The Journal of Infectious Diseases

MAJOR ARTICLE

Methylome-wide Analysis Reveals Epigenetic Marks Associated With Resistance to Tuberculosis in Human Immunodeficiency Virus–Infected Individuals From East Africa

Catherine M. Stein,[1,2,a] Penelope Benchek,[1,a] Jacquelaine Bartlett,[1] Robert P. Igo, Jr,[1,b] Rafal S. Sobota,[3] Keith Chervenak,[2] Harriet Mayanja-Kizza,[4] C. Fordham von Reyn,[5] Timothy Lahey,[5] William S. Bush,[1] W. Henry Boom,[2] William K. Scott,[6] Carmen Marsit,[7] Giorgio Sirugo,[8,c] and Scott M. Williams[1,9,c]

[1]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, USA, [2]Division of Infectious Disease and HIV Medicine, Department of Medicine, Case Western Reserve University, Cleveland, Ohio, USA, [3]Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA, [4]Department of Medicine and Mulago Hospital, School of Medicine, Makerere University, Kampala, Uganda, [5]Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire, USA, [6]John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida, USA, [7]Rollins School of Public Health, Emory University, Atlanta, Georgia, USA, [8]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Philadelphia, Pennsylvania, USA, and [9]Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio, USA

Background. Tuberculosis (TB) is the most deadly infectious disease globally and is highly prevalent in the developing world. For individuals infected with both *Mycobacterium tuberculosis* (*Mtb*) and human immunodeficiency virus (HIV), the risk of active TB is 10% or more annually. Previously, we identified in a genome-wide association study (GWAS) a region on chromosome 5 associated with resistance to TB, which included epigenetic marks that could influence gene regulation. We hypothesized that HIV-infected individuals exposed to *Mtb* who remain disease free carry epigenetic changes that strongly protect them from active TB.

Methods. We conducted a methylome-wide study in HIV-infected, TB-exposed cohorts from Uganda and Tanzania and integrated data from our GWAS.

Results. We identified 3 regions of interest that included markers that were differentially methylated between TB cases and controls with latent TB infection: chromosome 1 (*RNF220*, $P = 4 \times 10^{-5}$), chromosome 2 (between *COPS8* and *COL6A3*, $P = 2.7 \times 10^{-5}$), and chromosome 5 (*CEP72*, $P = 1.3 \times 10^{-5}$). These methylation results co-localized with associated single-nucleotide polymorphisms (SNPs), methylation QTLs, and methylation × SNP interaction effects. These markers were in regions with regulatory markers for cells involved in TB immunity and/or lung.

Conclusions. Epigenetic regulation is a potential biologic factor underlying resistance to TB in immunocompromised individuals that can act in conjunction with genetic variants.

Keywords. methylation; epigenetics; infectious disease; genetics; genomics; lung function; immunology.

Tuberculosis (TB), caused by infection with *Mycobacterium tuberculosis* (*Mtb*), results in approximately 1.2 million deaths per year [1], though most of the approximately 2 billion people infected with *Mtb* do not progress to disease. Although the risk of developing TB is low in most infected people, it is the most common cause of death in human immunodeficiency virus (HIV)–infected people living in TB-endemic countries [1, 2]. As many as 10% of coinfected people develop TB each year, illustrating how immunocompromise contributes to risk. People with HIV infection who do not develop TB, despite *Mtb* infection, offer a major opportunity to understand resistance to TB despite being immunocompromised, and possibly a key to how any *Mtb* infection leads to TB.

Several studies indicate that susceptibility (or resistance) to TB is partially due to genomic factors [3]. Genome-wide association studies (GWAS) have identified associating loci, but most show small effect sizes in HIV-uninfected subjects. In contrast, our GWAS [4], conducted in HIV-infected subjects, found a significant association with a region of chromosome 5, containing *IL12B*. Annotation of this region showed that the associated single-nucleotide polymorphisms (SNPs) resided in a histone mark, indicating that epigenetic marks may influence regulation of a nearby gene that we hypothesized to be *IL12B*. Because differences in genomic features, such as histone and other epigenetic marks, can impact gene expression patterns, thereby linking genetic and environmental risk

Received 18 August 2020; editorial decision 17 December 2020; accepted 4 January 2021; published online January 5, 2021.

[a]C. M. S. and P. B. contributed equally as first authors.

[b]Deceased.

[c]G. S. and S. M. W. contributed equally as senior authors.

Correspondence: Catherine M. Stein, PhD, Case Western Reserve University, 2103 Cornell Road, Wolstein Research Building Room 1316, Cleveland, OH 44106, USA (cmj7@case.edu).

The Journal of Infectious Diseases® 2021;224:695–704

DOI: 10.1093/infdis/jiaa785

factors, and since epigenetic marks can be inherited, they have been hypothesized to explain some of the "missing heritability" for complex diseases [5, 6]. Evidence for this model exists in cancer and autoimmune diseases, demonstrating an association between epigenetic marks and disease [7, 8]. We hypothesized that similar regulatory factors, including methylation marks, associate with TB susceptibility/resistance.

Protection from TB is affected not only by genetic factors, but by environmental factors from smoking to undernutrition [1], and one means of assessing the role of factors that are sensitive to the environment is to examine epigenetic variation in people who are exposed to *Mtb* but avoid disease. This is an indirect measure of environmental factors, but evidence for association can represent both environmental exposures, affecting epigenetic states, and the role of genes that respond to such exposures. To address this, we chose to study one class of epigenetic features, DNA methylation, which is impacted both directly or indirectly via DNA-level variation as well as physiologic state [9]. We used a methylome-wide analysis (MWAS) approach, examining association with methylation status and with nearby SNPs. We also investigated differential methylation near loci identified by GWAS of TB risk (Figure 1). Our underlying model was that both genetic and epigenetic factors are associated with TB risk and that the 2 can modify each other. To understand functional implications of identified loci, we annotated them using available databases. By integrating results across data types [10], we constructed a more complete assessment of genomic features, both epigenetic and genetic, and their interactions, and associations with TB pathogenesis.

