

**Methylome-wide analysis reveals epigenetic marks associated with resistance to tuberculosis in HIV-infected individuals from East Africa**

Catherine M. Stein<sup>1,2\*</sup>, Penelope Benchek<sup>1\*</sup>, Jacqueline Bartlett<sup>1</sup>, Robert P. Igo, Jr.<sup>1</sup>, Rafal S. Sobota<sup>4</sup>, Keith Chervenak<sup>2</sup>, Harriet Mayanja-Kizza<sup>5</sup>, C. Fordham von Reyn<sup>6</sup>, Timothy Lahey<sup>6</sup>, William S. Bush<sup>1</sup>, W. Henry Boom<sup>2</sup>, William K. Scott<sup>7</sup>, Carmen Marsit<sup>8</sup>, Giorgio Sirugo<sup>9+</sup>, Scott M. Williams<sup>1,3+</sup>

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

\*These authors contributed equally as first authors of this work.

+ These authors contributed equally as senior authors of this work

## ABSTRACT

**Background:** Tuberculosis (TB) is the most deadly infectious disease globally and highly prevalent in the developing world, especially sub-Saharan Africa. Even though a third of humans are exposed to *Mycobacterium tuberculosis* (Mtb), most infected immunocompetent individuals do not develop active TB. In contrast, for individuals infected with both TB and the human immunodeficiency virus (HIV), the risk of active disease is 10% or more per year. Previously, we identified in a genome-wide association study a region on chromosome 5 that was associated with resistance to TB. This region included epigenetic marks that could influence gene regulation so we hypothesized that HIV-infected individuals exposed to Mtb, who remain disease free, carry epigenetic changes that strongly protect them from active TB. To test this hypothesis, we conducted a methylome-wide study in HIV-infected, TB-exposed cohorts from Uganda and Tanzania.

**Results:** In 221 HIV-infected adults from Uganda and Tanzania, we identified 3 regions of interest that included markers that were differentially methylated between TB cases and LTBI controls, that also included methylation QTLs and associated SNPs: chromosome 1 (*RNF220*,  $p=4 \times 10^{-5}$ ), chromosome 2 (between *COPS8* and *COL6A3* genes,  $p=2.7 \times 10^{-5}$ ), and chromosome 5 (*CEP72*,  $p=1.3 \times 10^{-5}$ ). These methylation results colocalized with associated SNPs, methylation QTLs, and methylation x SNP interaction effects. These markers were in regions with regulatory markers for cells involved in TB immunity and/or lung.

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

**Keywords (3-10):** methylation, epigenetics, infectious disease, genetics, genomics, lung function, immunology

## BACKGROUND

*Mycobacterium tuberculosis* (Mtb), the cause of tuberculosis (TB), results in approximately 1.5 million deaths per year [1], but the vast majority of the almost 2 billion people infected with Mtb do not progress to active TB. Although the risk of developing active TB is low in most infected people, it is the most common cause of death in people with HIV infection who live in TB endemic countries [2, 3]. As many as 10% of co-infected people develop active disease each year, illustrating how strongly immune compromise contributes to risk. People with HIV infection, who do not develop active TB despite Mtb infection, offer a major opportunity to understand resistance to TB that persists despite being immunocompromised, and possibly a key to how any Mtb infection leads to active TB.

Several studies indicate that susceptibility (or resistance) to TB is at least partially due to genomic factors [4]. Indeed, several genome-wide association studies (GWAS) have been conducted, identifying loci of interest, but most of these studies show small effect sizes in HIV uninfected subjects. In contrast, our GWAS [5], conducted in HIV-infected subjects, found significant association with resistance to TB and a region of chromosome 5 containing the *IL12B* gene. Annotation of this region showed that the associated SNPs resided in a histone mark, indicating that epigenetic marks in this region 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 disease risk factors, and since epigenetic marks can be inherited, they have been hypothesized to explain some of the “missing heritability” for complex diseases [6-8]. Evidence for this model has been found in cancer and autoimmune diseases, where associations have been demonstrated between epigenetic marks and disease risk [9, 10]. Based on these studies and our own prior results on chromosome 5, we hypothesized that

regulatory factors, including methylation marks, associate with TB susceptibility/resistance, beyond the genetic influences encoded in the DNA.

Since disease risk may be influenced by epigenetic variation, either directly or indirectly via DNA-level variation [11], we examined loci identified through an methylome-wide analysis (MWAS) and their association with single nucleotide polymorphisms (SNPs) at those same loci. We also investigated differential methylation near loci initially identified through GWAS of TB risk (Figure 1). Our underlying model was that both genetic and epigenetic factors are associated with risk of TB and that the two may modify the effects of each other. To understand the functional implications of identified loci, we annotated these loci using available databases. By integrating results across multiple platforms and data types [12], we extended GWAS results to show that a framework containing both epigenetic changes and DNA variations, and their interactions, can associate with regulation of lung and immune cell function relevant to TB pathogenesis.

## RESULTS

### **Overall approach**

Our approach (Figure 1) was to first conduct an MWAS in the Ugandan and Tanzanian cohorts independently. 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 in the discovery cohort at  $p < 5 \times 10^{-5}$  to account for multiple testing, and  $p < 0.1$  in the replication cohort were considered for subsequent analyses. In the next step, we examined SNPs from our GWAS  $\pm 200$  kb of the methylation signal, and identified whether they were significantly ( $p < 0.05$ ) associated with TB; we did not

adjust for multiple testing because these and subsequent analyses were hypothesis-driven. In the third step, we conducted a meQTL analysis, where we examined SNPs  $\pm$  200 kb of the associated methylation marker to test for association with methylation of the associated CpG site. Fourth, we examined the interaction between methylation status and SNPs in association with TB. A comparison 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. Lastly, all loci significant in the first step underwent functional annotation as described in the Methods. In parallel, we examined the association between methylation level and TB in regions that were significantly associated with TB in our GWAS (Table 1 from Sobota et al.[5]).

