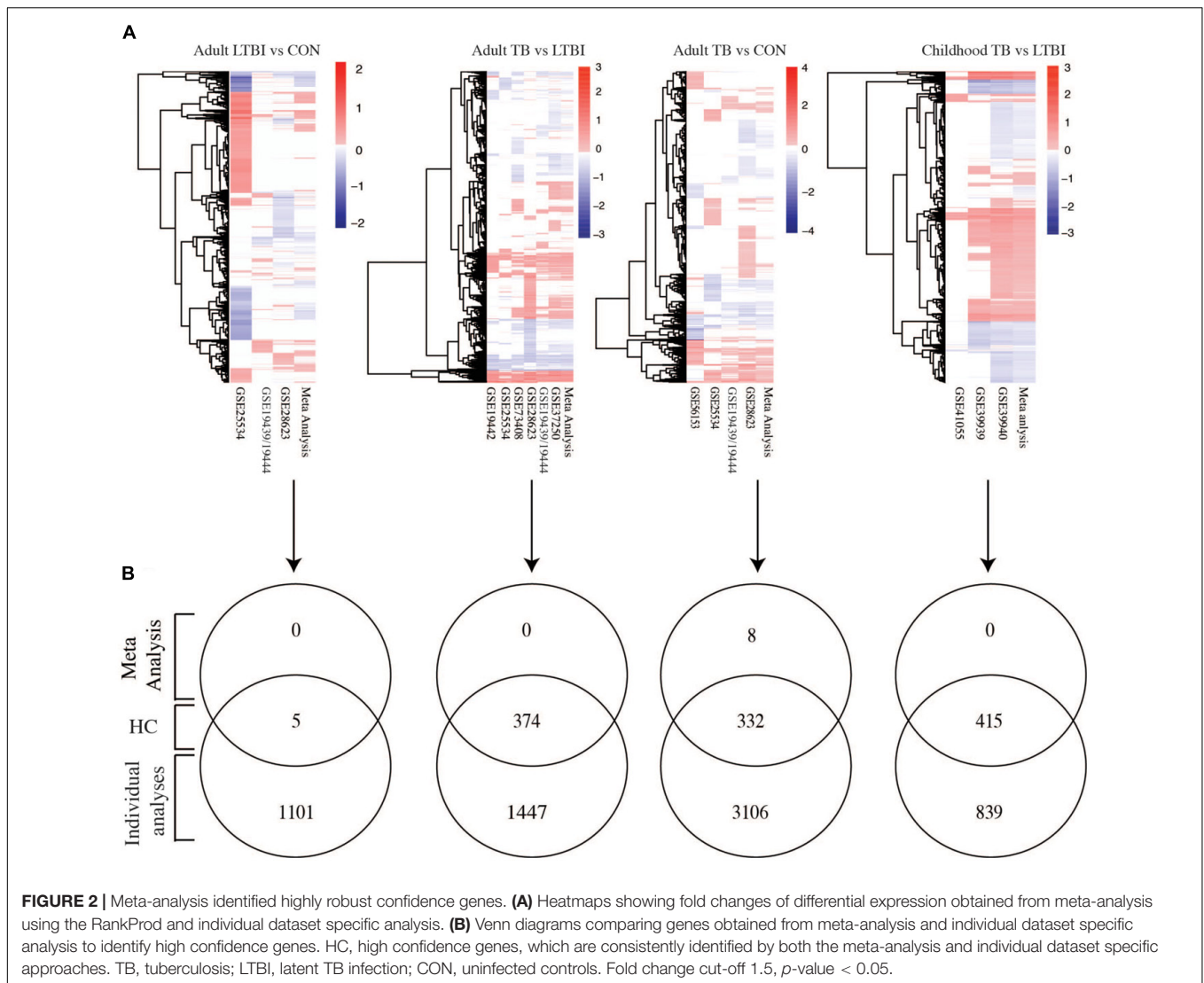
Meta-Analysis Identified High Confidence Differentially Expressed Genes Across Datasets

The analysis of individual studies is subject to study design and technical biases that affect reproducibility or relevance. Therefore, to identify with high confidence genes that were reproducibly regulated across different datasets, differentially expressed genes from the meta-analysis were compared with genes from the individual dataset specific analyses. Figure 2A shows heatmaps of the fold changes of significant genes (adjusted p -values < 0.05) obtained from the meta-analysis for each phenotypic comparison made and those obtained from independently analyzing each of the datasets. As expected from a meta-analysis, it can be seen that many genes with notable fold changes in individual studies lose this characteristic in meta-analysis (where the influence of individual studies

is reduced by size-based weighting and in context of all studies), thereby reducing the count of false positive results and increasing the number of high confidence genes with differential expression. Only five genes were identified as high confidence in LTBI versus uninfected controls further highlighting the gene expression similarities between these two groups (Supplementary Table S1). Active TB versus LTBI had 374 differentially expressed genes which were identified with high confidence while active TB versus uninfected controls had 332 genes. Active TB versus LTBI in children had more high confidence genes (415 genes) compared to adult TB (374 genes) (Figure 2B and Supplementary Table S1). Eight genes were identified by meta-analysis alone in adults with active TB versus uninfected controls and not in any individual dataset specific analysis. These are CMPK2, FCGR1A, HSPA1B, LOC440607, CCL3L1, LOC649853, LOC653980, and CCL4L2. In summary, while the meta-analysis failed to identify many new differentially expressed compared to the individual studies analyses, a highly robust high confident set of expression signals was developed with small number of previously undisclosed markers. Most notably, meta-analysis significantly reduced the number of false positive results, eliminating genes with low reproducibility in differential expression across studies. For all subsequent pathway analyses, we focused on these high confidence sets of genes.

Overlap Between Childhood and Adult Meta-Analyses Detected Genes and Previously Identified Diagnostic Signatures

Given that several studies have identified gene signatures potentially useable for novel TB diagnosis, we next compared differentially expressed genes (FC cut-off 1.5, adjusted p -values < 0.05) from the meta-analysis with gene signatures identified by previous studies. Berry et al. (2010) identified 380 transcripts (312 genes) as a TB diagnostic signature of which, only 55% (172 genes) were identified by the meta-analysis while the remaining 140 genes were not (Figure 3A). Kaforou et al. (2013) describe 27 transcripts (25 genes) as a classifier to diagnose adult TB, of which only 16 genes (64%) were confirmed as significant by the meta-analysis (Figure 3B). Another 51 transcripts (50 gene) signature was identified by Walter et al. (2016) and only 14% (Kaforou et al., 2013) of these genes were identified as confirmatory markers (Figure 3C). Finally, in the childhood TB diagnostic signature (42 transcript; 40 genes; Anderson et al., 2014), 33% (14 genes) were identified by the meta-analysis while the remaining 26 were not (Figure 3D). This reveals a disquietingly marked low confirmatory rate for the existing published expression signatures that range at best from 64% to as low as 14% identity. Our results thus identify those “unreliable” diagnostic signature markers that are likely a consequence of the heterogeneity and high variability of expression in the different populations. However, it is conceivable that the low concordance could be due to the differences in the fold change set for the meta-analysis (FC 1.5) in comparison to those set by the different studies. In Kaforou et al. (2013) and Anderson et al. (2014), the fold change was set to \log_2 0.5 (1.4 linear fold change), and in