METHODS

We first conducted an MWAS in the Ugandan and Tanzanian cohorts independently (Figure 1). Because the Ugandan cohort was larger, we considered it the discovery cohort and the Tanzanian cohort as the replication set, although we repeated analyses in the opposite direction. Markers that were differentially methylated at $P < 5 \times 10^{-5}$ in the discovery cohort and $P < .1$ in the replication cohort were considered for subsequent analyses. These thresholds were set to assign reasonable probability of association in the absence of a clearly recognized threshold. We required evidence for association in 2 cohorts and concordance between data types, for example, GWAS and methylation QTL (meQTL) evidence. Second, we examined SNPs from our GWAS within 200 kb of the methylation signal, and identified whether any were significantly ($P < .05$) associated with TB. Third, we conducted an meQTL analysis, examining SNPs within 200 kb of the associated methylation marker to test for association with methylation of the associated CpG site. Fourth, we examined interaction between methylation status and SNPs in association with TB. Synthesis of the results of the second through fourth steps addressed whether genetic variation at least partially influences TB susceptibility through effects on methylation, or if differential methylation and genetic variation acted independently. Last, all loci significant in the first step were annotated for function. In parallel, we examined the association between methylation level and TB in regions significantly associated with TB in our GWAS (Table 1) [4].

Study Participants

This study includes the same subjects as in our previous GWAS [4]. In brief, subjects were from a household contact study in Kampala, Uganda, enrolled during 2002–2009 [11], or from a clinical trial (ClinicalTrials.gov identifier NCT00052195) and observational study of TB in Dar es Salaam, Tanzania, enrolled during 2001–2005 [12]. All subjects were HIV infected and aged >15 years, and none received antiretroviral therapy (ART) due to unavailability at the time of enrollment

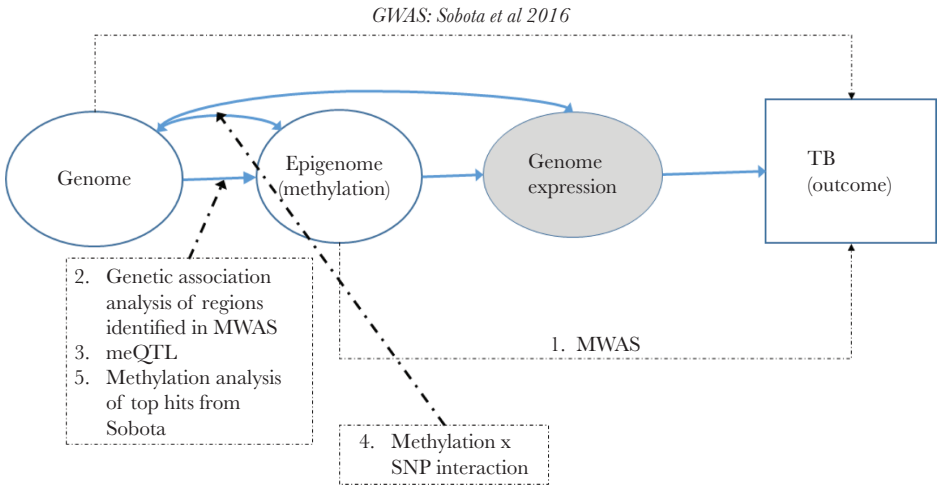

Figure 1. Analytical framework. Abbreviations: GWAS, genome-wide association study; meQTL, methylation QTL; MWAS, methylome-wide analysis; SNP, single-nucleotide polymorphism; TB, tuberculosis.

Table 1. Study Population Characteristics

Characteristic	TB	LTBI	Overall	*P* Value Across TB Groups (TB vs LTBI)
Uganda cohort	(n = 76)	(n = 67)	(n = 143)	
Proportion male sex (No.)	0.50 (38)	0.36 (24)	0.43 (62)	.12[a]
Mean age, y (SD)	31.4 (9.3)	26.1 (12.5)	28.9 (11.2)	.02[b]
Tanzania cohort	(n = 32)	(n = 46)	(n = 78)	
Proportion male sex (No.)	0.19 (6)	0.26 (12)	0.23 (18)	.63[a]
Mean age, y (SD)	35.0 (7.7)	36.0 (7.6)	35.6 (7.6)	.30[b]

LTBI determined based on a positive tuberculin skin test; TB diagnosis based on isolation of *Mycobacterium tuberculosis* in culture and symptoms consistent with TB.

Abbreviations: LTBI, latent tuberculosis infection; SD, standard deviation; TB, tuberculosis.

[a]Pearson χ^2 test of independence.

[b]Wilcoxon–Mann–Whitney test.

(confirmed in subjects' medical records). TB cases were defined as individuals with clinical symptoms and who were *Mtb* culture confirmed. Controls all had latent *Mtb* infection (LTBI) based on a positive tuberculin skin test (TST); interferon-γ release assays were not performed at the time of ascertainment. All subjects provided written informed consent. These analyses utilized samples from those subjects who had DNA remaining from the GWAS.

Molecular Methods

DNA came from buffy coat samples in the Uganda cohort and either buffy coat or whole blood in the Tanzania cohort, and prepared as described previously [4], then bisulfite converted according to specifications for the Illumina Methylation EPIC 850k chip. As blood carries marks of immune response and is generally the most available tissue, these studies should be reflective of many responses relevant to TB. GWAS data were available from our previous analysis [4].

Statistical Analysis

Quality Control and Principal Components Analysis

Quality control and principal components analysis of methylation data are described in the Supplementary Methods. We estimated cell proportions in the combined Uganda and Tanzania cohort for CD8, CD4, natural killer, B, monocyte, and neutrophil cells using the minfi package.

MWAS Analysis

Within each of the 2 cohorts, we tested for association between CpG β values (converted to M-values [$\log_2(\beta/1-\beta)$]) and TB status using limma in R [13]. We adjusted for significant methylation-based Principal Components, aforementioned estimated cell proportions, age, and sex in a linear model that compares methylation values between TB and LTBI subjects.

Genetic Association Analysis

Genetic association analysis was performed using appropriate data from [4]. Association between SNPs within 200 kb of the associated methylation markers and TB was conducted using PLINK [14]. A Bonferroni-corrected *P* value was derived based on the total number of SNPs tested across all 3 regions (Supplementary Methods).

Methylation QTL Analysis

We conducted a targeted cis-meQTL analysis around the replicated CpG sites in the combined Uganda/Tanzania sample that had overlapping methylation and genotype data (N = 188: 75 from Tanzania [32 with TB] and 113 from Uganda [66 with TB]). SNPs within 200 kb of the replicated CpG sites were examined within the GWAS dataset [4]. This analysis allowed determination of differential methylation associated with genetic variation at the MWAS loci. Using Matrix eQTL, we performed linear regression adjusting for sex, age, cohort, cell proportions (CD4, CD8, monocytes, neutrophils), and the first 2 genetic-based PCs [4] to test association between nearby SNPs and CpG sites.