### **Sample description**

This analysis included a total of 221 adult subjects with HIV infection (Table 1). The Uganda cohort included 143 subjects, of whom 76 had TB. The Tanzania cohort included 78 subjects, of whom 32 had TB. The subjects who did not have active TB all had latent Mtb infection (LTBI) based on a positive tuberculin skin test. No subjects received antiretroviral therapy (ART) for HIV because all were enrolled prior to the availability of ART in Africa. In the Uganda cohort, there was a greater proportion of males among TB cases than controls, reflecting the greater preponderance of TB among males in the general population. Thus, sex was considered as a covariate in all analyses.

### **Differentially methylated regions associated with TB**

In our analysis, we identified 3 regions, the first two on chromosomes 1 and 2, respectively, that were differentially methylated with  $p < 5 \times 10^{-5}$  in the Uganda sample and  $p < 0.10$  in the

Tanzania cohort with effect estimates in the same direction (Table 2) and a third region, on chromosome 5 that showed a significant ( $p=2 \times 10^{-5}$ ) differentially methylated marker detected in the Tanzanian cohort that replicated in Ugandan cohort ( $p=0.0398$ ) and had effect estimates in the same direction in both populations. A nearby chromosome 5 marker also showed association in the Ugandan cohort ( $p=2.08 \times 10^{-5}$ ), but did not replicate in the Tanzanian cohort. The two 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 or 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 interrogated the regions on chromosomes 1, 2, and 5 that contained differentially methylated CpG sites in the GWAS dataset that included the same subjects [5] (Supplemental Table 1). All three regions had a SNP associated with TB within 200 kb of the differentially methylated sites ( $p < 0.05$  unadjusted for multiple testing): chromosome 1, rs175222 ( $p=0.00016$ ) (Figure 3a), chromosome 2 rs7586225 (within *COL6A3*,  $p=0.0082$ ) (Figure 3b), and chromosome 5, rs12518227 ( $p=0.018$ ) (figure 3c). Associated SNPs in these regions are in regulatory regions, or in introns, and one on chromosome 5 is a missense variant (rs868649). One of these SNPs (rs175222 on chromosome 1) maintained significance after Bonferroni correction.

### **SNP-methylation QTL associations and SNP x methylation interaction**

Next we examined both association between SNPs in these regions and the level of methylation of the methylation site (meQTL analysis), and also examined the interaction between SNP genotype and methylation marker in its association with TB. These results are shown in Figure 3 to illustrate their co-localization with the original methylation findings and marginal SNP associations with TB. Again, each region had a significant meQTL effect and SNP x methylation interaction effect. On chromosome 1 (Figure 3a), rs928685 was a significant meQTL ( $p=0.014$ ), and there were 3 SNPs with significant interaction effects with the methylation site (rs270709,  $p=0.0108$ ; rs6664827,  $p=0.0405$ ; rs1890948,  $p=0.0412$ ), all within the *RNF220* gene. On chromosome 2 (Figure 3b), there were 2 meQTLs — rs2645771, within *COL6A3* ( $p=0.0019$ ), and rs10165956 ( $p=0.0318$ ) — and 1 methylation x SNP interaction at rs4530312 ( $p=0.0457$ ). On chromosome 5 (Figure 3c), there were 3 meQTLs: rs4956936 (within *AHRR*,  $p=0.0103$ ), rs1697952 ( $p=0.0316$ ), and rs6864158 (within *SLC9A3*,  $p=0.0392$ ). This region had one SNP that interacted with methylation status, rs12518227 (within *CEP72*,  $p=0.0405$ ). The SNPs that associate with methylation and those that associate with TB are in all cases different.

### **Functional implications of 3 differentially methylated regions**

The methylation site on chromosome 1 falls in a frequently interacting region (FIRE) and is a bivalent enhancer in a variety of cells involved in TB immunity, including T cells, monocytes, and B cells; bivalent enhancers have been linked to increased gene expression [13]. 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 functional regions in these regions are shown in Figure 2.

### **SNPs significant in original GWAS are also in differentially methylated regions**

We also examined loci that were genome-wide significant in our prior GWAS for TB [5], and found that they contained significantly differentially methylated sites. The MWAS did not identify these regions as having significant methylation effects after multiple testing correction.

However, we explored these based on our original hypothesis that regions associated with TB risk in GWAS might also act through epigenetic mechanisms. In most cases, the same methylation marker was not significant in both the Uganda and Tanzania datasets. For the methylation marks nearest *IL12B*, the most significant finding from our GWAS [5], differential methylation associated with TB in both Uganda (cg15353886,  $p=0.0015$ ) and Tanzania (cg11092268,  $p=0.0089$ ). Two other methylation markers that replicated in the two samples from these regions fell on chromosome 5 (cg09049927,  $p=0.00089$  in combined data) and chromosome 17 (cg24357302,  $p=0.00071$  in combined data). Both of these methylation markers fell in “open seas”, with the chromosome 5 marker falling near an miRNA, and the chromosome 17 in an intron of *ABCA8* (Supplemental Table 2).

### **DISCUSSION**

Host factors play an important role in the progression from LTBI to active TB disease, including genetic and transcriptomic factors [4, 14]. To our knowledge, only two small methylome-wide studies have been published [15, 16] in human cohorts, though *in vitro* studies have been conducted [17-19]. Our analysis of TB patient cohorts from Uganda and Tanzania revealed three regions that were differentially methylated in HIV-infected individuals who exhibited protection from active TB. These same regions contained nominally significant SNPs associated with TB, as well as 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, providing mechanistic plausibility of the results. In addition, regions that were previously shown to have genome-wide significant associations between SNPs and TB also demonstrated differential methylation. Although none of these analyses alone provide compelling evidence for association, the concordance of the different analyses does; these results indicate that epigenetic factors, in combination with genetic variation, can influence susceptibility to TB.

Cellular and humoral immunity are well established components of the immune response to *Mtb*, but our results newly establish that epigenetic regulation of T cell, B cell and monocytes can influence protection from active TB. This may define a path from SNP to regulation of gene expression to protection from TB disease as well as suggest new drug targets for prevention of active TB. As seen in Figure 2, the three regions identified through our analyses contain regulatory markers for monocytes, T cells, natural killer cells, and lung tissue, among others, suggesting a role for these specific loci in TB susceptibility.