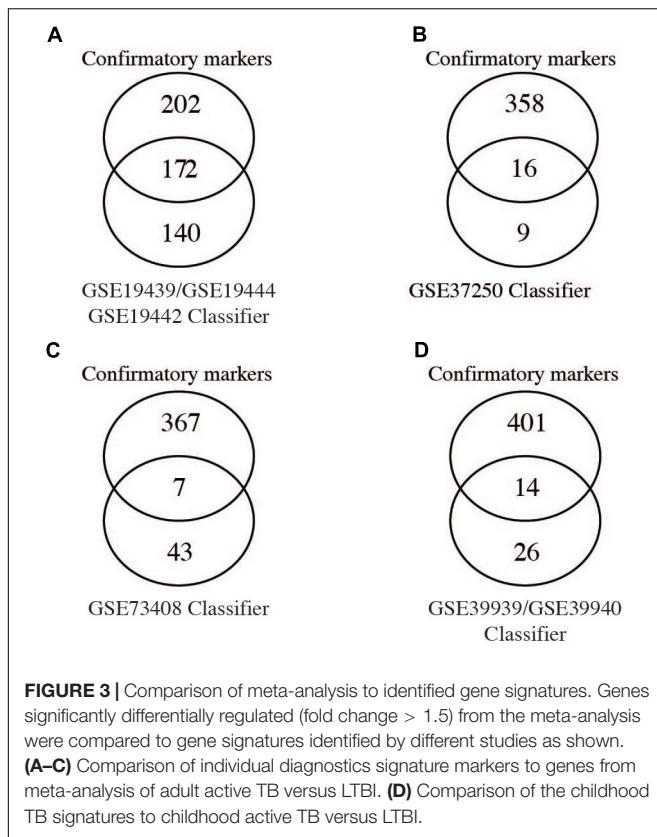
Walter et al. (2016), 1.2 linear fold change was set during feature selection.

Common Predictive Pathway Biology Response to Active Tuberculosis

Outputs of the meta-analyses can be further used to inform high confidence pathway analysis of active TB versus LTBI. Here, genes upregulated in both adults and children were enriched with pathways associated with innate immune responses including type I IFN alpha/beta signaling, TLR signaling, cytokine signaling, and numerous metabolic transporters and enzymes. Notably complement activation (*C1QB*, *C1QC*, *CRI*, *PROS1*), defensin (*SLPI*, *DEFA1*, *A3*, and *A4*) and antimicrobial factors (*S100A12*, *LCN2*, *DDX60L*, *LTF*, *TRRD9*, *ADM*, *APOBEC3A*), and inflammasome activation (*IL18R1*, *IL18RAP*, *CARD16* and *17*, *NLR4*, *AIM2*, *NAIP*, *CASP1*) were all significantly upregulated (**Figure 4B**). *TLR2* signaling was the most prominent TLR pathway and included its interaction with *TLR1* and *TLR6*

cascade signaling (**Figure 4B**). Other notable pattern recognition receptors equally regulated in adults and children included *TLR5*, *Ly96*, *PGLYRP-1*, and importantly *CLEC4D* and *E* that recognize the mycobacterial cell wall glycolipid trehalose 6,6'-dimycolate. Previously unrecognized innate immune pathways were also upregulated, in particular *OLM4*, WNT signaling involving *KREMEN1*, *LRRK2*, and *TSH23* and apparent immunosuppressive macrophage markers including *ARG1*, *MS4ALA* and related atypical chemokine receptors *CCRL2*. Further upregulated immune suppressive signaling includes *IRAK3*, *SIGLEC5*, *SLPI*, *PSTPIP2*, *PIK3AP1*, and the lymphoid co-inhibitory molecule *CD274* (*PD-L1*).

In connection with the upregulation of multiple inhibitory pathways, genes significantly downregulated in both adult and childhood TB were predominantly enriched with pathways including chemokine receptor binding, Ca^{2+} signaling, B-cell (*CD79A* and *B*, *BLR1*, *EBI2*, *VPREB3*, *ID3*, *FCRLA*, *FAIM3*, *FCGBP*, *CCR7*) and T-cell (*BACH2*, *CD6*, *CD5*, *BCL11B*, *TCL1A*, *TCF7*, *CCR7*, *LEF1*), and T/NK cell functions that showed



significant suppression of the CD3 zeta (*CD247*)-EPHA4/NCR3 axis, *CD96*, *UBASH3A*, *GZMK*, and critically *IL7R* that when suppressed leads to severe immunodeficiency (**Figure 4A**).

Thus, a common pathogenic predictive pathway biology response underpins TB across different ages and populations. In active TB, this comprises an upregulation of the inflammatory innate arm of the immune system (encompassing a marked inflammasome type I IFN-TLR2 axis, complement activation, and antimicrobial factors) and which is concomitant with immunosuppressive downregulation of the adaptive immune arm (encompassing predominantly B and T cell signaling and effector functions).

Correlation Between Active TB Versus LTBI to Active TB Versus Uninfected Controls

Individuals with LTBI and uninfected controls showed similar expression profiles and could not be distinctly clustered based on the 1000 most variable genes (**Figure 1B**). We therefore compared differentially expressed genes between active TB versus LTBI and those from active TB versus uninfected individuals to identify phenotype specific gene sets. There was a good correlation of fold changes obtained from the two comparisons (**Figure 5A**) and a good overlap between differentially expressed genes at 1.5-fold change cut-off (adjusted *p*-value < 0.05) while some genes were only expressed on either of the comparisons. Two hundred and sixteen genes were common

to both comparisons, 124 genes differentially expressed only in active TB versus healthy controls, and 158 only expressed in active TB versus LTBI as shown in **Figure 5B**. In summary, this suggests as expected that the lack of specific clustering of differentially expressed genes in either comparison is indicative of a less immune activated status in latent TB in comparison with active TB. However, it is also notable there are limited set of genes specific to latent TB phenotype and point to a subtle yet detectable difference.

