 doi:10.20944/preprints202110.0045.v1

Interaction Between *M. tuberculosis* Lineage and human genetic variants reveals novel pathway associations with severity of TB

Michael L McHenry[1], Eddie M Wampande[3], Moses L Joloba[3], LaShaunda L Malone[2], Harriet Mayanja-Kizza[4], William S. Bush[1], W. Henry Boom[2], Scott M. Williams[1]*, Catherine M. Stein[1,2]*

[1] Department of Population and Quantitative Health Sciences, and [2] Division of Infectious Diseases and HIV Medicine, Department of Medicine, Case Western Reserve University, Cleveland, OH; [3] Department of Medical Microbiology, College of Health Sciences, Makerere University, Kampala, Uganda and [4] Department of Medicine and Mulago Hospital, School of Medicine, Makerere University, Kampala, Uganda

*These two authors jointly contributed as senior authors of this work.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

Abstract

Tuberculosis (TB) remains a major public health threat globally, especially in sub-Saharan Africa. Both human and *Mycobacterium tuberculosis* (MTBC) genetic variation affect TB outcomes, but few studies have examined if and how the two genomes interact to affect disease. We hypothesize that long-term coexistence between human genomes and MTBC lineages modulate disease to affect its severity. We examined this hypothesis in our TB household contact study in Kampala, Uganda, in which we identified 3 MTBC lineages of which one, L4.6-Uganda, is clearly derived and hence recent. We quantified TB severity using the Bandim TBscore and examined the interaction between MTBC lineage and human single nucleotide polymorphisms (SNPs) genome-wide, in two independent cohorts of TB cases (N=149 and N=127). We found a significant interaction between a SNP in *PPIAP2* and the Uganda lineage (combined $p=4x10^{-8}$). *PPIAP2* is a pseudogene that is highly expressed in immune cells. Pathway and eQTL analyses indicated potential roles between coevolving SNPs and cellular replication and metabolism as well as platelet aggregation and coagulation. This finding provides further evidence that host-pathogen interactions affect clinical presentation differently than host and pathogen genetic variation independently, and that human-MTBC coevolution is likely to explain patterns of disease severity.

Keywords: tuberculosis severity, M. tuberculosis, population genetics, lineage-based GWAS, M. tuberculosis-human coevolution

 doi:10.20944/preprints202110.0045.v1

Introduction

Pulmonary tuberculosis (TB) creates a large global public health burden with 10 million active TB cases and 1.6 million deaths and in 2017 [1]. Prior to the COVID-19 pandemic, *Mycobacterium tuberculosis* (MTB) was the mostly deadly pathogen on earth and has been for centuries [2]. While the global incidence of TB has generally trended downwards, TB continues to be a major driver of infectious disease mortality, and is a re-emerging infectious disease in Southeast Asia and Sub-Saharan Africa [2]. Almost half (44%) of TB cases worldwide occur in Sub-Saharan Africa. In Uganda alone, the location of our study, the incidence has been increasing since 2015, where there are almost 25,000 deaths per year due to TB[2]. TB is also the leading cause of death among people infected with human immunodeficiency virus (HIV)[3]. The bacterium, *Mycobacterium tuberculosis*, causes most TB and is transmitted via airborne droplets from coughing and sneezing by people with active disease. Despite these troubling numbers, TB morbidity and mortality is not as injurious as it could be, as only one-fourth to one-third of the entire globe is thought to be latently (asymptomatically) infected but only a small fraction of this number presents with disease. Identifying what protects the vast majority of infected people is critical to our understanding of disease pathogenesis, treatment, and prevention.

There are many species within the *Mycobacterium* genus. The *Mycobacterium tuberculosis* complex (MTBC), which causes most human disease, is classified into 8 major lineages with different geographical boundaries and timelines of human exposure [4, 5]. Some of these are ancient MTBC (L1, L5, L6, L7, L8) while others are modern MTBC (L2, L3, and L4). Generally, ancient lineages of MTBC are less virulent than the modern ones, providing some support to the hypothesis that the emergence of newly evolved lineages leads to more virulent

 doi:10.20944/preprints202110.0045.v1

disease, a phenomenon referred to as disrupted coevolution[6-11]. MTBC Lineage 4 (L4), MTBC Lineage 2 (L2), and MTBC Lineage 3 (L3) are the most common, and MTBC L4 is the most widespread geographically. MTBC L4 is thought to have originated in Europe prior to its global spread. MTBC L3 is mostly found in the Middle East, India, and East Africa, while MTBC L2 is found predominantly in East Asia. In addition, there are numerous sub-lineages derived from these modern lineages that can be extremely limited in their geographical distribution, some to a single country or geographic region. Many of these sub-lineages are thought to be recently diverged, and many of them are thought to have come from the widespread MTBC L4 lineage [4].

While recent work has shown that several sub-lineages of MTBC L4 are more recently evolved than the major MTBC L4 lineage, the dates for these events have been difficult to determine. Over time, at least 10 genetically distinct sub-lineages of MTBC L4 arose in specific and limited geographic areas. Important to this study, there is a sub-lineage found solely in Uganda and neighboring countries known as the MTBC L4.6/Uganda . Two studies in Uganda have shown no association between this MTB sub-lineage and severity of disease, as measured by presence of cavitary disease and extent of lung involvement[12, 13]. These results present a major issue in assessing TB risk; namely, whether pathogenicity or virulence is a function of the host, MTB, or both. This is especially pertinent to TB, as most infections do not cause disease.

While both human and MTB genetic variants can affect TB risk, it is also possible that interactions between the two can impact severity. Such a model posits that human-MTB coevolution exists and that risk and/or severity is associated with specific combinations of human genes and MTB genes[11]. Studies examining possible interactions between host genotype and MTB lineage are sparse and studies of host-MTB genome interaction have been

 doi:10.20944/preprints202110.0045.v1

case-only studies that at best examine association between lineage and host genotypes as independent factors, but usually without explicit interactions [14-17]. Studying interactions is especially difficult for TB susceptibility as it is virtually impossible to know what MTB lineages an individual was exposed to prior to enrollment in a study; it is only possible to know what lineage is present in someone with active disease. Therefore, severity allows us to examine interactions between MTB lineage and human genetic variants. Due to the constraints of looking for interaction in the context of susceptibility, examining severity may currently be the only way to examine MTB-human interactions at the population level.

The theories of prudent exploitation and disrupted coevolution hypothesize that long-term coexistence between the human genome and an *M. tuberculosis* complex (MTBC) lineage may decrease the severity of disease, and the presence of newly evolved or introduced strains may cause more severe disease [11, 18-20]. While this has rarely been shown on a population level, the possibility of coevolution between humans and MTBC has previously been suggested as an important area of research[6, 7, 10, 11, 21-26]. Suspicion of coevolution is based on an extensive history of coexistence with humans, the observation that modern lineages of MTB are more virulent than ancient ones, and that certain lineages and sub-lineages of MTBC appear to be adapted to specific human populations[10]. Consistent with the coevolution hypothesis, a study of TB transmission in San Francisco showed that TB transmission was most likely to occur among its sympatric host population despite mixed exposure[24].

TB severity in the context of coevolution is especially important as evolutionary theories hypothesize that long-term coexistence between the human genome and MTBC lineage may decrease risk of developing active TB or minimize the severity of disease, if disease is present[11]. Coevolution implies concordant genetic variation of both MTB lineages and human

 doi:10.20944/preprints202110.0045.v1

variation as a product of long-term coexistence that promotes mutual adaptation and thereby modulates the effects of infection. In some cases, coevolution has been hypothesized to manifest as prudent exploitation or covert infection, such that exposure and latent infection does not necessarily lead to active disease and will cause less virulent disease, when present[19, 27]. There are other models of coevolution, but we think that prudent exploitation is the most likely scenario for the case of coevolution between humans and MTB because: 1.) most hosts who are infected do not develop active disease; 2.) even fewer die from TB; and 3.) the vast majority of TB infections worldwide are latent, where infection persists without symptoms or mortality for the host. A pathogen that is prudently exploiting the human host would be evolutionarily incentivized to persist and be transmitted, but not cause disease severe enough to cause rapid death that could lead to the extinction of the host population and ultimately the MTB population. [25]. Under the coevolutionary model of prudent exploitation, a newly divergent MTB lineage that did not historically coexist with the human population in question is more likely be associated with disease and also more likely to cause severe disease [20]. In practice, only the disease state can be explicitly studied as exposure histories are impossible to know in the absence of disease. Consistent with this possibility and the conditions that could lead to coevolution, humans and MTB have a very long history (between 6 and 80 thousand years) and most people exposed to MTB do not progress to active disease (90-95%), making MTB a likely prudent exploiter [28]. The potential for human-MTB coevolution has been explored in human and model systems, but few studies have yet to identify a clear effect at the population level[8, 14, 15, 29-32].