While the methylation marker on chromosome 2 resides in an “open sea”, the SNPs associated with TB and methylation level, as well as SNP x methylation interaction, are all within *COL6A3*. Collagen VI, as a component of the extracellular matrix, plays a role in innate immune defense against bacteria and regulates autophagy [20, 21], thus indicating that *COL6A3* influences protection from TB via influencing immune responses to TB antigens. Collagen VI-related myopathies are associated with decreased pulmonary function [22]. *COL6A3* has been associated with lung cancer [23]. Thus, we hypothesize that *COL6A3* may influence TB susceptibility through its effect on lung function. The fact 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 also both regulatory and result in coding changes. *RNF220* is in the middle of the chromosome 1 region, and many of the associated SNPs are within the *RNF220* coding region (Figure 3a). *RNF220* is known to enhance Wnt signaling [24], and thus may indicate a role for epigenetic modulation of Wnt signaling in the innate immune response to *Mtb* [25].

The associated methylation sites and SNPs on chromosome 5 cross multiple genes, but most results co-localize to *SLC9A3* (Figure 3c). Another methylome-wide study found association between *SLC9A3* and atopy and asthma [26], whereas SNPs in *SLC9A3* were also associated with lung function in cystic fibrosis patients [27, 28]. These data provide further support for our prior hypothesis [29] that lung immune responses associate with protection from TB such as previously observed in patients with asthma [30, 31]. In further support of this hypothesis, *SLC9A3* is also a component of a biomarker that predicts progression to TB [32]. A recent GWAS of lung function identified an association with *CEP72* [33], which is also in this region with associated methylation findings, so it possible that either, or both, gene(s) in this region are involved in TB susceptibility.

The results from our connected analyses (MWAS, genetic association, meQTL, and methylation x SNP interaction) are challenging to interpret, but do lay out a possible connection between genetic variation, its implications for methylation/gene regulation and TB. Nominally significant SNP association, meQTLs, and/or SNP x methylation interaction findings co-localize in regions that map to genes of interest, though different SNPs were associated in each of these terms. One explanation is because linkage disequilibrium patterns vary across African populations and

the true functional variants are tagged differently across cohorts [34]. Limited sample size, combined with variable allele frequencies and distributions of the methylation marker, are additional factors. A second potential explanation for non-exact replication across cohorts may be exposure to different Mycobacterial lineages and/or environmental differences, such as cooking method, that may affect cellular phenotype. This study does not distinguish between methylation differences induced by Mtb stimulation, which has been shown by other studies [17, 19, 35, 36] versus inherent differences between individuals with TB or LTBI. An alternative hypothesis may be that subjects who resist development of TB respond to Mtb-induced methylation differently; future studies are needed to distinguish these hypotheses. Larger sample sizes will be important to elucidate the effects of each of these factors. Nonetheless, our results taken as a whole indicate that methylation and genetic variation are both important factors in TB susceptibility, and the effects are not necessarily independent of each other.

It is important to emphasize that the study subjects in this analysis were not on ART or anti-TB treatment at the time that blood was drawn for this study, as methylation profiles may be influenced by ART [37-39]. 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. Another potential limitation of our study is the use of a broad array instead of bisulfite sequencing, which examines only select nucleotides 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 that are not generalizable to HIV-uninfected. Our MWAS results do not attain multiple testing correction, but the support of the aforementioned SNP association tests bolster confidence in these findings. Lastly, the source of DNA for this assay was buffy coat, which consists of a variety of cell types, including 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 future studies should explore the linkage between epigenetic regulation of cellular and humoral immune responses with protection from TB along with other genomic data.

It is also of interest to understand how differential methylation affects RNA expression.

However, this question in TB is not trivial. The most easily accessible tissue, peripheral blood mononuclear cells, may not be the relevant tissue for TB. Some TB transcriptomic studies [40] stimulate monocyte-derived macrophages and examine the RNA expression change after stimulation with Mtb. While this gets closer to the immune response, it is more difficult to measure than RNA expression level at a putative baseline that is usually what is studied [12].

The primary cell of interest would be alveolar macrophages. There are ongoing studies to determine whether gene expression in the lung is different than in circulating blood.

## CONCLUSION

In conclusion, this is the first study to identify methylation changes associated with protection from active TB in patients with HIV-associated susceptibility to TB disease. Methylation is of interest because it is a possible pharmacologic target that describes cellular response to MTB infection and disease progression. Future studies will need to extend our findings and examine the impact of methylation on differential RNA expression and how these can vary by patterns of genetic variation, especially in regulatory regions, and how methylation profiles differ by specific cell type. However, it is now clear that patterns of methylation and genetic variation synergize with each other and can identify important associations in TB risk.

## METHODS

### **Study participants**

This study includes the same subjects as in our previous GWAS [5]. Briefly, subjects were from a household contact study in Kampala, Uganda [41], or two clinical trials of TB in Dar es Salaam, Tanzania [42]. All subjects were HIV-infected, over the age of 15, and none received ART due to unavailability at the time of enrollment. TB cases were culture-confirmed. Controls all had LTBI based on a positive tuberculin skin test; interferon- $\gamma$  release assays were not performed at the time of subject ascertainment. All subjects provided written informed consent. The present analyses utilized samples from those subjects who had DNA remaining after the GWAS.

### **Molecular methods for MWAS and GWAS**

DNA came from buffy coat samples in the Uganda cohort and either buffy coat or whole blood in the Tanzania cohort, and was prepared as previously described [5] then bisulfite converted according to the specifications for the Illumina Methylation EPIC 850k chip. GWAS data were available from our previous analysis [5].

### **Quality control and statistical analysis**

#### *Quality control and principal components analysis*

Quality control and normalization of raw methylation data (as Illumina .idat files) were applied to the combined Uganda and Tanzania cohorts and carried out using the Bioconductor package minfi for R [43, 44]. Data for probes and samples with good detection p-values (less than 0.01) were retained. We normalized signal intensity by means of the BMIQ algorithm [45], which

adjusts for differences between Infinium I and II probes, adjusted for batch effects using ComBat [46] and adjusted for between-array technical variation, as well as, background variation using the functional normalization procedure [47].