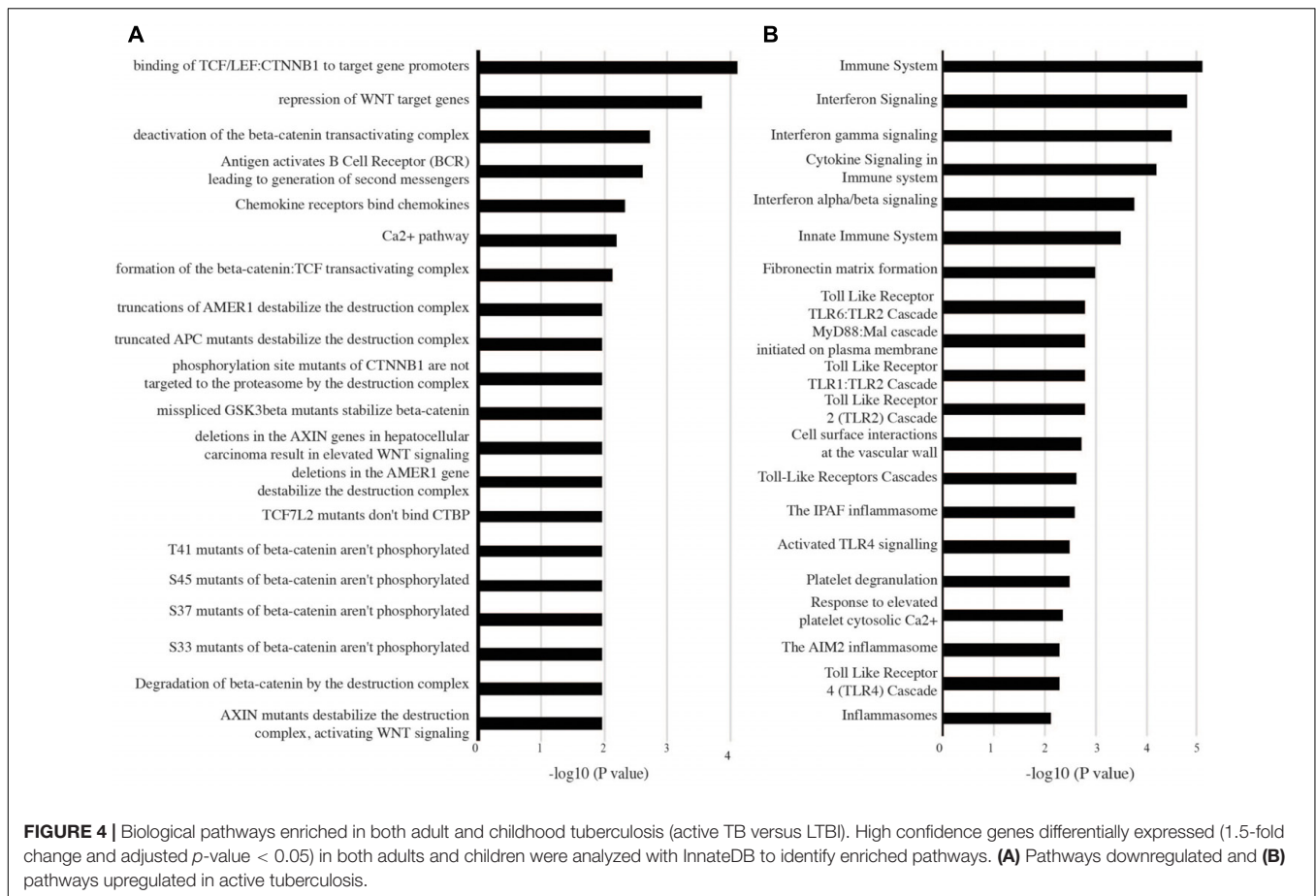
Comparison Between Childhood and Adult TB Host Responses

Childhood and adult TB clinical features and manifestations differ; in adults, the infection is largely restricted to the lungs while in children the infection can often disseminate to other parts of the body (Alcaïs et al., 2005) and hence could result in vastly different host systemic transcriptional responses. Therefore, we compared adult and childhood TB host systemic transcriptional responses and unexpectedly found good overlap of differentially expressed genes between them with a smaller subset of genes specific to either adults or children. **Figure 5C** shows there was a good correlation of fold changes obtained from adult and childhood active TB versus LTBI. However, some genes exhibit higher fold changes in adults and vice versa as shown in **Figure 5C**. At a 1.5 FC cut-off, 206 genes were differentially expressed in both adult and childhood TB, 168 differentially expressed only in adults, and more genes (209 genes) were differentially expressed only in childhood TB (**Figures 5C,D**) and their associated pathways are shown in **Figure 6**. In summary, we observed that childhood TB has about 20% more systemic differentially expressed genes than adults but shared the same differential pattern (representing over 55%) of the differential expressed genes in adult TB.

Nevertheless, genes could be identified that were significantly upregulated only in adult TB represented an extension of the common pathways noted above associated with innate immune inflammatory response including an increased number of IFN regulated genes (*PML*, *CXCL10*, *IFITM1*, *IRF7*, *OAS1*, *IFITM4P*, *ISG15*, *IFI35*, *STAT1*, *SOCS1*, *DUSP3*), *TLR4*, *TLR7/8*, and *TLR9* signaling cascades, and a further augmented significance for the first regulated step of the glycolytic pathway (*HK3*).

The adult selective downregulated gene signatures accentuate the suppressed adaptive immune pathways including T-cell cytotoxic granulysin molecule (*GNLY*) against TB, TCR alpha/beta or gamma delta signaling (in particular *CD2*, *EOMES*, *TRA*, *CD8A*, *TRBV5-4*, *ZAP70*, *SKAP1*, *CD3D*, *TARP*, *FGFB2*, *RASGRP1*, *DYRK2*, *EVL*, *SH2D1A*, *ITK*, *MATK*), B-cell functions (*BANK1*, *CD19*, *BLK*, *FAM129C*, *MS4A1*, *FCRL3*, *EBF1*, *SPIB*, *RASGRP1*, *SH2D1A*, *FCRL2*). NK/T-cell co-inhibitory molecules *KLRG*, *KLRB1*, and cytokine *IL32*; suppression of TGFbeta inhibitory-molecule *TGFBR3* (**Supplementary Figure S7**).

Signature sets differentially upregulated only in childhood TB were notably replete of IFN regulated genes but enriched with other immune pathways including elevation of alternative antimicrobial defense mechanisms (*H2AFJ*, *HISTIH2BG*, and *CTSG*) and multi-functional *IL27* that blocks *IL-17* and *IRF1*



signaling as well as counter T-reg functions, platelet adhesion, and metabolic pathways involved in transport of glucose and other sugars, erythrocytes uptake of carbon dioxide and release of oxygen and cell surface interactions at the vascular wall. There is a significant increase in pathways associated with metal ion homeostasis including *SELENBP1*, *MT2A*, *STEAP4*, *S100P*, and cell adhesion and interactions (notable markers are *GCA*, *ESAM*, *GPR97*, *COL17A1*, and *PSG3*). There are concurrent blockade and stimulatory pathways for IFN/MDA5 activity involving the R1OK3 kinase and DDX60 pathways, respectively. While inhibitory pathways for B-cell activation (via *SAMSN1*) and leukotriene-B (Bruchfeld et al., 2015; degraded by *CYP4F3*) for polymorphonuclear leucocyte chemoattractant activity are unchecked, elevation of *PSG9* that promotes T-reg functions is countered via increased *IL27*.