A recent study of ours has shown that interaction between host variants in *SLC11A1* and MTB lineage can affect severity of TB, measured by the TBscore[23]. Specifically, we found

 doi:10.20944/preprints202110.0045.v1

that the recently diverged Ugandan L4 sub-lineage (L4.6) caused more severe disease, but only in individuals with an ancestral *SLC11A1* genotype; those with the ancestral genotype and the older MTB lineages had less severe disease. In general, the combinations of host genotype and MTB lineage that had coexisted longer had less severe disease, lending support to the model of coevolution and prudent exploitation of humans by MTB. Relevant to *SLC11A1*, this gene was a well-studied candidate gene for TB susceptibility, but not all studies found an association. Our results indicated that failure to replicate was a function of MTB lineage among those with disease. Given how lineage can affect the relationship between host genotype and TB, and that interactions between the two genomes may drive disease processes, it will be important to refine the understanding of the role that MTB lineage - human genome interactions may play in TB. While our earlier study was focused on host candidate genes, this hypothesis has not been examined genome wide.

Coevolution will drive reciprocal changes in the genomes of the two species such that fitness changes occur simultaneously in the pathogen and host. As we have previously argued, coevolution can be shown via population genetic analyses when the outcome (fitness) depends on the interaction of two genetic variants, one in each species[33]. In this study, this outcome correlates with survival and therefore, reproductive fitness. If local adaptation exists, it should be possible to measure disease risk or severity in terms of human–MTB coevolution that has resulted from historical coexistence and demonstrated interaction of the two genomes [25, 33]. This can be assessed in a regression model by testing for multiplicative interaction (i.e. effect modification) between the genetic variation of the host and MTB lineage[11, 22]. Statistical interaction between the mycobacterial lineage and human genotype may provide evidence to support the theory of prudent exploitation/coevolution in which infection does not necessarily

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

lead to active disease and may cause less virulent disease, when present[19, 27]. Host-pathogen coevolution has been demonstrated in other organisms, such as *Helicobacter pylori*, using this approach [11, 22]. To study TB severity, we have employed the TBscore that examines clinical severity using symptoms and clinical examination[34].

We hypothesized that there is coevolution between human genotypes and pathogen lineage as shown through a significant statistical interaction between the host variants and MTB lineage. Based on the theoretical argument for coevolution, we hypothesized that the more ancient lineages of MTB should have reduced severity in the presence of ancient host alleles, and that the derived lineages should result in more severe disease, especially in the presence of the ancient host alleles. We tested this model explicitly in this study. This approach helped elucidate the degree to which the effects of lineage on severity are modified by and thus dependent on the genotype of the human host.

Results

The final study population included 276 subjects with data for human genotype, MTBC lineage, and the covariates of interest. We identified statistically significant differences between the two cohorts in the percentage of people who were HIV+ and the mean TBscore (Table 1). Males made up a larger proportion of subjects than females in the sample (which is consistent with prior studies showing that active TB is more prevalent among males), and most subjects were HIV-negative. All the models were adjusted for HIV status and sex. The distribution of MTB lineages showed no differences between cohorts, based on an ANOVA test ($p=0.66$). The L4.6/Uganda lineage was the most prevalent in both cohorts (Table 1). The analysis of how

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

lineage affects severity when considered singly showed that there was no association between lineage and severity (p=0.65) and the boxplot confirms this graphically (Figure 1).

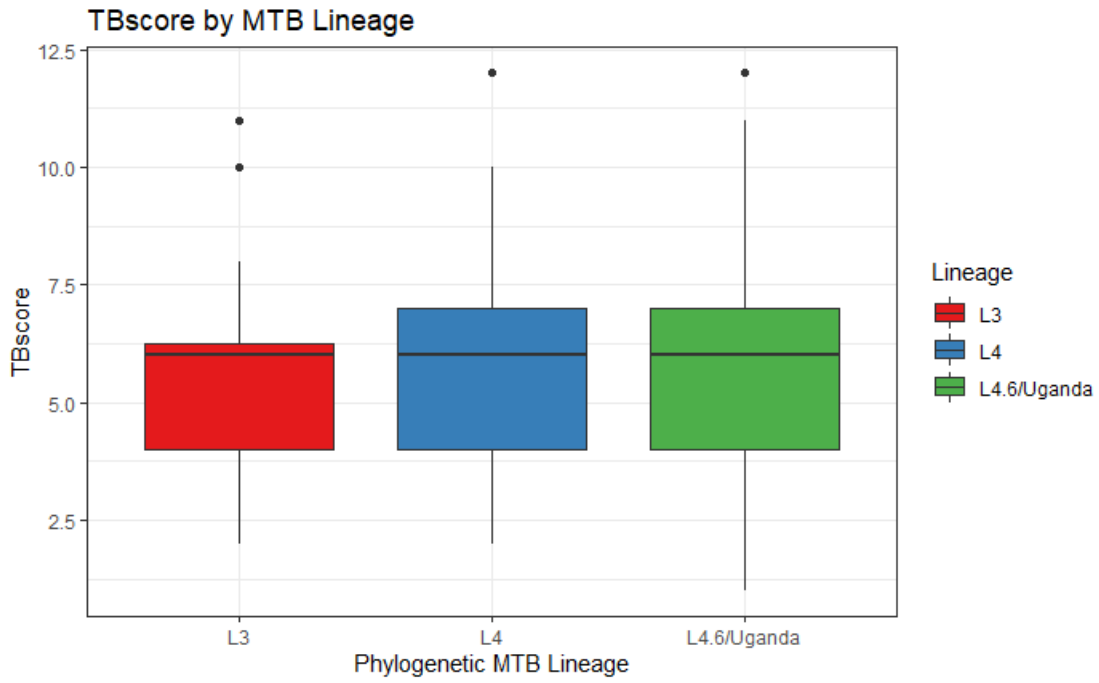
Table 1. Study Population

	Cohort 1 N=149	Cohort 2 N=127	Total	P
Mean Age (SD)	28.7 (8.1)	28.7(9.8)	28.7 (9.0)	0.99
# Male (%)	81 (54.4)	73 (57.5)	154 (55.8)	0.69
# HIV+ (%)	31 (24.4)	15 (10.1)	46 (16.7)	0.0025
Mean TBscore (SD)	6.2 (2.1)	5.4 (2.2)	5.8 (2.2)	0.0032
# L4.6/Ugandan (%)	93 (62.4)	75 (59.1)	168 (60.9)	0.66
# L4 (%)	15 (10.1)	17 (13.4)	32 (11.6)	
# L3 (%)	39 (26.2)	31 (24.4)	70 (25.3)	

Differences in age and TBscore were analyzed using a Student's t-test and differences in the percentage of males and HIV+ subjects were analyzed using Z-statistics. Differences in the distribution of lineages were analyzed using a Chi-squared test. For all tests, $P<0.05$ was considered a significant difference

Figure 1. Boxplot of Severity by Lineage

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

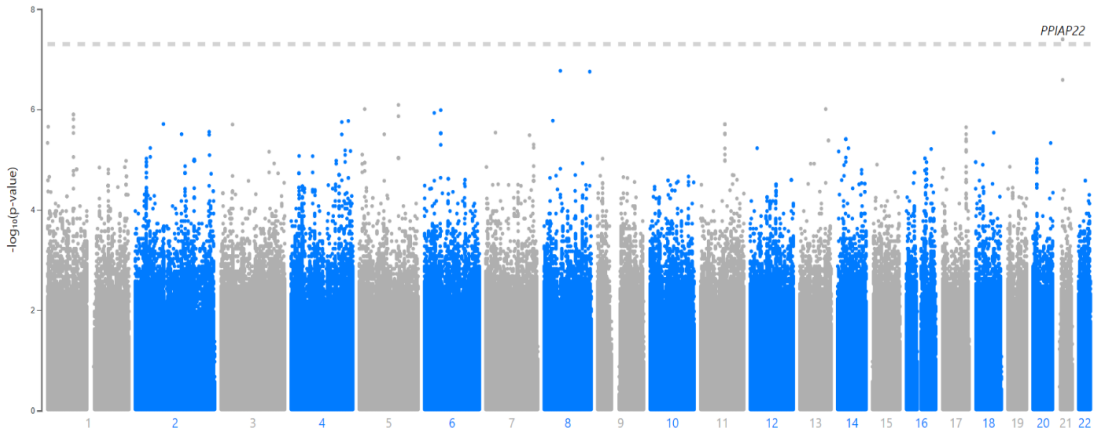


Box plot shows the TBscore values for subjects who are infected with each of the 3 lineages present in our combined cohort. The midline of the box is the median. The box represents the inter-quartile range (25^{th} to 75^{th} percentiles) of the TBscore and the dots represent outliers.

