1 **Resistance to TST/IGRA conversion in Uganda: Heritability and Genome-Wide**

2 **Association Study**

3

4 Michael L McHenry[1]*, Penelope Benchek[1]*, LaShaunda Malone[2], Mary Nsereko[3], Harriet

5 Mayanja-Kizza[3], W. Henry Boom[2], Scott M. Williams[1], Thomas R. Hawn[4], Catherine M. Stein[1,2]

6 *These authors jointly contributed as first authors of this work

7 1) Department of Population & Quantitative Health Sciences, Case Western Reserve
8 University, Cleveland, OH, USA.
9 2) Department of Medicine, Case Western Reserve University, Cleveland, OH, USA.
10 3) Department of Medicine, School of Medicine, Makerere University, Kampala, Uganda.
11 4) Department of Medicine, University of Washington, Seattle, WA, USA.
12

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

Abstract

Background: Pulmonary tuberculosis (TB) is one of the most deadly pathogens on earth. However, the majority of people have resistance to active disease. Further, some individuals, termed resisters (RSTRs), do not develop traditional latent tuberculosis (LTBI). The RSTR phenotype is important for understanding pathogenesis and preventing TB. The host genetic underpinnings of RSTR are largely understudied.

Methods: In a cohort of 908 Ugandan subjects with genome-wide data on single nucleotide polymorphisms, we assessed the heritability of the RSTR phenotype and other TB phenotypes using restricted maximum likelihood estimation (REML). We then used a subset of 263 RSTR and LTBI subjects with high quality phenotyping and long-term follow-up to identify DNA variants genome-wide associated with the RSTR phenotype relative to LTBI subjects in a case-control GWAS design, and annotated and enriched these variants to better understand their role in TB pathogenesis.

Results: The heritability of the TB outcomes was very high, at 55% for TB vs. LTBI and 50.4% for RSTR vs. LTBI among HIV- subjects, controlling for age and sex. We identified 27 loci associated with the RSTR phenotype (P<5e-05) and our annotation and enrichment analyses suggest an important regulatory role for many of them.

Interpretation: The heritability results show that the genetic contribution to variation in TB outcomes is very high and our GWAS results highlight variants that may play an important role in resistance to infection as well as TB pathogenesis as a whole.

Funding: This work was funded by grants from the Bill and Melinda Gates Foundation grant OPP1151836 (to T.R.H., W.H.B, C.M.S., H.M.K.); R01AI124348 (to W.H.B, T.R.H., C.M.S., H.M.K.), U01AI115642 (to W.H.B, T.R.H., C.M.S., H.M.K), 4R33AI138272 (WHB, TRH, SMC, HMK), and Tuberculosis Research Unit (grant N01-AI95383 and HHSN266200700022C/ N01-AI70022, to W.H.B, C.M.S, H.M.K, and T.R.H.). MLM was supported by grants T32 HL007567, T32 GM007250, and TL1 TR002549.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

Key words (4-6): resistance to tuberculosis infection, M. tuberculosis, GWAS, phenotype definition, TB outcomes

Research in Context

Evidence Before This Study: Pulmonary tuberculosis (TB) is an infectious respiratory disease caused by the bacterium *Mycobacterium tuberculosis*, one of the most deadly pathogens on earth. TB primarily affects people living in developing countries, and has a particularly high burden in Southeast Africa. Prior studies suggest that most people have resistance to developing symptoms and most of those infected do not show apparent signs of infection (despite harboring the bacteria). However, prior research suggests that some individuals are resistant to becoming infected at all, despite consistent exposure to the bacteria, and these people have been termed resisters (RSTR). Prior evidence suggests that genetic variation in the human hosts likely plays a role in this resistance to infection but the extent to which this resistance is genetically determined and which variants are most important is in need of greater study.

Added Value of This Study: Proper study design and follow-up is very important to characterizing RSTR individuals and prior studies have not always been performed with adequate follow-up and measurement and misclassification is of concern in RSTR studies. This study has used the longest follow-up and the most accurate way of determining resistance to quantify the genetic contribution to resistance and identify the DNA variants that may play a role in conferring it. Understanding why some individuals are resistant may inform efforts to prevent TB through vaccines and treat TB through more effective therapies.

Implications of All Available Evidence: We have quantified the genetic contribution to (i.e., the heritability of) resistance to infection, in addition to the heritability of other important TB outcomes. We have also identified specific DNA variants which may play an important role in

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

63 the biological processes that govern resistance, helping to improve our understanding of how to

64 prevent and combat TB infection and thus reduce the large global burden it imposes.

65

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

66 **Introduction**

67 Pulmonary tuberculosis (TB) is a major public health problem, as it causes more deaths
68 than any other single infectious agent prior to the COVID-19 pandemic [1]. It is also the leading
69 cause of death among people infected with human immunodeficiency virus (HIV) [2]. The
70 bacterium, *Mycobacterium tuberculosis* (MTB) that causes most TB is transmitted via airborne
71 droplets from coughing and sneezing by people with active disease. However, most people
72 exposed to MTB do not develop active disease. In 2017, only 10 million people developed active
73 disease and 1.6 million people died despite there being ~ 1.7 billion latently infected people in
74 2014 [1, 3]. These numbers are of interest because they demonstrate that the vast majority of
75 people have resistance to active disease.

76 In addition to resisting active disease, some individuals do not develop traditional LTBI
77 with a positive TST or IGRA immunologic response, even in the face of prolonged and persistent
78 exposure to an infectious TB case [4, 5]. These individuals, who may resist or clear MTB
79 infection, or acquire infection with a non-IFNγ−centric T-cell response, have been termed
80 resisters (RSTRs) [6, 7]. Estimates of prevalence of RSTR varies as a function of follow-up time
81 and method of diagnostics (i.e., tuberculin skin test (TST) and/or interferon gamma release assay
82 (IGRA)), with current estimates in high-exposure settings ranging from 7-25% [6, 8-10]. This
83 phenotype is important for understanding pathogenesis and is relevant in designing potential
84 vaccines and prevention strategies against TB, as it can provide insight into how individuals
85 respond to prevent infection by MTB, a critical first step to developing disease[7, 11]. The
86 present study can expand on our current understanding of TB heritability greatly by 1.) utilizing
87 a more accurate classification of the RSTR phenotype and 2.) analyzing a wider range of TB
88 phenotypes than previous studies.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

89 Several studies have shown that host genetic factors can play a role in
90 susceptibility/resistance to TB disease, but few have studied resistance to infection[12]. There
91 does appear to be a genetic component to resistance to infection as our previous study of
92 persistently TST-negative household contacts estimated heritability of this phenotype at 21.7%
93 [13]. Also, purified protein derivative (PPD) reactivity is correlated among siblings, but not
94 among unrelated children who live in the same household with similar exposures to MTB[14].
95 Such data are indicative of a genetic component to the RSTR phenotype. In addition, linkage
96 analyses, candidate gene studies, and genome-wide association studies (GWAS) have added to
97 the evidence that genetic variation associates with MTB infection[12, 13, 15-23], though very
98 few have examined RSTR based on long-term follow-up and using both TST and IGRA for
99 phenotype designation. Genetic influences on resistance to infection have been studied in two
100 ways. Some studies have examined LTBI as the trait of interest, using cross-sectional study
101 design, without any long-term follow-up. Other studies have conducted longitudinal follow-up
102 to identify individuals who started as uninfected (TST or IGRA negative), but eventually
103 converted to TST/IGRA positive or LTBI. We have argued that RSTRs cannot be defined based
104 on a single assessment without long-term follow-up, as conversion to test positivity can occur
105 later, and the clinical significance of individuals that are discordant on the TST and IGRA is not
106 well understood [5, 6, 8, 9].
107 Previous studies have calculated the heritability for TST positivity but the heritability
108 estimates for other possible outcomes in the pathogenesis of TB have not been established. There
109 are several different possible outcomes once an individual has been exposed to MTB. Post
110 exposure to MTB, some people exhibit signs of infection, while others never become infected
111 [11]. In a small number of people, there may be early clearance of the bacteria [8, 24, 25]. If the

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

112 infection is controlled, the host enters a stage of latent (i.e. asymptomatic) infection. This
113 contrast between resistance to/clearance of the infection is the focus of our association analysis.
114 In about 5-10% of those latently infected, the host progresses to active TB[26, 27]. The extent to
115 which genetic variation affects this transition can be examined by comparing active TB cases to
116 LTBI subjects with a heritability estimate. These two transitions post-exposure
117 (clearance/resistance relative to infection and active infection relative to latent) can be
118 understood by different contrasts in heritability estimates to determine the extent to which each is
119 genetically influenced. Additionally, there is a third contrast of interest that is not representative
120 of a single step in TB pathogenesis but is in line with much of the prior literature surrounding
121 genetic association studies of susceptibility to TB. Susceptibility is most often represented as a
122 binary outcome where people with active TB are compared to those without. Thus, we can
123 represent this contrast by comparing active TB cases to all others (including RSTRs, LTBI, and
124 those who cannot be definitively classified as either).