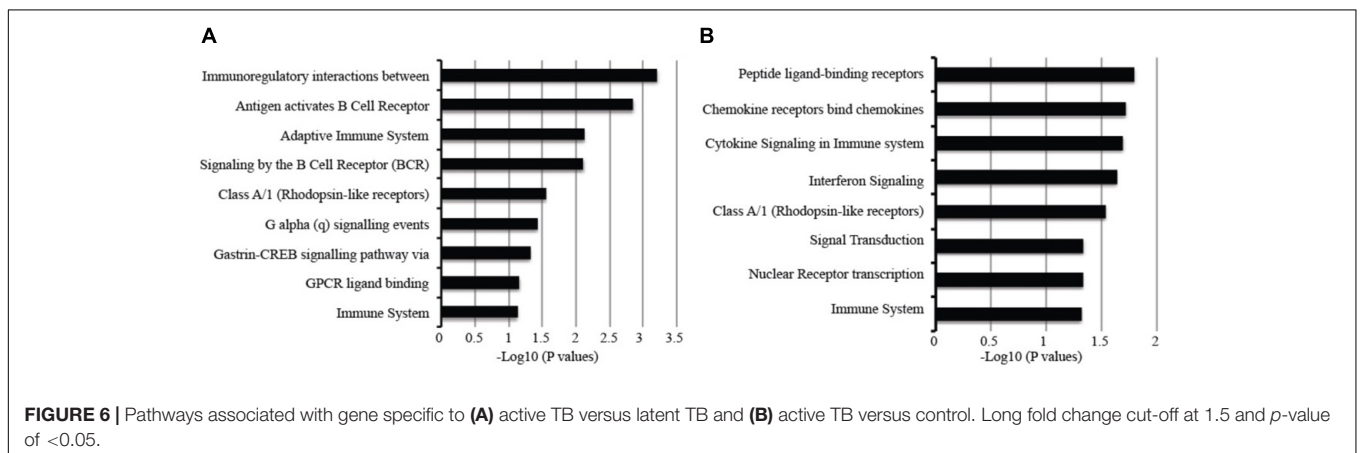
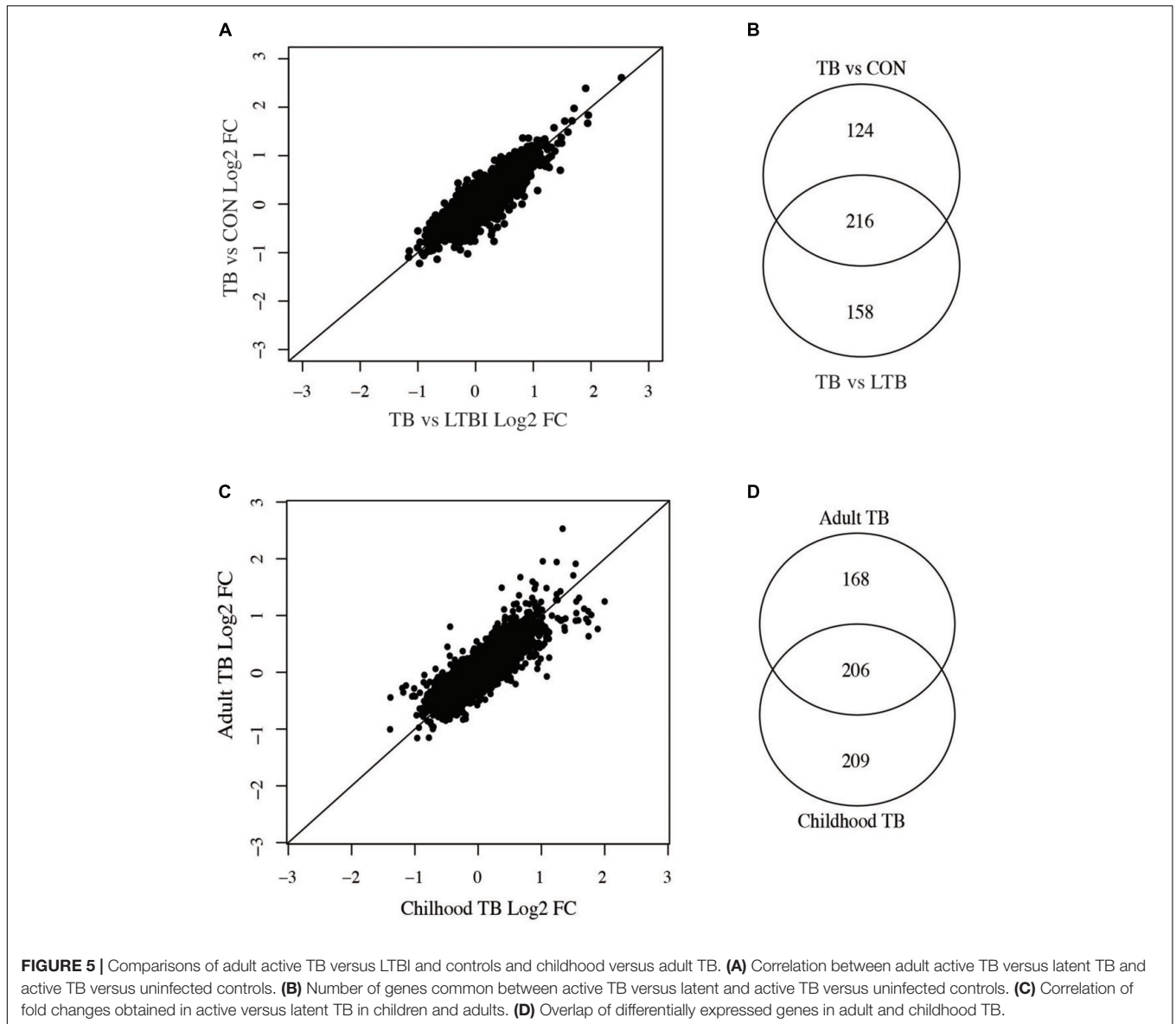
Notable downregulated genes specific for childhood TB are enriched with pathways mainly involved in stimulation of T/B cells (*CD40L*, *CD7*, *ICOS*, *FCER2*, *PTPRCAP*, *ADAM23*), dendritic cell development (*FLT3LG*), alternative promoters of inflammation (*CD248* and *EDAR*), and inhibition of neutrophil degranulation by *ADORA3*. Moreover, there is a significant suppression of *IL23a* that heterodimerizes with *IL12B* to activate the JAK-STAT-IFNG axis, and the OLIG2-SIGLEC8 eosinophil-axis is low in active but high in latent TB. Further eukaryotic translation such as translation initiation, elongation

and termination, metabolism of proteins, amino acid transport, and gene expression as further shown in **Supplementary Figure S7**.

In summary, a comparison between adult and childhood TB shows that over a third of the confirmatory response is shared across the different populations while the other third, significant genes are exclusively significant for either children or adult responses. Unexpectedly, our findings reveal a differential yet plastic response, showing a more significant qualitative and quantitative contribution of both type I and II IFN responses in adults over children and the use of different but functionally similar co-inhibitory molecules underscored by the upregulation of *HLA-G* and *PDCD1LG2* (*PD-L2*) in children and adults, respectively.

Active Tuberculosis Patients Present With Higher Proportions of Innate Immune Cells and Lower Proportions of Adaptive Immune Cells Compared to Controls

Whole blood is made up of different innate and adaptive immune cellular components which can change in response to internal and external challenges, such as inflammatory and infectious diseases including TB. As such, it could be expected that some



immune cellular components will be different between people with active TB and LTBI and uninfected controls and could contribute to observed gene signatures. Therefore, we performed whole blood cell deconvolution to determine differences in cell proportions in active TB, LTBI, and uninfected controls using the CellMix package (Gaujoux and Seoighe, 2013). There were higher proportions of innate immune cells in active TB compared to both latently infected and uninfected individuals. Specifically, proportions of neutrophils and monocytes were significantly higher in active TB compared to controls whereas T-helper (CD4) cells and B cells were higher in controls compared to active TB patients. **Figure 7A** and **Supplementary Figure S5** show differences in cell proportions between active TB and LTBI in adults and **Figure 7B** shows cell proportion differences between active TB and uninfected controls. Resting natural killer (NK) cells were significantly lower in active TB patients while the proportions of activate NK cells were significantly higher in active TB patients. No significant differences in immune cell proportions were observed between LTBI and uninfected controls.

Similar observations were made in childhood TB; children with active TB had higher proportions of neutrophils and other innate immune cells while the proportions of B and T-cells were lower compared to latently infected children. However, significant differences in cell proportions were not observed in the childhood TB dataset collected from the Warao Amerindian children in Venezuela as shown in **Figure 8** and **Supplementary Figure S6**.