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. Within each cohort, we conducted a methylome-wide principal components analysis, using the prcomp function in R. We regressed our TB outcome on each of the first 20 principal components (PCs) selecting PCs that were significantly associated (at p-value < 0.1) with TB in a model that included cell proportions. This resulted in multiple PCs and cell proportions (dependent on cohort) retained as covariates in each cohort specific MWAS.

#### *MWAS Analysis*

Within each of the two cohorts we tested for association between CpG beta values (converted to M values ( $\log_2(\beta/1-\beta)$ ) and TB status using limma in R [48]. Here we adjusted for significant methylation-based PCs, cell proportions, age and sex in a linear model that compares methylation values between TB and LTBI samples. Because Uganda was the discovery cohort, we sought to replicate top hits from the Uganda cohort (at p-value <  $5 \times 10^{-5}$ ) in the Tanzania cohort (with p-value < 0.10).

#### *Genetic association analysis*

Genetic association analysis was performed using the full data from Sobota et al. [5] Association between SNPs in the regions of interest and TB was conducted using PLINK [49]

for SNPs with minor allele frequency  $> 0.10$ . We examined SNPs within  $\pm 200$  kb of the associated methylation marker. A Bonferonni-corrected p-value was derived based on the total number of SNPs tested across all 3 regions (20 SNPs on chromosome 1, 33 on chromosome 2, and 22 on chromosome 3).

#### *meQTL analysis*

We conducted a targeted cis-methylation QTL analysis around the replicated CpG sites in the combined Uganda/Tanzania sample that had overlapping methylation and genotype data (total  $N = 188$ , with 75 from the Tanzania cohort (32 with TB) and 113 from the Uganda cohort (66 with TB)). SNPs within  $\pm 200$  kb of the replicated CpG sites were pulled from the GWAS dataset [5]. Using Matrix eQTL, we ran a linear regression model adjusting for gender, age, cohort, cell proportions (CD4, CD8, monocytes, neutrophils) and the first two genetic-based PCs [5] to determine if nearby SNPs were influencing the methylation levels of our replicated CpG sites.

#### *SNP x methylation interaction*

We conducted a targeted interaction analysis around the replicated CpG sites in the combined Uganda/Tanzania sample that had overlapping methylation and genotype data (sample sizes as above). As in the meQTL analysis, SNPs within 200 kb of the replicated CpG sites were pulled from the GWAS dataset [5]. Using glm in R we ran a logistic regression model adjusting for gender, cohort, CD4 cell proportions and the first two GWAS-based PCs [5] to determine if the difference in methylation levels between TB and LTBI samples are modified by genotypes of nearby SNPs.

## Functional annotation

HUGIn [50] and RegulomeDB [51] were used to examine chromatin state evidence predicting whether or not our methylation markers fell in promoter or enhancer regions, whether associated methylation markers were in regions of DNAase hypersensitivity, transcription factor binding sites, and promoter regions, and also to identify frequently interacting regions (FIREs).

## LIST OF ABBREVIATIONS

ARVs	anti-retroviral therapy
GWAS	genome-wide association study
HIV	Human Immunodeficiency virus
LTBI	latent <i>M. tuberculosis</i> infection
meQTL	methylation quantitative trait locus
Mtb	<i>Mycobacterium tuberculosis</i>
MWAS	methylome-wide association study
SNP	single nucleotide polymorphism
TB	Tuberculosis

## DECLARATIONS

### Ethics approval and consent to participate

Informed consent was obtained from all patients in the extended DarDar follow-up cohort. The research ethics committee at the Muhimbili University of Health and Allied Sciences and the Committee for the Protection of Human Subjects at Dartmouth College and the Dartmouth-Hitchcock Medical Center approved this study. Informed consent was obtained from all subjects in the Household Contact study in Kampala, Uganda. Ethics committees that approved this work were at Muhimbili University of Health and Allied Sciences, Committee for the Protection of

Human Subjects at Dartmouth College (#14606), Uganda Council for Science and Technology, and University Hospitals of Cleveland (10-01-25)

### **Consent for publication**

N/A

### **Availability of data and materials**

Summary data (marker names and p-values) will be provided at Dryad (<https://github.com/CDL-Dryad/dryad>). The Ugandan IRB did not grant approval for broad genetic data sharing, so individual-level data may be requested from the Data Access Committee, chaired by Dr. Sudha Iyengar ([ski@case.edu](mailto:ski@case.edu)). Methylation data from the Tanzanian cohort will be available at Dryad (<https://github.com/CDL-Dryad/dryad>).

### **Competing interests**

N/A

### **Funding**

This work was supported by NIH grants R56 AI130947 and R01AI124348. Recruitment and characterization of Ugandan study subjects was funded by grant N01-AI95383 and HHSN266200700022C/ N01-AI70022 from the NIAID.

### **Authors' contributions**

CMS, SMW, and GS conceived of the study. CS, SMW, CM, WSB, and WKS directed the statistical analyses. WKS supervised genotyping and methylation typing. PB, JB, RSS, and RPI, Jr conducted the statistical analyses. CMS, H M-K, WHB, CFvR, and TL oversaw the clinical characterization of the subjects. KC and RSS obtained DNA samples and supervised shipment of samples to the laboratory. CMS and PB drafted the initial manuscript. All authors read and approved the final manuscript.

### **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 would like to acknowledge the invaluable contribution made by the study medical officers, health visitors, 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 would like to acknowledge and thank Dr. Francis Adatu Engwau, Head of the Uganda National Tuberculosis and Leprosy Program, for his support of this project. We would like to 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, Wandegaya, for their contributions to this study. We would also like to 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 be possible without the generous participation of the Ugandan and Tanzanian patients and families.