125 The purpose of this study is: 1.) to establish the extent to which different TB phenotypes,
126 including the RSTR phenotype, are influenced by genetic variation (i.e. the genetic heritability)
127 and 2.) to identify individual variants that are associated with the RSTR phenotype relative to
128 LTBI based on the results of a recently published long-term follow-up study that stringently
129 characterized the LTBI and RSTR phenotypes.

130

131 **Methods**

132 **Subject ascertainment and characterization of phenotypes**

133 Subjects were ascertained as part of the Kawempe Community Health Study in Kampala,
134 Uganda; a subset that was limited to RSTR and LTBI subjects were included in a long-term

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

135 follow-up study, as previously described by Stein et al. [28, 29]. Subjects were initially enrolled
136 between 2002-2012, and the follow-up study was conducted from 2014-2017. All TB cases were
137 culture-confirmed based on isolation of MTB from clinical gastric or sputum samples in the
138 original study. Six subjects that were RSTR or LTBI in the original study but had developed
139 symptoms prior to follow-up were not enrolled in this study. Household contacts of index TB
140 cases were confirmed to have lived with the index case for at least 7 consecutive days during the
141 previous 3 months in the original (2002-12) study and were followed for at least 12 months as
142 part of the original study. Latent MTB infection (LTBI) was determined based on a positive
143 TST during the initial study and positive TST and IGRA during the follow-up study (and no
144 symptoms of active TB). Resistance to infection (i.e. the RSTR phenotype) was defined as
145 being consistently negative TST tests despite confirmed exposure in the original study and
146 remaining TST/IGRA negative during the follow-up study, an average of 9 years [29].
147 Individuals who were TST negative or TST positive during the original study but not included in
148 the follow-up study (which only included the RSTR and LTBI subjects) are referred to as "no
149 active TB" in this analysis, because they did not have the IGRA results necessary to confirm
150 their LTBI or RSTR status. This was done for two reasons. First, discordance between the TST
151 and IGRA tests has previously been demonstrated [29]. This introduces the possibility that TST+
152 subjects might actually be IGRA-negative. Secondly, persistently TST negative individuals may
153 have converted to LTBI since their initial negative TST and therefore could not be accurately
154 classified as RSTRs. Thus, the LTBI and RSTR individuals in this study were all confirmed as
155 such in both the original and follow-up studies mentioned above. This helps maintain a
156 consistent and accurate definition of the RSTR phenotype in our study which minimizes
157 misclassification [10].

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

158 In order to address our hypotheses, we used different but overlapping study samples from
159 the same population in our two types of analyses. The first set of analyses aimed to determine the
160 genetic contribution to various TB phenotypes (i.e. the heritability analyses), including but not
161 limited to RSTR and LTBI. We chose four different phenotypic comparisons because they
162 represent different biological processes in the pathogenesis of tuberculosis. Figure 1 shows the
163 number of subjects with each phenotype definition in the heritability estimates, which included
164 97 RSTR, 228 LTBI, 350 Active TB, and 233 "No Active TB" subjects. Our first comparison,
165 contrasting the 350 active TB subjects with all our other phenotypes, is similar to what has been
166 called TB susceptibility in prior genetic studies, i.e. the probability of developing active TB
167 relative to the rest of the population who do not have active TB. This is important to estimate as
168 it allows the quantification of the genetic component of disease susceptibility that has been the
169 major focus in TB genetics literature [11]. Our estimate is also unique in that we were able to
170 confirm similar exposure for all subjects in the study.

171 Our second comparison, representing the heritability of being resistant to infection by
172 MTB relative to those who are latently infected by MTB (but do not have active symptoms),
173 allows us to quantify the genetic component for those individuals whose immune systems are not
174 sensitized with positive TST/IGRA responses, the traditional markers of LTBI with an IFNγ
175 dominant response. These individuals may prevent or clear infection entirely, or perhaps mount
176 a IFN-γ independent response. This was done by contrasting the 97 RSTR subjects with the 228
177 LTBI subjects. The third contrast we made compared the 97 RSTR subjects with both the 228
178 LTBI subjects and the 350 active TB cases. This contrast groups the LTBI and active TB cases
179 together as they have both been infected with MTB (and are thus not RSTR). Most if not all of
180 the active TB cases are presumed to have gone through a stage of latent infection prior to having

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

181 active TB and thus should, in theory, represent a group was not able to resist initial infection or
182 clear the bacteria early.

183 Our fourth and final comparison, of active TB relative to LTBI, allows us to examine the
184 genetic contribution to developing TB symptoms subsequent to infection relative to those who
185 are infected but remain asymptomatic. This included our 350 active TB cases relative to the 228
186 LTBI subjects. Thus, the interpretation of each estimate sheds light on a potentially unique
187 biological process and important step in the pathogenesis of TB. The one comparison that does
188 not mirror one of the transitions in the pathogenesis of TB is the analysis of active TB cases vs.
189 all other phenotypes. However, this analysis mirrors prior studies of TB susceptibility (which
190 represent the majority of genetic association studies performed in TB), and allows us to produce
191 a heritability estimates that used the same phenotype as that in prior literature so we can
192 understand this heritability estimate in the context of prior literature.

193 Thus, these analyses included subjects who were enrolled in the follow-up study
194 described above but also required additional subjects from the original KC Health study. This
195 analysis included active TB cases and subjects who were not infected with active TB but could
196 not be confirmed as RSTR based on the reasons outlined above. These analyses also included
197 subjects who were genotyped with a different method (the Illumina Omni5 chip) than those
198 included in the GWAS analysis, as this was necessary to obtain more subjects with active TB
199 and those who were uninfected. The subjects from each chip included in the particular analysis
200 are highlighted in Figure 1.

201 The second analysis was a GWAS that only included the 74 RSTR and 189 LTBI
202 individuals who were genotyped on a single chip (the Illumina Infinium MEGA^{EX} chip) and
203 included in the recently published long-term follow-up study; using long-term follow-up

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

204 significantly reduced the chances of misclassification of the phenotype. This was done as we
205 believe this is the most stringently defined phenotype of any GWAS study to date for RSTRs and
206 as such may be best able to detect variants associated with the phenotype. This analysis examines
207 the contrast between subjects who had a persistent asymptomatic infection (LTBI) and were
208 concordantly TST+/IGRA+, and those who either resisted infection initially or were able to clear
209 the infection entirely (RSTR) who were concordantly TST-/IGRA-. We also considered the
210 possibility of operationalizing our outcome in the GWAS study as a case-control design where
211 RSTR subjects are cases and both active TB subjects and LTBI subjects are considered controls
212 (i.e. those who are not resistant to infection). However, we observed that the active TB sample
213 includes subjects that are not concordant with respect to TST and IGRA. Thus, there is a greater
214 possibility for misclassification which does not exist when we only examine the RSTR vs. LTBI
215 with high quality follow-up data. Further, our heritability estimates revealed that the heritability
216 of RSTR vs. LTBI is greater than that of RSTR vs. LTBI and active TB cases. We believe this is
217 due to the higher potential for misclassification when the active TB cases are included.

218 **Genotyping and QC**

219 For the GWAS analysis, DNA samples were typed for the ~2 million markers on the
220 MEGA^{EX} chip. This chip was used for the RSTR and LTBI subjects included in the follow-up
221 study as well as a set of active TB cases from the original study. Genotypes were called using
222 Illumina's Genome Studio, version 2.0, with the SNP no call threshold set to 0.15. Duplicates
223 and indels were removed prior to QC. Data cleaning was done to remove samples with a call rate
224 <0.93 SNPs, a minor allele frequency < 0.05, or that deviated from Hardy Weinberg Equilibrium
225 ($p < 10^{-6}$ for deviation from HWE across all samples). After the QC thresholds described above
226 and prior to analysis 733,040 markers remained.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

227 Gender was estimated for all subjects by the Genome Studio Software. Subjects were if
228 there were irreconcilable differences between the reported gender and the gender as determined
229 from genotype analysis. A check was made for subject-identifier mismatch by means of expected
230 genetic kinship within and between pedigrees and in some cases helped to resolve gender-
231 mismatch issues.

232 For our heritability analysis, there were two sets of subjects genotyped on different chips.
233 One set was comprised of the RSTR, LTBI, and active TB subjects genotyped on the MEGA^{EX}
234 chip described above. However, in order to include the phenotypes we needed to examine the
235 full range of our TB comparisons representing different transitions in pathogenesis, we had to
236 combine the genotype data from the aforementioned subjects with a set of subjects that used a
237 different genotyping chip into a single large genetic relatedness matrix (GRM) for use in our
238 heritability estimates. These set of subjects included 201 additional active TB cases, 15 RSTR
239 subjects, 32 LTBI subjects, and 233 "No Active TB" subjects. This second genotyping chip was
240 the Illumina HumanOmni5 microarray, which comprises ~4.3 million genome wide markers, and
241 offers high coverage of common genetic variation even for African populations[30]. Genotype
242 calling and quality control were performed as previously described [30]. After QC and filtering
243 by minor allele frequency (MAF > 0.05), genotyping call rate of 98%, and Hardy-Weinberg
244 equilibrium threshold of $p < 10^{-6}$, a total of 337, 566 SNPs passed QC and overlapped between
245 the two genotyping chips; QC criteria differed slightly because of the difference in genotyping
246 chips. These 337, 566 SNPs were used for the heritability analyses.