In summary, this analysis suggests that active TB patients have significantly higher proportions of innate immune cells and lower proportions of adaptive immune cells compared to controls. These alterations in cell proportions are in good agreement with alterations observed at the molecular functional pathway level involved in the pathogenesis of TB.

Modulation of Co-inhibitory and Co-stimulatory Signaling Molecules in Tuberculosis

To understand further the relationship between cellular and molecular pathway biology responses, we next investigated the modulation of host co-inhibitory and stimulatory immune signaling molecules in TB. *CD 27* (*CD27*), *CD274*, and suppressor of cytokine signaling 3 (*SOC3*) were upregulated while *LILRA5* and *TNFRSF25* were downregulated, respectively, in both adult and childhood TB (**Figure 9**). *CD27* is required for generation and maintenance of long-term memory by transducing the signal to activate the *NF-kB* and *MAPK8/JNK*, *CD274* encodes an immune inhibitory signal expressed in immune cells including T and B-cells, which inhibit T cell activation and cytokine production and *SOSC3* also negatively regulates immune signaling through inhibition of *JAK2* kinase. *LILRA5* and *TNFRSF25*, which were downregulated, are also important in signal transduction and generation of a strong immune response.

Only two co-stimulatory genes; *PDCD1LG2* and *CD2* were up and downregulated, respectively, exclusively in adults (**Figure 7**).

CD2 is a cell surface receptor important for antigen presentation and *PDCD1LG2* encodes a programmed cell death 1 ligand 2.

More immune stimulatory molecules were modulated in childhood TB; four upregulated (*HLA-G*, *LILRA6*, *IL1RN*, and *IL1R2*) and five downregulated (*CD40LG*, *HLA-DOB*, *CD28*, and *HLA-DQB1*). *HLA-G* encodes histocompatibility antigen, class 1G and *LILRA6* encodes for leukocyte immunoglobulin like receptor 6, both of which are important immune regulators of interaction between lymphoid and non-lymphoid cells. Importantly, *HLA-G* is known to play a critical role for suppressing maternal immunity in fetal development. *IL1RN* encodes an *IL1* receptor antagonist that inhibits *IL1* mediated signaling during the immune response and *IL1R2* is a decoy receptor which binds to *IL1* alpha/beta inhibiting its activity with its ligand. *CD40LG* is a *CD40* ligand expressed on the surface of T cells and involved in the regulation of B cell function. *HLA-DOB* and *HLA-DQB1* are members of the major histocompatibility complex important for antigen processing and presentation. *CD28* is also expressed on the surface of T cells and important for T-cell proliferation, cytokine production, and Th2 cell development. Overall this data indicate upregulation of inhibitory immune signaling molecules and downregulation of co-stimulatory signaling molecules important for immune signaling during active TB.

DISCUSSION

In the present study, we used a systematic meta-analysis approach to comprehensively reanalyze publicly available microarray datasets to identify high confidence differentially regulated signatures in adult and childhood TB across multiple continents and a spectrum of circulating endemic pathogens and HIV infection status. Compared to the individual analyses of each dataset, the meta-analysis provided increased power to detect only a limited number of differentially regulated genes but most critically identified high confidence differentially regulated gene sets, reducing a large number of false positive markers detected in the individual study analyses. This demonstrated a small overlap of potential biomarkers across multiple studies conducted in different geographical locations, and which agrees with another very recently published meta-analysis of TB (Sambarey et al., 2017). In adults with active TB versus LTBI, 374 genes were differentially regulated and in active TB versus uninfected controls, 332 genes were differentially regulated. As observed in previous studies, the expression profiles of latently infected individuals and non-infected controls were very similar (Maertzdorf et al., 2011a) but critically identifying for the first time five significantly latent specific biomarkers. When we compared genes regulated in active TB versus LTBI or active TB versus uninfected controls, there was a good overlap (216 genes). By contrast, 124 genes and 158 genes were only regulated in active TB versus uninfected controls and active TB versus LTBI, respectively. This therefore clearly indicates a much-reduced immune altered state in latent TB but which retains a minimal detectable systemic response. In relation to age, childhood TB developed a more differentially regulated