SNP-Methylation Interaction

We conducted a targeted interaction analysis around replicated CpG sites in the combined Uganda/Tanzania sample that had overlapping methylation and genotype data (described above). As in the meQTL analysis, SNPs within 200 kb of replicated CpG sites were determined from the GWAS [4]. Using glm in R, we ran logistic regression adjusting for sex, cohort, CD4 cell proportions, and the first 2 GWAS-based PCs [4] to determine if the difference in methylation levels between TB and LTBI samples is modified by nearby genotypes.

Functional Annotation

We examined whether loci are associated with related traits in the GWAS catalog (https://www.ebi.ac.uk/gwas/). We also examined whether meQTLs associated with differential gene expression (eg, eQTLs) in GTex (https://gtexportal.org/home/). HUGIn [15], and RegulomeDB [16] were used to examine chromatin state evidence predicting whether methylation markers fell into promoter or enhancer regions, whether associated methylation markers were in DNAase hypersensitivity regions or transcription factor binding sites, and to identify frequently interacting regions (FIREs).

RESULTS

This analysis included 221 adult subjects with HIV infection (Table 1). The Uganda cohort included 143 subjects (76 with TB), and the Tanzania cohort included 78 subjects (32 with TB). The subjects who did not have TB were all LTBI based on TST positivity. In the Uganda cohort, there were more males among TB cases than controls, reflecting the preponderance of TB among males in the general population. Thus, sex was included as a covariate in all analyses.

Differentially Methylated Regions Associated With TB

We identified 3 differentially methylated regions, the first 2 on chromosomes 1 and 2, respectively, with $P < 5 \times 10^{-5}$ in the Uganda sample and $P < .10$ in the Tanzania cohort with the same direction of effect (Table 2), and a third region on chromosome 5 that showed a significant ($P = 2 \times 10^{-5}$) differentially methylated marker detected in the Tanzania cohort that was replicated in the Uganda cohort ($P = .0398$) with effect estimates in the same direction in both populations. Another chromosome 5 marker also showed association in the Uganda cohort ($P = 2.08 \times 10^{-5}$), but did not replicate in the Tanzania cohort. The 2 methylation markers on chromosome 5 were uncorrelated in both cohorts ($r = 0.03$ in Uganda, $r = 0.18$ in Tanzania). The marker on chromosome 1 fell in a methylation island in *RNF220* (Figure 2A), the marker on chromosome 2 fell in an "open sea" (CpG sites not associated with a CpG island [CGI]) flanked by *COPS8* and *COL6A3* (Figure 2B), and the markers on chromosome 5 were on the "south shore" and "north shore" (regions up to 2 kb from a CGI), respectively, of the *CEP72* gene (Figure 2C).

Validation of Differentially Methylated Regions in GWAS

We then assessed whether SNPs from the GWAS associated with TB from the chromosomes 1, 2, and 5 differentially methylated CpG sites, in the same subjects [4] (Supplementary Table 1). All 3 regions had an SNP associated with TB within 200 kb of the differentially methylated sites ($P < .05$ unadjusted for multiple testing): chromosome 1, rs175222 ($P = .00016$; Figure 3A), chromosome 2, rs7586225 (within *COL6A3*, $P = .0082$; Figure 3B), and chromosome 5, rs12518227 ($P = .018$; Figure 3C). Associated SNPs in these regions were in regulatory regions, or in introns, and one on chromosome 5 is a missense variant (rs868649). One SNP, rs175222 on chromosome 1, maintained significance after Bonferroni correction.

meQTL Associations and SNP-Methylation Interaction

Next we examined both association between SNPs in these regions and methylation level of the CpG marker (meQTL analysis), and interaction between SNP genotype and methylation marker in its association with TB. Figure 3 illustrates their co-localization with the original methylation findings and marginal SNP associations with TB. Each region had a significant meQTL effect and SNP-methylation interaction effect. On chromosome 1 (Figure 3A), rs928685 was a significant meQTL ($P = .014$), and 3 SNPs significantly interacted with the methylation sites (rs270709, $P = .0108$; rs6664827, $P = .0405$; rs1890948, $P = .0412$), all within *RNF220*. On chromosome 2 (Figure 3B), there were 2 meQTLs—rs2645771, within *COL6A3* ($P = .0019$), and rs10165956 ($P = .0318$)—and one methylation-SNP interaction at rs4530312 ($P = .0457$). In GTex, one meQTL (rs2645771) is an eQTL for *COPS8* in cell-cultured fibroblasts (Table 3). On chromosome 5 (Figure 3C), there were 3 meQTLs: rs4956936 (within *AHRR*, $P = .0103$), rs1697952 ($P = .0316$), and rs6864158 (within *SLC9A3*, $P = .0392$). All 3 of these were also eQTLs, for multiple genes in both blood and lung, including *SLC9A3* and *CEP72* (Table 3). This region had one SNP that interacted with methylation status, rs12518227 (within *CEP72*, $P = .0405$).

Functional Implications of Differentially Methylated Regions

The methylation site on chromosome 1 falls in a FIRE and is a bivalent enhancer in multiple cells involved in TB immunity, including T cells, monocytes, and B cells; bivalent enhancers have been linked to increased gene expression [17]. The chromosome 2 region contains a histone modifier and is an enhancer in lung tissue; a FIRE crosses the *COL6A3* gene. The differentially methylated region on chromosome 5 falls within an active transcription start site in T cells and lung tissue, and flanks the transcription start site in monocytes, B cells, and neutrophils. Tracks showing these regions and functions are in Figure 2.

Table 2. Significantly Associated Methylation Markers

Methylation Marker Name	chr	Position	Relation to Island	UCSC_RefGene_Name	Uganda *P* Value	Tanzania *P* Value	Combined *P* Value
Methylation markers that are replicated in both cohorts							
cg19382731	chr1	44883990	Island	*RNF220*	4.62×10^{-5}	.07	4.63×10^{-5}
cg16974832	chr2	238188707	OpenSea	Flanking genes (*COPS8* and *COL6A3*)	4.47×10^{-5}	.04	2.69×10^{-5}
Different methylation markers in same region by population attaining significance							
cg18730862	chr5	611926	N_Shore	*CEP72*	2.08×10^{-5}	9.21×10^{-1}	2.28×10^{-4}
cg03602880	chr5	612961	S_Shore	*CEP72*	.0398	2.09×10^{-5}	1.25×10^{-5}

Abbreviations: chr, chromosome; UCSC, University of California, Santa Cruz Genome Browser.