247 **Kinship and PC Generation**

248 Kinship and principal component (PC) generation was performed separately for the two
249 analyses. For the GRM-based heritability analysis, we computed principal components but did

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

250 not include them based on: 1.) prior studies using the same genotype data showed that there is
251 not substantial population sub-structure or batch effects between the two chips; and 2.) the plots
252 of PC1 vs. PC2 that showed a lack of clustering in our combined genotype data used in the
253 heritability estimation. We also performed sensitivity analyses to show that the inclusion of these
254 PC's did not change our final estimates (Supplemental Methods). Previous literature has shown
255 that it is not necessary to include kinship estimation in a genetic related matrix based heritability
256 estimation [31]. In the GWAS analysis described above, to correct for relatedness among
257 individuals in our sample due to population and family structure, we carried out a principal
258 components analysis (PCA) and estimated kinship among the 263 RSTR or LTBI individuals
259 passing QC during the association analysis (see below). A genome-wide panel of 135,859
260 common (MAF ≥ 0.05) independent (pairwise $r^2 < 0.1$) variants passing marker QC from the
261 MEGA^{EX} panel was chosen for the PC and kinship generation. We calculated PCs using PC-AiR
262 and kinship using PC-Relate in the R, Genesis package[32]. The Genesis package pipeline for
263 generating PCs and kinship first determines relatedness via an initial KING estimated kinship
264 matrix. It then determines the population structure from the unrelated subset of the sample
265 (determined by a KING-estimated kinship threshold of 0.0221) and projects the loadings onto the
266 subset of related subjects to obtain PCs that account for population structure with family
267 structure removed. Genesis then re-calculates kinship using PC-Relate by adjusting for the PCs
268 generated from PC-AiR to get an estimate of family structure with population structure removed.
269 The total number of PCs used in the GWAS was determined by examining the elbow plot of PCs
270 and selecting the count where the variation explained by additional PCs was minimal
271 (Supplemental Figure 1).

272 **Statistical Analyses**

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

273 For our GWAS we employed a score-based association test in a mixed model framework
274 that allows for the inclusion of a polygenic random effect (e.g. a genetic relationship matrix). We
275 used the assocTestSingle function in the Genesis package to conduct the genome-wide
276 association analysis. Because our phenotype is binary (RSTR/LTBI) Genesis uses the penalized
277 quasi-likelihood (PQL) approximation to the generalized linear mixed model (GLMM) to fit the
278 specified model, following the procedure of GMMAT[33]. We controlled for population
279 structure and family structure in our model including the first two principal components (see
280 above) as fixed effects and the genetic relationship matrix (GRM, see above) as a random effect,
281 respectively. Two PCs were selected based on an elbow plot of the variation explained as a
282 function of the number of clusters in our data. In order to interrogate the role of variants that did
283 not meet the traditional GWAS p-value threshold but could nonetheless be truly associated, we
284 used the $p=5x10^{-5}$ threshold as "suggestive" of significance thereby allowing us to examine
285 variants that may have important regulatory or biological function, despite not reaching the
286 typical GWAS threshold of significance (i.e. $p=5x10^{-8}$)[34]. In previous studies, SNPs below the
287 GWAS threshold in one study have successfully replicated when tested again using the same
288 phenotype in subsequent studies, revealing additional variants that are associated with the
289 phenotype of interest. Specifically, SNPs that meet the suggestive threshold but not the GWAS
290 threshold in some cases have at times been shown to be GWAS significant (i.e. $p<5x10^{-8}$) after
291 further study with a larger sample size [35, 36]. Thus, it may be useful and informative to
292 examine the biological role of SNPs that are in this range of p values, especially with regard to a
293 potential regulatory role and keeping in mind that we had a relatively small sample size [34].
294 For the heritability analyses, a genetic relatedness matrix was constructed for restricted
295 maximum likelihood estimation (REML) [37] in GCTA software to estimate the genome-wide

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

Preprint not peer reviewed

296 heritability of three different phenotypes: 1.) susceptibility to active tuberculosis (compared to
297 latent TB and uninfected); 2.) RSTR (compared to LTBI); and 3.) of progression to active TB
298 (compared to latent TB). GCTA estimates the variance explained by all the SNPs across the
299 whole genome for a complex trait[37, 38]; all references to heritability estimates derived for this
300 paper used this methodology and are simply referred to as "heritability" hereafter.

301 In order to account for the impact of age, sex, and HIV status on these heritability
302 estimates we first ran all heritability estimates unadjusted. We then computed heritability
303 estimates that were adjusted for age, sex, and HIV status as covariates. Further, as it is possible
304 that the inclusion of HIV+ subjects might affect the heritability estimates in ways that are not
305 accurately reflected by the inclusion of the three aforementioned covariates, we ran a stratified
306 analysis that included only HIV- subjects. This allows us to observe how these covariates affect
307 our estimates and to compute estimates that are less affected by bias.

308 **Functional annotation**

309 We prioritized loci based on a scoring scheme that combines various forms of evidence
310 regarding putative functionality. This was done in order to provide semi-quantitative evidence
311 that may indicate a potential functional role for our SNPs that did meet the GWAS threshold but
312 showed low enough p-values to warrant further interrogation[39]. The scheme assigned 1 point
313 to each locus for each of the following categories: having >1 SNP below the suggestive threshold
314 at the locus with an extra point for loci with SNPs having a p-value $< 5x10^{-6}$; evidence for a
315 regulatory role as shown by RegulomeDB; serving as an eQTL based on FUMA; and biological
316 relevance to TB based on prior literature. Because the majority of our findings are intergenic
317 and/or fall in noncoding regions, we relied on the annotation tool FUMA version 1.36 for
318 mapping our variants to genes based on genomic proximity, eQTL evidence and chromatin

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

319 interaction evidence[40]. Default settings in FUMA were used, except for tissue specificity.
320 FUMA does not distinguish between cis and trans-eQTL's in their annotations. However, we
321 considered any locus within 1 Mb of the gene being regulated (and on the same chromosome) as
322 a cis-eQTL, a definition based on prior literature surrounding eQTL function[41, 42]. We
323 hypothesized that gene expression and regulation would be most relevant in lung, immune cells
324 and blood and thus, focused on eQTL and chromatin interaction evidence in these target tissues.
325 FUMA and RegulomeDB v2 were used for eQTL identification and to further examine
326 chromatin state evidence and specify enhancer or transcription evidence within lung, immune
327 cells, and blood. GeneCards was used to elucidate gene function and evidence from the literature
328 was used to elucidate a potential biological role for genes in the context of resistance to TB[43].
329 To further enrich our results and yield greater biological insight, we utilized FUMA
330 GWAS' GENE2FUNC feature for the genes represented in our GWAS summary statistics. This
331 feature maps GWAS summary statistics to genes, and then provides gene set or pathway
332 enrichments based on gene sets from MsigDB, KEGG, WikiPathways, and the GWAS Catalog.
333 This function tests our mapped genes for enrichment using pre-established databases of gene sets
334 from prior gene set enrichment analysis (GSEA) analyses using hypergeometric mean pathway
335 analysis and adjusts the p-values for significance based on the Benjamini-Hochberg method (i.e.
336 FDR) using the number of data sources of tested gene sets. FUMA reports gene sets with
337 adjusted P-value ≤ 0.05 and the number of genes that overlap with the gene set > 1 [40].
338 Additionally, STRING network analyses were used to assess if there were protein-protein
339 interactions between the downstream products of the genes identified in our analyses. This
340 allows us to look for common networks or interactions that provide further biological insight into
341 our results. STRING can determine if there is a greater degree of relatedness than expected

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

342 among our results and use this to determine overall protein-protein interaction enrichment as
343 well as enrichment for specific networks, gene ontologies, and previously published works[44].

344 **Sample size estimation**

345 There was no power calculation or sample size calculation done prior to the onset of this
346 study. The sample was determined by which study participants had sufficient DNA and data
347 available for the planned analyses.

348 **Ethics**

349 The study was approved by the National AIDS Research Committee, The Uganda
350 National Council on Science and Technology, and the institutional review board at University
351 Hospitals Cleveland Medical Center. Written informed consent was obtained from all
352 individuals in the study.

353 **Role of the Funding source**

354 The funding source had no role in the study design, statistical analysis, or interpretation
355 of data.

356

357 **Results**

358 **Study Population**

359 We examined 908 people in the final combined sample (Table 1), including 263 RSTR
360 and LTBI individuals (74 RSTR and 189 LTBI) with high quality genotyping data and follow-up
361 clinical information (and thus used in the GWAS analysis), who were all HIV-negative and were
362 similar with respect to age and sex (Table 2)[45]. 350 active TB cases and 233 "No active TB"
363 subjects were in the heritability analysis, but not included in the association (GWAS) analysis.
364 The RSTR and LTBI categories for our heritability estimation included a larger number of

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

365 subjects (97 and 228, respectively) than the GWAS analysis. The differences are due to the
366 inclusion of HIV+ subjects and subjects who were genotyped using a different chip (i.e. the
367 Omni5 chip) in the heritability analysis.

