

RESEARCH ARTICLE

# Investigating the Role of Gene-Gene Interactions in TB Susceptibility

Michelle Daya, Lize van der Merwe, Paul D. van Helden, Marlo Möller, Eileen G. Hoal\*

SA MRC Centre for TB Research, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

\* [egvh@sun.ac.za](mailto:egvh@sun.ac.za)



**OPEN ACCESS**

**Citation:** Daya M, van der Merwe L, van Helden PD, Möller M, Hoal EG (2015) Investigating the Role of Gene-Gene Interactions in TB Susceptibility. PLoS ONE 10(4): e0123970. doi:10.1371/journal.pone.0123970

**Academic Editor:** Francesc Calafell, Universitat Pompeu Fabra, SPAIN

**Received:** October 10, 2014

**Accepted:** February 24, 2015

**Published:** April 28, 2015

**Copyright:** © 2015 Daya et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Study participants did not consent to have their data publicly shared, and data sharing is governed by the authors' ethics review board. Data requests may be sent to the corresponding author.

**Funding:** South African National Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

Tuberculosis (TB) is the second leading cause of mortality from infectious disease worldwide. One of the factors involved in developing disease is the genetics of the host, yet the field of TB susceptibility genetics has not yielded the answers that were expected. A commonly posited explanation for the missing heritability of complex disease is gene-gene interactions, also referred to as epistasis. In this study we investigate the role of gene-gene interactions in genetic susceptibility to TB using a cohort recruited from a high TB incidence community from Cape Town, South Africa. Our discovery data set incorporates genotypes from a large number of candidate gene studies as well as genome-wide data. After limiting our search space to pairs of putative TB susceptibility genes, as well as pairs of genes that have been curated in online databases as potential interactors, we use statistical modelling to identify pairs of interacting SNPs. We attempt to validate the top models identified in our discovery data set using an independent genome-wide TB case-control data set from The Gambia. A number of models were successfully validated, indicating that interplay between the *NRG1 - NRG3*, *GRIK1 - GRIK3* and *IL23R - ATG4C* gene pairs may modify susceptibility to TB. Gene pairs involved in the NF-κB pathway were also identified in the discovery data set (*SFTPD - NOD2*, *ISG15 - TLR8* and *NLRC5 - IL12RB1*), but could not be tested in the Gambian study group due to lack of overlapping data.

## Introduction

Tuberculosis (TB) is a serious global health problem, with 8.6 million new infections and 1.2 million deaths reported in 2012 [1]. In South Africa, it is the fourth leading cause of mortality [2]. The South African Coloured population (SAC) is the largest demographic in metropolitan areas of Cape Town that have some of the highest reported incidences of TB worldwide, despite extensive BCG vaccination and low HIV prevalence [3].

Although up to a third of the world's population has latent TB infection [1], only about 10% of immunocompetent individuals progress to disease. Many studies have established that host genetic factors are involved in the disease [4]. As is the case for other complex diseases, only a

small proportion of the posited heritability has been found [5–7]. The results of TB association studies are furthermore often inconsistent between studies [8, 9].

One of the common explanations for the missing heritability of complex disease is gene-gene interactions, a.k.a. epistasis [5, 10–12]. It has also been postulated that failure to validate genetic associations in independent studies may be ascribed to epistasis [13]. Epistasis can be defined as the effect of a genetic locus on a phenotype being modified by one or more other loci. The term was first used by Bateson based on his experiments with flower color in pea plants, showing that the effects of one gene can be masked by another gene [14, 15]. A similar term, “epistacy”, was later coined by Fisher, referring to the interaction term in regression models that attempts to encapsulate the relationship between two genetic loci and an outcome variable [16]. Based on context, the term epistasis can thus either refer to biological interaction, where effects are mutually dependent and describe a state of nature, or statistical interaction, alluding to the interaction term of two or more variables in a regression model [12, 16, 17]. The notion of biological interaction has often been demonstrated experimentally in model organisms such as yeast, bacteria and animal models [11, 18–27], but this has been less successfully demonstrated in humans [26, 28]. It should also be noted that absence of detectable statistical interaction does not necessarily imply lack of biological interaction [12, 17].

The immune system is complex and comprises many intricate elements, thus progression to active TB may be elucidated by identifying the interplay of gene products in the host defence against TB infection. Only a small number of TB susceptibility gene-gene interaction studies have been published to date [29–35] and these were limited to a small number of candidate genes. In this study, we use a large sample bank of TB case and control samples collected from SAC individuals residing in areas of high TB incidence to detect gene-gene interactions that may underlie TB susceptibility. The data set constitutes genotype data collected from a large number of candidate gene studies (76 genes and 214 SNPs), as well as a large micro-array (chip) data set (388 654 SNPs, 642 cases and 91 controls). We also incorporate correction for ancestry.

A large variety of software packages have been developed to detect gene-gene interactions [36–52]. Approaches implemented in these and other packages can be broadly classed as “traditional” regression based approaches, Bayesian frameworks, testing for allelic association, machine learning and pathway or network based approaches. In traditional regression based methods, interactions are identified by a linear model with phenotype as outcome variable and genotypes as predictor variables. These models includes interaction term(s) which measures the departure of two or more loci from additivity. Interaction models can also be identified using Bayesian frameworks. A prior distribution for the unknown parameter(s), such as the number of predictors to use in a regression model, or the type of effect markers have on the phenotype (no, main or interaction) is specified. The posterior distribution of the parameter(s) is then estimated using simulation techniques such as Markov chain Monte Carlo (MCMC). Due to the large dimensionality of especially genome-wide data sets, an initial filtering step is sometimes employed prior to testing for association using regression. A simple technique employed initially was one proposed by Marchini et al. [53], where tests for interaction are limited to loci that are marginally associated with the phenotype. A more recent strategy is to limit association testing to loci based on curated biological knowledge [54–57]. Another popular approach is the use of test statistics that can be computed efficiently [36, 52]. A particularly intuitive test statistic that measures interaction and that can be computed efficiently is a test for allelic association [48], which can be computed in cases only, or used to test for differing allelic association between cases and controls. Machine learning and data mining techniques use computationally efficient algorithms to identify a set of variables that can be used to predict or classify an outcome. These techniques are especially useful for identifying multiple predictors,

and often use the notion of training and testing data sets to first train and then test models on different subsets of data. Pathway and network based approaches have also recently become popular, and describes complex networks of interactions that may affect a phenotype [26, 38, 41]. Graph theory is used to find subnetworks of genes that represent a common pathway and that are enriched for association with the outcome of interest. In this way underlying disease pathways are identified, rather than specific variants that may be interacting.

With the advent of new genome editing technologies such as CRISPR, detecting pairs of SNPs rather than pathways or networks lends itself to experimental validation. Given the size of our study group, limiting our search to pairs of SNPs only is also appropriate. Our SAC case-control data set comprises sample sets that were genotyped in a number of different studies which did not always overlap well, resulting in a relatively sparse data set. Age, gender and ancestry also differ between cases and controls [58]. We therefore used statistical modelling rather than data mining techniques to identify interactions between pairs of SNPs. Statistical modelling allows for the adjustment of known confounders, and utilizes all available data for each test, without requiring imputation or other complex strategies to deal with missing data. We also limit our search space to pairs of genes that have been identified as TB susceptibility candidate genes, and pairs of genes that have been curated in online databases as potential interactors, a strategy that has previously been used successfully [55, 56]. Finally we attempt to validate our findings in an independent Gambian TB case-control data set.

## Subjects and Methods

### Sample Collection and Ethical Approval

Individuals from the Ravensmead and Uitsig suburbs in Cape Town, who self-identified as South African Coloured, were recruited to participate in this study. (The collective term for people of mixed ancestry in southern Africa is “Coloured” and is recognized and used officially in South Africa. Whilst we acknowledge that in some cultures this term may have acquired a derogatory connotation, this is certainly not intended here.) These suburbs have a homogenous socio-economic environment, low prevalence of HIV and high incidence of TB [3]. TB patients were diagnosed using bacterial confirmation (smear positive/culture positive). Healthy individuals with no prior history of TB were selected as controls. All participants were HIV negative. Our previous study of healthy children and young adults from the control community found that 80% of children older than 15 years had positive tuberculin skin tests (TST), an indication of latent infection with *Mycobacterium tuberculosis* (*M. tuberculosis*) [59]. The majority of the control population is therefore TST positive, and with the average age of the controls in this study being 31 years, we estimate a TST positivity of 80% or above. These healthy individuals had no previous history of TB disease or treatment and were unrelated to all others included in the study.

This study was approved by the Ethics Committee of the Faculty of Health Sciences, Stellenbosch University (project registration numbers 95/072, NO6/07/132 and N11/07/210). Blood samples for DNA were collected with written informed consent, and written informed consent was obtained from the next of kin, caretakers, or guardians on behalf of the minors enrolled in the study. The research was conducted according to the principles expressed in the Declaration of Helsinki.

Sample collection and ethical approval of a Gambian tuberculosis study group, obtained from the Wellcome Trust Case Control Consortium (WTCCC), are described by Thye et al. (2012, Supplementary Information) [60].

## Genotyping and Quality Control

A total of 955 case and 521 control SAC samples were collected between 1994 and 2007. The samples were used to perform a number of candidate gene studies, using unrelated individuals, summarized in [S1 Table](#). Single nucleotide polymorphisms (SNPs) that were genotyped in these studies were used in the present study, and SNPs with a minor allele frequency (MAF) lower than 0.01 and a Hardy Weinberg equilibrium (HWE) p-value (exact test) lower than 0.01 were discarded, leaving 214 SNPs from 76 genes. The SAC sample bank was also used to genotype 969 samples on the Affymetrix GeneChip Human Mapping 500K Array set (Affymetrix 500K chip set). A total of 642 cases, 91 controls and 388 654 SNPs was retained in the data set after SNP calling [61], quality control and removal of related individuals [62]. The data set was also aligned to the Genome Reference Consortium Human genome build 37 (GRCh37). The SAC candidate gene and Affymetrix data sets were then combined, and used to identify pairs of SNPs that jointly modify the odds of having TB.

The Gambian tuberculosis data set was used to validate the top interaction models found in the SAC data set. A total of 1 498 cases and 1 496 controls was genotyped on the Affymetrix 500K chip set (more detail can be found in Thye et al., 2010, Supplementary Methods [63]). After SNP quality control (removal of SNPs with calling probability < 0.95, HWE p-value < 0.0001, MAF < 0.01, missing rate > 0.05) and alignment to GRCh37, 402 856 SNPs remained in the data set. Individuals with excess heterozygosity, outlying individuals and related individuals (degree of relatedness  $\leq 2$ , according to identity by state estimates) were also removed. The final data set were composed of 1 156 cases and 1 206 controls.