The analyses performed separately in each cohort did not show any significant interactions based on the GWAS threshold (Supplemental Figures 1 and 2). Neither analysis showed any evidence of genome-wide inflation based on their Q-Q plots or genomic inflation parameters ($\lambda<1.0$ for both) (Supplemental Figures 3 and 4). The meta-analytic summary statistics also showed no signs of inflation in either the Q-Q plot nor the genomic control parameter, $\lambda=0.989$, indicating little genome-wide inflation of the test statistics (Figures 2 and 3).

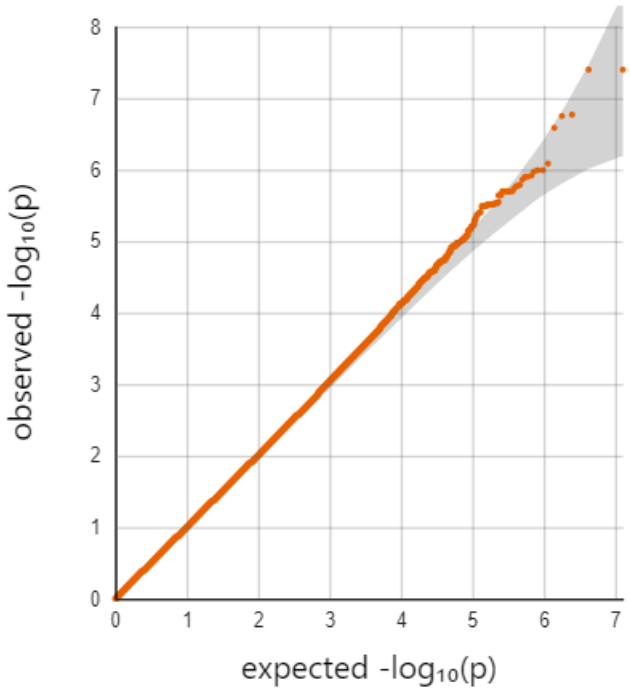
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

Figure 2. Manhattan Plot for Meta-Analytic P-values of Interaction Between SNP and L4.6/Ugandan Lineage



The Manhattan plot shows the inverse log(10) of the p-values for the association between interaction of each SNP and the L4.6/Ugandan lineage and TBscore on the y-axis and the x-axis represent the physical location of each SNP on the chromosomes, which are in order from 1-22.

Figure 3. Q-Q Plot for Meta-Analytic P-Values for Interaction Between SNP and L4.6 Sub-Lineage



The quantile-quantile (Q-Q) plot shows the inverse log(10) of the observed p-values on the Y-axis relative to what is expected if there was no association on the x-axis. Deviations above the line indicate an association with the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

outcome. If the line deviates at the low quantiles, then this is considered evidence to suggest genome-wide inflation of the test statistics, which typically indicates unmeasured confounding.

In the meta-analysis, one SNP by lineage interaction met statistical significance at the GWAS threshold, rs114945555, a SNP on chromosome 21 that maps to *PPIAP22* (Beta= -4.13; p=4.01e-08)(Table 2, Figure 2).As this SNP did not show association with TBscore in the absence of an interaction term and the lineages did not show any association with TBscore, this result provided evidence that the interaction is driving the effect we see rather than either the first order SNP or lineage effects. This is also evident in Table 3, which shows that the strongest and most significant effects in the regression equations (which also includes the first order effects of SNP and lineage) are the interaction terms. This most significant SNP (rs114945555) was not significant in the model without interaction term (p=0.28). The interaction term and the analysis were operationalized such that the L4.6/Ugandan lineage and the derived allele were both coded as 1 while the other lineages and ancestral alleles were coded as 0. Thus, the interaction term indicated that in the simultaneous presence of both the derived allele and the L4.6/Ugandan lineage there is a 4.1-point decrease on the TBscore when the derived allele is present along with the derived MTB lineage relative to the derived allele in the presence of the generalist MTB lineages (Figure 4). This combination of lineage and genotype was observed in 25 of our subjects in this analysis. This is greater than the number of subjects (n=12) who have the same genotype and the generalist lineages, but still represents a relatively small subset of our overall sample, as the derived allele has a relatively low frequency overall. There was an additional SNP on chromosome 21 (rs113863482) in near complete linkage disequilibrium (LD) (with completely identical beta and P-values) with rs114945555 that was also GWAS significant. A third SNP in LD (rs112560854) was just below the GWAS threshold (P= 2.583e-07)(Figure 5).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

Table 2. GWAS Significant Loci for Interaction between SNP and L4.6 Sub-Lineage

SNP	CHR:BP	Ref/Alt	Gene*	Location	MAF (LWK)	P-Value
rs114945555	21:20187488	C/T	*PPIAP22*	Intergenic	7%	4.0e-08

***** rs114945555 mapped to *PPIAP22* in LocusZoom and FUMA GWAS web applications. Minor allele frequencies were ascertained from the 1000G project using Ensembl v104.

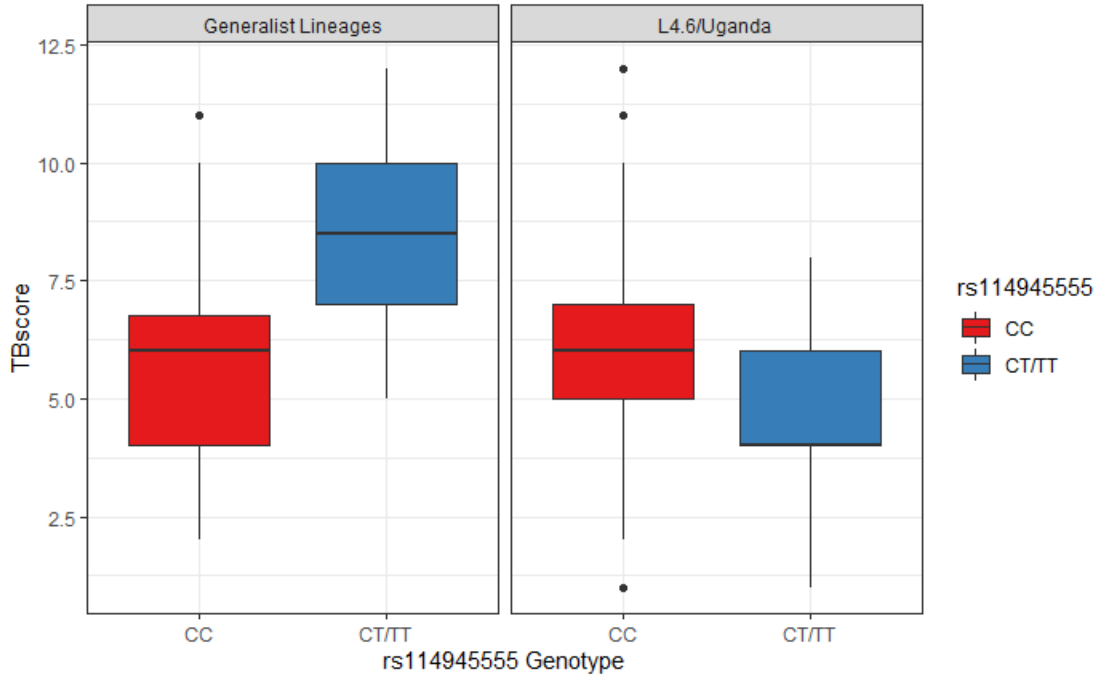
Table 3. rs114945555 x Lineage Interaction across Cohorts 1 and 2

	Cohort 1		Cohort 2		Combined	
	β	p	β	p	β	p
rs114945555 (β_3)†	1.98 (0.28, 3.68)	0.024	2.94 (1.27, 4.61)	7.70E-04	2.47	4.80E-05
L4.6/Ugandan(β_2)	0.59 (-0.13, 1.3)	0.11	0.57 (-0.24, 1.38)	0.17	0.58	0.03
Sex(β_4)	0.49 (-0.17, 1.14)	0.15	1.09 (0.32, 1.85)	0.0061	0.75	0.012
HIV+ Status(β_1)	0.60 (-0.48, 1.69)	0.28	-0.14 (-1.02, 0.74)	0.75	0.16	0.66
rs114945555*L4.6/Ugandan (combination of CT/TT and L4.6/Ugandan) (β_5)	-3.77 (-5.79, -1.75)	3.62E-04	-4.53 (-6.68, -2.38)	6.79E-05	-4.13	4.00E-08