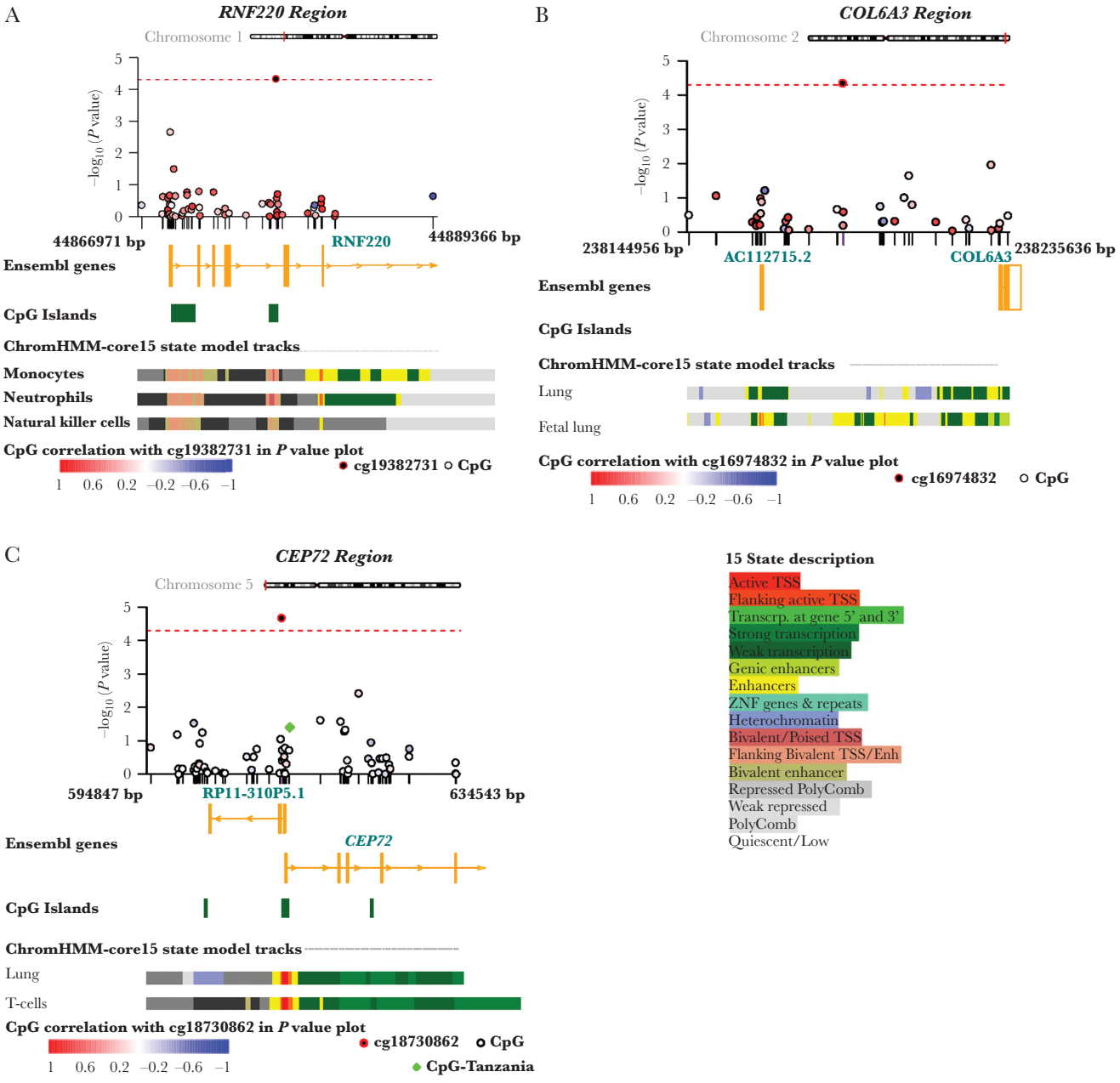

Figure 2. Regional plots of epigenetic-phenotype association results. CoMET plots of the top CpG methylation association results are shown. The top track shows the methylation-phenotype association *P* values across the region for the Uganda cohort, with CpG sites depicted as circles. Also shown in the top track are the estimated co-methylation values (correlation in DNA methylation values) between the reference (top) CpG site and others in the region, depicted by the fill of the circle, as a gradient from red to blue (correlation mapping shown in legend at the bottom of each panel). Additional tracks include Ensembl genes, CpG islands, and regulatory information for the specified tissues via the ChromHMM-core 15 state model. Ensembl genes shown may be truncated if they extend outside of the viewing window. *A*, *RNF220* gene CpG site cg1938271 shown ± 25 kb. *B*, *COL6A3* gene CpG site cg16974832 shown ± 50 kb. *C*, *CEP72* gene CpG site cg18730862 shown ± 25 kb. The CpG-Tanzania indicated site is for cg03602880 ($P = 2 \times 10^{-5}$ in the Tanzania cohort, $P = .04$ in the Uganda cohort); Spearman correlation between cg030602880 and cg18730862 was 0.11 overall (0.03 in Uganda only and 0.18 in Tanzania only). The 15 chromatin states and corresponding color mappings are shown at bottom right.