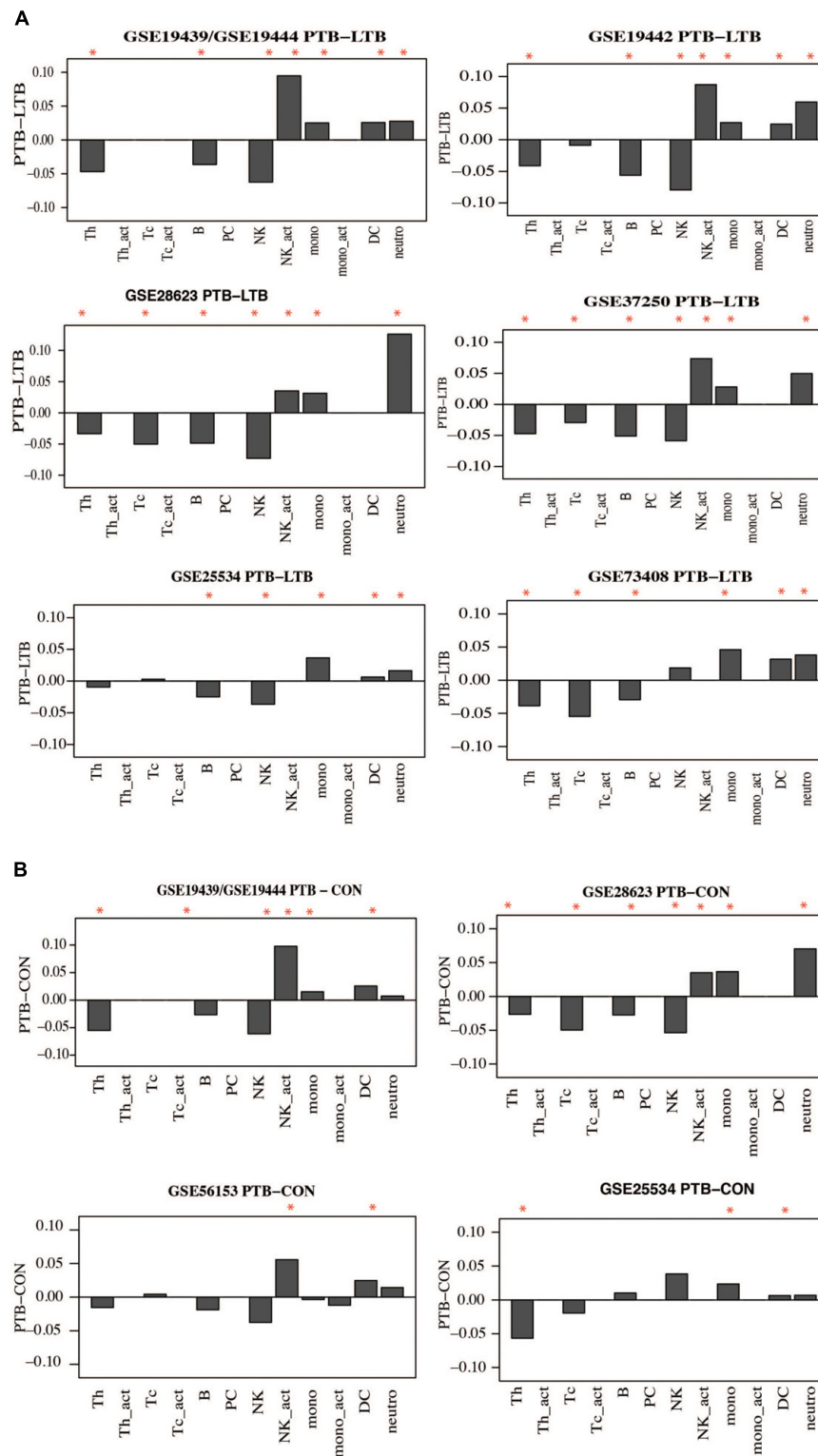
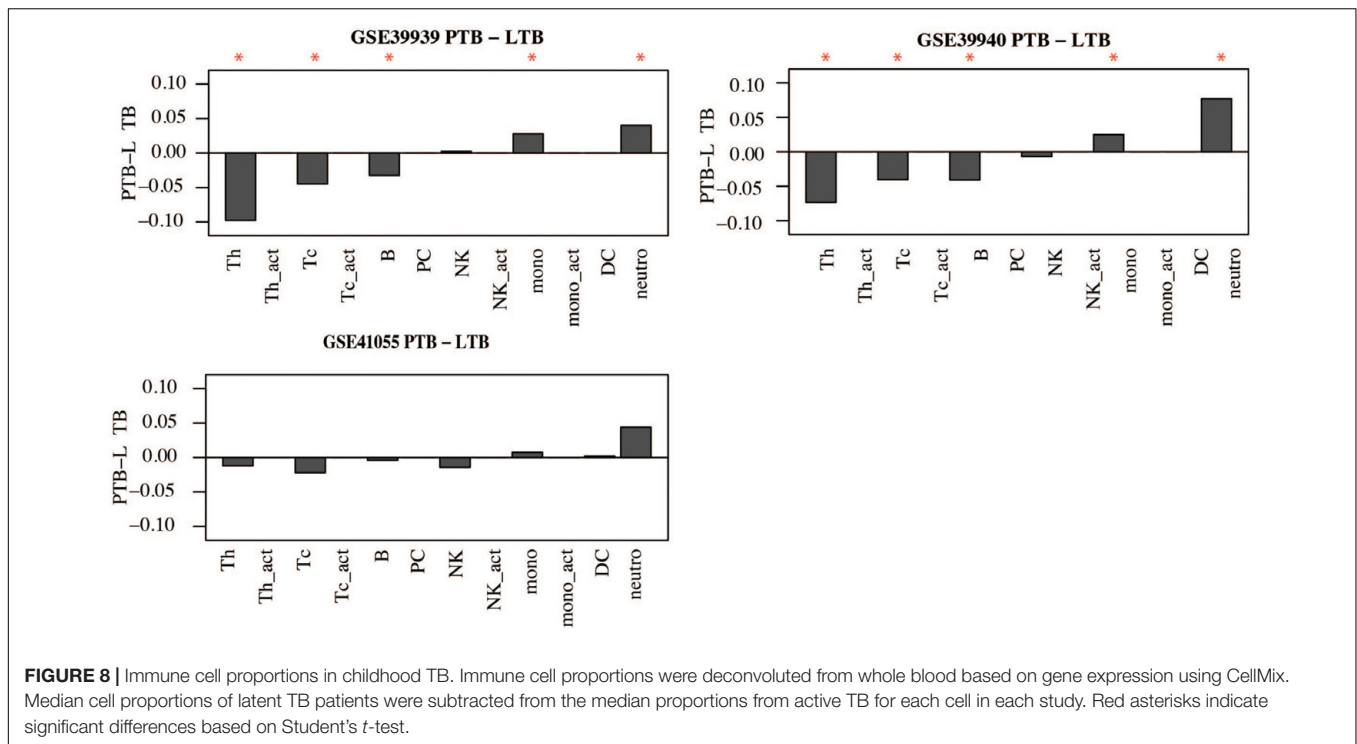


FIGURE 7 | Difference in Immune cell proportions in adult tuberculosis. Gene expression signals were used to deconvolute cell proportions in whole blood collected from TB and controls using the CellMix package. The median for each cell component per clinical phenotype was determined. Median cell proportions from controls were subtracted from medians from active TB. **(A)** Immune cell proportion difference between active TB and latent TB. **(B)** Difference in active TB and uninfected controls. Red asterisks indicate those that are significant ($p < 0.05$) based on a Student's t -test.



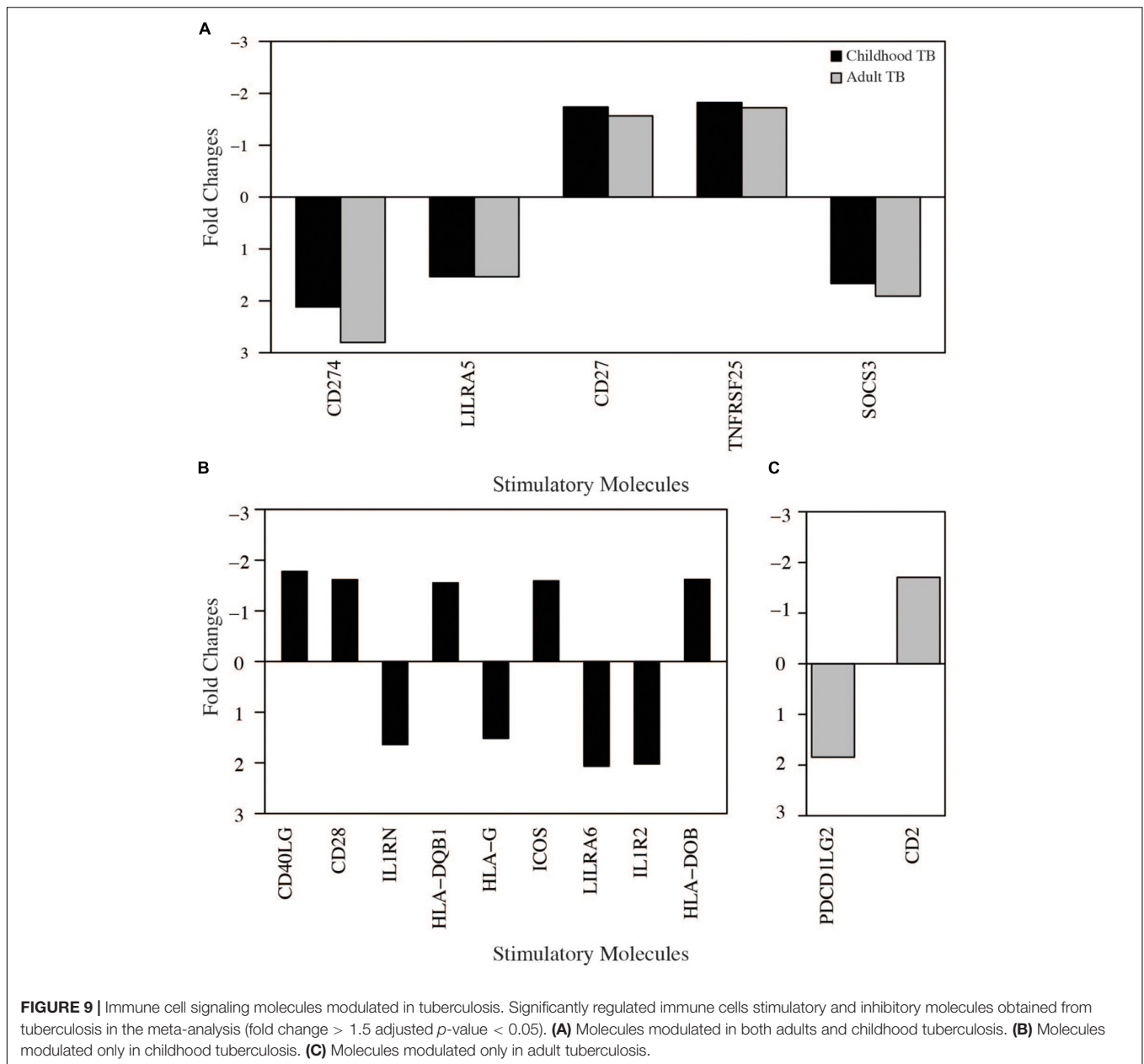
genes (415) than to adult active TB versus LTBI of which approximately half (206) were also significant in adults, revealing a common core pathogenic network response as well as the anticipated differences in response to TB in adults and children that likely relate to the different clinical manifestations (Alcais et al., 2005).