Regression Model:
$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 (X_2 X_3) + \varepsilon$
X_1 = HIV+ Status, X_2=L4.6/Ugandan lineage, X_3 = rs114945555, X_4 = Sex
†Model of inheritance (CT/TT vs CC as referent)

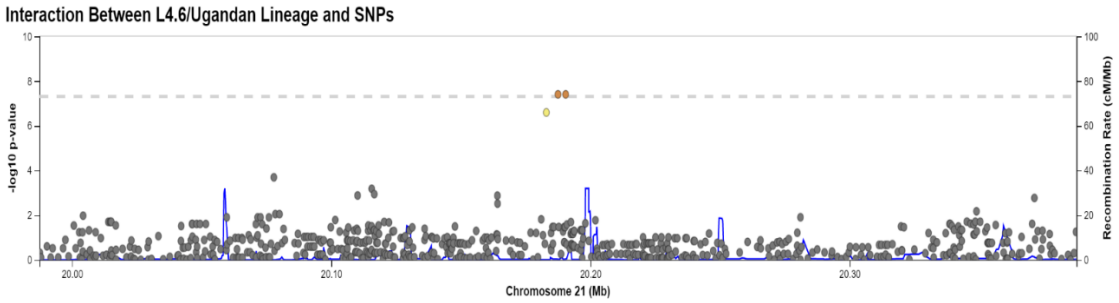
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

Figure 4. Interaction between rs114945555 and L4.6/Ugandan Lineage



There were also 6 SNP by lineage interactions with P<1e-06 (Table 4). Though none of these were below the GWAS threshold for significance, all were nominally significant (P<0.05) and had consistent directions of effect (i.e. the beta value had the same sign) in both cohorts.

Figure 5. LocusZoom Plot for Region surrounding rs114945555



The LocusZoom plot shows the region surrounding rs1848553 on chromosome 5, using an LD panel and reference genome from the AFR super-population in the 1000G project. Yellow and orange indicate higher levels of LD.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

Table 4. SNPs Above GWAS Threshold with P<1e-06 (Meta-analytic P-values and Betas)

CHR	BP	SNP	P	BETA	Gene	Location
5	17775271	rs369093426	9.82E-07	3.2095	*LOC105374666*	Intron
5	121258204	rs76190408	8.14E-07	3.0447	None	Intergenic
8	50958714	rs203964	1.70E-07	2.7085	*SNTG1*	Intron
8	141085471	rs56990580	1.77E-07	-3.3244	*TRAPPC9*	Intron
13	98589842	rs8000063	9.82E-07	2.6203	None	Intergenic
21	20182990	rs112560854	2.58E-07	-3.7805	None	Intergenic

Gene and location was ascertained using Ensembl v.104

PPIAP22 is a pseudogene and thus does not get translated into a functional protein. It is a pseudogene for Cyclophilin A, a protein that is an important mediator of the inflammation response [35]. While *PPIAP22* is not translated into protein, it is transcribed into RNA, and evidence from the DICE database shows that it is expressed in a number of immune cells, with CD4+ T cells showing the highest expression (Figures 6-7)[36].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

Figure 6. *PPIAP22* Expression in Immune Cells from DICE Database of Cell-specific Gene Expression

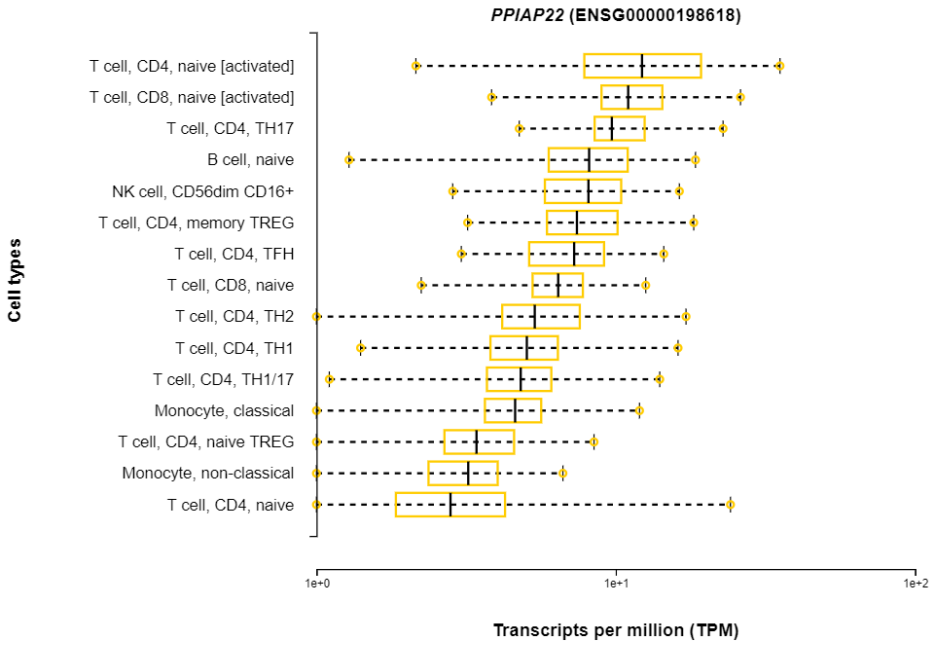
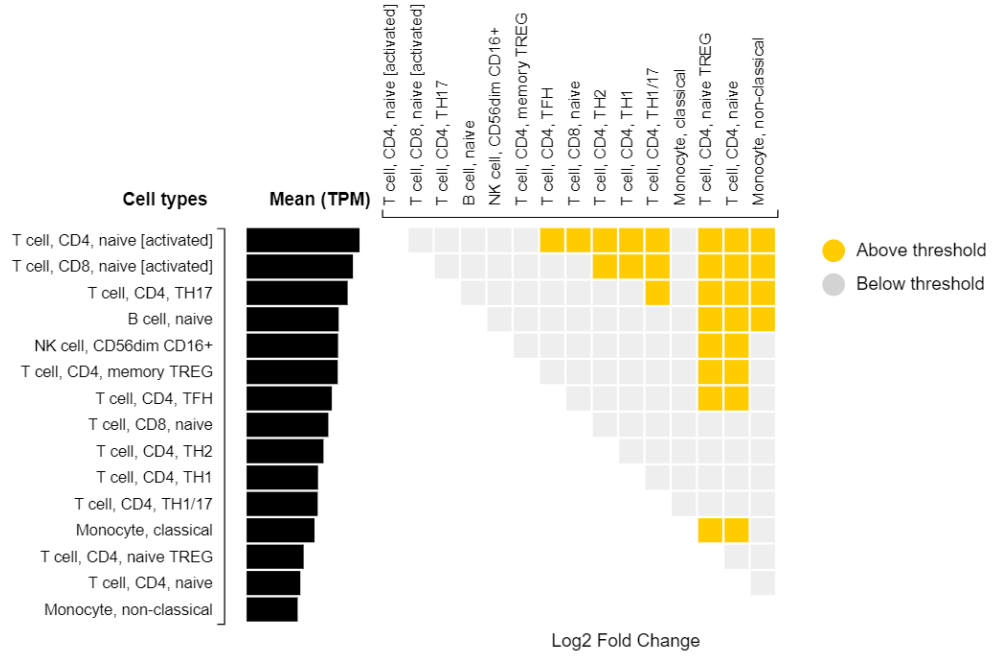


Figure 7. Differential Expression of PPIAP22 in Immune Cells from DICE database of cell-specific gene expression



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

The minor allele for rs114945555 is almost exclusively found in East Africa based on the 1000G project's data, showing a 7% MAF in the LWK population (a population based in Kenya). Outside of Africa, the MAF is 1% or 0% in every population and the LWK have the highest MAF in any African sub-population. In our data, the MAF was 9%, and this did not differ substantially with the East African data from 1000G. The direction of the interaction in the L4.6/Ugandan lineage is consistent with the hypothesis concerning effect modification of the relationship between genotype and severity by MTB lineage. Specifically, in the presence of the ancestral generalist MTB lineage (i.e. L4) with which the population has had a longer exposure history, we hypothesized that the derived allele associates with more severe disease relative to the ancestral allele. In the presence of the more newly emergent specialist MTB lineage (i.e. L4.6/Uganda), we expected that the derived allele would associate with less severe disease relative to the ancestral allele. This was true for the interaction between rs114945555 (the only SNP showing a GWAS significant interaction) and MTB lineage (Figure 4).