SNPs Significant in Original GWAS Are Also in Differentially Methylated Regions

We also examined loci that were genome-wide significant in our prior TB GWAS [4]; they contained significantly differentially methylated sites. The MWAS did not identify these regions as having significant methylation effects after multiple testing correction, but based on our original hypothesis that regions associated with TB risk in GWAS might also act through epigenetic mechanisms proved interesting. In most cases, the same methylation mark was not significant in both Uganda and Tanzania. For the methylation marks nearest *IL12B*, the most significant finding from our GWAS [4], differential methylation

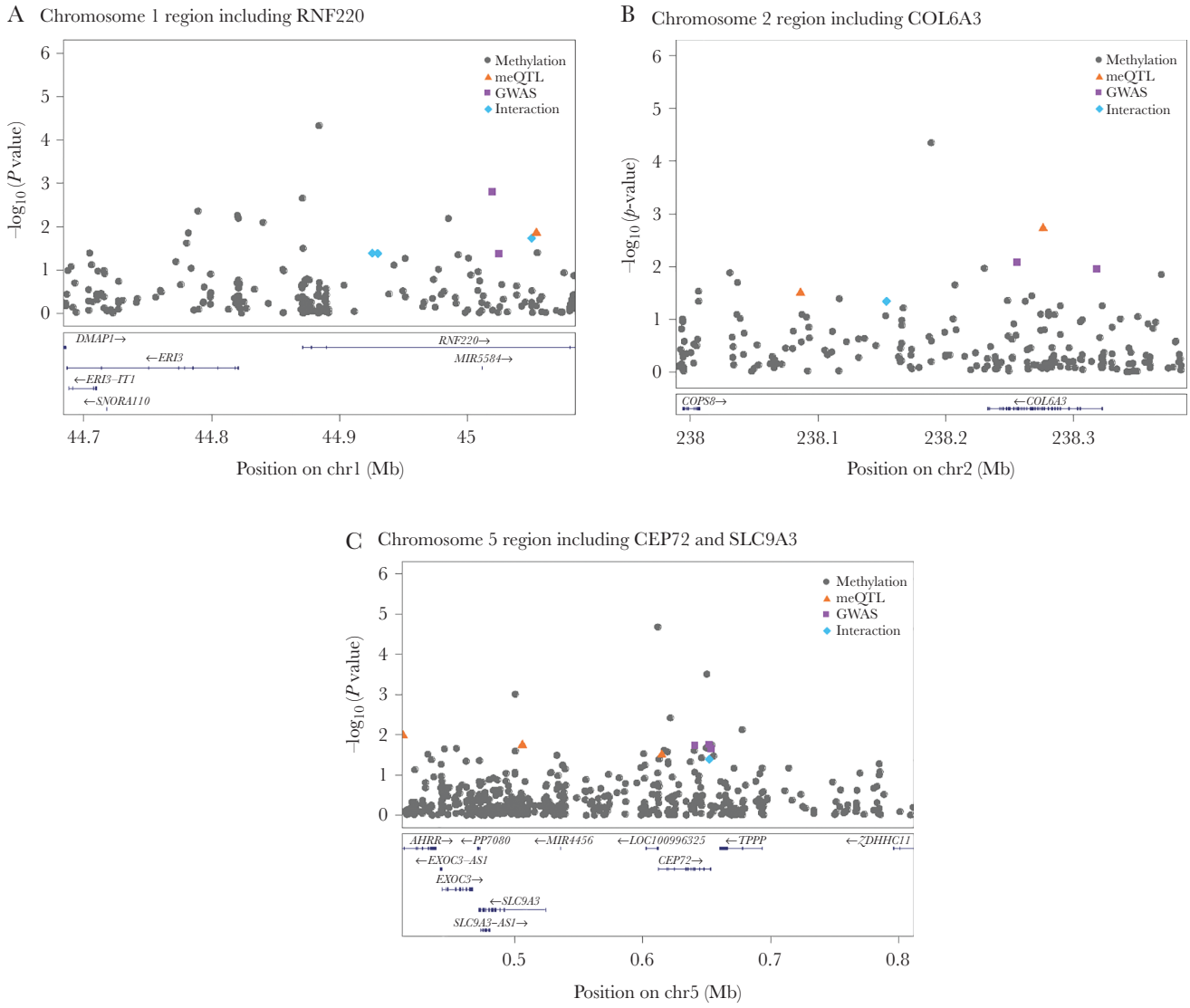

Figure 3. Plots of association *P* values from methylome-wide analysis (MWAS), genome-wide association study (GWAS), methylation QTL (meQTL), and methylation × single-nucleotide polymorphism (SNP) interaction analyses. MWAS *P* values come from the Uganda cohort, as it was the discovery cohort, and GWAS, meQTL, and methylation × SNP *P* values come from the combined cohort. Results from each type of analysis are represented by different colors and shapes, with −log(*P* value) plotted on the y-axis, and chromosomal location on the x-axis. *A*, Chromosome 1 (20 SNPs). *B*, Chromosome 2 (33 SNPs). *C*, Chromosome 5 (22 SNPs).

was associated with TB in both Uganda (cg15353886, $P = .0015$) and Tanzania (cg11092268, $P = .0089$). Two other methylation markers were replicated in the 2 samples from these regions, chromosome 5 (cg09049927, $P = .00089$ in combined data) and chromosome 17 (cg24357302, $P = .00071$ in combined data). Both of these methylation markers fell in "open seas," with the chromosome 5 marker falling near an microRNA, and the chromosome 17 in an intron of *ABCA8* (Supplementary Table 2).

DISCUSSION

Host factors play an important role in progression from LTBI to active TB disease, including genetic and transcriptomic factors [3, 18]. To our knowledge, only 2 small methylome-wide studies have been published [19, 20] in human cohorts, though in vitro studies have been conducted [21–23]. Our analysis of TB cohorts from Uganda and Tanzania revealed that 3 regions were differentially methylated in HIV-infected individuals who were protected from TB. These same regions contained nominally significant SNPs associated with TB, SNPs associated with methylation level, and SNPs interacting with methylation level in association with TB. Functional annotation revealed that these loci have regulatory effects on cells involved in the TB immune response and roles in gene expression, providing mechanistic scenarios of the results. In addition, regions previously shown to have genome-wide significant associations between SNPs and TB demonstrated differential methylation. Although none of these analyses alone provide compelling evidence for association, the concordance of the different analyses

Table 3. Functional Annotation of Significant Loci Identified From Methylome-wide Analysis With Follow-up Findings From Genome-wide Association Study (GWAS), Methylation QTL, and Functional Annotation From GTex and GWAS Catalog