## REFERENCES CITED

1. Organization. WH: **Global tuberculosis control : epidemiology, strategy, financing : WHO report 2009**. Geneva: World Health Organization,; 2009.
2. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, Dye C: **The growing burden of tuberculosis: global trends and interactions with the HIV epidemic**. *Arch Intern Med* 2003, **163**(9):1009-1021.
3. Organization. WH: **Global tuberculosis report**. In.; 2012.
4. Stein CM, Sausville L, Wejse C, Sobota RS, Zetola NM, Hill PC, Boom WH, Scott WK, Sirugo G, Williams SM: **Genomics of human pulmonary tuberculosis: from genes to pathways**. *Curr Genet Med Rep* 2017.
5. Sobota RS, Stein CM, Kodaman N, Scheinfeldt LB, Maro I, Wieland-Alter W, Igo RP, Jr., Magohe A, Malone LL, Chervenak K *et al*: **A Locus at 5q33.3 Confers Resistance to Tuberculosis in Highly Susceptible Individuals**. *Am J Hum Genet* 2016, **98**(3):514-524.
6. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH: **Missing heritability and strategies for finding the underlying causes of complex disease**. *Nat Rev Genet* 2010, **11**(6):446-450.
7. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A *et al*: **Finding the missing heritability of complex diseases**. *Nature* 2009, **461**(7265):747-753.
8. Cavalli G, Heard E: **Advances in epigenetics link genetics to the environment and disease**. *Nature* 2019, **571**(7766):489-499.
9. Weichenhan D, Plass C: **The evolving epigenome**. *Hum Mol Genet* 2013, **22**(R1):R1-R6.
10. Liu Y, Li H, Xiao T, Lu Q: **Epigenetics in immune-mediated pulmonary diseases**. *Clin Rev Allergy Immunol* 2013, **45**(3):314-330.
11. Richardson TG, Zheng J, Davey Smith G, Timpson NJ, Gaunt TR, Relton CL, Hemani G: **Mendelian Randomization Analysis Identifies CpG Sites as Putative Mediators for Genetic Influences on Cardiovascular Disease Risk**. *Am J Hum Genet* 2017, **101**(4):590-602.
12. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M: **Linking disease associations with regulatory information in the human genome**. *Genome Res* 2012, **22**(9):1748-1759.
13. Bernhart SH, Kretzmer H, Holdt LM, Juhling F, Ammerpohl O, Bergmann AK, Northoff BH, Doose G, Siebert R, Stadler PF *et al*: **Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer**. *Sci Rep* 2016, **6**:37393.

14. 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**(1):88-102.
15. Chen YC, Hsiao CC, Chen TW, Wu CC, Chao TY, Leung SY, Eng HL, Lee CP, Wang TY, Lin MC: **Whole Genome DNA Methylation Analysis of Active Pulmonary Tuberculosis Disease Identifies Novel Epigenotypes: PARP9/miR-505/RASGRP4/GNG12 Gene Methylation and Clinical Phenotypes.** *International journal of molecular sciences* 2020, **21**(9).
16. Esterhuysen MM, Weiner J, 3rd, Caron E, Loxton AG, Iannaccone M, Wagman C, Saikali P, Stanley K, Wolski WE, Mollenkopf HJ *et al*: **Epigenetics and Proteomics Join Transcriptomics in the Quest for Tuberculosis Biomarkers.** *mBio* 2015, **6**(5):e01187-01115.
17. DiNardo AR, Rajapakshe K, Nishiguchi T, Grimm SL, Mtetwa G, Dlamini Q, Kahari J, Mahapatra S, Kay A, Maphalala G *et al*: **DNA hypermethylation during tuberculosis dampens host immune responsiveness.** *J Clin Invest* 2020, **130**(6):3113-3123.
18. Sharma P, Garg G, Kumar A, Mohammad F, Kumar SR, Tanwar VS, Sati S, Sharma A, Karthikeyan G, Brahmachari V *et al*: **Genome wide DNA methylation profiling for epigenetic alteration in coronary artery disease patients.** *Gene* 2014, **541**(1):31-40.
19. Asaad M, Abo-Kadoum MA, Nzungize L, Uae M, Nzaou SAE, Xie J: **Methylation in Mycobacterium-host interaction and implications for novel control measures.** *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2020, **83**:104350.
20. Abdillahi SM, Tati R, Nordin SL, Baumgarten M, Hallgren O, Bjermer L, Erjefalt J, Westergren-Thorsson G, Singh B, Riesbeck K *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.** *Frontiers in immunology* 2018, **9**:1988.
21. Schnoor M, Cullen P, Lorkowski J, Stolle K, Robenek H, Troyer D, Rauterberg J, Lorkowski S: **Production of type VI collagen by human macrophages: a new dimension in macrophage functional heterogeneity.** *J Immunol* 2008, **180**(8):5707-5719.
22. Foley AR, Quijano-Roy S, Collins J, Straub V, McCallum M, Deconinck N, Mercuri E, Pane M, D'Amico A, Bertini E *et al*: **Natural history of pulmonary function in collagen VI-related myopathies.** *Brain* 2013, **136**(Pt 12):3625-3633.
23. Duan Y, Liu G, Sun Y, Wu J, Xiong Z, Jin T, Chen M: **COL6A3 polymorphisms were associated with lung cancer risk in a Chinese population.** *Respiratory research* 2019, **20**(1):143.
24. Ma P, Yang X, Kong Q, Li C, Yang S, Li Y, Mao B: **The ubiquitin ligase RNF220 enhances canonical Wnt signaling through USP7-mediated deubiquitination of beta-catenin.** *Molecular and cellular biology* 2014, **34**(23):4355-4366.
25. Villasenor T, Madrid-Paulino E, Maldonado-Bravo R, Urban-Aragon A, Perez-Martinez L, Pedraza-Alva G: **Activation of the Wnt Pathway by Mycobacterium tuberculosis: A Wnt-Wnt Situation.** *Frontiers in immunology* 2017, **8**:50.
26. Forno E, Wang T, Qi C, Yan Q, Xu CJ, Boutaoui N, Han YY, Weeks DE, Jiang Y, Rosser F *et al*: **DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study.** *The Lancet Respiratory medicine* 2019, **7**(4):336-346.
27. Dorfman R, Taylor C, Lin F, Sun L, Sandford A, Pare P, Berthiaume Y, Corey M, Durie P, Zielenski J: **Modulatory effect of the SLC9A3 gene on susceptibility to infections and pulmonary function in children with cystic fibrosis.** *Pediatric pulmonology* 2011, **46**(4):385-392.
28. Corvol H, Blackman SM, Boelle PY, Gallins PJ, Pace RG, Stonebraker JR, Accurso FJ, Clement A, Collaco JM, Dang H *et al*: **Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis.** *Nature communications* 2015, **6**:8382.