368

369 **Heritability Analysis**

370 For our analysis of active TB relative to all other phenotypes (i.e. LTBI, RSTR and no
371 infection), the heritability (i.e. the percentage of variation in the outcome explained by all
372 variants present on the GWAS chip) without adjustment was 28.7% (Table 3). The heritability
373 was 50.9% for RSTR vs. LTBI, 45.2% for RSTR vs. LTBI and active TB was 45.2%, and 56.0%
374 for active TB vs. LTBI. Adjusting for age, sex, and HIV status increased the estimated
375 heritability to 37.1% for active TB relative to LTBI or no infection and lowered the heritability
376 for RSTR vs. LTBI slightly to 48.3%. The adjustments had no effect on the heritability for
377 RSTR vs. LTBI and active TB (46.2% unadjusted vs. 45.2% adjusted), and of active TB vs.
378 LTBI (56.4% adjusted vs 56.0% unadjusted).

379 To examine whether the inclusion of HIV+ subjects had an effect on the estimates, we
380 estimated heritability among the HIV- persons only. Our analysis stratified by HIV status (and
381 adjusted for age and sex) showed that the heritability estimates were only slightly different when
382 HIV+ subjects are removed but the effect was not large. For the HIV- subjects, the estimates
383 were 38.7% for TB vs. all other phenotypes, 50.4% for RSTR vs. LTBI, 45.8% for RSTR vs.
384 LTBI and active TB, and 56% for active TB vs. LTBI. Thus, our stratified analysis showed that
385 the inclusion of HIV+ subjects in the heritability estimates did not change our outcome
386 drastically compared to our adjusted estimates, implying that these estimates are not particularly
387 sensitive to HIV status.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

388

389 **Association Analysis**

390 There were 40 SNPs spanning 27 loci with a suggestive association with RSTR vs LTBI

391 ($p < 5 \times 10^{-5}$) (Figure 2, Supplemental Table 1). Supplemental Table 1 shows one line for each

392 loci with at least one SNP showing a p-value below 5e-05; the rsID and position is for the SNP

393 with the lowest p-value within each locus. The plurality of associated SNPs were intronic (48%)

394 or intergenic (38%). Sixty percent of these suggestive loci overlapped regions showing strong

395 evidence for transcription and/or enhancer activity in lung, immune cell or blood tissues (as

396 determined by RegulomeDB) and 30% harbored eQTLs in these relevant tissues (Supplemental

397 Table 2). Using our scoring scheme for prioritizing SNPs, there was one locus with 4 points and

398 five loci with 3 points (Supplemental Table 1). We have described these 6 loci (Table 4, Figure

399 3) in greater detail below.

400 The locus with the highest score in our scheme was a region on chromosome 3, including

401 the *ABHD6* gene (Figure 3A). The top SNP in the region, rs9848072 (OR=4.8, $p=2\times10^{-5}$), falls

402 in the last intron of *ABHD6* and the locus is a cis-eQTL for this gene. The region also harbors

403 eQTLs for several other genes, *RPP14*, *ABHD6*, *PXK*, *DNASE1L*, *FLNB* and *PDHB*, in lung,

404 immune cell and blood tissues. Analyzing the 6 genes in a STRING network shows 2 edges and

405 enrichment for endonuclease activity, actin binding, and hydrolase activity, acting on ester bonds

406 (PPI enrichment $p=4.2\text{x}10^{-5}$) (Supplemental Figure 2). It is possible that these genes are co-

407 expressed as part of a gene network, which is why one SNP shows an association with all 6,

408 despite mapping to only 1 of them (i.e. being a cis-eQTL for *ABHD6*).

409 The first locus with 3 points mapped to the *THRB* gene on chromosome 3, and encodes a

410 thyroid hormone receptor[43] (Figure 3B). This locus showed strong evidence for regulation of

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

411 transcription in the lung and contains an enhancer for *THRB* in lung tissue, according to
412 RegulomeDB [46]. The second locus with 3 points was located on chromosome 5 and mapped to
413 *LINC01861*, a long intergenic non-coding RNA gene (Figure 2C). This locus contains an cis-
414 eQTL that is located within 0.1 Mb of and on the same chromosome as *MFAP3* and *FAM114A2*,
415 two genes that previously showed evidence for co-expression in a micro-array analysis [47]. It is
416 also an enhancer in immune cells [46]. The third locus with 3 points was located on chromosome
417 20 and is flanked by *ZHX3* and *LPIN3* (Figure 3D). It shows evidence of strong transcription in
418 the lung and is a cis-eQTL for *CHD6*, a gene within 0.1 Mb on the same chromosome that
419 encodes a DNA-dependent ATPase and is active in chromatin remodeling in response to
420 oxidative stress[43]. The fourth locus was on chromosome 6 and maps to *KIF6* (Figure 3E). It
421 also includes a cis-eQTL for *KIF6* that encodes a protein in the kinesins family. This family of
422 proteins is responsible for intracellular transport of protein complexes[43].

423 A 5th locus with 3 points was on chromosome 5 and mapped to *FAF2* (Figure 3F). The
424 top SNP in the region, rs2963672, falls in the UTR3 within *FAF2* (OR=3.1, $p=4 \times 10^{-5}$) and is a
425 cis-eQTL for *FAF2*. The region also harbors eQTLs for several other genes, *ARL10*, *NOP16*,
426 *HIGD2A*, *RNF44*, *CDHR2* and *GPRIN1* in blood[46]. This region shows strong evidence for
427 regulation of transcription of *FAF2* in lung, blood and immune tissue in RegulomeDB. A
428 STRING network analysis showed a number of edges and similar to our top SNP, this may
429 represent a network of genes with associated expression (PPI enrichment $p=2.6x10^{-7}$)
430 [44](Supplemental Figure 3).

431 We also examined regions that have been previously associated with TST reactivity or
432 LTBI in previous genome-wide and candidate gene studies (Supplemental Table 3). While we
433 were unable to strictly replicate association within specific candidate genes because we did not

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

434 type the same SNPs, we did observe association with the RSTR phenotype in chromosomes 2

435 and 5 originally identified by a genome-wide linkage study in Ugandans ($p=1x10^{-4}$, both

436 regions), the TST2 locus originally linked to TST reactivity ($p=1.66x10^{-4}$), and the *IL9* region

437 associated with TST reactivity in an HIV+ cohort ($6.99x10^{-4}$).

438 In our gene enrichment analysis of SNPs that were P<1e-05 for association, our SNPs

439 showed enrichment for only one biological pathway, the anti-citrullinated protein antibody

440 positive rheumatoid arthritis (ACPA positive RA) pathway from the GWAS catalog, with a p-

441 value of $9.36x10^{-6}$ (FDR-adjusted $p=1.70x10^{-2}$) as determined by GENE2FUNC. This was based

442 on four genes that showed association with the RSTR phenotype and are also found in the ACPA

443 positive RA pathway: *CD40*, *DNASE1L3*, *RPP14*, *PXK* (Figure 4).

444

445 **Discussion**

446 Overall, our results show that the RSTR phenotype has a higher heritability than

447 previously reported and that specific variants associated with RSTR likely serve regulatory roles

448 in lung and immune cells. We posit that the higher heritability estimate seen in this study

449 compared to previous studies is due to a purer phenotypic definition, with less misclassification.

450 Further, we were able to confirm exposure to MTB infection, greatly reducing false negatives in

451 our data. We previously found that the prevalence of the RSTR phenotype varies as a function of

452 follow-up time and use of TST as compared to IGRA [6, 8, 9, 48]. Importantly, studies without a

453 long duration of follow-up will likely misclassify subjects due to subjects converting to TST

454 positivity after the follow-up has concluded. In addition, subjects can revert from positive to

455 negative with TST/IGRA tests which can lead to misclassification. As the RSTRs included in

456 this study had an average of 9 years follow-up after initial TB exposure and both TST and IGRA

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

457 were utilized in the clinical definition we were able to assign the RSTR status with little or no
458 ambiguity. This is the first genetic analysis to utilize this strict definition of resistance to
459 infection. The importance of strict definition for heritability estimation was also seen in our TB
460 heritability estimates. When including subjects without long-term follow-up and/or without
461 confirmed RSTR status, the heritability estimate was lower. Our strict TB vs. LTBI contrast
462 yielded a higher heritability estimate, likely reflecting strong genetic influence on progression;
463 because MTB exposure is documented, the influence of misclassification by lack of exposure on
464 heritability can be minimized.

465 While our GWAS results did not identify any single variant that is significant at the
466 traditional "GWAS threshold," our results are consistent with a number of putatively associating
467 variants with regulatory roles consistent with the RSTR phenotype. The SNPs that were
468 significant at a $p<1x10^{-5}$ level are mostly in regions that have evidence of a regulatory function
469 in RegulomeDB. Many of these SNPs were eQTLs and some exhibit biological relevance as
470 judged by tissue specific effects.