MWAS Loci	Results From GWAS Data	meQTLs	GTex Findings for meQTLs	Significant SNP × Methylation Interaction	GWAS Catalog Results for Locus and Putative Function
Chromosome 1 *RNF220*	rs175222	rs928685	None	3 interactions	COPD, IL-6 levels
Chromosome 2 *COPS8* and *COL6A3*	rs7586225	rs2645771 rs10165956	rs2645771 for **COPS8** in cell-cultured fibroblasts	1 interaction	*COL6A3*: lung function (FEV/FCV); *COPS8*: leukocyte, granulocyte, and neutrophil counts
Chromosome 5 *CEP72* and *SLC9A3*	rs12518227	rs4956936 rs6864158 rs1697952	rs4956936: EXOC3 and C5orf55 in lung, and EXOC3, C5orf55, **SLC9A3**, CCDC127, CTD-228K2.5, PDCD6, and SDHA in blood; rs6864158: EXOC3, C5orf55, CTD-228K2.5, **SLC9A3** in blood rs1697952: **SLC9A3**, and AC026740.1 in lung, and EXOC3, C5orf55, TPPP, **SLC9A3, CEP72,** AC026740.1 in blood	1 interaction	*SLC9A3*: lung severity in cystic fibrosis
cg15353886 (Uganda) cg11092268 (Tanzania)	**IL12B**	NA	NA	NA	TB resistance and TB severity
cg24357302	**ABCA8**	NA	NA	NA	CRP levels

Boldface indicates that the same gene that showed a genetic association signal also shows differential gene expression effect.

Abbreviations: COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; FCV, forced vital capacity; FEV, forced expiratory volume; GWAS, genome-wide association study; IL6, interleukin 6; meQTL, methylation QTL; MWAS, methylome-wide analysis; NA, not applicable; SNP, single-nucleotide polymorphism; TB, tuberculosis.

does, indicating that epigenetic factors, together with genetic variation, can influence TB susceptibility.

The ability to resist active TB has significant genetic and environmental components, but how these features affect disease is unclear. Here, we examined the methylome to assess whether there are clear paths from exposure to disease, delineated by definable molecular mechanisms. Each locus we identified has been associated previously with either lung disease or regulation of cell types or cytokine levels relevant to lung function and/or TB. At least one of these loci has also been shown to associate with monocyte levels in GWAS (Table 3). These results indicate that TB susceptibility can operate through both genetic factors and response to environment mediated by methylation and/or interaction with genetic variants. Our results indicate that genetics and environment act together to regulate genes and pathways that affect the likelihood of exposed individuals progressing to TB. For example, cellular and humoral immunity are well-established components of the immune response to *Mtb*, but our results newly establish that epigenetic regulation of T cells, B cells, and monocytes can influence protection from disease. This may define a path from SNP to regulation of gene expression to protection from TB as well as suggest new drug targets for prevention of TB.

While the methylation marker on chromosome 2 resides in an "open sea," the SNPs associated with TB and methylation level, as well as SNP-methylation interaction, are within *COL6A3*. Collagen VI, a component of the extracellular matrix, plays a role in innate immune defense against bacteria and regulates autophagy [24, 25], thus indicating that *COL6A3* influences protection from TB via immune responses to TB antigens. Collagen VI–related myopathies are also associated with decreased pulmonary function [26]. *COL6A3* has been associated with lung cancer [27]. Thus, we hypothesize that *COL6A3* may influence TB susceptibility through its effect on both immune response and lung function. That the methylation site we identified falls in an enhancer region for lung tissue reinforces this hypothesis. The mechanism of action may be diverse as the associated SNPs are both regulatory and result in coding changes. *RNF220* is in the middle of the chromosome 1 region, and many associated SNPs are within the *RNF220* coding region. RNF220 enhances Wnt signaling [28], and thus may indicate a role for epigenetic modulation of Wnt signaling in the innate immune response to *Mtb* [29]. *RNF220* SNPs have also been associated with chronic obstructive pulmonary disease and interleukin 6 levels, the first relating to lung function and the second to TB and treatment response, as well as having SNPs previously associated with TB [30–35].

The associated methylation sites and SNPs on chromosome 5 cross multiple genes, but most results co-localize to *SLC9A3*. Another methylome-wide study found that *SLC9A3* was associated with atopy and asthma [36], and SNPs in *SLC9A3* were also associated with lung function in patients with cystic fibrosis [37,

38]. These data provide further support for our prior hypothesis [39] that lung immune responses are associated with protection from TB as previously observed in asthma [40, 41]. The meQTLs we identified are associated with *SLC9A3* and *CEP72* expression in lung, further supporting a role for these genes in lung function in TB. In further support, *SLC9A3* is also a component of a biomarker that predicts progression to TB [42]. A recent GWAS showed lung function associated with *CEP72* [43], another gene in this region with methylation association, so it possible that either or both gene(s) in this region are involved in TB susceptibility.

The results from our integrated analyses (MWAS, genetic association, meQTL, and methylation-SNP interaction) are challenging to interpret, but lay out connections between genetic variation, its implications for methylation/gene regulation, and TB. Nominally significant SNP associations, meQTLs, and SNP-methylation interaction findings together map to regions of interest, although different SNPs are associated. One explanation is that linkage disequilibrium patterns vary across African populations and true functional variants are tagged differently [44]. Limited sample sizes, variable allele frequencies, and distributions of the methylation marker may be additional factors. A second potential explanation for nonexact replication across cohorts may be exposure to different mycobacterial lineages and/or environments, such as cooking method and smoking, that may affect cellular phenotype. This study does not distinguish between methylation differences induced by *Mtb* stimulation, examined by other studies [21, 23, 45, 46]. Alternatively, subjects may resist development of TB via *Mtb*-induced methylation differently; future studies with larger sample sizes are needed to distinguish these hypotheses. Nonetheless, our results taken together indicate that methylation and genetic variation are both important factors in TB susceptibility that can be studied together to infer mechanisms, and effects of different types of genomic markers are not necessarily independent of each other.

It is important that our study subjects were not on ART or anti-TB treatment at the time of recruitment, as ART may influence methylation profiles [47–49]. This is a strength of our study, although given the ubiquity of ART, future replication studies will only identify differentially methylated regions that are robust in patients on ART. Because ART was unavailable in these countries during the study, CD4 counts were not routinely measured; these missing data would have been informative for analysis. Another potential limitation of our study is the use of a broad array instead of bisulfite sequencing; we only examined select sites for epigenetic modification. HIV may have an impact on differential methylation, but since all subjects in this study were HIV infected, that potential confounder was controlled. This may result in findings not generalizable to HIV-uninfected individuals. Our MWAS results do not attain significance after very stringent multiple testing correction, but support from SNP associations bolsters confidence in these loci. Last, the source of DNA for this assay was buffy coat, which consists of a variety of cell types, including some of the major cells involved in the TB immune response, potentially enriching findings significant to those cell types compared to others. Nonetheless, these findings clearly indicate that future studies should explore the role of epigenetic regulation of cellular and humoral immune responses with protection from TB along with other genomic data.