The MAGMA gene-based analysis included a total 19,229 protein coding genes represented by the SNPs in our data, and thus the threshold for significance was 2.6e-06. Our results did not show any GWAS significant effects for gene-level analysis, gene set enrichment, or tissue specificity, but the gene with the strongest interactions was *CA12* (p=6.7e-03), a carbonic anhydrase gene. The GSEA from the GENE2FUNC analyses showed significant enrichment for two of the chemical and genetic perturbation pathway gene sets (MsigDB c2) and two gene sets reported in the GWAS catalog (Table 5). Specifically, a breast cancer gene set containing "genes within amplicon 16p13 identified in a study of 191 breast tumor samples" and a gene set containing "Genes with copy number gains in primary neuroblastoma tumors" were

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

significantly enriched in MsigDB c2. In the GWAS catalog, the nephrolithiasis and urolithiasis gene sets were significantly enriched.

Table 5. Gene Set Enrichment for SNPs by Lineage Interactions

GeneSet	N	n	P-value	adjusted P	Genes	Database
Breast Cancer Amplicon	329	24	3.5E-23	1.15E-19	*GPRC5C, TMEM104, GRIN2C, FADS6, CDR2L, KCTD2, ATP5H, SLC16A5, ARMC7, NT5C, HN1, SUMO2, NUP85, GGA3, MRPS7, MIF4GD, SLC25A19, GRB2, KIAA0195, CASKIN2, TSEN54, LLGL2, RECQL5, EXOC7*	Chemical and Genetic Perturbations (MsigDB c2)
Neuroblastoma Copy Number Up	181	9	8.92E-08	0.000147	*GPRC5C, ATP5H, ARMC7, HN1, SUMO2, NUP85, MRPS7, SLC25A19, TSEN54*	
Nephrolithiasis	8	3	4.33E-06	7.87E-03	*INMT, FAM188B, AQP1, AQP1*	GWAS Catalog
Urolithiasis	17	3	5.12E-05	4.64E-02	*TFAP2B, INMT, FAM188B*	

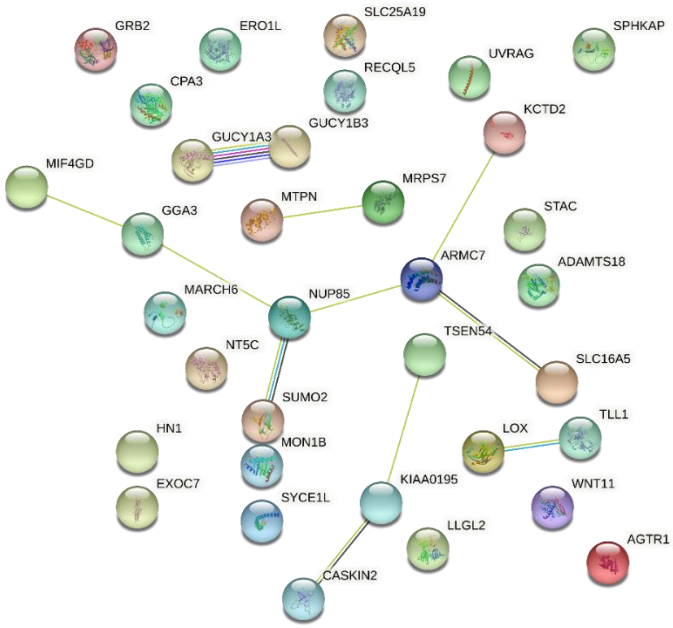
N indicates the total number of genes in the set while n shows the number of genes that SNPs mapped to in our GWAS data. Gene sets were queried using FUMA GWAS. The adjusted P is an FDR-corrected P-value based on the number of gene sets examined.

For the SNPs showing interaction with the L4.6 Sub-lineage (p<1e-05), we found that 21 of the 81 SNPs were eQTL's spanning 36 different genes for a total of 528 eQTL effects across the different tissues and databases (Figure 8, Supplemental Table 1). A STRING DB analysis for PPIs among the genes showed significant enrichment for protein-protein interactions (PPI P-value=0.0089; 11 edges) and was significantly enriched in 6 KEGG pathways (Table 6). From the STRING diagram and the table of KEGG pathways showing significant enrichment, we can see that many of these results are driven by the same proteins downstream of the genes from the eQTL analysis. Specifically, *GUCY1B3* and *GUCY1A3* are in every single pathway that was enriched, and the diagram shows interaction between them. *AGTR1* is also present in 4/6 of these KEGG pathways. Several of the enriched pathways appear to be related to renal function but are

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

also closely tied to regulation of vascular tone, endothelial permeability, and platelet aggregation. It has previously been suggested that these processes play a role in the response to infection and/or in the infection-induced inflammatory response [37-41].

Figure 8. STRING Network for PPI's of eQTL Response Genes



This figure is a STRING diagram showing protein-protein interactions among the genes identified in my analysis. Lines represent an experimentally determined protein-protein interaction and multiple lines between the same two proteins indicates multiple interactions that have been identified but multiple interactions between the same two proteins is still considered to be one edge.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

Table 6. Significantly Enriched KEGG Pathways for eQTL Response Genes

KEGG Pathway	# Genes	# In set	Strength	FDR	Matching Proteins
Renin secretion	3	63	1.44	0.0166	GUCY1B3,GUCY1A3,AGTR1
Gap junction	3	87	1.3	0.0207	GUCY1B3,GUCY1A3,GRB2
Renin-angiotensin system	2	23	1.7	0.0234	CPA3,AGTR1
Vascular smooth muscle contraction	3	119	1.16	0.0251	GUCY1B3,GUCY1A3,AGTR1
cGMP-PKG signaling pathway	3	160	1.03	0.0459	GUCY1B3,GUCY1A3,AGTR1
Purine metabolism	3	173	1	0.0474	NT5C,GUCY1B3,GUCY1A3

Our previously published analysis of how the interaction between the L4.6/Ugandan sub-lineage and host genotype affects severity identified rs17235409, an exonic SNP within *SLC11A1*, a gene that has been well-studied in the context of TB susceptibility. Cohort 2 is comprised of the same subjects utilized in our prior study so we examined the interaction between rs17235409 and the L4.6/Ugandan sub-lineage in Cohort 1 only, as replication in a distinct cohort. We did not have data available for rs17235409, so we examined SNPs within +/- 5kb of *SLC11A1* on chromosome 2. We identified 2 SNPs with a p-value below 0.05: rs13390257 and rs116577076. Both are located within introns of *DIRC3* (which stands for disrupted in renal carcinoma gene 3), a long non-coding RNA gene characterized by its role in renal carcinoma (data not shown).

Discussion

Overall, our results showed that there are interactions between the human genome and MTB phylogenetic lineage that are associated with TBscore (i.e. active TB severity). The strongest interaction showed the same directionality of effect modification that we hypothesized

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

and have previously published[23]. In these data, the difference in TB score between the two MTB lineage categories among those with the derived allele is very large. The beta value for the interaction term is -4.13. A TBscore difference of this magnitude is very clinically meaningful and represents a substantial difference in disease experience, level of disability, and risk of mortality. As the correlation with eventual mortality is a big driver of selection, our finding adds to the plausibility of the argument that coevolution is at play.

The GWAS significant SNP, rs114945555, shows the highest MAF in the 1000G population (LWK) that is closest to Uganda in geographical proximity. The fact that the derived allele, which associates with relatively less severe disease in the presence of the more recently derived MTB lineage, is more prevalent in a TB endemic region where the L4.6/Ugandan lineage is found, and is consistent with positive selection for this variant only where the new lineage exists, supports the coevolutionary hypothesis. If there is an evolutionary advantage to this allele in this specific population, then it is logical that it would increase in frequency. *PPIAPP22* is a pseudogene, and thus it is difficult to discern its functional role in the immune response to active TB. However, the evidence from the DICE database shows that it is differentially expressed in immune cells that are relevant to TB. Further, studies have shown that pseudo-genes may play an important regulatory role in human genetic diseases, with the potential to regulate protein-coding genes[42]. Some pseudogenes show a tissue-specific pattern of activation, and some pseudogene transcripts can be processed into short interfering RNAs (siRNA's) capable of regulating protein-coding genes[42]. In cancer and pharmacology, pseudogenes have shown to have an important regulatory role in pathogenesis and are even considered therapeutic targets[43]. Thus, while *PIAPP22* is unlikely to play a direct role in the immune response, it may have important regulatory functions relevant to its interaction with MTB lineage.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

The gene set analyses were also difficult to interpret, as breast cancer, kidney cancer, and stones found in the kidneys and urinary tract bear little obvious relevance to tuberculosis. However, it is possible that the pleiotropic effects of the gene pathways have not yet been elucidated, and there may yet be functional relevance of these gene sets. Many genes associated with cancer phenotypes may have general functions in cellular replication and metabolism. Thus, while they were shown as being important in cancer, they might have functions that affect a variety of phenotypes. Cancer-related gene sets have previously been shown to be enriched in several non-cancer phenotypes, and in previous studies a lack of apparent connection to TB does not necessarily indicate that there is no biological function[44]. Rather, it will be important to determine to what extent these genes may have functions outside of known associations. For the gene sets related to kidney function and kidney stones, pleiotropic effects may also be possible. Kidney complications in pulmonary TB are not uncommon and pulmonary TB has systemic effects on numerous organs because of generalized inflammation. Further, the kidneys play an important role in regulating homeostasis, particularly with respect to blood pressure and fluid levels. These may act to affect TBscore related phenotypes. As such, it is possible that these results might yield interesting connections to TB severity upon further examination.