## Limiting the Search Space based on Biological Evidence

We limited tests for interaction to SNP pairs of genes that have been identified as TB susceptibility candidate genes (which we refer to as candidate gene SNP pairs), and pairs of genes that are known to interact based on experimental evidence, or that are found in the same biological pathway, ontological category or protein family (which we refer to as biofilter SNP pairs, after the software program used to identify the pairs). A total of 76 candidate genes previously genotyped by our group, as well as 33 additional tuberculosis and pulmonary tuberculosis candidate genes curated in the HGV&TB database based on literature reviews (<http://genome.igib.res.in/cgi-bin/hgvtb/inter.cgi>), with at least one SNP genotyped in the SAC chip data, was used to generate 5 886 candidate gene-gene pairs, composed of 1 278 unique candidate gene SNPs. Seventeen of these SNPs were genotyped on both the Affymetrix chip and another platform, and the strand orientation of 10 of these SNPs were flipped when combining the data sets. Duplicate genotypes were available for 4 686 SNPs, of which 280 genotypes mismatched (error rate of 0.06). The mismatched genotypes were discarded. Another 2 438 interacting gene-gene pairs were identified, comprised of 28 936 unique SNPs. After discarding SNP pairs with less than 60 genotypes available for either cases or controls, 854 703 candidate gene SNP pairs and 1 040 161 biofilter SNP pairs were identified for testing.

## Statistical Analyses

Logistic regression was used to identify pairs of SNPs that jointly modify the odds of having TB. The genotypes of SNPs were encoded as factor variables, and SNPs on chromosome X of male individuals were encoded as homozygotes. Covariates, the main effects of each SNP and an interaction term were included in each model (Case/Control  $\sim$  Covariates + SNP1 + SNP2 + SNP1 $\times$ SNP2).

The p-value of the interaction term was used to detect and report the significance of interactions (4 degrees of freedom test). Reported p-values were not corrected for multiple testing. To aid interpretation of the results, the nature of the association is illustrated by graphs of the observed genotype combination proportions in the data, as interaction effects such as odds ratios are difficult to describe and interpret. Furthermore, reliable estimates of odds ratios could often not be calculated, as some of the genotype combinations include zero counts. Graphs of allele combination frequencies in cases and controls are also provided. An expectation-maximisation (EM) algorithm was used to infer allele combinations per subject. The particular algorithm was originally designed to infer haplotypes, but does not assume the physical coupling of SNPs, and is therefore also appropriate for estimating allele combination frequencies. We note that the only uncertainty in inferring these allele pairs is double heterozygotes. The logits of the possible genotype combinations are also illustrated. This demonstrates the differing direction or magnitude that a SNP has on the odds of having disease, depending on the genotype of the second SNP; non-parallel lines being indicative of an interaction effect. The effects were estimated by absorbing the marginal effects of the SNPs into the SNP  $\times$  SNP interaction term, and adjusting for the covariates included in the model by averaging over them [64].

After the top interacting pairs of SNPs were identified, the individual effects of each of the identified SNPs were tested separately in the SAC and Gambian cohorts using logistic regression. SNPs were encoded as factor variables and covariates were included in each of the models (Case/Control  $\sim$  Covariates + SNP).

Allelic interaction of the identified top SNP pairs was also tested in the SAC cohort. SNPs were encoded as numeric variables, according to the number of copies of the rare variant, as follows: 0, 1 or 2 copies of the rare variant for additive encoding, 0 or 1 for dominant encoding, with 1 representing heterozygotes and rare homozygotes, and 0 or 1 for recessive encoding, with 1 representing rare homozygotes. Each of the nine possible allelic encoding combinations were then tested for each of the identified top SNP pairs.

Age and gender are differentially distributed in the SAC TB cases and controls and gender is differentially distributed between the Gambian TB cases and controls (S2 Table, age not available for the Gambian data). Age and gender were therefore included as covariates in the SAC study group models, and gender was included as a covariate in the Gambian study group models.

Previous work has shown that TB cases have a higher proportion of African ancestry compared to controls in the SAC study group [58, 62], necessitating adjustment for ancestry. Ancestry proportions for each of the 5 source ancestries of the SAC (African San, African non-San, European, South Asian and East Asian) were estimated using a panel of 116 AIMs, as described previously [58]. Ancestry proportions were estimated in a similar manner but using genome-wide data, for those individuals that were also genotyped on the Affymetrix chip. These ancestry proportions were included as covariates in the SAC study group models.

Quality control of the Gambian data set revealed that missing genotypes were associated with having TB for a relatively large proportion of SNPs, which may be indicative of batch effects [65]. As this can be mitigated by the inclusion of principal components in statistical models, principal components were used to adjust the analysis, rather than ancestry proportions as was done for the SAC cohort. Principal components would adjust the models for both differences in ancestry and batch effects between cases and controls [66]. Principal component analysis of the Gambian study group showed associations between having TB and principal components 1, 2, 5, 6, 8, 9 and 10 (p-values  $<$  0.05). These principal components were included as covariates in the Gambian study group models.



## Validation

Statistical modelling was used to identify gene pairs that most likely jointly modify the odds of having TB, and not to quantify the achieved level of statistical significance. Due to the large number of tests done in the SAC study group, and the limited size of the study group (especially the limited number of controls that were available for many of the tests), none of the interaction associations would be statistically significant if adjusted for multiple testing. In addition, many of the multiple testing methods that have been suggested in the literature have severe shortcomings. The straightforward Bonferroni adjustment is too stringent when several genetic associations are tested in the same study group due to correlation (LD) between markers [67, 68]. Alternative methods of correcting for multiple tests were also not feasible for this study. Firstly, roughly 2 million tests were done in differing subsets of individuals from the same study group, which complicates the use of multiple testing correction methods that do not rely on the simple adjustment of p-values by for example dividing by the number of tests done. Bayesian methods require a priori probability of association, which is not known. Due to the large number of tests that were done, permutation testing is also not feasible. Permutation testing is also inappropriate in the context of gene-gene interactions, as permutation based methods do not account for correlation between genotypes [69]. Furthermore, a large proportion of the tests were done on an unbalanced number of cases and controls, which may result in biased permutation-based calculation of p-values. A method to determine the number of effective independent tests when testing pairs of SNPs for interaction in a genome-wide context has also been proposed [55, 70]. This number of effective tests can then be used in a Bonferroni adjustment or to control the false discovery rate. The method does however not take into account that a gene may be tested in multiple gene-pair models, and the accuracy of the original method was evaluated using permutation testing, which may be inappropriate for interaction tests. Due to these reasons an appropriate alpha level was not determined, and we simply selected the top 20 unique gene pairs for validation in the Gambian study group. A similar strategy has been suggested by Kerr [71], albeit in the context of unbalanced microarray gene expression data. The selected models would be the most likely true positives, if any exist.

As patterns of linkage disequilibrium (LD) differ between populations, tag SNPs of causal variants may vary between the SAC and Gambian populations. A SNP associated in the SAC study group points to a region of LD, and any SNP within this region may be the causal SNP [72]. The 20 models that were selected for validation were therefore tested using all possible combinations of SNP pairs found in the region of the SNP tested in the SAC study group. Using a strategy similar to that of Shriner et al. [72] and Ramos et al. [73], SNPs used for validation of a SNP tested in the SAC study group was selected based on the following criteria: the SNPs were found in the same gene region, within 250 000 base pair positions of the SNP, and having a pairwise LD  $r^2$  value of at least 0.3 with the SNP in SAC controls. Although some of the SNPs genotyped in candidate gene studies were selected for their putative functional effects, we note that all the variants in the top twenty models that were genotyped in candidate gene studies were originally selected as they were variants in a gene of interest, and not for their functional effects per se.

After selecting SNP pairs to test using this strategy, a resulting total of 245 regression models were fitted to the Gambian study group. P-values smaller than 0.05 were described as statistically significant.

## Software

Version information, web URLs and important parameter settings of the software packages used in this study are summarized in [S3 Table](#).

PLINK was used for quality control of the SAC chip data set and Gambian chip data set [45]. The SAC and Gambian chip data sets were aligned to GRCh37 using a script and Affymetrix SNP information files available at <http://www.well.ox.ac.uk/~wrayner/strand/>.

ADMIXTURE was used to estimate ancestry proportions of the SAC study group [74]. For the Gambian study group, Eigenstrat was used to infer the top 10 principal components and test for association between these principal components and disease outcome [66]. Prior to estimating ancestry proportions and inferring principal components in the SAC chip data set and Gambian data set, PLINK was used to remove SNPs from the data set that were in LD, as this may lead to biased inference.

Biofilter was used to generate SNP pair combinations of genes that are known to interact based on experimental evidence, or that are found in the same biological pathway, ontological category or protein family. Only those combinations having three or more sources were used for testing interaction in the SAC chip data set. Biofilter was also used to find SNPs within gene regions that are available in the Gambian data set for validation of the top SAC gene-gene models.

The freely available R programming environment was used for statistical analyses, quality control of the SAC candidate genes and graphing [75]. The R *genetics* package was used to test for HWE in the SAC candidate genes and was also used to calculate pairwise LD  $r^2$  and  $D'$  values [76]. The R *haplo.stats* package was used to estimate allele combination frequencies in cases and controls [77]. The adjusted logits of the genotype combinations were estimated using the *effects* package [64]. Figures were created using the R *ggplot2* package [78].

## Results

The top 20 unique gene pair models discovered in the SAC cohort are summarized in [Table 1](#). These models were identified using logistic regression, that tests whether the effect of a SNP on disease outcome is modified by the effect of another SNP, after taking into account (adjusting for) the main effects of the two SNPs. When encountering the same gene-gene model but with differing SNP pairs, only the gene-gene model with the smallest p-value is shown (4 models were excluded for this reason). SNP pairs and p-values of the corresponding highest scoring Gambian models are also reported in the table. As no suitable SNPs were available for some of the genes, some of the models could not be tested in the Gambian data set. Results of the top 250 SAC models and all 245 Gambian models are available in [S4 Table](#). The effects of each of the SNPs in [Table 1](#) were also tested individually in the relevant cohorts, and these single SNP association results are reported in [S5 Table](#). Only two of the SNPs are individually associated with having TB in the SAC cohort (rs15842 and rs3740107 of models 7 and 20), but with a much lower level of significance than the interaction effect of the models (single SNP p-values of  $1.49 \times 10^{-2}$  and  $1.64 \times 10^{-2}$ , interaction p-values of  $6.23 \times 10^{-6}$  and  $1.37 \times 10^{-5}$ , respectively). Only one of the genes reported in [Table 1](#), GRIK1, was identified by the top 36 single SNP associations from a previous genome-wide association study of the cohort [62]. By evaluating combinations of genes, a number of genes were identified that may play a role in TB pathogenesis, which would not have been evident if their effects were assessed individually.