It is also important to understand how differential methylation affects RNA expression. However, in TB, this is not trivial. The most easily accessible tissue, peripheral blood mononuclear cells, may not be the most relevant tissue for TB. Some TB transcriptomic studies [50] stimulated monocyte-derived macrophages and examined the RNA expression change after *Mtb* stimulation. While this better approximates the immune response, it is more difficult to measure than RNA expression in blood as usually studied [10]. The relevant cell of interest is alveolar macrophages, and studies are ongoing to determine whether gene expression in the lung differs from that in circulating blood.

In conclusion, this study is the first to identify methylation changes associated with protection from active TB in HIV-infected patients. Our observation that differentially methylated sites are concordant with meQTLs and SNPs from GWAS indicates that multiple –omic levels can converge to identify loci of interest that can lead to treatment at several levels. Future studies will need to extend our findings and examine the impact of methylation on differential RNA expression and how they vary by patterns of genetic variation, especially in regulatory regions, and how methylation profiles differ by specific cell type. However, it is clear that patterns of methylation and genetic variation synergize to identify important associations in TB risk.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University. We acknowledge the invaluable contributions made by the study medical officers, health visitors, and laboratory and data personnel: Dr Lorna Nshuti, Dr Roy Mugerwa, Dr Sarah Zalwango, Allan Chiunda, Bonnie Thiel, Mark Breda, Dennis Dobbs, Hussein Kisingo, Mary Rutaro, Albert Muganda, Richard Bamuhimbisa, Yusuf Mulumba, Deborah Nsamba, Barbara Kyeyune, Faith Kintu, Dr Mary Nsereko, Gladys Mpalanyi, Janet Mukose, Grace Tumusiime, Pierre Peters, Annet Kawuma, Saidah Menya, Joan Nassuna, Alphonse Okwera, LaShaunda Malone, Keith Chervenak, Karen Morgan,

Alfred Etwom, Micheal Angel Mugerwa, Dr Brenda Okware, and Lisa Kucharski. We acknowledge and thank Dr Francis Adatu Engwau, Head of the Uganda National Tuberculosis and Leprosy Program, for his support of this project. We acknowledge the medical officers, nurses, and counselors at the National Tuberculosis Treatment Centre, Mulago Hospital, the Ugandan National Tuberculosis and Leprosy Program, and the Uganda Tuberculosis Investigation Bacteriological Unit, Wandegeya, for their contributions to this study. We also thank the staff at the Infectious Disease Center in Dar es Salaam, Tanzania, and Betty Mchaki and Mecky Matee at Muhimbili University of Health and Allied Sciences. This study would not have been possible without the generous participation of the Ugandan and Tanzanian patients and families.

Financial support. This work was supported by the National Institutes of Health, National Institute of Allergy and Infectious Diseases (grant numbers R56 AI130947 and R01AI124348). Recruitment and characterization of Ugandan study subjects was funded by the National Institute of Allergy and Infectious Diseases (grant numbers N01-AI95383 and HHSN266200700022C/ N01-AI70022).

Potential conflicts of interest. All authors: No reported conflicts of interest.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. World Health Organization. Global tuberculosis report. Geneva, Switzerland: WHO, **2020**.
2. Corbett EL, Watt CJ, Walker N, et al. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch Intern Med **2003**; 163:1009–21.
3. Stein CM, Sausville L, Wejse C, et al. Genomics of human pulmonary tuberculosis: from genes to pathways. Curr Genet Med Rep **2017**; 5:149–66.
4. Sobota RS, Stein CM, Kodaman N, et al. A locus at 5q33.3 confers resistance to tuberculosis in highly susceptible individuals. Am J Hum Genet **2016**; 98:514–24.
5. Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet **2010**; 11:446–50.
6. Cavalli G, Heard E. Advances in epigenetics link genetics to the environment and disease. Nature **2019**; 571:489–99.
7. Weichenhan D, Plass C. The evolving epigenome. Hum Mol Genet **2013**; 22:R1–6.
8. Liu Y, Li H, Xiao T, Lu Q. Epigenetics in immune-mediated pulmonary diseases. Clin Rev Allergy Immunol **2013**; 45:314–30.
9. Richardson TG, Zheng J, Davey Smith G, et al. Mendelian randomization analysis identifies CpG sites as putative mediators for genetic influences on cardiovascular disease risk. Am J Hum Genet **2017**; 101:590–602.
10. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. Genome Res **2012**; 22:1748–59.
11. Stein CM, Zalwango S, Malone LL, et al. Resistance and susceptibility to *Mycobacterium tuberculosis* infection and disease in tuberculosis households in Kampala, Uganda. Am J Epidemiol **2018**; 187:1477–89.
12. von Reyn CF, Mtei L, Arbeit RD, et al; DarDar Study Group. Prevention of tuberculosis in bacille Calmette-Guérin-primed, HIV-infected adults boosted with an inactivated whole-cell mycobacterial vaccine. AIDS **2010**; 24:675–85.
13. 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.
14. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet **2007**; 81:559–75.
15. Martin JS, Xu Z, Reiner AP, et al. HUGIn: Hi-C unifying genomic interrogator. Bioinformatics **2017**; 33:3793–5.
16. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res **2012**; 22:1790–7.
17. Bernhart SH, Kretzmer H, Holdt LM, et al. Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. Sci Rep **2016**; 6:37393.
18. Cliff JM, Kaufmann SH, McShane H, van Helden P, O'Garra A. The human immune response to tuberculosis and its treatment: a view from the blood. Immunol Rev **2015**; 264:88–102.
19. Chen YC, Hsiao CC, Chen TW, et al. Whole genome DNA methylation analysis of active pulmonary tuberculosis disease identifies novel epigenotypes: PARP9/miR-505/RASGRP4/GNG12 gene methylation and clinical phenotypes. Int J Mol Sci **2020**; 21:3180.
20. Esterhuyse MM, Weiner J 3rd, Caron E, et al. Epigenetics and proteomics join transcriptomics in the quest for tuberculosis biomarkers. mBio **2015**; 6:e01187–15.
21. DiNardo AR, Rajapakshe K, Nishiguchi T, et al. DNA hypermethylation during tuberculosis dampens host immune responsiveness. J Clin Invest **2020**; 130:3113–23.
22. Sharma P, Garg G, Kumar A, et al. Genome wide DNA methylation profiling for epigenetic alteration in coronary artery disease patients. Gene **2014**; 541:31–40.
23. Asaad M, Abo-Kadoum MA, Nzungize L, Uae M, Nzaou SAE, Xie J. Methylation in *Mycobacterium*-host interaction and implications for novel control measures. Infect Genet Evol **2020**; 83:104350.
24. Abdillahi SM, Tati R, Nordin SL, et al. The pulmonary extracellular matrix is a bactericidal barrier against *Haemophilus influenzae* in chronic obstructive pulmonary