A subset of the eQTL SNPs that showed interaction with lineage appear to collectively be involved in the biological processes that regulate coagulation and vascular tone. The enrichment for processes shown in Table 6 appear to be driven primarily by the same 3 genes, *GUCY1B3*, *GUCY1A3*, and *AGTR1*. Five of these processes (renin, the renin-angiotensin system, gap junctions, vascular smooth muscle contraction, and the cGMP pathway) are involved in the regulation of blood pressure and vascular tone. Under normal physiological conditions, these pathways are part of a system that regulates blood pressure in response to factors such as fluid

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

levels, electrolyte concentrations, stress, and cardiovascular output[45]. The vasculature also plays an important role in the response to infection, making changes that allow immune effector cells to move to where they are needed[39, 40]. Under conditions of acute stress, and particularly in the context of infection-induced inflammation, this system can be perturbed and lead to acute hypertension (increase in blood pressure) that can reduce blood flow to vital organs[46]. In some cases, such as a cytokine storm in the context of septic shock, there can also be damage or death caused by hypotension, i.e. a decrease in blood pressure[47]. More specifically, the renin-angiotensin system (RAS) has been implicated as an important driver of the inflammation response in studies of lung damage and pulmonary vascular disease, where the changes in vascular tone of the micro-vasculature within the lungs can dictate the extent of alveolar damage and potential for recovery[48].

The cGMP pathway is also important to platelet homeostasis and the coagulation cascade and has previously been shown to be an important regulator of the cell migration and T cell polarization in the host immune response [49, 50]. Platelets play an important role in the immune and inflammatory response to infection[37, 41]; the dysregulation of the system as a result of infection has been well-documented in other infectious diseases[38], and active TB patients experience a pro-coagulatory state as a result of systemic inflammation[51, 52]. cGMP pathways have also been directly implicated as a driver of inflammation in the context of infectious disease, and specifically in the context of pulmonary infectious diseases. For example, they may represent a therapeutic target that can mitigate the most severe forms of COVID-19[49, 53, 54]. Thus, the enrichment results indicate that the consequences of systemic inflammation may be driving more severe manifestations of TB disease. This is important, as prior studies of TB

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

susceptibility have primarily identified immune response genes (though there is some overlap), indicating that different biological processes may be at play in susceptibility and severity.

This study was not without limitations. This sample was limited to 279 subjects across the two cohorts, which is relatively small compared to many modern GWAS studies, which can include thousands of subjects. While we were able to detect one GWAS significant interaction, there were 6 other SNP by lineage interactions with P<1e-06 and consistency across our two cohorts. Thus, it is possible that with a larger sample size, these interactions might have achieved GWAS significance. Nonetheless, we were able to detect one interaction that may play a role in TB severity, and we generated a list of 91 SNPs with p<1e-05 that showed enrichment for multiple gene sets. Finally, we were able to provide further evidence for the phenomenon of human-MTB co-evolution and demonstrate how this can be done in genetic epidemiology studies.

Materials and Methods

Study participants

The data in this study were drawn from the Kawempe Community Health Study (KC Health Study) in Kampala, Uganda [55]. The KC Health study enrolled 3,818 total participants that included 1045 active TB cases, from which the study population in this paper is derived. Our analyses included only individuals with available data on human genotype, clinical symptoms related to active TB severity, and information on the specific lineage/sub-lineage of MTBC with which the patient is infected. The TBscore was developed for adults and may not appropriate for individuals under 15. Thus, our sample was limited to subjects 15 years old and older. We

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

examined two samples collected separately as part of this study that will be referred to as Cohort 1 and Cohort 2 (N=149 and N=127, respectively). The two cohorts were recruited at different times and genotyping of the humans was performed using different arrays for them, as described below. Ascertainment criteria did not differ.

All TB cases were culture-confirmed positive based on isolation of MTB in sputum or gastric samples and the clinical characteristics were collected as part of the visit during which subjects were diagnosed with active TB. Chest radiographs (X-rays) were performed at the Uganda Cancer Institute on subjects with confirmed active TB. The study protocol was approved by the National HIV/AIDS Research Committee of Makerere University and the institutional review board at University Hospitals Cleveland Medical Center. Final clearance was given by the Uganda National Council for Science and Technology. All participants provided written informed consent. Additional details about the original study protocol are described elsewhere [55]. The two cohorts differed in percentage of HIV positive individuals (Table 1); therefore, HIV status was used as a covariate in all regression models. Previous analyses of microsatellite data from these cohorts indicated no substantial population substructure, as previous principal components (PC) analyses have corroborated [23, 56].

Human Genotyping and QC

Cohort 1 was genotyped on the Illumina Infinium Mega^{EX} chip, comprising 2.1M markers genome wide. For Cohort 2, we used the Illumina HumanOmni5 microarray comprising 4.3M markers genome-wide, offering high genome wide coverage of common genetic variation even within African populations[57]. The difference in choice of genotyping chips was based only on commercial availability at the time of genotyping. Only SNPs that had a call rate greater than 0.98, MAF> 0.05, and did not show deviation from Hardy-Weinberg equilibrium ($p<10^{-6}$)

 doi:10.20944/preprints202110.0045.v1

prior and subsequent to imputation in both samples were used in the analysis. The genotype data for Cohort 1 initially included 2,036,060 SNPs. After QC and prior to imputation there were 717,705 SNPs remaining. After imputation and QC we had 8,146,092 SNPs. The genotype data for Cohort 2 included 2,989,642 SNPs to start. After QC, 1,931,961 SNPs remained prior to imputation. After imputation and another round of QC thresholds, there were 9,626,100 SNPs. These were the final numbers of SNPs in each cohort prior to meta-analysis. The total number of SNPs that overlapped between the two cohorts was 6,421,278. Principal components were computed using Plink v1.9[58].

Determination of MTB Lineage

MTB was isolated from sputum of each of these subjects, and lineages were classified according to lineage-identifying SNPs using real-time PCR and validated with long sequence polymorphism (LSP) PCR and sequencing [12, 59]. Lineage was determined from three SNPs that accurately distinguish the MTBC L4.6 Uganda, MTBC L3, and MTBC L4 lineages, as previously described [4, 6, 7]. The classifications delineated by these SNPs were then compared to previously established LSP based lineages to validate these distinctions. In the context of this study setting, the relevant MTBC lineages were MTBC Lineage 4 (referred to in this paper as L4/Non-Uganda), MTBC Lineage 3 (L3 also known as Central Asian Strains), and MTBC Lineage 4.6/Uganda, which is a sub-lineage of MTBC L4 that is only found in Uganda and the countries immediately surrounding it [4, 6, 7, 9, 12]. The MTBC L4.6/Uganda is the most recently evolved of the three, a sub-lineage of the MTBC L4 generalist lineage, and is unique to this part of Africa[4, 9].

SNP and LSP-based phylogeny have been proven to be consistent in multiple studies of MTBC sub-lineages, and the body of literature on MTBC lineages indicates that this is an

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

excellent method for differentiating the MTBC L4 sub-lineages[6, 24, 60]. Low sequence variation and lack of horizontal transfer make SNPs and LSPs a method well-suited to distinguish lineages, and this approach has been previously validated and published[6, 12, 13]. MTBC L4.6 is the most common lineage among active TB cases in this cohort[6, 7, 9, 12]. This sub-lineage has been shown to have highly conserved T-cell epitopes (i.e. a lower proportion of variable epitopes) and a much smaller geographic range than non-specialized lineages, indicating that it may be adapted to a specific host population(s)[4].