Interaction effects observed in the SAC study group are illustrated in [Figs 1–3](#) for validated models (p-value < 0.05 in the Gambian data set), as well as models that could not be validated due to lack of data, but that have interesting functional interpretations. Note that due to the differing SNP pairs used in the validation, as well as different allele frequencies and LD patterns in the two cohorts, the trend observed in a “validated” Gambian model may not necessarily reflect that of the corresponding SAC model, and we use the term here to imply that there is evidence in both cohorts that the gene pair jointly modifies the odds of having TB. [Figs 1 and 2](#)

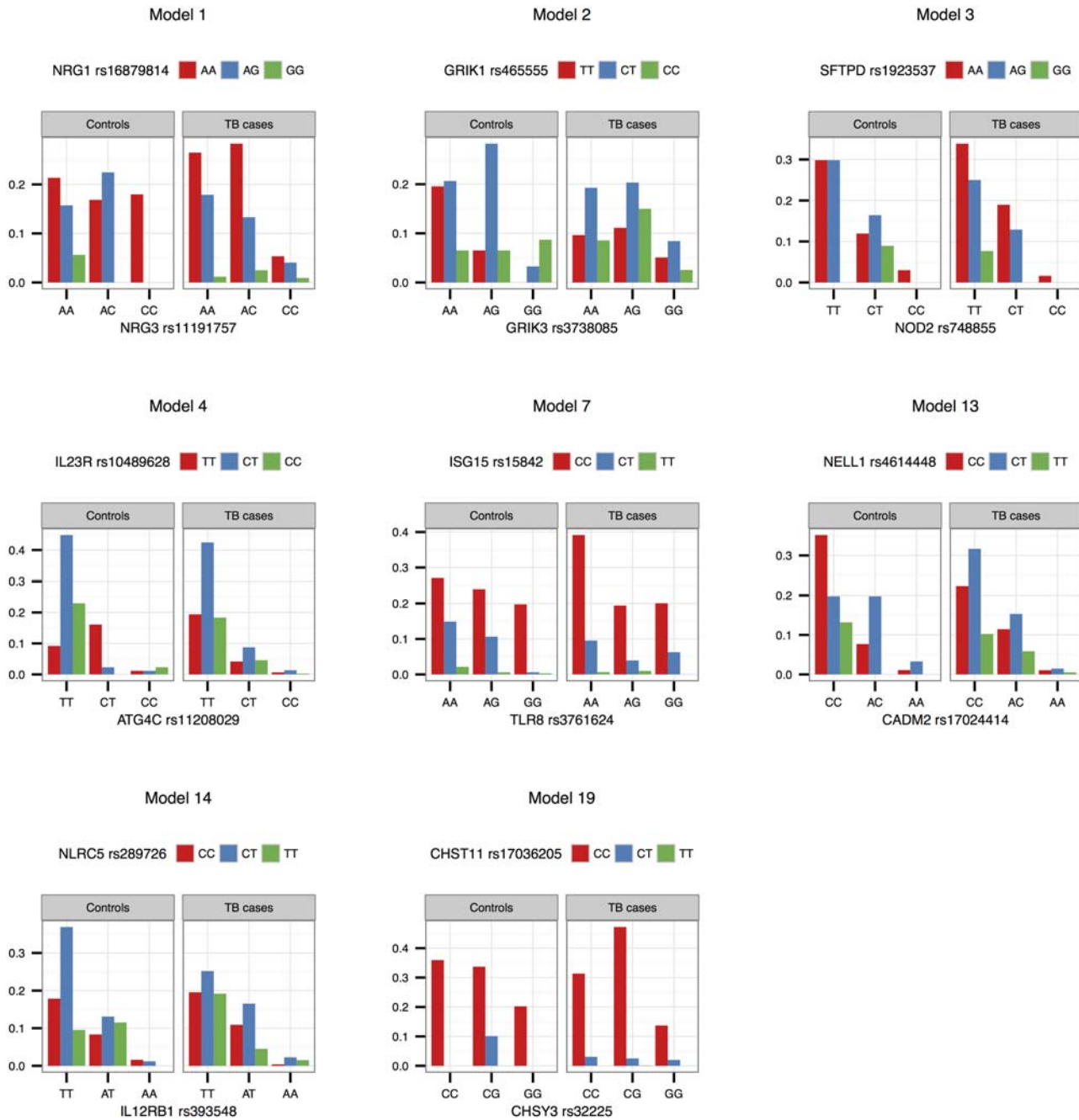
**Table 1. Top twenty interaction models.** This table summarizes the top twenty interaction models identified in the SAC study group. P-values reflect the overall significance of the association between the genotype combinations and having TB, after adjusting for the main effects of the SNPs and covariates. A model of type C indicates a candidate gene pair, and a model of type B indicates a biofilter gene pair. These models were validated in the Gambian study group set using multiple SNPs found within the same gene regions, and the SNP pairs and p-values of the highest scoring Gambian models are reported. For some of the models, no SNPs were available in the Gambian data set for one or both of the genes (blank entries).

	Gene 1	Gene 2	Type	SAC				Gambian			
				Nr cases	Nr controls	SNP 1	SNP 2	SNP 1	SNP 2	P-value	Nr tests
<b>Model 1</b>	NRG1	NRG3	B	634	87	rs16879814	rs11191757	rs16879814	rs2224109	<b>0.0389</b>	1 × 12 = 12
<b>Model 2</b>	GRIK1	GRIK3	B	620	90	rs465555	rs3738085	rs460583	rs476894	<b>0.0476</b>	5 × 11 = 55
<b>Model 3</b>	SFTPD	NOD2	C	216	65	rs1923537	rs748855				
<b>Model 4</b>	IL23R	ATG4C	C	613	85	rs10489628	rs11208029	rs10489628	rs11208029	<b>0.0350</b>	1 × 3 = 3
<b>Model 5</b>	FUT8	B4GALT1	B	627	90	rs17102844	rs12342831	rs9323464	rs10758189	0.1399	4 × 7 = 28
<b>Model 6</b>	EXT1	EXT2	B	626	91	rs6469713	rs903509				
<b>Model 7</b>	ISG15	TLR8	C	271	321	rs15842	rs3761624				
<b>Model 8</b>	NCAM2	IRF8	C	620	87	rs8134735	rs8054065	rs8132838	rs147968	0.0794	4 × 2 = 8
<b>Model 9</b>	ANK1	ANK3	B	606	91	rs2102360	rs2393618				
<b>Model 10</b>	NELL1	NOS2	C	639	91	rs1377741	rs2297516	rs1377741	rs2314809	0.4098	1 × 2 = 2
<b>Model 11</b>	CADM3	SLC22A4	C	224	67	rs16841729	rs13179900				
<b>Model 12</b>	ANK2	ANK3	B	636	90	rs1354679	rs10821731	rs1354679	rs10761481	0.1544	5 × 18 = 90
<b>Model 13</b>	NELL1	CADM2	C	625	89	rs4614448	rs17024414	rs4614448	rs17024876	<b>0.0329</b>	3 × 7 = 21
<b>Model 14</b>	NLRC5	IL12RB1	C	231	245	rs289726	rs393548				
<b>Model 15</b>	PLCB1	PLCE1	B	633	91	rs708914	rs4918082	rs1703634	rs4918082	0.3165	2 × 1 = 2
<b>Model 16</b>	C1QA	TMEFF2	C	263	79	rs12033074	rs4077949				
<b>Model 17</b>	NELL1	CADM3	C	621	84	rs11025887	rs862991	rs12577018	rs862991	0.2107	2 × 1 = 2
<b>Model 18</b>	PDE2A	PDE4B	B	626	87	rs171021	rs536025	rs3781931	rs17423910	0.1169	2 × 4 = 8
<b>Model 19</b>	CHST11	CHSY3	B	623	87	rs17036205	rs32225	rs17036205	rs244745	<b>0.0401</b>	1 × 10 = 10
<b>Model 20</b>	SLC22A4	ALOX5	C	629	91	rs2306772	rs3740107	rs3792880	rs3780909	0.1117	4 × 1 = 4

doi:10.1371/journal.pone.0123970.t001

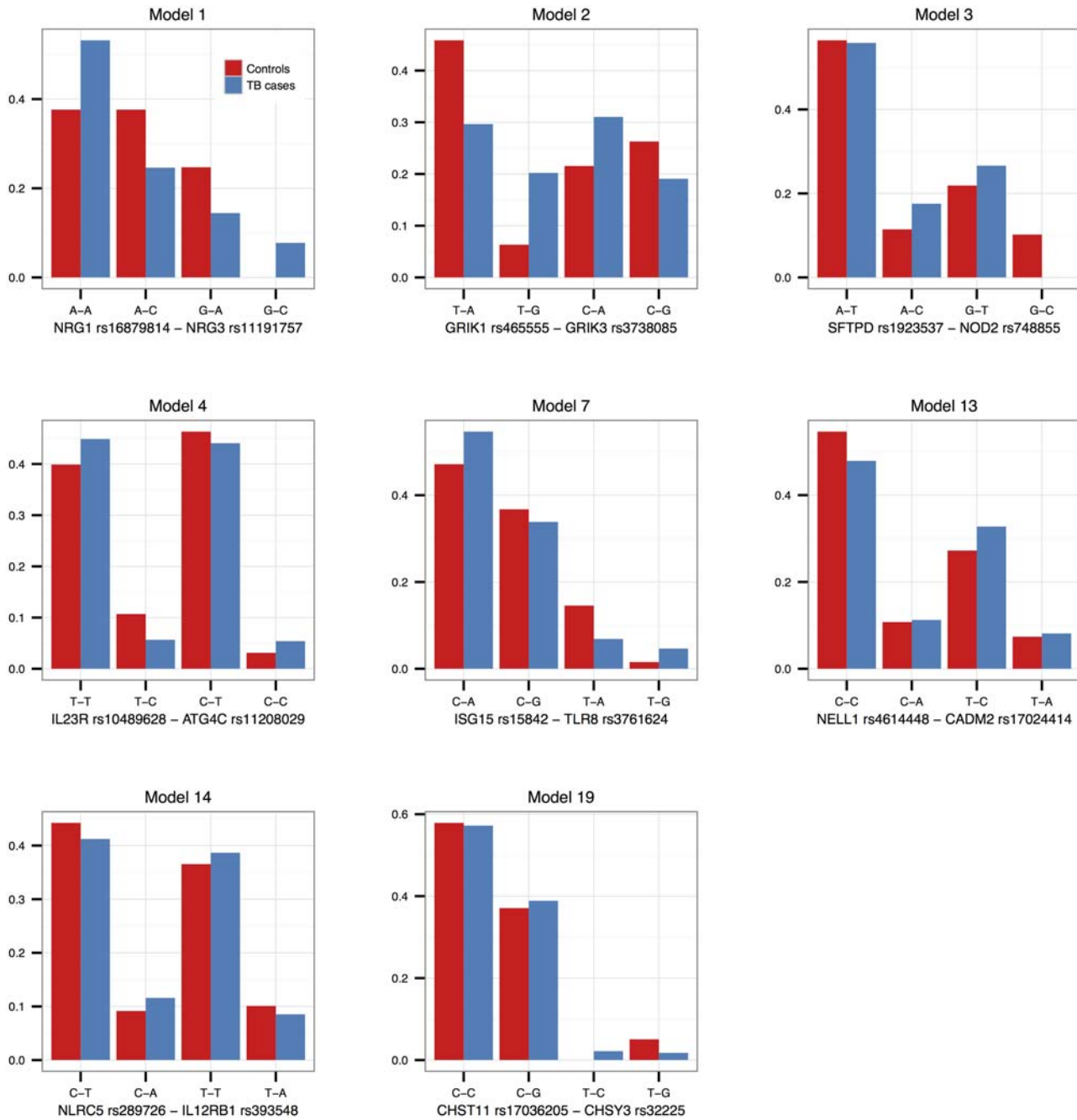
show the frequencies of the genotype and allele combinations in cases and controls. As per the definition of interaction, the allele combination graphs demonstrate the reversal of effects in cases and controls, e.g. if the SNP 1 allele 1—SNP 2 allele 1 combination has a lower frequency in controls compared to cases, then the SNP 1 allele 1—SNP 2 allele 2 combination has a higher frequency in controls compared to cases, i.e. the effect of allele 1 of SNP 1 is modified by the SNP 2 allele. Fig 3 depicts the joint effect that genotype combinations have on the odds of having TB, after adjustment for covariates; non-parallel lines being indicative of interaction effects. For example, model 7 in Fig 3 shows that compared to the CT-AG genotype combination, the CT-GG combination increases the odds of having TB, whereas compared to the TT-AG combination, the TT-GG combination decreases the odds of having TB. Put another way, depending on whether the first SNP has one or two copies of the rare allele T, the effect of having two instead of one copies of the rare allele G for that SNP may increase or decrease the odds of having disease. The frequencies and effects in the SAC study group for the remaining top models are depicted similarly in supplementary figures (S1–S3 Figs), and Figs 4–6 show the frequencies and effects of the validated Gambian models. Below we highlight models that were validated in the Gambian data set as well as three models that could not be tested, but that have interesting functional effects.