471 *ABHD6*, our top locus using the scoring scheme we developed, produces a lipase that can
472 degrade bis monoacylglycerol phosphate (BMP) and constitutes the major enzyme for BMP
473 catabolism. BMP is expressed in the late endosomes and lysosomes of phagocytosing
474 macrophages [12]. In most mammalian cells, BMP levels are low, comprising only about 1–2%
475 of total phospholipids. However, BMP constitutes 16% of the total phospholipids in lung
476 alveolar macrophages [13]. Additionally, the gene *DNASE1L*, a gene for which the *ABHD6* SNP
477 is a trans-eQTL, codes for a protein that plays a key role in degrading neutrophil extracellular
478 traps (NETs). NETs are mainly composed of DNA fibers and are released by neutrophils to bind
479 pathogens during inflammation. NETs may play a key role in the pathway responsible for non-

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

480 specific inflammation and tissue destruction in pulmonary TB [14]. Therefore, the SNP in
481 *ABHD6* is potentially relevant to TB development through two distinct biological processes.

482 *FAF2*, one of our second highest scoring loci, is part of the innate immune system and
483 regulates endoplasmic reticulum-associated degradation (ERAD), a system forubiquitin-
484 dependent degradation of misfolded proteins. FAF2 controls the steady-state expression of the
485 IGF1R receptor, thus indirectly regulates the insulin-like growth factor receptor signaling
486 pathway. IGF-I has been shown to contribute to the maintenance of *Mycobacterium leprae*
487 persistence in the host, reinforcing a key role for IGF-I in leprosy pathogenesis. It has been found
488 that blocking IGF-I signaling rescues antimicrobial activity in *M. leprae* (ML)-infected
489 macrophages and furthermore, knockdown of IGF-1R rescues antimicrobial activity in ML-
490 infected human macrophages [23]. These cytokines and processes are similar between the *M.*
491 *leprae* response and the response to MTB infection and adds plausibility that these genes could
492 play important roles in the RSTR phenotype.

493 Four of our loci mapped to genes that show enrichment for the rheumatoid arthritis (RA)
494 pathway based on the GWAS catalog. The component genes of the RA pathway as well as the
495 pathway itself are known to be important in susceptibility to infectious disease. RA is an
496 inflammatory disease characterized by increased levels of pro-inflammatory cytokines
497 (particularly TNF-α) that are important to host immune response in TB[49]. Many RA drugs are
498 anti-TNF biologics, and depression of normally important cytokines among RA patients
499 receiving treatment leads to a susceptibility to developing active TB or a re-activation among
500 LTBI patients [50]. Previous immunological studies have shown the importance of TNF (and the
501 other Th1 cytokines) in the TB response [51]. A previous meta-analysis of hundreds of TB
502 susceptibility studies showed enrichment for the RA pathway when the meta-analytic summary

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

503 statistics were analyzed for pathway enrichment [22]. Further, the RA pathway has been shown
504 to be an important part of the alveolar inflammation response to infection; alveolar macrophages
505 are the first target of infection for MTB, and they have previously been discussed as a possible
506 mediator of the RSTR phenotype [52, 53]. Thus, the results presented in this study in conjunction
507 with previous literature are consistent with genetic variants in the pro-inflammatory Th1
508 cytokine response, which are also grouped together in the RA pathway, being associated with the
509 RSTR phenotype and resistance to MTB infection.

510 This study was not without limitations. We were not able to detect any variants that were
511 associated with the RSTR phenotype at the GWAS threshold of $p<5x10^{-8}$ but this may be due to
512 small sample size and the inability to replicate findings due to the lack of an additional cohort
513 with a similarly strict phenotype definition. Further, while family-based association testing may
514 have a lower power to detect an association, we were able to control for exposure between RSTR
515 and LTBI individuals whereas typical case-control studies do not. Despite these limitations, our
516 results identified eQTLs, enhancers, and other regulatory functions that are biologically plausible
517 in the context of resistance to infection by MTB. Additionally, we were able to utilize a larger
518 sample size for our heritability estimates and we were able to demonstrate that the RSTR
519 phenotype has a high heritability despite a lack of individually associating SNPs at the GWAS
520 level. This may imply that the RSTR phenotype is influenced by a variety of genes and variants
521 rather than individual SNPs or genes.

522 Overall, these findings demonstrate how a purer clinical phenotype can yield higher, and
523 probably more accurate, heritability estimates and identify potential new candidate genes for
524 resistance to MTB infection. Our observation that many of these associated SNPs have
525 regulatory functions add to our existing understanding of how genetic variants, including those

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

526 from previous studies of the RSTR phenotype, influence resistance to MTB infection and build
527 on our previous knowledge of TB pathogenesis.
528

529 **Contributors:** C.M.S, M.L.M, and T.R.H conceived of the design and analysis. W.H.B, H.M-
530 K, M.N. and C.M.S. oversaw the subject recruitment and characterization. M.L.M. and P.B.
531 conducted the statistical analysis, L.M. oversaw the data collection and constructed clinical
532 datasets, and C.M.S. and S.M.W. supervised the statistical analyses. M.L.M., P.B., and C.M.S.
533 verified the underlying data. M.L.M. and P.B. drafted the initial version of the manuscript, and
534 all authors had full access to all of the data in this paper and edited and approved the final
535 version of the manuscript.
536

537 **Declaration of Interests**: The authors have no conflicts of interest to report.
538

539 ACKNOWLEDGEMENTS
540 We want to acknowledge the contributions made by senior physicians, medical officers, health
541 visitors, laboratory and data personnel: Drs. Alphonse Okwera and Moses Joloba, Ms. Dorcas
542 Lamunu, Deborah Nsamba, Annet Kawuma, Saidah Menya, Joan Nassuna, Joy Beseke, Michael
543 Odie, Henry Kawoya, Shannon Pavsek, Dr. E. Chandler Church, Anna Duewiger, Keith
544 Chervenak, and Bonnie Thiel. This study would not be possible without the generous
545 participation of the Ugandan patients and families. Lastly, we would like to acknowledge the
546 invaluable contributions to early data analyses by Dr. Robert Igo, Jr, who passed away prior to
547 the preparation of this manuscript.
548

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

549 **Data Sharing Statement:** Because of restrictions on the study placed by the Ugandan ethical
550 review boards, the GWAS data are not available for broad sharing. Investigators who are
551 interested in obtaining the data must apply to the Data Access Committee, chaired by Dr. Sudha
552 Iyengar (ski@case.edu) with a data analysis plan, data security plan, and plan to protect human
553 subjects.

554

555 **References**

556 1. Organization, W.H., *Global Tuberculosis Report 2018 Factsheet*. 2018.
557 2. Bell, L.C.K. and M. Noursadeghi, *Pathogenesis of HIV-1 and Mycobacterium tuberculosis co-*
558 *infection.* Nature Reviews Microbiology, 2017. **16**: p. 80.
559 3. Houben, R.M.G.J. and P.J. Dodd, *The Global Burden of Latent Tuberculosis Infection: A Re-*
560 *estimation Using Mathematical Modelling.* PLoS medicine, 2016. **13**(10): p. e1002152-e1002152.
561 4. Ma, N., et al., *Clinical and epidemiological characteristics of individuals resistant to M.*
562 *tuberculosis infection in a longitudinal TB household contact study in Kampala, Uganda.* BMC
563 Infectious Diseases, 2014. **14**(1): p. 352.
564 5. Stein, C.M., et al., *Resistance and Susceptibility to Mycobacterium tuberculosis Infection and*
565 *Disease in Tuberculosis Households in Kampala, Uganda.* Am J Epidemiol, 2018. **187**(7): p. 1477-
566 1489.
567 6. Simmons, J.D., et al., *Immunological mechanisms of human resistance to persistent*
568 *Mycobacterium tuberculosis infection.* Nat Rev Immunol, 2018. **18**(9): p. 575-589.
569 7. Lu, L.L., et al., *IFN-γ-independent immune markers of Mycobacterium tuberculosis exposure.*
570 Nature Medicine, 2019. **25**(6): p. 977-987.
571 8. Verrall, A.J., et al., *Early Clearance of Mycobacterium tuberculosis: The INFECT Case Contact*
572 *Cohort Study in Indonesia.* J Infect Dis, 2019.
573 9. Mave, V., et al., *Infection free "resisters" among household contacts of adult pulmonary*
574 *tuberculosis.* PLoS One, 2019. **14**(7): p. e0218034.
575 10. Gutierrez, J., et al., *Phenotype Definition for "Resisters" to Mycobacterium tuberculosis Infection*
576 *in the Literature—A Review and Recommendations.* Frontiers in Immunology, 2021. **12**(173).
577 11. McHenry, M.L., S.M. Williams, and C.M. Stein, *Genetics and evolution of tuberculosis*
578 *pathogenesis: New perspectives and approaches.* Infection, Genetics and Evolution, 2020: p.
579 104204.
580 12. McHenry, M.L., S.M. Williams, and C.M. Stein, *Genetics and evolution of tuberculosis*
581 *pathogenesis: New perspectives and approaches.* Infection, Genetics and Evolution, 2020. **81**: p.
582 104204.
583 13. Igo, R.P., et al., *Fine-mapping analysis of a chromosome 2 region linked to resistance to*
584 *Mycobacterium tuberculosis infection in Uganda reveals potential regulatory variants.* Genes &
585 Immunity, 2019. **20**(6): p. 473-483.
586 14. Sepulveda, R.L., et al., *Evaluation of tuberculin reactivity in BCG-immunized siblings.* Am J Respir
587 Crit Care Med, 1994. **149**(3 Pt 1): p. 620-4.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