A comparison of confirmatory differentially regulated genes with previously identified TB diagnostic signatures (Berry et al., 2010; Kaforou et al., 2013; Anderson et al., 2014) showed a low confirmatory rate (ranging from 14 to 64%) as shown in **Figure 3**. While this could be possibly explained by the different fold change cut-off set used in the independent studies, the meta-analysis approach reduces false positives and therefore the low correspondence most probably reflects the underlying heterogeneity and overall poor reproducibility for those markers. This highlights the urgent need for potential diagnostic markers to be validated across different and larger patient sample size populations, and critically the standardization of statistical methods and significant cut-offs across studies.

Pathway analyses indicated that genes up and downregulated in both adult and childhood TB were enriched with innate and adaptive immune response pathways, respectively, suggesting strong upregulation of innate inflammatory immune responses and downregulation of adaptive responses, which corroborates previous findings (Berry et al., 2010; Maertzdorf et al., 2011a). Adults show more IFN driven innate immune pathways and downregulated adaptive pathways. On the other hand, childhood specific upregulated genes were associated with inflammasome IL1RN-IL1R2 suppression axis, glucose transport, CO₂ and O₂ release, and cell surface interaction pathways while

downregulated genes were associated with mRNA translation, protein metabolism, and amino acid transport. Indeed, recent studies have indicated that *M. tuberculosis* imports host amino acids and relies on them to thrive in the host (Gouzy et al., 2014). Most importantly, in childhood TB, the immune inhibitory molecules, *IL1RN* and *IL1R2* which inhibit functional *IL1* signaling and molecules involved in generation of an adaptive immune response (*CD40LG*, *HALDOB*, *CD28*) that were downregulated is consistent with emerging evidence linking the cross talk between *IL1* and type 1 IFN to TB and which provide potential targets for host directed therapy (Mayer-Barber et al., 2014). This could explain the observation that children have less pronounced adaptive immune responses to TB compared to adults, potentially resulting in disseminated forms of the diseases seen more frequently in childhood TB. Our findings further highlight the differences between adult and childhood TB and most critically a previously unrecognized potential of using the same common host signature for novel diagnostics in both adult and childhood TB.

Furthermore, whole blood deconvolution indicated high proportions of neutrophils in active TB compare to latently infected individuals and uninfected individuals which is reflective of observed pathway biology. This corroborates previous studies indicating the important role of neutrophils in TB, previously identified as major contributors to observed gene expression signatures (Martineau et al., 2007; Berry et al., 2010; Eum et al., 2010; Mcnab et al., 2011). This meta-analysis which combined studies from different locations further strengthens evidence for the important contribution of neutrophils in TB. Proportions of monocytes and activated NK cells were also higher



in active TB compared to controls, and NK cells have been shown to play an important role in combating mycobacterial infections (Allen et al., 2015; Choreño Parra et al., 2017). However, proportions of resting NK cells were lower in active TB compared to LTBI individuals and uninfected controls which agrees with previous studies (Kee et al., 2012). On the other hand, proportions of T and B cells which make up the adaptive arm of the immune system were downregulated, which corroborated results observed in the pathway analysis. This is in line with recent studies showing that patients with active TB have reduced frequency of circulating B cell and T cell responses and higher in successfully TB treated individuals compared to patients recently diagnosed with TB (Joosten et al., 2016).

CONCLUSION

By using meta-analysis to re-analyze publicly available microarray datasets, we identified with high confidence signature genes differentially expressed across studies removing a large number of false positive markers. Most importantly, from translational perspective, we reveal a previously unrecognized potential of using a common host immune-metabolic signature network as a novel diagnostic tool for use in both adult and childhood TB. We believe that the use of the confirmatory expression signatures identified here will help in guiding optimized diagnostic, prognostic, and therapeutic monitoring modalities for TB patients, but which will still necessitate high quality prospective randomized trials to assess clinical efficacy.

ETHICS STATEMENT

This was meta-analysis of already existing publicly available online dataset from gene expression data repositories.

AUTHOR CONTRIBUTIONS

SB and PG conceived and designed the study. SB conducted the analyses. TF and PD helped with analysis. SB, BK, and PG wrote the manuscript.

FUNDING

SB was supported by joint doctoral scholarship from MRCG and Darwin Trust Scholarship award (Edinburgh) and his postdoctoral fellowship was supported by the Wellcome Trust

REFERENCES

- Abbas, A. R., Baldwin, D., Ma, Y., Ouyang, W., Gurney, A., Martin, F., et al. (2005). Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun.* 6, 319–331.
- Alcais, A., Fieschi, C., Abel, L., and Casanova, J.-L. (2005). Tuberculosis in children and adults: two distinct genetic diseases. *J Exp Med.* 202, 1617–1621.
- Allen, M., Bailey, C., Cahatol, I., Dodge, L., Yim, J., Kassissa, C., et al. (2015). Mechanisms of control of *Mycobacterium tuberculosis* by NK cells: role of glutathione. *Front. Immunol.* 6:508. doi: 10.3389/fimmu.2015.00508
- Anderson, S. T., Kaforou, M., Brent, A. J., Wright, V. J., Banwell, C. M., Chagaluka, G., et al. (2014). Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N. Engl. J. Med.* 370, 1712–1723. doi: 10.1056/NEJMoa1303657
- Berry, M. P. R., Graham, C. M., McNab, F. W., Xu, Z., Bloch, S. A. A., Oni, T., et al. (2010). An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 466, 973–977. doi: 10.1038/nature09247
- Bloom, C. I., Graham, C. M., Berry, M. P. R., Wilkinson, K. A., Oni, T., Rozakeas, F., et al. (2012). detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS One* 7:e46191. doi: 10.1371/journal.pone.0046191
- Bruchfeld, J., Correia-Neves, M., and Källenius, G. (2015). Tuberculosis and HIV Coinfection. *Cold Spring Harb. Perspect. Med.* 5:a017871. doi: 10.1101/cshperspect.a017871
- Choreño Parra, J. A., Martínez Zúñiga, N., Jiménez Zamudio, L. A., Jiménez Álvarez, L. A., Salinas Lara, C., and Zúñiga, J. (2017). Memory of natural killer cells: a new chance against *Mycobacterium tuberculosis*? *Front. Immunol.* 8:967. doi: 10.3389/fimmu.2017.00967
- Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V., et al. (1994). Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA* 271, 698–702.
- Cooper, A. M. (2009). Cell-mediated immune responses in tuberculosis. *Annu. Rev. Immunol.* 27, 393–422. doi: 10.1146/annurev.immunol.021908.132703
- Dunkelberger, J. R., and Song, W. (2010). Complement and its role in innate and adaptive immune responses. *Cell Res.* 20, 34–50. doi: 10.1038/cr.2009.139
- Dwivedi, V. P., Bhattacharya, D., Chatterjee, S., Prasad, D. V. R., Chattopadhyay, D., Van Kaer, L., et al. (2012). Mycobacterium tuberculosis directs T helper 2 cell differentiation by inducing interleukin-1 β production in dendritic cells. *J. Biol. Chem.* 287, 33656–33663.
- Eum, S. Y., Kong, J. H., Hong, M. S., Lee, Y. J., Kim, J. H., Hwang, S. H., et al. (2010). Neutrophils are the predominant infected phagocytic cells in the airways of patients with active pulmonary TB. *Chest* 137, 122–128. doi: 10.1378/chest.09-0903

DELTA through WACCBIP (DEL-15-007: Awandare). These studies were in part supported by MRC (G0701289) and Welsh government and EU ERDF funds to PG.