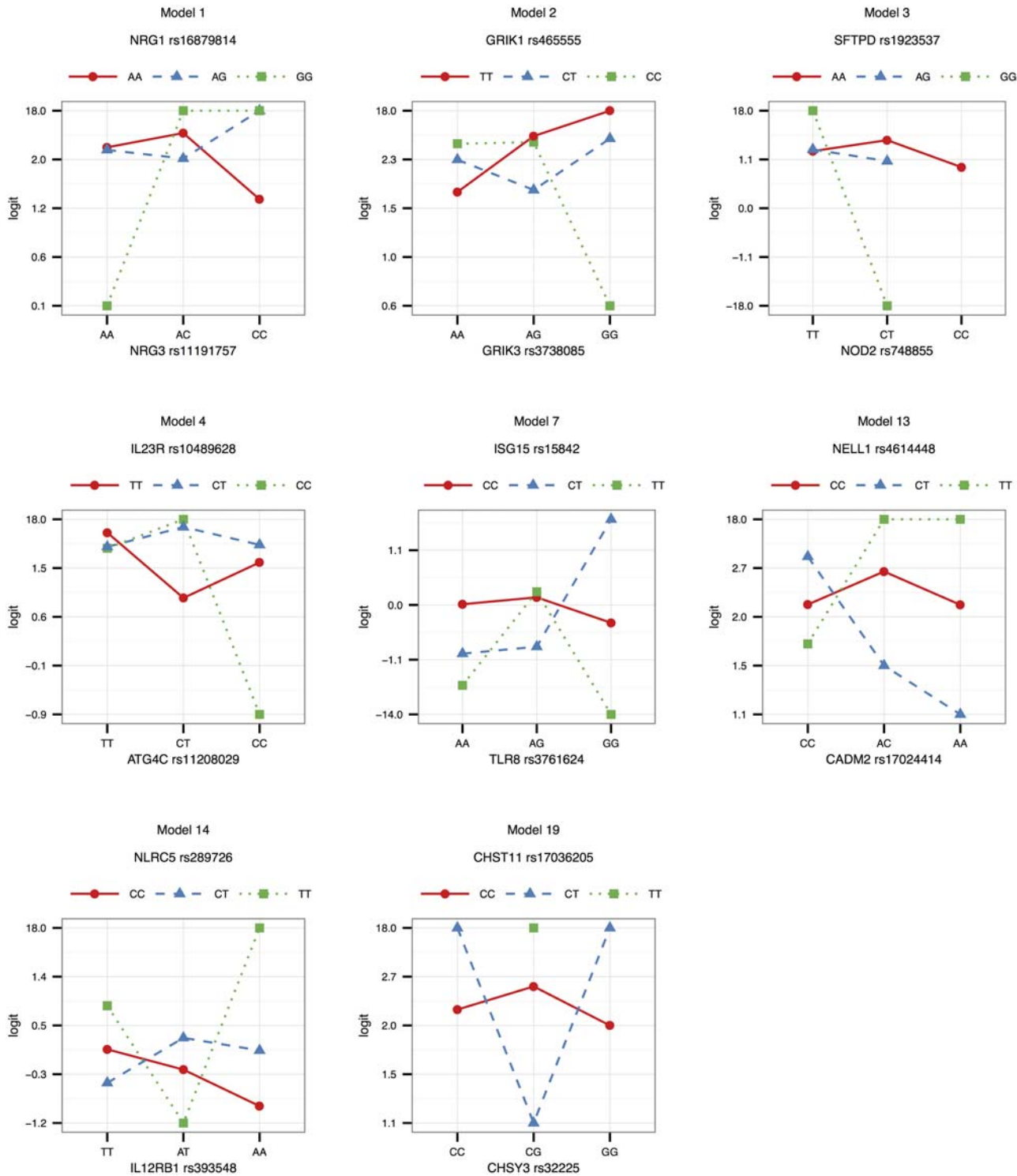
**Fig 1. Genotype combination proportions in the SAC study group.** The observed proportions of the nine possible SNP pair genotype combinations from models 1, 2, 3, 4, 7, 13, 14 and 19 are depicted in this figure, per cases and controls. Genotypes are ordered according to minor allele frequency, with the wildtype homozygote appearing first, and the rare homozygote appearing last.

doi:10.1371/journal.pone.0123970.g001



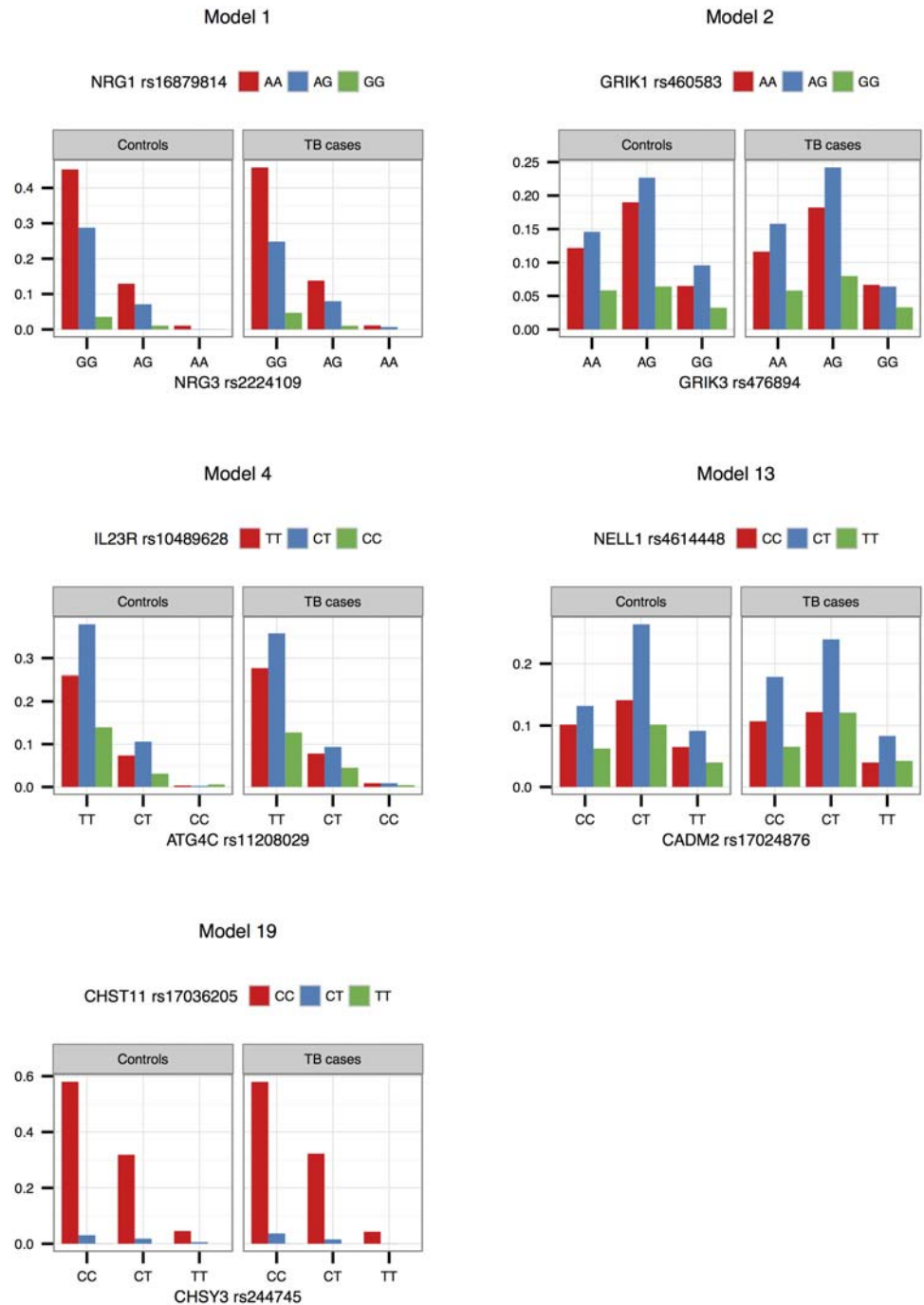
**Fig 2. Allele combination frequencies in the SAC study group.** The frequencies of the four possible SNP pair allele combinations from models 1, 2, 3, 4, 7, 13, 14 and 19 are depicted in this figure, per cases and controls. The frequencies were estimated using an EM-algorithm.