29. Sobota RS, Stein CM, Kodaman N, Maro I, Wieland-Alter W, Igo RP, Jr., Magohe A, Malone LL, Chervenak K, Hall NB *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**(6):e1006710.
30. von Hertzen L, Klaukka T, Mattila H, Haahtela T: **Mycobacterium tuberculosis infection and the subsequent development of asthma and allergic conditions.** *The Journal of allergy and clinical immunology* 1999, **104**(6):1211-1214.
31. von Mutius E, Pearce N, Beasley R, Cheng S, von Ehrenstein O, Bjorksten B, Weiland S: **International patterns of tuberculosis and the prevalence of symptoms of asthma, rhinitis, and eczema.** *Thorax* 2000, **55**(6):449-453.
32. Duffy FJ, Weiner J, 3rd, Hansen S, Tabb DL, Suliman S, Thompson E, Maertzdorf J, Shankar S, Tromp G, Parida S *et al*: **Immunometabolic Signatures Predict Risk of Progression to Active Tuberculosis and Disease Outcome.** *Frontiers in immunology* 2019, **10**:527.
33. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, Batini C, Fawcett KA, Song K, Sakornsakolpat P *et al*: **New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries.** *Nat Genet* 2019, **51**(3):481-493.
34. Tishkoff SA, Williams SM: **Genetic analysis of African populations: human evolution and complex disease.** *Nat Rev Genet* 2002, **3**(8):611-621.
35. Tarashi S, Badi SA, Moshiri A, Ebrahimzadeh N, Fateh A, Vaziri F, Aazami H, Siadat SD, Fuso A: **The inter-talk between Mycobacterium tuberculosis and the epigenetic mechanisms.** *Epigenomics* 2020, **12**(5):455-469.
36. Sharma G, Sowpati DT, Singh P, Khan MZ, Ganji R, Upadhyay S, Banerjee S, Nandicoori VK, Khosla S: **Genome-wide non-CpG methylation of the host genome during M. tuberculosis infection.** *Sci Rep* 2016, **6**:25006.
37. Ay E, Banati F, Mezei M, Bakos A, Niller HH, Buzas K, Minarovits J: **Epigenetics of HIV infection: promising research areas and implications for therapy.** *AIDS Rev* 2013, **15**(3):181-188.
38. Groschel B, Hover 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**(4-7):487-492.
39. 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):1-9.
40. Thuong NT, Dunstan SJ, Chau TT, Thorsson V, Simmons CP, Quyen NT, Thwaites GE, Thi Ngoc Lan N, Hibberd M, Teo YY *et al*: **Identification of tuberculosis susceptibility genes with human macrophage gene expression profiles.** *PLoS Pathog* 2008, **4**(12):e1000229.
41. Stein CM, Zalwango S, Malone LL, Thiel B, Mupere E, Nsereko M, Okware B, Kisingo H, Lancioni C, Whalen CC *et al*: **Resistance and susceptibility to Mycobacterium tuberculosis infection and disease in tuberculosis households in Kampala, Uganda.** *Am J Epidemiol* 2018, **187**:1477-1489.
42. von Reyn CF, Mtei L, Arbeit RD, Waddell R, Cole B, Mackenzie T, Matee M, Bakari M, Tvaroha S, Adams LV *et al*: **Prevention of tuberculosis in Bacille Calmette-Guerin-primed, HIV-infected adults boosted with an inactivated whole-cell mycobacterial vaccine.** *Aids* 2010, **24**(5):675-685.
43. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA: **Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.** *Bioinformatics* 2014, **30**(10):1363-1369.
44. Fortin JP, Triche TJ, Jr., Hansen KD: **Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi.** *Bioinformatics* 2017, **33**(4):558-560.

45. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S: **A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.** *Bioinformatics* 2013, **29**(2):189-196.
46. Johnson WE, Li C, Rabinovic A: **Adjusting batch effects in microarray expression data using empirical Bayes methods.** *Biostatistics* 2007, **8**(1):118-127.
47. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, Greenwood CM, Hansen KD: **Functional normalization of 450k methylation array data improves replication in large cancer studies.** *Genome Biol* 2014, **15**(12):503.
48. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015, **43**(7):e47.
49. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ *et al*: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**(3):559-575.
50. Martin JS, Xu Z, Reiner AP, Mohlke KL, Sullivan P, Ren B, Hu M, Li Y: **HUGIn: Hi-C Unifying Genomic Interrogator.** *Bioinformatics* 2017, **33**(23):3793-3795.
51. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S *et al*: **Annotation of functional variation in personal genomes using RegulomeDB.** *Genome Research* 2012, **22**(9):1790-1797.