Statistical Analysis

To assess the association between variants and TBscore, we utilized a linear regression model with sex and HIV status as covariates in Plink v1.9 software. For the interaction term between lineage and human genotype, we chose to operationalize lineage as a binary variable. Each subject is coded as 1 for the L4.6/Ugandan lineage or as 0, which encompasses the L4/Non-Uganda and MTBC L3/Central Asian Strains together. As the L4/Non-Uganda and L3/Central Asian lineages have a longer history of human contact compared to the L4.6/Uganda, which is a newer sub-lineage, we can examine coevolution as we are contrasting a lineage that is more recent , relative to the two older lineages (MTBC L3 and MTBC L4). As we expected a longer historical coexistence to associate with lesser severity and the introduction of a newer sub-lineage to associate with greater severity, we grouped the two older lineages together. This also affords greater power to detect a difference than if we were to examine all 3 lineages independently. A power calculation showed that we have 70% power to detect a 2.5 point change in TBscore due to the interaction (the beta value from our previous paper) at a $p=5x10^{-6}$. The regression equation for modeling co-evolution was: $Y= \beta_0 +\beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5 (X_2X_3) + \varepsilon$, where X_1 = HIV-positive Status, X_2=L4.6/Uganda lineage, X_3 = SNP, X_4 = Sex. In

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0045.v1

these analyses, we used a dominant model of inheritance (Homozygous minor allele and heterozygotes as effect vs the homozygous major allele as referent).

We then combined the summary statistics from the two cohorts to generate meta-analytic p-values. To determine meta-analytic p-values and beta coefficients across the two cohorts, we utilized random effects meta-analysis with inverse variance weighting. Based on the Cochrane handbook recommendations, all variants with an $I^2 > 40\%$ were excluded from the analysis to reduce heterogeneity between the cohorts[61]. To be considered GWAS significant, the interaction term between a SNP and the L4.6 Sub-Lineage had to have a $p<0.05$ in both cohorts, the sign of the beta value had to be the same in both cohorts, and the meta-analytic P-value had to be below 5e-08, the canonical GWAS threshold. To be included in further enrichment and annotation analyses, the meta-analytic P-value had to be below 1e-05. We chose this threshold because previous studies have shown that some variants that do not meet the GWAS threshold may still have important regulatory or biological functions in some cases and may be worthy of further study and follow-up, especially in the context of gene regulation [62-64]. We used FUMA GWAS to annotate and enrich our SNPs below the threshold. Analyses performed through FUMA included gene mapping, regulatory annotation, tissue specificity, MAGMA analysis (gene-based analysis), gene set enrichment, and pathway analyses [65]. In addition to FUMA, we utilized GeneCards, Ensembl, DICE, and STRING DB to help annotate and enrich our results with respect to function, expression, and downstream protein interactions [35, 36, 66, 67].

 doi:10.20944/preprints202110.0045.v1

Supplementary Materials (in separate file)

Supplemental Figure 1. Manhattan Plot for Interaction Between SNP and L4.6/Ugandan Lineage in Cohort 1

Supplemental Figure 2. Manhattan Plot for Interaction Between SNP and L4.6/Ugandan Lineage in Cohort 2

Supplemental Figure 3. Quantile-Quantile Plot for Interaction Between SNP and L4.6/Ugandan Lineage in Cohort 1

Supplemental Figure 4. Quantile-Quantile Plot for Interaction Between SNP and L4.6/Ugandan Lineage in Cohort 2

Supplemental Table 1. SNPs in eQTL analysis

Funding

Funding for this work was provided by the Tuberculosis Research Unit (grant N01-AI95383 and HHSN266200700022C/ N01-AI70022 from the NIAID) and grants RO1AI124348 and R56AI130947 from the NIAID. MLM was supported by T32 HL007567, T32 GM007250, and TL1 TR002549.

Acknowledgments

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, Dr. Christopher Whalen, Dr. Deo Mulindwa, Dr. Christina Lancioni, Allan Chiunda, Denise Johnson, Bonnie Thiel, Mark Breda, Dennis Dobbs, Hussein Kisingo, Mary Rutaro, Albert Muganda, Richard Bamuhimbisa, Yusuf Mulumba, Deborah Nsamba, Barbara

 doi:10.20944/preprints202110.0045.v1

Kyeyune, Faith Kintu, Mary Nsereko, Gladys Mpalanyi, Janet Mukose, Grace Tumusiime, Pierre Peters, Annet Kawuma, Saidah Menya, Joan Nassuna, Alphonse Okwera, Keith Chervenak, Karen Morgan, Alfred Etwom, Micheal Angel Mugerwa, 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, Wandegeya, for their contributions to this study. This study would not be possible without the generous participation of the Ugandan patients and families.

Author Contributions

Conceptualization, MLM, CMS, and SMW; Methodology, SMW and MLM.; Formal Analysis, MLM and EMM.; Investigation, CMS, HM-K, MLJ, WHB.; Resources, HM-K, MLJ.; Data Curation, LLM, EMM.; Writing – Original Draft Preparation, MLM, CMS.; Writing – Review & Editing, all authors.; Visualization, MLM.; Supervision, CMS and SMW.; Project Administration, CMS, WHB, HM-K.; Funding Acquisition, WHB, CMS, SMW.

Institutional Review Board Statement

The study protocol was approved by the National HIV/AIDS Research Committee of Makerere University and the institutional review board at University Hospitals Cleveland Medical Center. Final clearance was given by the Uganda National Council for Science and Technology. All participants provided written informed consent.

 doi:10.20944/preprints202110.0045.v1

Data Availability Statement

Because of the IRB restriction on the data from Uganda, individual level data are only available upon request from the Uganda Genetics of TB Data Access Committee by contacting Dr. Sudha Iyengar (ski@case.edu).

Conflicts of Interest

The authors have no conflicts of interest to report.

References

1. *WHO tuberculosis fact sheet*. https://www.who.int/news-room/fact-sheets/detail/tuberculosis. 2017.
2. Organization, W.H., *Global Tuberculosis Report 2018 Factsheet*. 2018.
3. Bell, L.C.K. and M. Noursadeghi, *Pathogenesis of HIV-1 and Mycobacterium tuberculosis co-infection.* Nature Reviews Microbiology, 2017. **16**: p. 80.
4. Stucki, D., et al., *Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages.* Nat Genet, 2016. **48**(12): p. 1535-1543.
5. Ngabonziza, J.C.S., et al., *A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region.* Nat Commun, 2020. **11**(1): p. 2917.
6. Gagneux, S. and P.M. Small, *Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development.* Lancet Infect Dis, 2007. **7**(5): p. 328-37.
7. Gagneux, S., *Ecology and evolution of Mycobacterium tuberculosis.* Nat Rev Microbiol, 2018. **16**(4): p. 202-213.
8. Coscolla, M. and S. Gagneux, *Consequences of genomic diversity in Mycobacterium tuberculosis.* Seminars in immunology, 2014. **26**(6): p. 431-444.
9. Comas, I., et al., *Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans.* Nat Genet, 2013. **45**(10): p. 1176-82.
10. Gagneux, S., *Host-pathogen coevolution in human tuberculosis.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2012. **367**(1590): p. 850-859.
11. Kodaman, N., et al., *Disrupted human-pathogen co-evolution: a model for disease.* Front Genet, 2014. **5**: p. 290.