ACKNOWLEDGMENTS

The authors wish to thank the staff of the Division of Infection and Pathway Medicine and MRCG training and staff development and Darwin Trust Scholarship award (Edinburgh) for supporting SB during his Ph.D. studies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00457/full#supplementary-material>

- Gaujoux, R., and Seoighe, C. (2013). CellMix: a comprehensive toolbox for gene expression deconvolution. *Bioinformatics* 29, 2211–2212. doi: 10.1093/bioinformatics/btt351
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5:R80.
- Gouzy, A., Poquet, Y., and Neyrolles, O. (2014). Amino acid capture and utilization within the *Mycobacterium tuberculosis* phagosome. *Future Microbiol.* 9, 631–637. doi: 10.2217/fmb.14.28
- Hong, F., Breitling, R., McEntee, C. W., Wittner, B. S., Nemhauser, J. L., and Chory, J. (2006). RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 22, 2825–2827.
- Joosten, S. A., van Meijgaarden, K. E., del Nonno, F., Baiocchi, A., Petrone, L., Vanini, V., et al. (2016). Patients with tuberculosis have a dysfunctional circulating b-cell compartment, which normalizes following successful treatment. *PLoS Pathog.* 12:e1005687. doi: 10.1371/journal.ppat.1005687
- Kaforou, M., Wright, V. J., Oni, T., French, N., Anderson, S. T., Bangani, N., et al. (2013). Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med.* 10:e1001538. doi: 10.1371/journal.pmed.1001538
- Karonga Prevention Trial Group. (1996). Randomised controlled trial of single BCG, repeated BCG, or combined BCG and killed *Mycobacterium leprae* vaccine for prevention of leprosy and tuberculosis in Malawi, Karonga Prevention Trial Group. *Lancet* 348, 17–24.
- Kauffmann, A., Gentleman, R., and Huber, W. (2009). Array quality metrics - A bioconductor package for quality assessment of microarray data. *Bioinformatics* 25, 415–416. doi: 10.1093/bioinformatics/btn647
- Kee, S. J., Kwon, Y. S., Park, Y. W., Cho, Y. N., Lee, S. J., Kim, T. J., et al. (2012). Dysfunction of natural killer T cells in patients with active *Mycobacterium tuberculosis* infection. *Infect Immun.* 80, 2100–2108. doi: 10.1128/IAI.06018-11
- Lowe, D. M., Redford, P. S., Wilkinson, R. J., O'Garra, A., and Martineau, A. R. (2012). Neutrophils in tuberculosis: friend or foe? *Trends Immunol.* 33, 14–25. doi: 10.1016/j.it.2011.10.003
- Maertzdorf, J., Ota, M., Reipsilber, D., Mollenkopf, H. J., Weiner, J., Hill, P. C., et al. (2011a). Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis. *PLoS One* 6:e26938. doi: 10.1371/journal.pone.0026938
- Maertzdorf, J., Reipsilber, D., Parida, S. K., Stanley, K., Roberts, T., Black, G., et al. (2011b). Human gene expression profiles of susceptibility and resistance in tuberculosis. *Genes Immun.* 12, 15–22. doi: 10.1038/gene.2010.51
- Maertzdorf, J., Weiner, J., Mollenkopf, H.-J., Network, T., Bauer, T., Prasse, A., et al. (2012). Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proc Natl Acad Sci. U.S.A.* 109, 7853–7858. doi: 10.1073/pnas.1121072109

- Martineau, A. R., Newton, S. M., Wilkinson, K. A., Kampmann, B., Hall, B. M., Nawroly, N., et al. (2007). Neutrophil-mediated innate immune resistance to mycobacteria. *J. Clin. Invest.* 117, 1988–1994.
- Mayer-Barber, K. D., Andrade, B. B., Oland, S. D., Amaral, E. P., Barber, D. L., Gonzales, J., et al. (2014). Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. *Nature* 511, 99–103. doi: 10.1038/nature13489
- McNab, F. W., Berry, M. P., Graham, C. M., Bloch, S. A., Oni, T., Wilkinson, K. A., et al. (2011). Programmed death ligand 1 is over-expressed by neutrophils in the blood of patients with active tuberculosis. *Eur. J. Immunol.* 41, 1941–1947. doi: 10.1002/eji.201141421
- Ottenhoff, T. H. M., Dass, R. H., Yang, N., Zhang, M. M., Wong, H. E. E., Sahiratmadja, E., et al. (2012). Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. *PLoS One* 7:e45839. doi: 10.1371/journal.pone.0045839
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 409.
- Sambarey, A., Devaprasad, A., Baloni, P., Mishra, M., Mohan, A., Tyagi, P., et al. (2017). Meta-analysis of host response networks identifies a common core in tuberculosis. *NPJ Syst. Biol. Appl.* 3:4. doi: 10.1038/s41540-017-0005-4
- Tientcheu, L. D., Maertzdorf, J., Weiner, J., Adetifa, I. M., Mollenkopf, H.-J., Sutherland, J. S., et al. (2015). Differential transcriptomic and metabolic profiles of *M. africanum*- and *M. tuberculosis*-infected patients after, but not before, drug treatment. *Genes Immun.* 16, 347–355. doi: 10.1038/gene.2015.21
- Verhagen, L. M., Zomer, A., Maes, M., Villalba, J. A., del Nogal, B., Eleveld, M., et al. (2013). A predictive signature gene set for discriminating active from latent tuberculosis in Warao Amerindian children. *BMC Genomics* 14:74. doi: 10.1186/1471-2164-14-74
- Walter, N. D., Miller, M. A., Vasquez, J., Weiner, M., Chapman, A., Engle, M., et al. (2016). Blood transcriptional biomarkers for active tuberculosis among patients in the United States: a case-control study with systematic cross-classifier evaluation. *J. Clin. Microbiol.* 54, 274–282. doi: 10.1128/JCM.01990-15
- Wang, Z., Arat, S., Magid-Slav, M., and Brown, J. R. (2018). Meta-analysis of human gene expression in response to *Mycobacterium tuberculosis* infection reveals potential therapeutic targets. *BMC Syst. Biol.* 12:3. doi: 10.1186/s12918-017-0524-z
- WHO (2015). *Global Tuberculosis Report 2015*. Available at: http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf?ua=1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Bah, Forster, Dickinson, Kampmann and Ghazal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.