**Table 1. Study population characteristics**

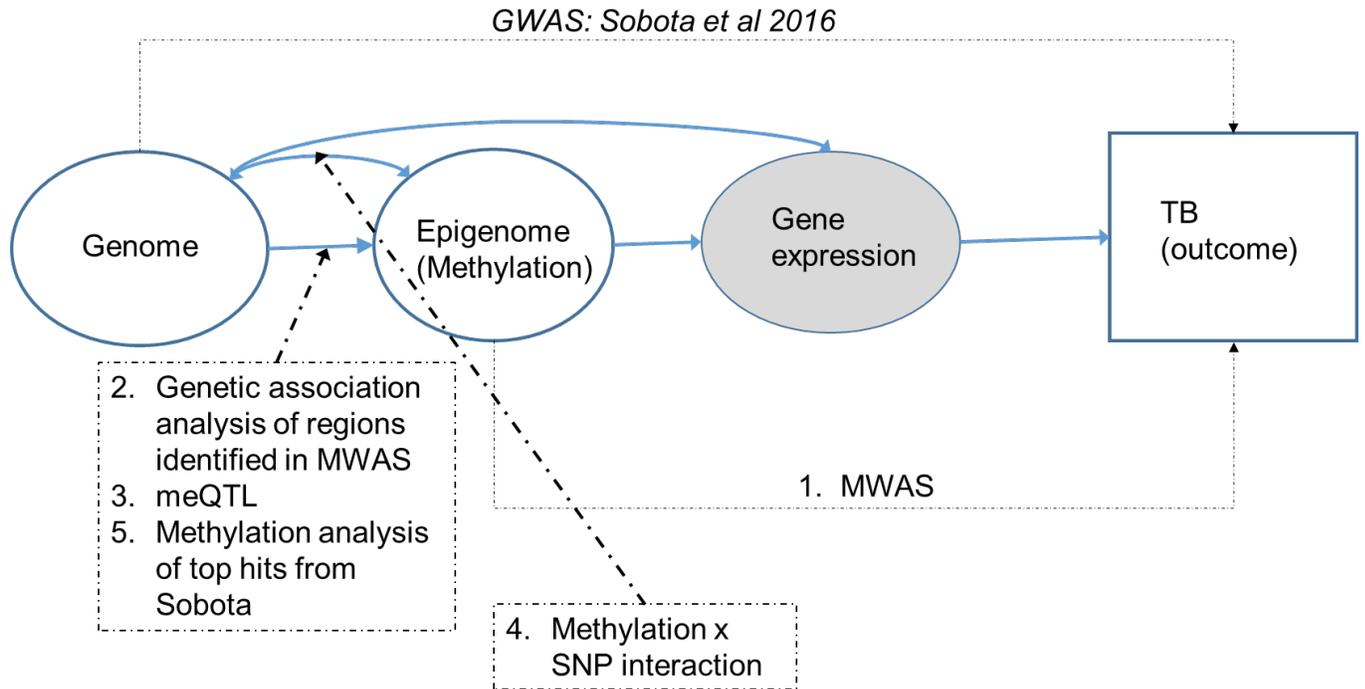
Cohort				
<b>Uganda</b>				
	TB (n = 76)	No TB (n = 67)	Overall (n = 143)	P-value across TB groups (TB vs No TB)
Proportion Male Sex (n)	0.50 (38)	0.36 (24)	0.43 (62)	0.12 <sup>1</sup>
Mean Age (SD)	31.4 (9.3)	26.1 (12.5)	28.9 (11.2)	0.02 <sup>2</sup>
<b>Tanzania</b>				
	TB (n = 32)	No TB (n = 46)	Overall (n = 78)	<i>p</i>
Proportion Male Sex (n)	0.19 (6)	0.26 (12)	0.23 (18)	0.63 <sup>1</sup>
Mean Age (SD)	35.0 (7.7)	36.0 (7.6)	35.6 (7.6)	0.30 <sup>2</sup>

<sup>1</sup> Pearson's chi-squared test of independence used. <sup>2</sup> Wilcoxon-Mann-Whitney test used.

**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
<i>Methylation markers that are replicated in both cohorts</i>							
cg19382731	chr1	44883990	Island	RNF220	4.62 E-05	0.07	4.63E-05
cg16974832	chr2	238188707	OpenSea	Flanking genes (COPS8 and COL6A3)	4.47 E-05	0.04	2.69E-05
<i>Different methylation markers in same region by population attaining significance</i>							
cg18730862	chr5	611926	N_Shore	CEP72	2.08E-05	9.21E-01	2.28E-04
cg03602880	chr5	612961	S_Shore	CEP72	0.0398	2.09E-05	1.25E-05

**Figure 1. Analytical framework**



**Figure 2 legend.** CoMET plots of our 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 viewing window. A) RNF220 gene CpG site cg1938271 shown +/- 25kb. B) COL6A3 gene CpG site cg16974832 shown +/- 50kb. C) CEP72 gene CpG site cg18730862 shown +/- 25kb. The CpG-Tanzania indicated site is for cg03602880 ( $p=2 \times 10^{-5}$  in the Tanzania cohort,  $p=0.04$  in the Uganda cohort); Spearman's correlation between cg03602880 and cg18730862 was 0.11 overall (0.03 in Uganda only and 0.18 in Tanzania only). D) Legend for panels A-C. The 15 chromatin states and corresponding color mappings.

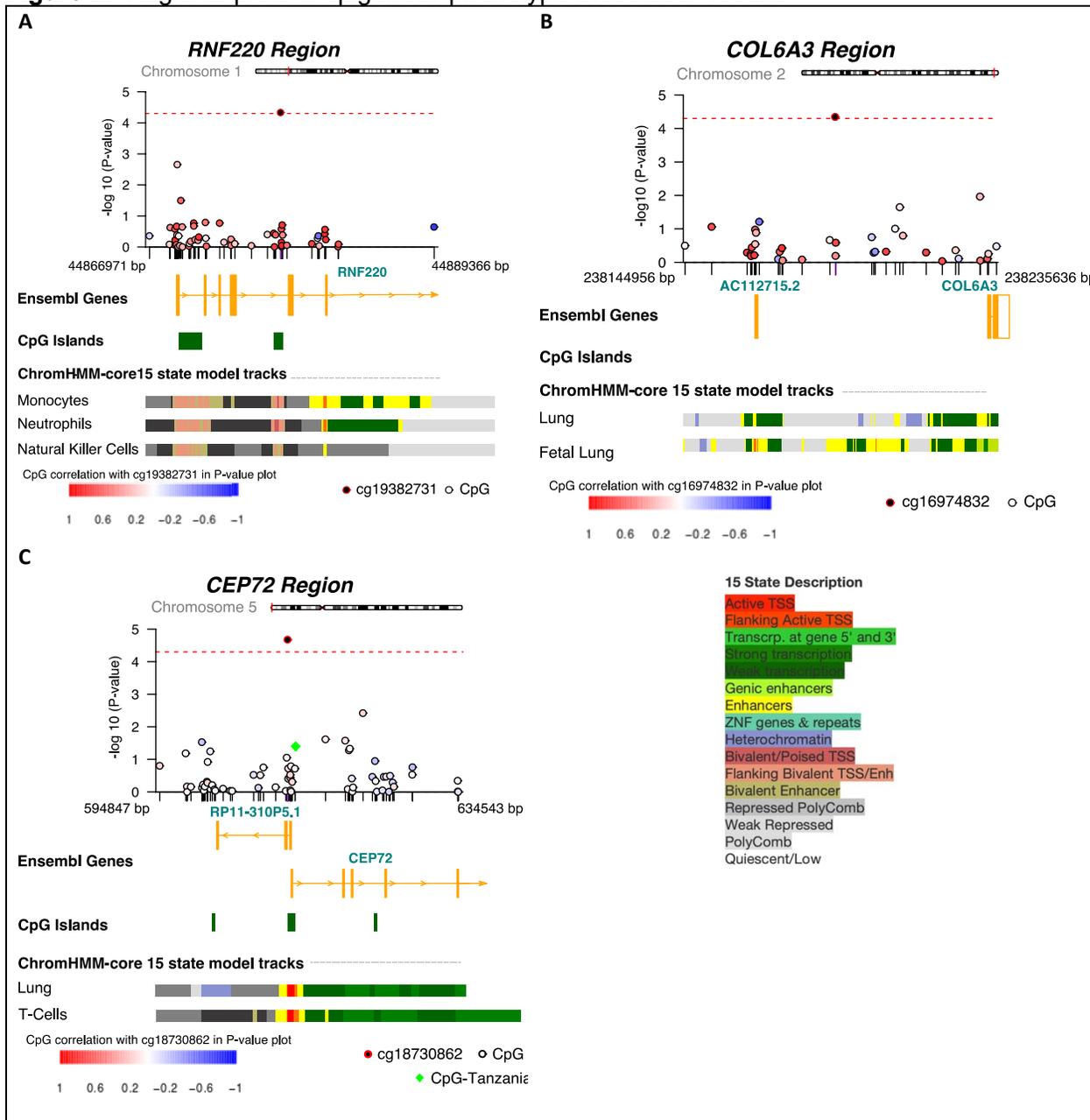
**Figure 3 legend.** Plots of association p-values from MWAS, GWAS, meQTL, and methylation x SNP interaction analyses. MWAS p-values come from Uganda cohort, as it was the discovery cohort, and GWAS, meQTL, and methylation x SNP p-values come from the combined cohort. Results from each type of analysis are represented by different colors and shapes, with  $-\log(p\text{-value})$  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)