disease (COPD): implications for an in vivo innate host defense function of collagen VI. Front Immunol **2018**; 9:1988.
25. Schnoor M, Cullen P, Lorkowski J, et al. Production of type VI collagen by human macrophages: a new dimension in macrophage functional heterogeneity. J Immunol **2008**; 180:5707–19.
26. Foley AR, Quijano-Roy S, Collins J, et al. Natural history of pulmonary function in collagen VI-related myopathies. Brain **2013**; 136:3625–33.
27. Duan Y, Liu G, Sun Y, et al. COL6A3 polymorphisms were associated with lung cancer risk in a Chinese population. Respir Res **2019**; 20:143.
28. Ma P, Yang X, Kong Q, et al. The ubiquitin ligase RNF220 enhances canonical Wnt signaling through USP7-mediated deubiquitination of β-catenin. Mol Cell Biol **2014**; 34:4355–66.
29. Villaseñor T, Madrid-Paulino E, Maldonado-Bravo R, Urbán-Aragón A, Pérez-Martínez L, Pedraza-Alva G. Activation of the Wnt pathway by mycobacterium tuberculosis: a Wnt-Wnt situation. Front Immunol **2017**; 8:50.
30. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al; SpiroMeta Consortium; International COPD Genetics Consortium. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nat Genet **2019**; 51:494–505.
31. Tekola Ayele F, Doumatey A, Huang H, et al. Genome-wide associated loci influencing interleukin (IL)-10, IL-1Ra, and IL-6 levels in African Americans. Immunogenetics **2012**; 64:351–9.
32. Clifford V, Tebruegge M, Zufferey C, et al. Cytokine biomarkers for the diagnosis of tuberculosis infection and disease in adults in a low prevalence setting. Tuberculosis (Edinb) **2019**; 114:91–102.
33. Wang JY, Chang HC, Liu JL, et al. Expression of Toll-like receptor 2 and plasma level of interleukin-10 are associated with outcome in tuberculosis. Eur J Clin Microbiol Infect Dis **2012**; 31:2327–33.
34. Bekker LG, Maartens G, Steyn L, Kaplan G. Selective increase in plasma tumor necrosis factor-alpha and concomitant clinical deterioration after initiating therapy in patients with severe tuberculosis. J Infect Dis **1998**; 178:580–4.
35. Mattos AM, Almeida Cde S, Franken KL, et al. Increased IgG1, IFN-gamma, TNF-alpha and IL-6 responses to *Mycobacterium tuberculosis* antigens in patients with tuberculosis are lower after chemotherapy. Int Immunol **2010**; 22:775–82.
36. Forno E, Wang T, Qi C, et al. DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study. Lancet Respir Med **2019**; 7:336–46.
37. Dorfman R, Taylor C, Lin F, et al; Members of Canadian Consortium for CF Genetic Studies. Modulatory effect of the SLC9A3 gene on susceptibility to infections and pulmonary function in children with cystic fibrosis. Pediatr Pulmonol **2011**; 46:385–92.
38. Corvol H, Blackman SM, Boëlle PY, et al. Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. Nat Commun **2015**; 6:8382.
39. Sobota RS, Stein CM, Kodaman N, et al. A chromosome 5q31.1 locus associates with tuberculin skin test reactivity in HIV-positive individuals from tuberculosis hyper-endemic regions in East Africa. PLoS Genet **2017**; 13:e1006710.
40. von Hertzen L, Klaukka T, Mattila H, Haahtela T. *Mycobacterium tuberculosis* infection and the subsequent development of asthma and allergic conditions. J Allergy Clin Immunol **1999**; 104:1211–4.
41. von Mutius E, Pearce N, Beasley R, et al. International patterns of tuberculosis and the prevalence of symptoms of asthma, rhinitis, and eczema. Thorax **2000**; 55:449–53.
42. Duffy FJ, Weiner J 3rd, Hansen S, et al; GC6-74 Consortium. Immunometabolic signatures predict risk of progression to active tuberculosis and disease outcome. Front Immunol **2019**; 10:527.
43. Shrine N, Guyatt AL, Erzurumluoglu AM, et al; Understanding Society Scientific Group. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nat Genet **2019**; 51:481–93.
44. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet **2002**; 3:611–21.
45. Tarashi S, Badi SA, Moshiri A, et al. The inter-talk between *Mycobacterium tuberculosis* and the epigenetic mechanisms. Epigenomics **2020**; 12:455–69.
46. Sharma G, Sowpati DT, Singh P, et al. Genome-wide non-CpG methylation of the host genome during *M. tuberculosis* infection. Sci Rep **2016**; 6:25006.
47. Ay E, Banati F, Mezei M, et al. Epigenetics of HIV infection: promising research areas and implications for therapy. AIDS Rev **2013**; 15:181–8.
48. Gröschel B, Höver G, Doerr HW, Cinatl J Jr. Zidovudine (AZT) resistance in H9 cells due to decreased TK expression is associated with hypermethylation of TK gene. Nucleosides Nucleotides Nucleic Acids **2001**; 20:487–92.
49. Wu S, Liu X, Solorzano MM, Kwock R, Avramis VI. Development of zidovudine (AZT) resistance in Jurkat T cells is associated with decreased expression of the thymidine kinase (TK) gene and hypermethylation of the 5′ end of human TK gene. J Acquir Immune Defic Syndr Hum Retrovirol **1995**; 8:1–9.
50. Thuong NT, Dunstan SJ, Chau TT, et al. Identification of tuberculosis susceptibility genes with human macrophage gene expression profiles. PLoS Pathog **2008**; 4:e1000229.