588 15. Horne, D.J., et al., *Human ULK1 Variation and Susceptibility to Mycobacterium tuberculosis*
589 *Infection.* The Journal of infectious diseases, 2016. **214**(8): p. 1260-1267.
590 16. Shah, J.A., et al., *A Functional Toll-Interacting Protein Variant Is Associated with Bacillus*
591 *Calmette-Guérin-Specific Immune Responses and Tuberculosis.* Am J Respir Crit Care Med, 2017.
592 **196**(4): p. 502-511.
593 17. Hall, N.B., et al., *Polymorphisms in TICAM2 and IL1B are associated with TB.* Genes & Immunity,
594 2015. **16**(2): p. 127-133.
595 18. Stein, C.M., et al., *Genome scan of M. tuberculosis infection and disease in Ugandans.* PLoS One,
596 2008. **3**(12): p. e4094.
597 19. Intemann, C.D., et al., *Autophagy Gene Variant IRGM −261T Contributes to Protection from*
598 *Tuberculosis Caused by Mycobacterium tuberculosis but Not by M. africanum Strains.* PLOS
599 Pathogens, 2009. **5**(9): p. e1000577.
600 20. Cobat, A., et al., *Two loci control tuberculin skin test reactivity in an area hyperendemic for*
601 *tuberculosis.* Journal of Experimental Medicine, 2009. **206**(12): p. 2583-2591.
602 21. Cobat, A., et al., *Identification of a major locus, TNF1, that controls BCG-triggered tumor necrosis*
603 *factor production by leukocytes in an area hyperendemic for tuberculosis.* Clinical infectious
604 diseases : an official publication of the Infectious Diseases Society of America, 2013. **57**(7): p.
605 963-970.
606 22. Stein, C.M., et al., *Genomics of human pulmonary tuberculosis: from genes to pathways.* Curr
607 Genet Med Rep, 2017. **5**(4): p. 149-166.
608 23. Sobota, R.S., et al., *A chromosome 5q31.1 locus associates with tuberculin skin test reactivity in*
609 *HIV-positive individuals from tuberculosis hyper-endemic regions in east Africa.* PLOS Genetics,
610 2017. **13**(6): p. e1006710.
611 24. Nemes, E., et al., *Prevention of M. tuberculosis Infection with H4:IC31 Vaccine or BCG*
612 *Revaccination.* N Engl J Med, 2018. **379**(2): p. 138-149.
613 25. Andrews, J.R., et al., *Serial QuantiFERON testing and tuberculosis disease risk among young*
614 *children: an observational cohort study.* Lancet Respir Med, 2017. **5**(4): p. 282-290.
615 26. Philips, J.A. and J.D. Ernst, *Tuberculosis pathogenesis and immunity.* Annu Rev Pathol, 2012. **7**: p.
616 353-84.
617 27. O'Garra, A., et al., *The Immune Response in Tuberculosis.* Annual Review of Immunology, 2013.
618 **31**(1): p. 475-527.
619 28. Stein, C.M., et al., *Resistance and susceptibility to Mycobacterium tuberculosis infection and*
620 *disease in tuberculosis households in Kampala, Uganda.* Am J Epidemiol, 2018. **187**: p. 1477-
621 1489.
622 29. Stein, C.M., et al., *Long-term Stability of Resistance to Latent Mycobacterium tuberculosis*
623 *Infection in Highly Exposed Tuberculosis Household Contacts in Kampala, Uganda.* Clinical
624 Infectious Diseases, 2018. **68**(10): p. 1705-1712.
625 30. Igo, R.P., Jr., et al., *Fine-mapping analysis of a chromosome 2 region linked to resistance to*
626 *Mycobacterium tuberculosis infection in Uganda reveals potential regulatory variants.* Genes
627 Immun, 2018.
628 31. Hall, J.B. and W.S. Bush, *Analysis of Heritability Using Genome-Wide Data.* Current Protocols in
629 Human Genetics, 2016. **91**(1): p. 1.30.1-1.30.10.
630 32. Conomos, M.P., M.B. Miller, and T.A. Thornton, *Robust inference of population structure for*
631 *ancestry prediction and correction of stratification in the presence of relatedness.* Genet
632 Epidemiol, 2015. **39**(4): p. 276-93.
633 33. Chen, H., et al., *Control for Population Structure and Relatedness for Binary Traits in Genetic*
634 *Association Studies via Logistic Mixed Models.* The American Journal of Human Genetics, 2016.
635 **98**(4): p. 653-666.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

636 34. Wang, X., et al., *Discovery and validation of sub-threshold genome-wide association study loci*
637 *using epigenomic signatures.* eLife, 2016. **5**: p. e10557.
638 35. Hammond, R.K., et al., *Biological constraints on GWAS SNPs at suggestive significance thresholds*
639 *reveal additional BMI loci.* eLife, 2021. **10**: p. e62206.
640 36. Williams, S.M. and J.L. Haines, *Correcting away the hidden heritability.* Ann Hum Genet, 2011.
641 **75**(3): p. 348-50.
642 37. Yang, J., et al., *Concepts, estimation and interpretation of SNP-based heritability.* Nature
643 Genetics, 2017. **49**(9): p. 1304-1310.
644 38. Yang, J., et al., *GCTA: a tool for genome-wide complex trait analysis.* American journal of human
645 genetics, 2011. **88**(1): p. 76-82.
646 39. Ciesielski, T.H., et al., *Diverse convergent evidence in the genetic analysis of complex disease:*
647 *coordinating omic, informatic, and experimental evidence to better identify and validate risk*
648 *factors.* BioData Mining, 2014. **7**(1): p. 10.
649 40. Watanabe, K., et al., *Functional mapping and annotation of genetic associations with FUMA.*
650 Nature Communications, 2017. **8**(1): p. 1826.
651 41. Yao, C., et al., *Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits.*
652 The American Journal of Human Genetics, 2017. **100**(4): p. 571-580.
653 42. Stranger, B.E., et al., *Patterns of Cis Regulatory Variation in Diverse Human Populations.* PLOS
654 Genetics, 2012. **8**(4): p. e1002639.
655 43. Stelzer, G., et al., *The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence*
656 *Analyses.* Current Protocols in Bioinformatics, 2016. **54**(1): p. 1.30.1-1.30.33.
657 44. von Mering, C., et al., *STRING: known and predicted protein-protein associations, integrated and*
658 *transferred across organisms.* Nucleic Acids Res, 2005. **33**(Database issue): p. D433-7.
659 45. Stein, C.M., et al., *Long-term Stability of Resistance to Latent Mycobacterium tuberculosis*
660 *Infection in Highly Exposed Tuberculosis Household Contacts in Kampala, Uganda.* Clinical
661 Infectious Diseases, 2019. **68**(10): p. 1705-1712.
662 46. Boyle, A.P., et al., *Annotation of functional variation in personal genomes using RegulomeDB.*
663 Genome research, 2012. **22**(9): p. 1790-1797.
664 47. Veerla, S. and M. Höglund, *Analysis of promoter regions of co-expressed genes identified by*
665 *microarray analysis.* BMC Bioinformatics, 2006. **7**: p. 384.
666 48. Stein, C.M., et al., *Importance of Study Design and Phenotype Definition in Ongoing Studies of*
667 *Resistance to Latent Mycobacterium tuberculosis Infection.* J Infect Dis, 2020. **221**(6): p. 1025-
668 1026.
669 49. Farrugia, M. and B. Baron, *The role of TNF-α in rheumatoid arthritis: a focus on regulatory T cells.*
670 Journal of clinical and translational research, 2016. **2**(3): p. 84-90.
671 50. Dixon, W.G., et al., *Drug-specific risk of tuberculosis in patients with rheumatoid arthritis treated*
672 *with anti-TNF therapy: results from the British Society for Rheumatology Biologics Register*
673 *(BSRBR).* Annals of the Rheumatic Diseases, 2010. **69**(3): p. 522.
674 51. Cavalcanti, Y.V.N., et al., *Role of TNF-Alpha, IFN-Gamma, and IL-10 in the Development of*
675 *Pulmonary Tuberculosis.* Pulmonary Medicine, 2012. **2012**: p. 745483.
676 52. Simmons, J.D., et al., *Immunological mechanisms of human resistance to persistent*
677 *Mycobacterium tuberculosis infection.* Nature Reviews Immunology, 2018. **18**(9): p. 575-589.
678 53. Cobat, A., et al., *Identification of a major locus, TNF1, that controls BCG-triggered tumor necrosis*
679 *factor production by leukocytes in an area hyperendemic for tuberculosis.* Clinical infectious
680 diseases, 2013. **57**(7): p. 963-970.
681
682

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

683

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

684 **TABLES**

685 **Table 1. Cohort Characteristics for Heritability** `

Age	23.4 (12.0)
Sex	
Male	442 (48.7%)
Female	466 (51.3%)
HIV Status	
Positive	98 (10.9%)
Negative	801 (89.1%)
Classification	
LTBI	228 (25.1%)
RSTR	97 (9.6%)
TB	350 (38.5%)
Not active TB	233 (26.8%)