### **Supplemental Tables.**

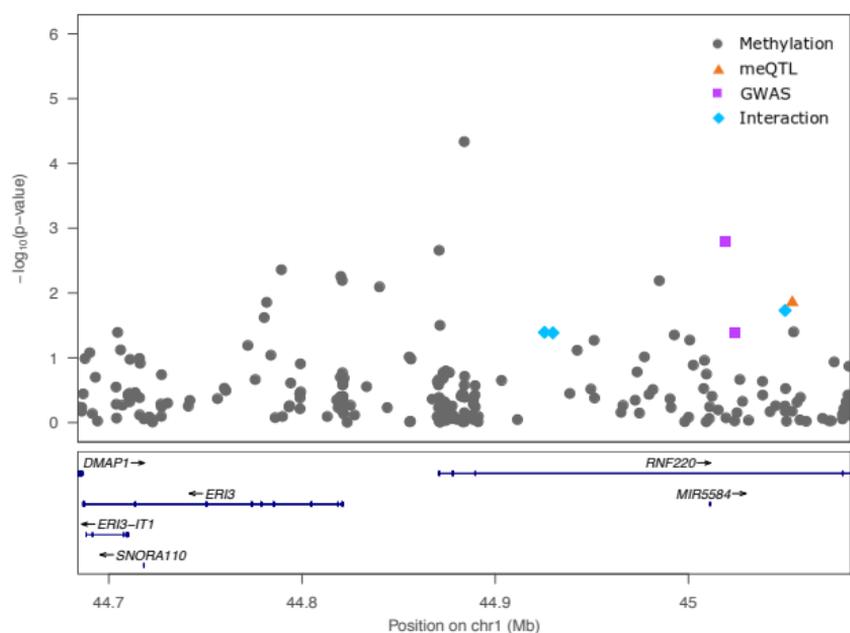
**Supplemental Table 1:** SNP association results within regions identified through MWAS.

**Supplemental Table 2:** Follow-up of regions identified in Sobota et al. [5] within MWAS data. MWAS p-values are provided separately for Ugandan and Tanzanian cohorts, and combined p-values are indicated as (\*p-value) within the cells with both Ugandan and Tanzanian cohort p-values.

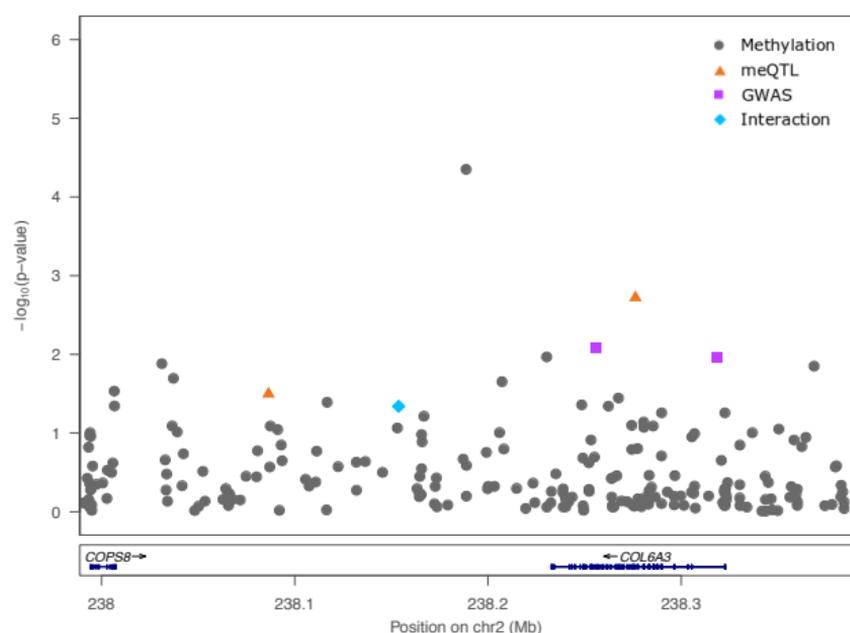
**Figure 2.** Regional plots of epigenetic-phenotype association results



a) Chromosome 1 region including RNF220



b) Chromosome 2 region including COL6A3



c) Chromosome 5 region including CEP72 and SLC9A3