doi:10.1371/journal.pone.0123970.g002



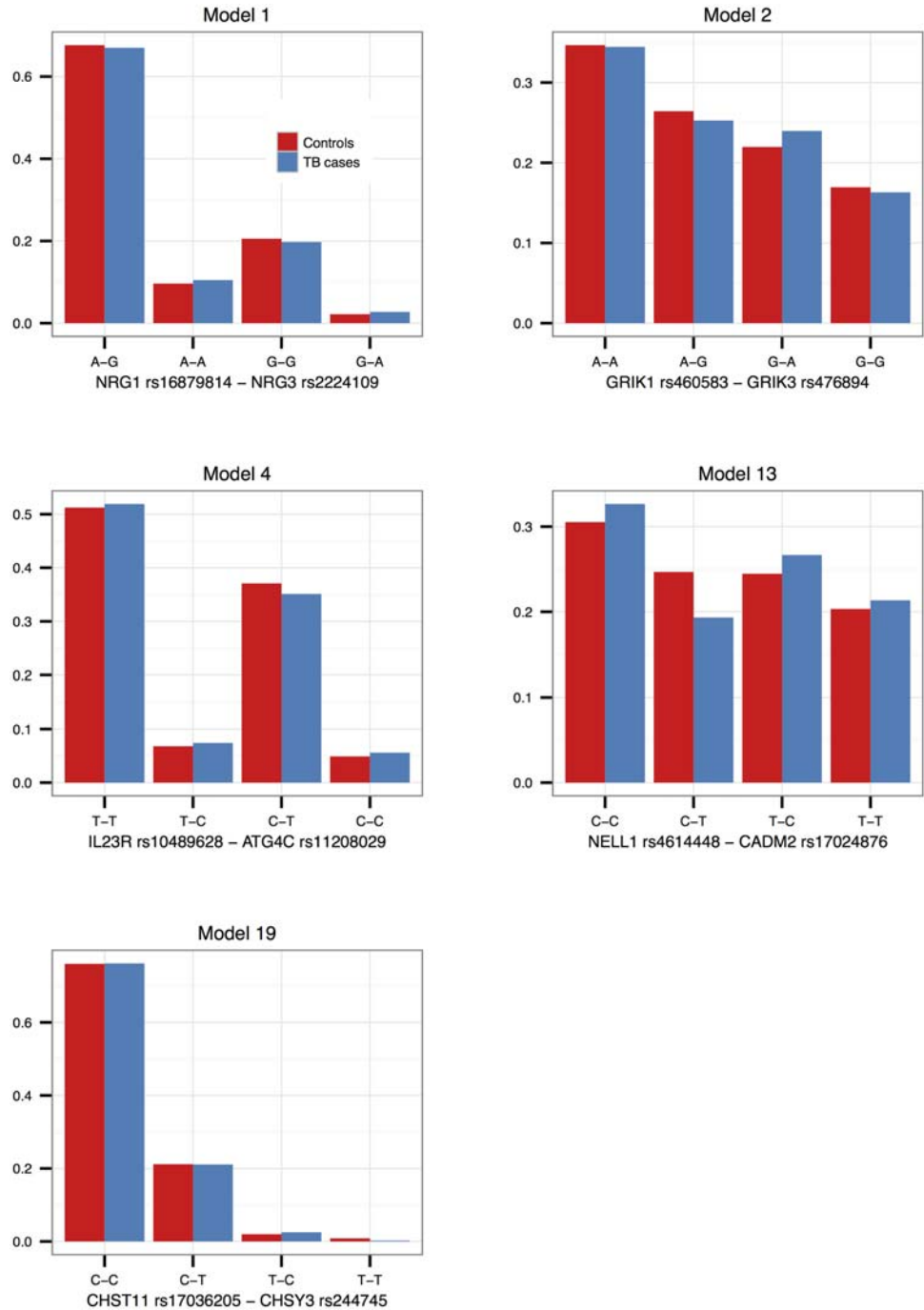
**Fig 3. Effects in the SAC study group.** The logits of genotype combinations from models 1, 2, 3, 4, 7, 13, 14 and 19 are depicted in this figure. Genotypes are ordered according to minor allele frequency, with the wildtype homozygote appearing first, and the rare homozygote appearing last. Non-parallel lines are indicative of interaction effects. The effects were estimated by absorbing the marginal effects of the SNPs into the SNP × SNP interaction term, and adjusting for the covariates included in the model by averaging over them.

doi:10.1371/journal.pone.0123970.g003



**Fig 4. Genotype combination proportions in the Gambian study group.** The observed proportions of the nine possible SNP pair genotype combinations from models 1, 2, 4, 13 and 19 are depicted in this figure, per cases and controls. Genotypes are ordered according to minor allele frequency, with the wildtype homozygote appearing first, and the rare homozygote appearing last.

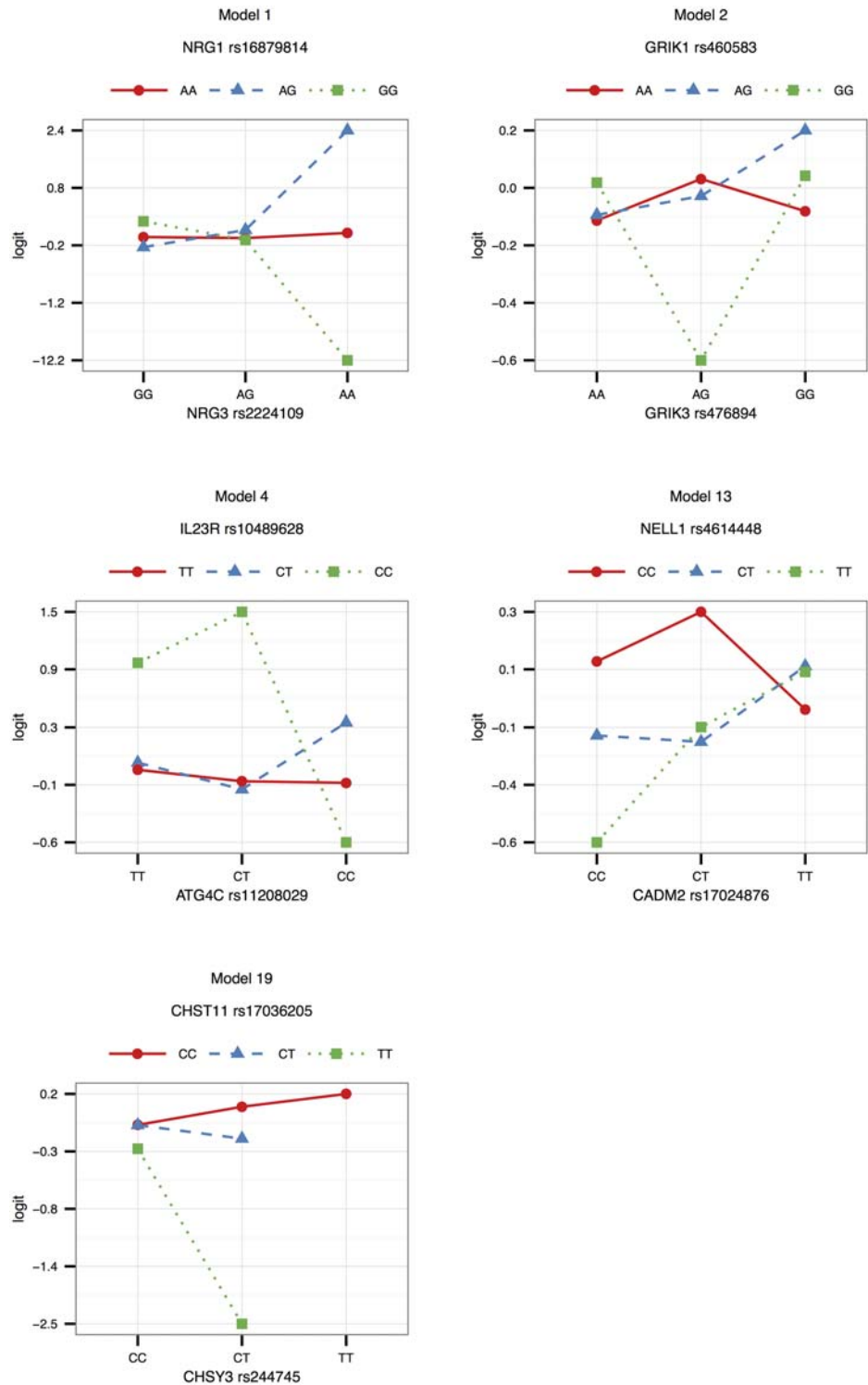
doi:10.1371/journal.pone.0123970.g004



**Fig 5. Allele combination frequencies in the Gambian study group.** The frequencies of the four possible SNP pair allele combinations from models 1, 2, 4, 13 and 19 are depicted in this figure, per cases and controls. The frequencies were estimated using an EM-algorithm.

doi:10.1371/journal.pone.0123970.g005





**Fig 6. Effects in the Gambian study group.** The logits of genotype combinations from models 1, 2, 4, 13 and 19 are depicted in this figure. Genotypes are ordered according to minor allele frequency, with the wildtype homozygote appearing first, and the rare homozygote appearing last. Non-parallel lines are indicative of interaction effects. The effects were estimated by absorbing the marginal effects of the SNPs into the SNP × SNP interaction term, and adjusting for the covariates included in the model by averaging over them.

doi:10.1371/journal.pone.0123970.g006

The *NRG1*—*NRG3* (Neuregulin 1 and 3) interaction effect observed in the SAC study group (model 1, p-value  $8.32 \times 10^{-7}$ ) was also detected in the Gambian study group (p-value 0.0389). The SAC and WTCCC *NRG1* SNP is the same, and the Gambian *NRG3* SNP is located 66 235 base pairs upstream from the SAC *NRG3* SNP. In both the SAC and Gambian study groups, compared to the GG-AC/GG-AG combination, the GG-AA combination decreases the odds of having TB and the AG-AA combination increases the odds of having TB. The same pattern is thus observed, albeit using a different SNP in the second gene in the Gambian study group. Studies investigating the link between *NRG1* and schizophrenia have demonstrated that *NRG1* has a functional effect on the immune system by influencing immune cell adhesion [79] and the concentration of autoantibodies and pro-inflammatory cytokines in plasma [80]. Gene-gene interaction between *NRG1* and *NRG3* has also been observed in a schizophrenia study [81], and according to the NCBI BioSystems database, *NRG3* may also be involved in the immune system.

An interaction between the *GRIK1* and *GRIK3* (glutamate receptor 1 and 2) genes was also detected in both study groups (model 2, SAC p-value  $1.62 \times 10^{-6}$  and Gambian p-value 0.0476). The TT-GG genotype combination was observed only in cases (5%) and the T-G allelic combination is more frequent in cases compared to controls (21% vs. 7%). The Gambian *GRIK1* SNP is located 3 478 base pairs downstream from the SAC *GRIK1* SNP and the Gambian *GRIK3* SNP is located 1 286 base pairs downstream from the SAC *GRIK3* SNP. *GRIK1* has been associated with susceptibility to diabetes [82], and according to T1DBase (a database focused on the genetics and genomics of type 1 diabetes susceptibility, <http://www.t1dbase.org>), *GRIK3* is also a putative diabetes susceptibility gene. Having diabetes increases susceptibility to TB [83], and this may explain the *GRIK1*—*GRIK3* interaction association we observed in the data.

Another model that was observed in both study groups is the interaction between *IL23R* (interleukin-23 receptor) and *ATG4C* (autophagy related 4C, cysteine peptidase) (model 4, SAC p-value  $2.18 \times 10^{-6}$  and Gambian p-value 0.0350). T helper 17 (Th17) cells are subsets of activated CD4+ (cluster of differentiation 4 plus) T cells that mediates the recruitment of macrophages to infected tissues. The Th17 response to *M. tuberculosis* infection is largely dependent on interleukin-23 [84]. *ATG4C* is thought to play a role in autophagy [85] and is up-regulated when TRPV1 (transient receptor potential cation channel subfamily V member 1) channels are expressed on CD4+ T cells [86, 87]. The *IL23R* and *ATG4C* gene products may therefore both be involved in the Th17 response to *M. tuberculosis*. The same SNPs are used in both models, with the CC-CC genotype combination decreasing the odds of having TB in both cohorts, compared to the CC-CT and CC-TT combinations. Both of the SNPs are located on chromosome 1p31 and are 6 centimorgans (4 451 385 base pairs) apart. Linkage disequilibrium between the SNPs is high in SAC controls ( $D' = 0.5451$ ) but not in SAC cases ( $D' = 0.0136$ ), and low in both Gambian controls and cases ( $D' = 0.0011$  and  $D' = 0.0386$  respectively).