686 Not active TB refers to subjects who did not have long-term follow up data including IGRA.
687 This includes subjects who were TST+ or TST- during the original study, but could not be
688 categorized without IGRA and long-term data.
689
690 **Table 2. Cohort Characteristics for GWAS**

Variable	Total	RSTR	LTBI	p*
Sample Size	263	74	189	-
Female (0)	125 (47.5%)	38 (51.4%)	87 (46.0%)	0.52
Age, years	23.58 ± 8.9	22.5 ± 8.8	24.0 ± 8.9	0.06
Age range, years	2-55	14-66	14-66	-

691 Values are shown as N (%) or as mean ± SD. *Comparisons between RSTR and LTBI were
692 made using Pearson's χ^2 test and for continuous variables using the non-parametric Wilcoxon
693 rank sum test.
694
695

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

696 **Table 3. Heritability Estimates**

Phenotype Definition	Unadjusted	Adjusted⁑	HIV-Only‡
Active TB vs. LTBI and uninfected*	28.7%	37.1%	38.7%
RSTR vs. LTBI	50.9%	48.3%	50.4%
RSTR vs. LTBI and Active TB	45.2%	45.2%	45.8%
TB vs. LTBI	56.0%	56.4%	55.2%

697 * "uninfected" refers to both RSTRs and subjects who did not have long-term follow up data
698 including IGRA. This includes subjects who were TST+ or TST- during the original study, but
699 could not be categorized without IGRA and long-term data and are referred to as "no active TB"
700 earlier in the document.
701 ⁑ "adjusted" refers to estimates that were adjusted for age, sex, and HIV status. For the RSTR vs.
702 LTBI, the age at which they were distinguished as RSTR or LTBI in a sub-study was used. For
703 the other estimates, the age at first presentation was used.
704 ‡These estimates only included the HIV- subjects and were adjusted for age and sex
705

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

Table 4. GWAS results for featured loci*

Chr:Pos	Top SNP	Associated Gene	MAF	OR	95% CI for OR	Score Stat	P	Effect Allele
3:58272639	rs9848072	ABHD6	0.07	4.81	[2.34, 9.92]	4.26	$2.06x10^{-5}$	G
3:24304291	rs78813564	THRB	0.057	8.12	[3.35, 19.51]	4.65	$3.39x10^{-6}$	G
5:153271156	rs919222	LINC01861	0.388	0.385	[0.26, 0.58]	-4.65	$3.38x10^{-6}$	A
20:39968188	rs6072343	ZHX3; LPIN3	0.0856	5.02	[2.43, 10.26]	4.38	$1.19x10^{-5}$	A
6:39530428	rs10484824	KIF6	0.0875	4.57	[2.28, 9.18]	4.27	$1.93x10^{-5}$	G
5:175935316	rs2963672	FAF2	0.19	3.06	[1.79, 5.23]	4.09	$4.22x10^{-5}$	G

*Featured loci all had P-values < $5x10^{-5}$ and were selected based on biological relevance to RSTR phenotype

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

FIGURES

Figure 1. Source of Subjects and Genotype Data

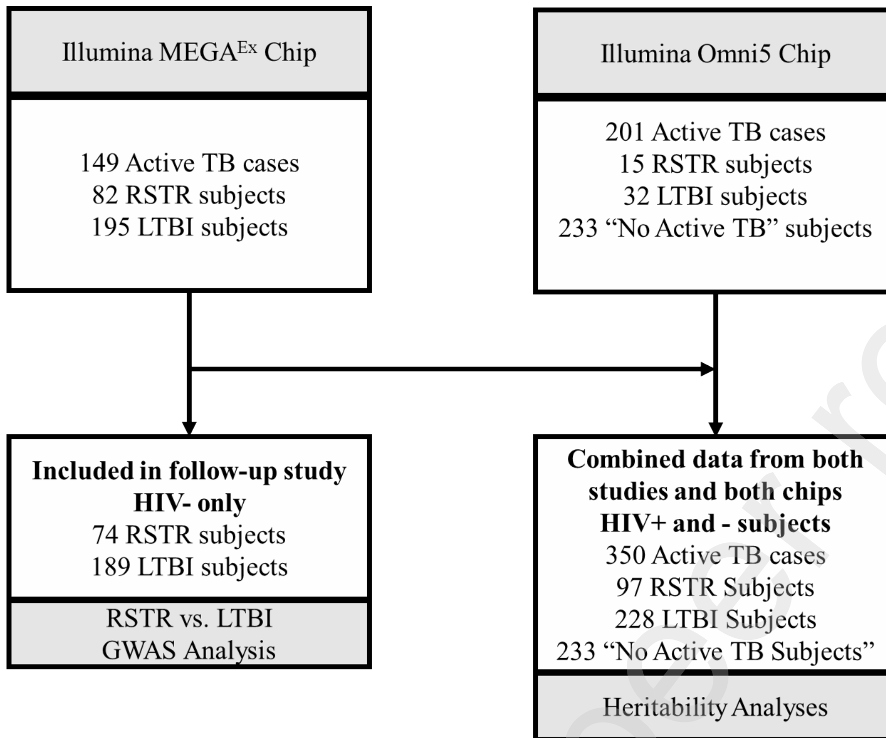


Figure 2. Manhattan Plot of P-Values for association between SNPs and RSTR phenotype (relative to LTBI)

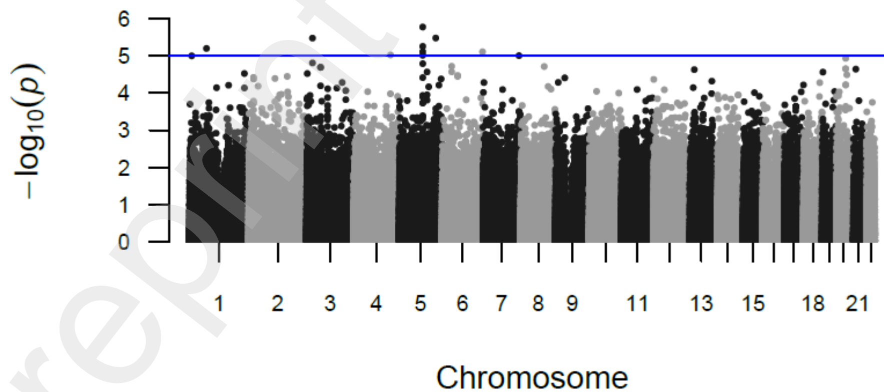
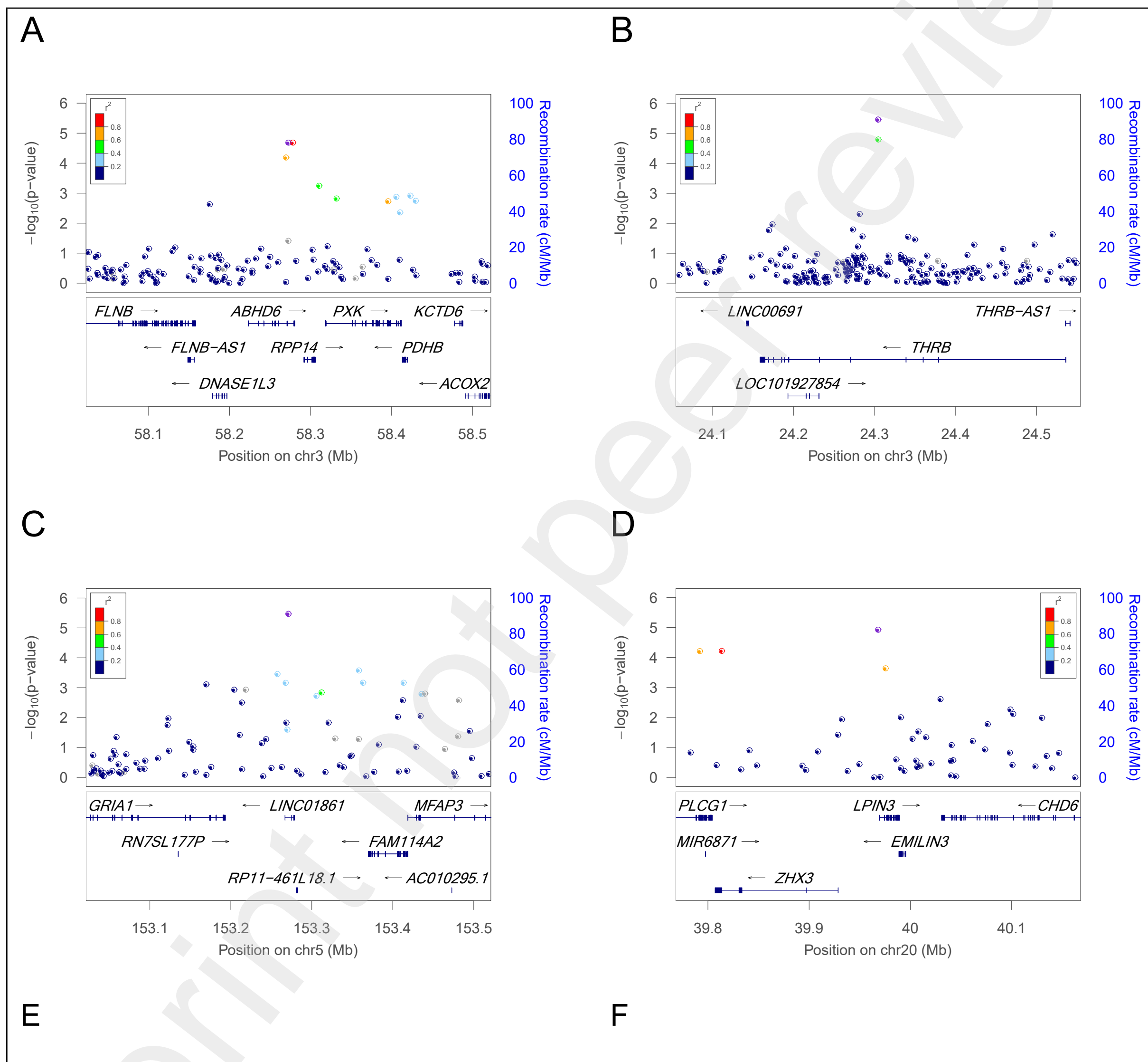