12. Wampande, E.M., et al., *Long-term dominance of Mycobacterium tuberculosis Uganda family in peri-urban Kampala-Uganda is not associated with cavitary disease.* BMC Infect Dis, 2013. **13**: p. 484.
13. Wampande, E.M., et al., *Genetic variability and consequence of Mycobacterium tuberculosis lineage 3 in Kampala-Uganda.* PLOS ONE, 2019. **14**(9): p. e0221644.
14. Nahid, P., et al., *Interplay of strain and race/ethnicity in the innate immune response to M. tuberculosis.* PLoS One, 2018. **13**(5): p. e0195392.
15. Salie, M., et al., *Associations between human leukocyte antigen class I variants and the Mycobacterium tuberculosis subtypes causing disease.* J Infect Dis, 2014. **209**(2): p. 216-23.
16. van Crevel, R., et al., *Infection with Mycobacterium tuberculosis Beijing genotype strains is associated with polymorphisms in SLC11A1/NRAMP1 in Indonesian patients with tuberculosis.* J Infect Dis, 2009. **200**(11): p. 1671-4.
17. Caws, M., et al., *The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis.* PLoS Pathog, 2008. **4**(3): p. e1000034.
18. Schmid Hempel, P., *Evolutionary parasitologythe integrated study of infections, immunology, ecology, and genetics*. 2011.
19. Frank, S.A., *Models of parasite virulence.* The Quarterly review of biology, 1996. **71**(1): p. 37-78.
20. Anderson, R.M. and R.M. May, *Coevolution of hosts and parasites.* Parasitology, 1982. **85 (Pt 2)**: p. 411-26.
21. Möller, M., et al., *Genetic Resistance to Mycobacterium tuberculosis Infection and Disease.* Frontiers in Immunology, 2018. **9**(2219).
22. Kodaman, N., et al., *Human and Helicobacter pylori coevolution shapes the risk of gastric disease.* Proc Natl Acad Sci U S A, 2014. **111**(4): p. 1455-60.
23. McHenry, M.L., et al., *Interaction between host genes and Mycobacterium tuberculosis lineage can affect tuberculosis severity: Evidence for coevolution?* PLOS Genetics, 2020. **16**(4): p. e1008728.
24. Gagneux, S., et al., *Variable host-pathogen compatibility in Mycobacterium tuberculosis.* Proc Natl Acad Sci U S A, 2006. **103**(8): p. 2869-73.
25. Abel, L., et al., *Genetics of human susceptibility to active and latent tuberculosis: present knowledge and future perspectives.* Lancet Infect Dis, 2018. **18**(3): p. e64-e75.
26. Di Pietrantonio, T. and E. Schurr, *Host-pathogen specificity in tuberculosis.* Adv Exp Med Biol, 2013. **783**: p. 33-44.
27. Sorrell, I., et al., *The evolution of covert, silent infection as a parasite strategy.* Proceedings of the Royal Society B: Biological Sciences, 2009. **276**(1665): p. 2217-2226.
28. O'Garra, A., et al., *The immune response in tuberculosis.* Annu Rev Immunol, 2013. **31**: p. 475-527.
29. Hoal, E.G., et al., *The arms race between man and Mycobacterium tuberculosis: Time to regroup.* Infect Genet Evol, 2017.
30. Orlova, M. and E. Schurr, *Human Genomics of Mycobacterium tuberculosis Infection and Disease.* Current genetic medicine reports, 2017. **5**(3): p. 125-131.
31. Thuong, N.T., et al., *MARCO variants are associated with phagocytosis, pulmonary tuberculosis susceptibility and Beijing lineage.* Genes Immun, 2016. **17**(7): p. 419-425.
32. Tientcheu, L.D., et al., *Immunological consequences of strain variation within the Mycobacterium tuberculosis complex.* Eur J Immunol, 2017. **47**(3): p. 432-445.
33. Woolhouse, M.E.J., et al., *Biological and biomedical implications of the co-evolution of pathogens and their hosts.* Nature Genetics, 2002. **32**(4): p. 569-577.

34. Wejse, C., et al., *TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have predictive value and may be used to assess clinical course.* Scand J Infect Dis, 2008. **40**(2): p. 111-20.
35. Stelzer, G., et al., *The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses.* Current Protocols in Bioinformatics, 2016. **54**(1): p. 1.30.1-1.30.33.
36. Ha, B., et al., *Database of Immune Cell EQTLs, Expression, Epigenomics.* The Journal of Immunology, 2019. **202**(1 Supplement): p. 131.18-131.18.
37. Portier, I. and R.A. Campbell, *Role of Platelets in Detection and Regulation of Infection.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2021. **41**(1): p. 70-78.
38. McDonald, B. and M. Dunbar, *Platelets and Intravascular Immunity: Guardians of the Vascular Space During Bloodstream Infections and Sepsis.* Frontiers in immunology, 2019. **10**: p. 2400-2400.
39. Pober, J.S. and W.C. Sessa, *Inflammation and the blood microvascular system.* Cold Spring Harbor perspectives in biology, 2014. **7**(1): p. a016345-a016345.
40. Pober, J.S. and G. Tellides, *Participation of blood vessel cells in human adaptive immune responses.* Trends in immunology, 2012. **33**(1): p. 49-57.
41. Speth, C., et al., *Platelets as immune cells in infectious diseases.* Future Microbiol, 2013. **8**(11): p. 1431-51.
42. Pink, R.C., et al., *Pseudogenes: pseudo-functional or key regulators in health and disease?* RNA (New York, N.Y.), 2011. **17**(5): p. 792-798.
43. Roberts, T.C. and K.V. Morris, *Not so pseudo anymore: pseudogenes as therapeutic targets.* Pharmacogenomics, 2013. **14**(16): p. 2023-2034.
44. Seshadri, C., et al., *Transcriptional networks are associated with resistance to Mycobacterium tuberculosis infection.* PLoS One, 2017. **12**(4): p. e0175844.
45. Oliveira, C.J., et al., *Nitric oxide and cGMP activate the Ras-MAP kinase pathway-stimulating protein tyrosine phosphorylation in rabbit aortic endothelial cells.* Free Radic Biol Med, 2003. **35**(4): p. 381-96.
46. Isidori, A.M., et al., *Targeting the NO-cGMP-PDE5 pathway in COVID-19 infection. The DEDALO project.* Andrology, 2021. **9**(1): p. 33-38.
47. Wentowski, C., N. Mewada, and N.D. Nielsen, *Sepsis in 2018: a review.* Anaesthesia & Intensive Care Medicine, 2019. **20**(1): p. 6-13.
48. Alfano, G., et al., *The Role of the Renin-Angiotensin System in Severe Acute Respiratory Syndrome-CoV-2 Infection.* Blood Purification, 2021. **50**(2): p. 263-267.
49. Friebe, A., P. Sandner, and A. Schmidtko, *cGMP: a unique 2nd messenger molecule – recent developments in cGMP research and development.* Naunyn-Schmiedeberg's Archives of Pharmacology, 2020. **393**(2): p. 287-302.
50. Gnipp, S., et al., *Nitric oxide dependent signaling via cyclic GMP in dendritic cells regulates migration and T-cell polarization.* Scientific Reports, 2018. **8**(1): p. 10969.
51. Caccamo, N. and F. Dieli, *Inflammation and the coagulation system in tuberculosis: Tissue Factor leads the dance.* Eur J Immunol, 2016. **46**(2): p. 303-6.
52. Kothari, H., et al., *Role of tissue factor in Mycobacterium tuberculosis-induced inflammation and disease pathogenesis.* PloS one, 2014. **9**(12): p. e114141-e114141.
53. Baker, D.A., et al., *A potent series targeting the malarial cGMP-dependent protein kinase clears infection and blocks transmission.* Nature Communications, 2017. **8**(1): p. 430.
54. Isidori, A.M., et al., *Targeting the NO-cGMP-PDE5 pathway in COVID-19 infection. The DEDALO project.* Andrology, 2021. **9**(1): p. 33-38.

55. Stein, C.M., et al., *Resistance and susceptibility to Mycobacterium tuberculosis infection and disease in tuberculosis households in Kampala, Uganda.* Am J Epidemiol, 2018. **187**: p. 1477-1489.
56. Stein, C.M., et al., *Genome scan of M. tuberculosis infection and disease in Ugandans.* PloS one, 2008. **3**(12): p. e4094.
57. Igo, R.P., Jr., et al., *Fine-mapping analysis of a chromosome 2 region linked to resistance to Mycobacterium tuberculosis infection in Uganda reveals potential regulatory variants.* Genes Immun, 2018.
58. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* American journal of human genetics, 2007. **81**(3): p. 559-575.
59. Wampande, E.M., et al., *A single-nucleotide-polymorphism real-time PCR assay for genotyping of Mycobacterium tuberculosis complex in peri-urban Kampala.* BMC Infect Dis, 2015. **15**: p. 396.
60. Kato-Maeda, M., et al., *Strain classification of Mycobacterium tuberculosis: congruence between large sequence polymorphisms and spoligotypes.* The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease, 2011. **15**(1): p. 131-133.
61. Deeks, J.J., J.P. Higgins, and D.G. Altman, *Analysing Data and Undertaking Meta-Analyses*, in *Cochrane Handbook for Systematic Reviews of Interventions*. 2008. p. 243-296.
62. Wang, X., et al., *Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures.* eLife, 2016. **5**: p. e10557.
63. Hammond, R.K., et al., *Biological constraints on GWAS SNPs at suggestive significance thresholds reveal additional BMI loci.* eLife, 2021. **10**: p. e62206.
64. Williams, S.M. and J.L. Haines, *Correcting away the hidden heritability.* Ann Hum Genet, 2011. **75**(3): p. 348-50.
65. Watanabe, K., et al., *Functional mapping and annotation of genetic associations with FUMA.* Nature communications, 2017. **8**(1): p. 1826-1826.
66. von Mering, C., et al., *STRING: known and predicted protein-protein associations, integrated and transferred across organisms.* Nucleic Acids Res, 2005. **33**(Database issue): p. D433-7.
67. Flicek, P., et al., *Ensembl 2011.* Nucleic acids research, 2011. **39**(Database issue): p. D800-D806.