Interaction between the *NELL1* (neural epidermal growth factor-like 1) and *CADM2* (cell adhesion molecule 2) genes is also evident in both the SAC and WTCCC study groups (model 13, SAC p-value  $1.14 \times 10^{-5}$  and Gambian p-value 0.0329), as well as interaction between the *NELL1* and *CADM3* (cell adhesion molecule 3) genes, although the latter was not validated in the Gambian study group (model 17, SAC p-value  $1.26 \times 10^{-5}$ ). The same trend between the effects of the heterozygote genotype combinations is observed in both study groups for model 13, compared to the pairing with the wildtype homozygote genotype (CC) of the second SNP in the pair. A large degree of homology exists between the *CADM1*, *CADM2* and *CADM3* genes [88, 89] and *CADM1* has been shown to affect the expression of interleukin-22 [90].

*NELL1* is expressed in pre-B cell development [91] and has been associated with inflammatory bowel disease, a complex auto-immune disorder [92]. The link between the interplay of these genes and TB susceptibility is however not clear. We also note that in the Gambian cohort, the *NELL1* single SNP association signal is stronger than the interaction effect (single SNP p-value of 0.0022, see S5 Table, vs. interaction p-value of 0.0329).

The *CHST11*—*CHSY3* (carbohydrate (chondroitin 4) sulfotransferase 11—chondroitin sulfate synthase 3) gene pair interaction was also detected in both study groups (model 19, SAC p-value  $1.34 \times 10^{-5}$  and Gambian p-value 0.0401). Uhlin et al. [93] showed that expression of *CSPG* (chondroitin sulfate proteoglycan) decreased when monocyte-derived macrophages are treated with interferon-gamma. *CSPG* is composed of a protein core and a chondroitin sulfate side chain. According to the NCBI BioSystems database, both the *CHST11* and *CHSY3* genes are involved in the chondroitin sulfate pathway. The CT-CG genotype combination has a higher frequency in SAC controls compared to cases (10% vs. 3%). The *CHST11* SNP is the same in the SAC and Gambian models, and the *CHSY3* Gambian SNP is 8 683 base pairs upstream from the SAC *CHSY3* SNP. The CT-TT combination was observed in 7 Gambian controls and in 1 case.

NF- $\kappa$ B (nuclear factor kappa-light-chain-enhancer of activated B cells) signalling plays an important role in the host defense against *M. tuberculosis* infection [94]. Both the *SFTPD* (surfactant protein D) and *NOD2* (nucleotide-binding oligomerization domain containing 2) genes are involved in this pathway [95, 96]. Interaction between these genes was identified in the SAC study group (model 3), but no suitable SNPs were available for validation in the Gambian data set. The GG-CT genotype combination is present only in SAC controls (9%), whereas the GG-TT genotype combination is present only in SAC cases (8%).

The *ISG15* (*ISG15* ubiquitin-like modifier) and *TLR8* (Toll-like receptor 8) gene products may also affect NF- $\kappa$ B signalling. *ISG15* stimulates interferon-gamma production [97] which in turn activates NF- $\kappa$ B signalling [98]. It has also been postulated that *TLR8* activates NF- $\kappa$ B signalling [99]. This gene pair showed interaction in the SAC study group (model 7) but could not be validated in the Gambian study group due to lack of suitable SNPs. The CT-GG genotype combination occurs in 7% of SAC cases, but in only 1% of controls.

Another model of interest that could not be validated in the Gambian cohort due to lack of suitable SNPs is the interaction between *NLRC5* (NLR family, CARD domain containing 5) and *IL12RB1* (interleukin 12 receptor, beta 1 genes) (model 14). The interferon-gamma/interleukin-12 pathway is an important component of the immune defense against mycobacterial infections [100]. *IL12RB1* is a receptor of interleukin-12, and it has been shown that the *NLRC5* promoter region is responsive to interferon-gamma, which implies that *NLRC5* may function as a molecular switch of interferon-gamma activation [101]. The TT-AT genotype combination has a higher frequency in SAC controls compared to cases (12% vs. 5%), whereas the TT-TT combination has a higher frequency in SAC cases compared to controls (19% vs. 10%).

Finally, we explored whether allelic encoding of the SNPs may better explain the interactions detected in the SAC cohort. S6 Table summarizes the p-values of the four degrees of freedom genotypic tests for interaction that was used to select the top 20 models, as well as the p-values of the corresponding allelic models that attained the highest level of significance. It is evident from these results that dominant/recessive effects may in some cases better encapsulate the interaction effects observed in the data, and this is depicted in S4 Fig. We note that for all of the five models that were successfully validated in the Gambian cohort, the genotypic test for interaction achieved the highest level of significance.

## Discussion

The South African Coloured population is an ideal cohort for the discovery of TB susceptibility genetic variants, since they received genetic contributions from diverse source populations that may differ in their susceptibility to TB. Seldin et al. [102] has argued that it is important to study the role of complex disease epistasis in such admixed populations, and that this may well uncover novel interactions that are not detectable in the source populations. In this study we used SAC genome-wide data as well as genotypes from a large number of candidate gene studies to discover genetic variants that may jointly modify the odds of having TB. We limited our search space to biologically plausible gene pair models and used statistical modelling to detect interactions, allowing us to adjust for known differences between cases and controls (age, gender and ancestry). Our study does however have a number of limitations, which we discuss below.

Genotypes available for testing the gene pair models were limited to SNPs that were genotyped on the Affymetrix 500K SNP chip as well as candidate gene studies performed in our group. The Affymetrix chip was originally designed based on LD patterns in European populations, and as a result the proportion of variants that are tagged in African populations may be much reduced [103].

Minor allele frequencies of the SNPs representing genes in the top models are in general quite different between the SAC and Gambian cohorts (S5 Table). Of the five models that were successfully validated, only one of the models was validated using exactly the same SNPs, three models were validated with one SNP in common, and one model was validated with completely different SNPs. Patterns of LD are also likely to differ between the SAC and Gambian cohorts, and according to NCBI, none of the SNPs in the top result set have functional effects, implying that the SNPs may all be tagging causative variants. Due to these factors, it is difficult to compare the effect sizes between the two cohorts directly. Indeed, two studies of the association between rs1024611 and TB susceptibility found that the association was statistically significant, but that the G allele of the SNP was protective in the one population, and increased susceptibility in the other population. The true causal variant that rs1024611 was in LD with was later identified, which may explain the opposite effects observed in the two populations [4]. The complexity of disentangling such different effects would be exacerbated in the context of interaction modelling. This could be alleviated to some degree if a higher density of markers was available, which would better capture causative variants. A denser marker panel could be imputed, but in our opinion, this exercise would likely be error-prone. Additional uncertainty would be introduced through imputation, and the proportion of genotype inaccuracies could potentially be large. Imputation relies on linkage disequilibrium between markers, which may not be captured accurately by the Affymetrix 500K SNP for our study cohorts, as a result of the chip's European-centric design. In addition, the San has contributed a large amount of genetic material to the SAC [61, 104–108], and due to the lack of large high density reference panels for the San, this may contribute to additional inaccuracies in imputation of the SAC data set.

It is difficult to quantify the precise levels of significance of our results, due to the large number of tests in the SAC data set, and the limitations of methods available to correct for multiple testing. If we were to use a multiple testing correction similar to the one used by Emily et al. [55], despite its limitations, we would have to show that the number of effective independent tests was 60 000, for the topmost model to be significant at a Bonferroni adjusted alpha level of 0.05 ( $0.05/60000 = 8.33 \times 10^{-7}$ , p-value of model 1 was  $8.33 \times 10^{-7}$ ). Given that roughly 2 million models were tested, and that Emily et al. found that the number of effective independent tests was approximately six times less than the actual number of tests, it is unlikely that we would be able to demonstrate this (a 33 times reduction would be required). The SAC genome-

wide data set was originally genotyped with a view to perform a case-only admixture mapping study, and for this reason, a limited number of controls was available for many of the two SNP interaction tests. Whilst the available group sizes are sufficient to detect two-SNP interactions, it is unlikely that any of the results would achieve statistical significance. For validation models fitted to the Gambian data set, we describe tests with  $p$ -values  $< 0.05$  as statistically significant. As 20 gene pair tests were done (ranging between 2 to 90 SNP pair tests per gene pair, see [Table 1](#)), none of these results would survive correction for multiple testing, and we note that the results should be interpreted with caution. We do however argue that both the SAC and Gambian data sets do not merely constitute random data, and that our results may contain actual associations that should not be dismissed [17]. Given the complex nature of the immune system defence against TB and the role that gene-gene interactions might play in this, it is plausible that some of our top results represent real biological phenomena worthy of further investigation.

Seven of the top models identified in the SAC study group could not be tested in the Gambian study group due to the absence of suitable SNPs. Eight of the models were not successfully validated. These results could be false positives, but their validation failure could also be ascribed to a number of other reasons. Differing patterns of LD between the SAC and Gambian populations and lack of SNP coverage by the Affymetrix 500K SNP chip, which our SNP selection strategy could not fully compensate for, could result in unsuccessful validation. The *M. tuberculosis* genome varies substantially across geographic regions [109], including between South and West Africa, and it has been hypothesised that interactions between host and pathogen differ between population groups [8, 109, 110]. Due to the heterogeneity of the source populations that contributed to the formation of the SAC, it is also possible that some interactions involved in TB susceptibility are unique to the SAC [102]. In spite of the limitations discussed above, five of the top twenty models were indeed validated in an independent Gambian case-control data set, although the levels of significance of the validation models were not very small ( $p$ -value  $< 0.05$  but  $> 0.01$ ). These models indicate that TB susceptibility is modified by interplay between the *NRG1—NRG3*, *GRIK1—GRIK3* and *IL23R—ATG4C* gene pairs, and the fact that the validation population is ethnically very different could imply that the interactions found have universal relevance.

The frequencies and effects are depicted graphically to aid interpretation, but as the SNPs used in the models are tag variants that may not have causative functional effects, the biological implications of the models are not yet fully understood. Validation in other populations, fine-mapping of the causal variants and functional studies will be required to elucidate our findings.

## Conclusion

In this study we investigated the role of gene-gene interactions in TB susceptibility in the South African Coloured population. To our knowledge, in terms of number of genetic loci considered, this is the largest study of gene-gene interactions and TB susceptibility that has been reported to date. We report a number of interesting results, five of which were validated in an independent cohort from the Gambia.