Figure 3. Locus Zoom plots for featured loci

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

Locus Zoom plots for featured, significant loci. Region shown is the gene for which the top variant falls in, plus 200kb flanking. LD shown is for 1000 Genomes, November 2014, African population. A) ABHD region with top SNP rs9848072 (3:58272639). B) CDH22 region with top SNP rs1321001 (20:44816736). C) TMEM13L (alias KIAA0922) region with top SNP rs17369958 (4:154459958). D) FAF2 region with top SNP rs2963672 (5:175935316).



This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

Preprint not peer reviewed

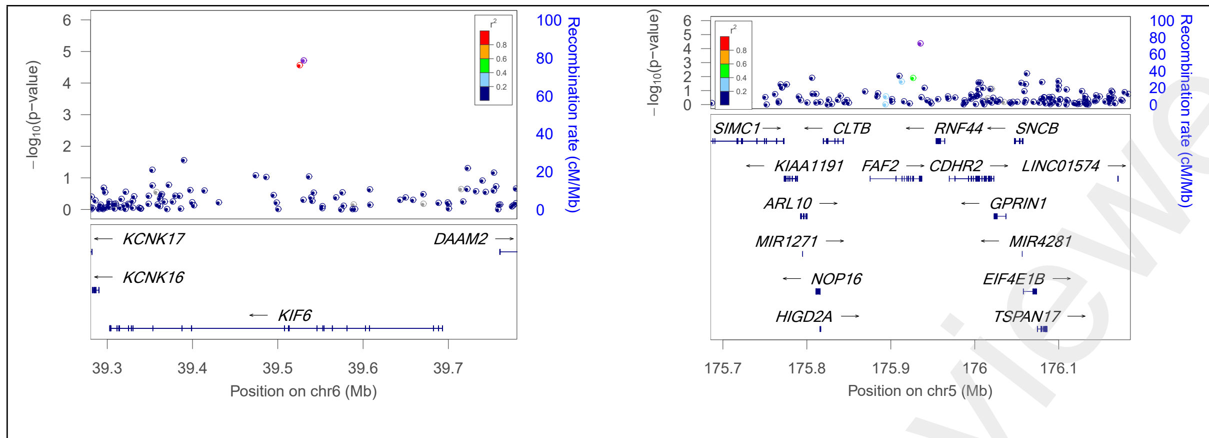
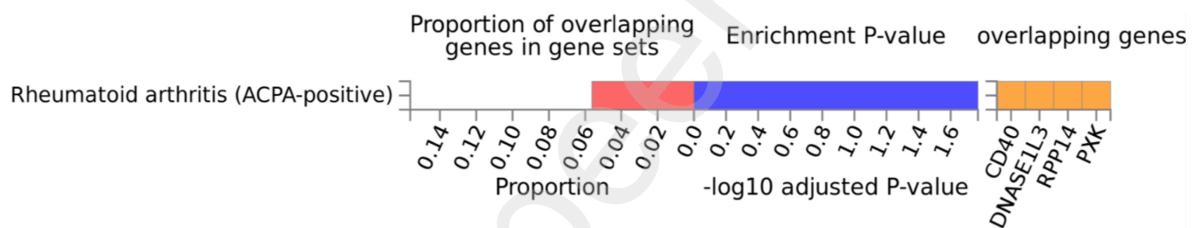


Figure 4. Gene Set Enrichment for GWAS Summary Statistics



Supplemental Tables.

Supplemental Table 1. Loci and lead SNPs with P<5e-05 for Association with RSTR Phenotype

Supplemental Table 2. eQTL's Located in Loci with P<5e-05 for Association with RSTR Phenotype

Supplemental Table 3. Loci Previously Associated with RSTR Phenotype and Accompanying Summary Statistics in Current Study

Supplemental Figures

Supplemental Figure 1. Principal Component Plots

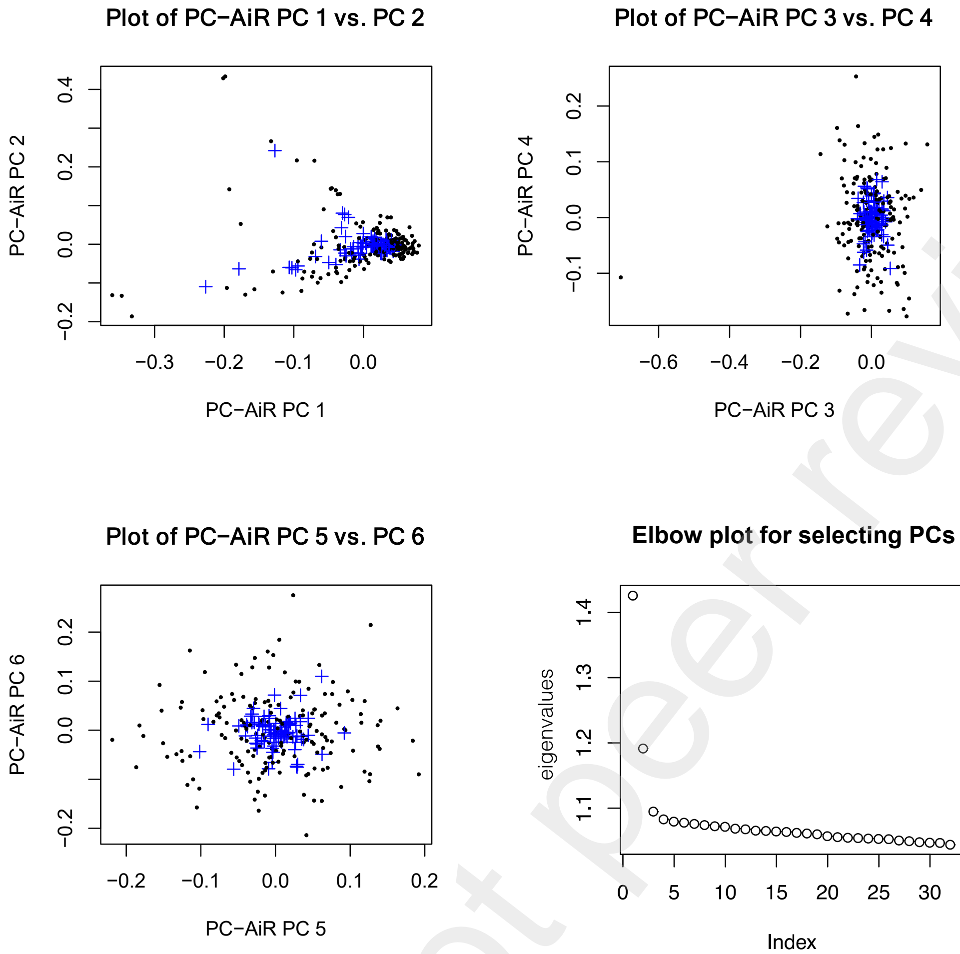
Supplemental Figure 2. STRING diagram for eQTLs

Supplemental Figure 3. STRING diagram for eQTLs

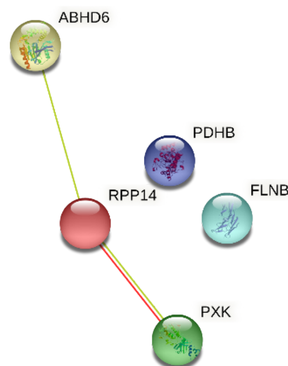
Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

Supplemental Figure 1. Principal Component Plots



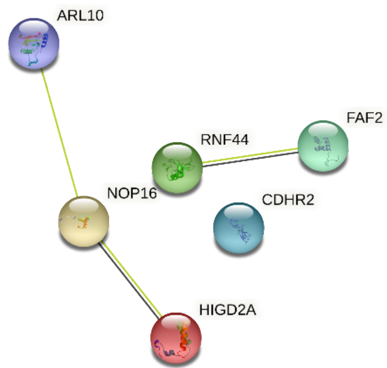
Supplemental Figure 2. STRING diagram for eQTLs



Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430

Supplemental Figure 3. STRING diagram for eQTLs



Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3891430