## Supporting Information

**S1 Fig. Genotype combination proportions in the SAC study group.** The observed proportions of the nine possible SNP pair genotype combinations from models 5, 6, 8, 9, 10, 11, 12, 15, 16, 17, 18 and 20 are depicted in this figure, per cases and controls. Genotypes are ordered according to minor allele frequency, with the wildtype homozygote appearing first, and the



rare homozygote appearing last.  
(PDF)

**S2 Fig. Allele combination frequencies in the SAC study group.** The frequencies of the four possible SNP pair allele combinations from models 5, 6, 8, 9, 10, 11, 12, 15, 16, 17, 18 and 20 are depicted in this figure, per cases and controls. The frequencies were estimated using an EM-algorithm.  
(PDF)

**S3 Fig. Effects in the SAC study group.** The logits of genotype combinations from models 5, 6, 8, 9, 10, 11, 12, 15, 16, 17, 18 and 20 are depicted in this figure. Genotypes are ordered according to minor allele frequency, with the wildtype homozygote appearing first, and the rare homozygote appearing last. Non-parallel lines are indicative of interaction effects. The effects were estimated by absorbing the marginal effects of the SNPs into the SNP  $\times$  SNP interaction term, and adjusting for the covariates included in the model by averaging over them.  
(PDF)

**S4 Fig. Dominant/recessive combination proportions in the SAC study group.** The observed proportions of SNP pair genotype combinations from models 3, 5, 7, 8, 9, 16, 17 and 18 are depicted in this figure, per cases and controls. Recessive/dominant effects in these models may better explain the interactions observed in the cohort (smaller p-values were achieved compared to the genotypic models, and the best models with 1 or more recessive or dominant encodings listed in [S5 Table](#) are presented in this figure). Rare homozygotes and heterozygotes are combined to represent dominant encoding of alleles, and wild type homozygotes and heterozygotes are combined to represent recessive encoding of alleles. For dominant and recessive allelic encodings of SNPs, the last genotype presented therefore reflects an encoding of 1.  
(PDF)

**S1 Table. TB susceptibility candidate gene association studies.** The table summarizes the total number of samples that were successfully genotyped in each candidate gene study and how many samples have complete confounder information (age, gender and ancestry).  
(PDF)

**S2 Table. Age and gender in the tuberculosis study groups.** P-values were calculated using logistic regression.  
(PDF)

**S3 Table. Software used in this study.** A summary listing web URLs, version information and important parameter settings of software used in this study.  
(PDF)

**S4 Table. Results of the top 250 SAC models and all 245 Gambian models.** A spreadsheet with two worksheets, showing the results of the top 250 SAC models and the 245 Gambian models that were used for validation.  
(XLS)

**S5 Table. Single SNP summary of the top model SNPs in the SAC and Gambian cohorts.** This table provides a summary of each SNP's individual minor allele frequency (MAF) and association with having TB.  
(PDF)

**S6 Table. P-values of the top models in the SAC cohort.** The genotypic model p-values, which were used to select the top 20 models, are presented in this table. The p-values of the

corresponding allelic interaction models that achieved the smallest p-values are also shown. (PDF)

## Acknowledgments

We thank all participants and field workers in this study. We also thank the developers of the open source software we used in our analyses.

## Author Contributions

Conceived and designed the experiments: MM PDVH EGH. Performed the experiments: MD. Analyzed the data: MD LVDM. Contributed reagents/materials/analysis tools: PDVH. Wrote the paper: MD LVDM EGH.

## References

1. WHO. Global tuberculosis report. World Health Organization (WHO); 2013. 2013.11.
2. MRC. Revised burden of disease estimates for the comparative risk factor assessment, South Africa 2000. Medical Research Unit, South Africa; 2006. June 2006.
3. Hoal EG, Lewis LA, Jamieson SE, Tanzer F, Rossouw M, Victor T, et al. SLC11A1 (NRAMP1) but not SLC11A2 (NRAMP2) polymorphisms are associated with susceptibility to tuberculosis in a high-incidence community in South Africa. *The International Journal of Tuberculosis and Lung Disease*. 2004; 8(12):1464–1471. PMID: [15636493](#)
4. Möller M, Hoal EG. Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. *Tuberculosis*. 2010; 90(2):71–83. doi: [10.1016/j.tube.2010.02.002](#) PMID: [20206579](#)
5. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics*. 2010; 11(6):446–450. doi: [10.1038/nrg2809](#) PMID: [20479774](#)
6. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics*. 2009; 10(4):241–251. doi: [10.1038/nrg2554](#) PMID: [19293820](#)
7. Wei Z, Wang K, Qu HQ, Zhang H, Bradfield J, Kim C, et al. From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. *PLoS Genetics*. 2009; 5(10):e1000678. doi: [10.1371/journal.pgen.1000678](#) PMID: [19816555](#)
8. Stein CM. Genetic epidemiology of tuberculosis susceptibility: impact of study design. *PLoS Pathogens*. 2011; 7(1):e1001189. doi: [10.1371/journal.ppat.1001189](#) PMID: [21283783](#)
9. Velez DR, Hulme WF, Myers JL, Stryjewski ME, Abbate E, Estevan R, et al. Association of SLC11A1 with tuberculosis interactions with NOS2A and TLR2 in African-Americans and Caucasians. *The International Journal of Tuberculosis and Lung Disease*. 2009; 13(9):1068. PMID: [19723394](#)
10. Flores-Villanueva PO, Ruiz-Morales JA, Song CH, Flores LM, Jo EK, Montañó M, et al. A functional promoter polymorphism in monocyte chemoattractant protein-1 is associated with increased susceptibility to pulmonary tuberculosis. *The Journal of Experimental Medicine*. 2005; 202(12):1649–1658. doi: [10.1084/jem.20050126](#) PMID: [16352737](#)
11. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*. 2012; 109(4):1193–1198. doi: [10.1073/pnas.1119675109](#)
12. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nature Reviews Genetics*. 2009; 10(6):392–404. doi: [10.1038/nrg2579](#) PMID: [19434077](#)
13. Greene CS, Penrod NM, Williams SM, Moore JH. Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS ONE*. 2009; 4(6):e5639. doi: [10.1371/journal.pone.0005639](#) PMID: [19503614](#)
14. Phillips PC. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*. 2008; 9(11):855. doi: [10.1038/nrg2452](#) PMID: [18852697](#)
15. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics*. 2002; 11(20):2463–2468. doi: [10.1093/hmg/11.20.2463](#) PMID: [12351582](#)
16. Cordell HJ, Clayton DG. Genetic association studies. *The Lancet*. 2005; 366(9491):1121–1131. doi: [10.1016/S0140-6736\(05\)67424-7](#)

17. Rothman KJ. Six Persistent Research Misconceptions. *Journal of General Internal Medicine*. 2014;p. 1–5.
18. Beltrao P, Cagney G, Krogan NJ. Quantitative genetic interactions reveal biological modularity. *Cell*. 2010; 141(5):739–745. doi: [10.1016/j.cell.2010.05.019](https://doi.org/10.1016/j.cell.2010.05.019) PMID: [20510918](https://pubmed.ncbi.nlm.nih.gov/20510918/)
19. Brem RB, Storey JD, Whittle J, Kruglyak L. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature*. 2005; 436(7051):701–703. doi: [10.1038/nature03865](https://doi.org/10.1038/nature03865) PMID: [16079846](https://pubmed.ncbi.nlm.nih.gov/16079846/)
20. Chou HH, Chiu HC, Delaney NF, Segrè D, Marx CJ. Diminishing returns epistasis beneficial mutations decelerates adaptation. *Science*. 2011; 332(6034):1190–1192. doi: [10.1126/science.1203799](https://doi.org/10.1126/science.1203799) PMID: [21636771](https://pubmed.ncbi.nlm.nih.gov/21636771/)
21. Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C. Systematic mapping of genetic interaction networks. *Annual Review of Genetics*. 2009; 3:601–625. doi: [10.1146/annurev.genet.39.073003.114751](https://doi.org/10.1146/annurev.genet.39.073003.114751)
22. Ferguson EL, Horvitz HR. The multivulva phenotype of certain *Caenorhabditis elegans* mutants results from defects in two functionally redundant pathways. *Genetics*. 1989; 123(1):109–121. PMID: [2806880](https://pubmed.ncbi.nlm.nih.gov/2806880/)
23. Huang W, Richards S, Carbone MA, Zhu D, Anholt RR, Ayroles JF, et al. Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proceedings of the National Academy of Sciences*. 2012; 109(39):15553–15559. doi: [10.1073/pnas.1213423109](https://doi.org/10.1073/pnas.1213423109)
24. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*. 2011; 332(6034):1193–1196. doi: [10.1126/science.1203801](https://doi.org/10.1126/science.1203801) PMID: [21636772](https://pubmed.ncbi.nlm.nih.gov/21636772/)
25. Lindén RO, Eronen VP, Aittokallio T. Quantitative maps of genetic interactions in yeast-Comparative evaluation and integrative analysis. *BMC Systems Biology*. 2011; 5(1):45. doi: [10.1186/1752-0509-5-45](https://doi.org/10.1186/1752-0509-5-45) PMID: [21435228](https://pubmed.ncbi.nlm.nih.gov/21435228/)
26. Okser S, Pahikkala T, Aittokallio T. Genetic variants and their interactions in disease risk prediction-machine learning and network perspectives. *BioData Mining*. 2013; 6:5. doi: [10.1186/1756-0381-6-5](https://doi.org/10.1186/1756-0381-6-5) PMID: [23448398](https://pubmed.ncbi.nlm.nih.gov/23448398/)
27. Shao H, Burrage LC, Sinasac DS, Hill AE, Ernest SR, O'Brien W, et al. Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proceedings of the National Academy of Sciences*. 2008; 105(50):19910–19914. doi: [10.1073/pnas.0810388105](https://doi.org/10.1073/pnas.0810388105)
28. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nature Reviews Genetics*. 2010; 11(1):17–30. doi: [10.1038/nrg2698](https://doi.org/10.1038/nrg2698) PMID: [19953080](https://pubmed.ncbi.nlm.nih.gov/19953080/)
29. Collins RL, Hu T, Wejse C, Sirugo G, Williams SM, Moore JH. Multifactor dimensionality reduction reveals a three-locus epistatic interaction associated with susceptibility to pulmonary tuberculosis. *BioData Mining*. 2013; 6(4). doi: [10.1186/1756-0381-6-4](https://doi.org/10.1186/1756-0381-6-4) PMID: [23418869](https://pubmed.ncbi.nlm.nih.gov/23418869/)
30. De Wit E, van der Merwe L, van Helden P, Hoal E. Gene-gene interaction between tuberculosis candidate genes in a South African population. *Mammalian Genome*. 2010;p. 1–11.
31. Edwards DRV, Tacconelli A, Wejse C, Hill PC, Morris GA, Edwards TL, et al. MCP1 SNPs and pulmonary tuberculosis in cohorts from West Africa, the USA and Argentina: lack of association or epistasis with IL12B polymorphisms. *PloS ONE*. 2012; 7(2):e32275. doi: [10.1371/journal.pone.0032275](https://doi.org/10.1371/journal.pone.0032275) PMID: [22384203](https://pubmed.ncbi.nlm.nih.gov/22384203/)
32. Motsinger-Reif AA, Antas PR, Oki NO, Levy S, Holland SM, Sterling TR. Polymorphisms in IL-1 $\beta$ , vitamin D receptor Fok1, and Toll-like receptor 2 are associated with extrapulmonary tuberculosis. *BMC Medical Genetics*. 2010; 11(1):37. doi: [10.1186/1471-2350-11-37](https://doi.org/10.1186/1471-2350-11-37) PMID: [20196868](https://pubmed.ncbi.nlm.nih.gov/20196868/)
33. Olesen R, Wejse C, Velez DR, Bisseye C, Sodemann M, Aaby P, et al. DC-SIGN (CD209), pentraxin 3 and vitamin D receptor gene variants associate with pulmonary tuberculosis risk in West Africans. *Genes and Immunity*. 2007; 8(6):456–467. doi: [10.1038/sj.gene.6364410](https://doi.org/10.1038/sj.gene.6364410) PMID: [17611589](https://pubmed.ncbi.nlm.nih.gov/17611589/)
34. Ravikumar M, Dheenadhayalan V, Rajaram K, Shanmuga Lakshmi S, Paul Kumaran P, Paramasivan CN, et al. Associations of HLA-DRB1, DQB1 and DPB1 alleles with pulmonary tuberculosis in south India. *Tubercle and Lung Disease*. 1999; 79(5):309–317. doi: [10.1054/tuld.1999.0213](https://doi.org/10.1054/tuld.1999.0213) PMID: [10707259](https://pubmed.ncbi.nlm.nih.gov/10707259/)
35. White MJ, Tacconelli A, Chen JS, Wejse C, Hill PC, Gomes VF, et al. Epiregulin (EREG) and human V-ATPase (TCIRG1): genetic variation, ethnicity and pulmonary tuberculosis susceptibility in Guinea-Bissau and The Gambia. *Genes and Immunity*. 2014;. doi: [10.1038/gene.2014.28](https://doi.org/10.1038/gene.2014.28) PMID: [24898387](https://pubmed.ncbi.nlm.nih.gov/24898387/)
36. Brinza D, Schultz M, Tesler G, Bafna V. RAPID detection of gene-gene interactions in genome-wide association studies. *Bioinformatics*. 2010 Nov; 26(22):2856–2862. doi: [10.1093/bioinformatics/btq529](https://doi.org/10.1093/bioinformatics/btq529) PMID: [20871107](https://pubmed.ncbi.nlm.nih.gov/20871107/)

37. Chen G, Yuan A, Zhou J, Bentley AR, Adeyemo A, Rotimi CN. Simple F Test Reveals Gene-Gene Interactions in Case-Control Studies. *Bioinformatics and Biology Insights*. 2012; 6:169. doi: [10.4137/BBI.S9867](https://doi.org/10.4137/BBI.S9867) PMID: [22837643](https://pubmed.ncbi.nlm.nih.gov/22837643/)
38. Hu T, Sinnott-Armstrong NA, Kiralis JW, Andrew AS, Karagas MR, Moore JH. Characterizing genetic interactions in human disease association studies using statistical epistasis networks. *BMC Bioinformatics*. 2011; 12(1):364. doi: [10.1186/1471-2105-12-364](https://doi.org/10.1186/1471-2105-12-364) PMID: [21910885](https://pubmed.ncbi.nlm.nih.gov/21910885/)
39. Kam-Thong T, Czamara D, Tsuda K, Borgwardt K, Lewis CM, Erhardt-Lehmann A, et al. EPIBLA-STER-fast exhaustive two-locus epistasis detection strategy using graphical processing units. *European Journal of Human Genetics*. 2011 Apr; 19(4):465–471. doi: [10.1038/ejhg.2010.196](https://doi.org/10.1038/ejhg.2010.196) PMID: [21150885](https://pubmed.ncbi.nlm.nih.gov/21150885/)
40. Liu Y, Xu H, Chen S, Chen X, Zhang Z, Zhu Z, et al. Genome-wide interaction-based association analysis identified multiple new susceptibility loci for common diseases. *PLoS Genetics*. 2011; 7(3): e1001338. doi: [10.1371/journal.pgen.1001338](https://doi.org/10.1371/journal.pgen.1001338) PMID: [21437271](https://pubmed.ncbi.nlm.nih.gov/21437271/)
41. McKinney BA, Crowe JE Jr, Guo J, Tian D. Capturing the spectrum of interaction effects in genetic association studies by simulated evaporative cooling network analysis. *PLoS Genetics*. 2009; 5(3): e1000432. doi: [10.1371/journal.pgen.1000432](https://doi.org/10.1371/journal.pgen.1000432) PMID: [19300503](https://pubmed.ncbi.nlm.nih.gov/19300503/)
42. Moore JH, Williams SM. New strategies for identifying gene-gene interactions in hypertension. *Annals of Medicine*. 2002; 34(2):88–95. doi: [10.1080/07853890252953473](https://doi.org/10.1080/07853890252953473) PMID: [12108579](https://pubmed.ncbi.nlm.nih.gov/12108579/)
43. Moore JH, White BC. Tuning ReliefF for genome-wide genetic analysis. In: *Evolutionary computation, machine learning and data mining in bioinformatics*. Springer; 2007. p. 166–175.
44. Motsinger-Reif AA, Dudek SM, Hahn LW, Ritchie MD. Comparison of approaches for machine learning optimization of neural networks for detecting gene-gene interactions in genetic epidemiology. *Genetic Epidemiology*. 2008; 32(4):325–340. doi: [10.1002/gepi.20307](https://doi.org/10.1002/gepi.20307) PMID: [18265411](https://pubmed.ncbi.nlm.nih.gov/18265411/)
45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*. 2007; 81(3):559–575. doi: [10.1086/519795](https://doi.org/10.1086/519795)
46. Robnik-Šikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*. 2003; 53(1–2):23–69. doi: [10.1023/A:1025667309714](https://doi.org/10.1023/A:1025667309714)
47. Turner SD, Dudek SM, Ritchie MD. ATHENA: A knowledge-based hybrid backpropagation-grammatical evolution neural network algorithm for discovering epistasis among quantitative trait Loci. *BioData Mining*. 2010; 3(1):5. doi: [10.1186/1756-0381-3-5](https://doi.org/10.1186/1756-0381-3-5) PMID: [20875103](https://pubmed.ncbi.nlm.nih.gov/20875103/)
48. Ueki M, Cordell HJ. Improved Statistics for Genome-Wide Interaction Analysis. *PLoS Genetics*. 2012 Apr; 8(4):e1002625. doi: [10.1371/journal.pgen.1002625](https://doi.org/10.1371/journal.pgen.1002625) PMID: [22496670](https://pubmed.ncbi.nlm.nih.gov/22496670/)
49. Wellek S, Ziegler A. A genotype-based approach to assessing the association between single nucleotide polymorphisms. *Human Heredity*. 2008; 67(2):128–139. doi: [10.1159/000179560](https://doi.org/10.1159/000179560) PMID: [19077429](https://pubmed.ncbi.nlm.nih.gov/19077429/)
50. Wu X, Dong H, Luo L, Zhu Y, Peng G, Reveille JD, et al. A novel statistic for genome-wide interaction analysis. *PLoS Genetics*. 2010; 6(9):e1001131. doi: [10.1371/journal.pgen.1001131](https://doi.org/10.1371/journal.pgen.1001131) PMID: [20885795](https://pubmed.ncbi.nlm.nih.gov/20885795/)
51. Zhang Y, Liu JS. Bayesian inference of epistatic interactions in case-control studies. *Nature Genetics*. 2007; 39(9):1167–1173. doi: [10.1038/ng2110](https://doi.org/10.1038/ng2110) PMID: [17721534](https://pubmed.ncbi.nlm.nih.gov/17721534/)
52. Zhang X, Huang S, Zou F, Wang W. Tools for efficient epistasis detection in genome-wide association study. *Source Code for Biology and Medicine*. 2011; 6(1):1. doi: [10.1186/1751-0473-6-1](https://doi.org/10.1186/1751-0473-6-1) PMID: [21205316](https://pubmed.ncbi.nlm.nih.gov/21205316/)
53. Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature Genetics*. 2005; 37(4):413–417. doi: [10.1038/ng1537](https://doi.org/10.1038/ng1537) PMID: [15793588](https://pubmed.ncbi.nlm.nih.gov/15793588/)
54. Bush WS, Dudek SM, Ritchie MD. Biofilter: A Knowledge-Integration System for the Multi-Locus Analysis of Genome-Wide Association Studies. *Pacific Symposium on Biocomputing*. 2009;p. 368–379. PMID: [19209715](https://pubmed.ncbi.nlm.nih.gov/19209715/)
55. Emily M, Mailund T, Hein J, Schausser L, Schierup MH. Using biological networks to search for interacting loci in genome-wide association studies. *European Journal of Human Genetics*. 2009; 17(10):1231–1240. doi: [10.1038/ejhg.2009.15](https://doi.org/10.1038/ejhg.2009.15) PMID: [19277065](https://pubmed.ncbi.nlm.nih.gov/19277065/)
56. Ma L, Brautbar A, Boerwinkle E, Sing CF, Clark AG, Keinan A. Knowledge-driven analysis identifies a gene-gene interaction affecting high-density lipoprotein cholesterol levels in multi-ethnic populations. *PLoS Genetics*. 2012; 8(5):e1002714. doi: [10.1371/journal.pgen.1002714](https://doi.org/10.1371/journal.pgen.1002714) PMID: [22654671](https://pubmed.ncbi.nlm.nih.gov/22654671/)
57. Pattin KA, Moore JH. Exploiting the Proteome to Improve the Genome-Wide Genetic Analysis of Epistasis in Common Human Diseases. *Human Genetics*. 2008 Aug; 124(1):19–29. doi: [10.1007/s00439-008-0522-8](https://doi.org/10.1007/s00439-008-0522-8) PMID: [18551320](https://pubmed.ncbi.nlm.nih.gov/18551320/)

58. Daya M, van der Merwe L, van Helden PD, Miller M, Hoal EG. The role of ancestry in TB susceptibility of an admixed South African population. *Tuberculosis*. 2014; 94(4):413–420. doi: [10.1016/j.tube.2014.03.012](https://doi.org/10.1016/j.tube.2014.03.012) PMID: [24832562](https://pubmed.ncbi.nlm.nih.gov/24832562/)
59. Gallant CJ, Cobat A, Simkin L, Black GF, Stanley K, Hughes J, et al. Impact of age and sex on mycobacterial immunity in an area of high tuberculosis incidence. *The International Journal of Tuberculosis and Lung Disease*. 2010; 14(8):952–959. PMID: [20626938](https://pubmed.ncbi.nlm.nih.gov/20626938/)
60. Thye T, Owusu-Dabo E, Vannberg FO, van Crevel R, Curtis J, Sahiratmadja E, et al. Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nature Genetics*. 2012; 44(3):257–259. doi: [10.1038/ng.1080](https://doi.org/10.1038/ng.1080) PMID: [22306650](https://pubmed.ncbi.nlm.nih.gov/22306650/)
61. De Wit E, Delport W, Rugamika CE, Meintjes A, Möller M, van Helden PD, et al. Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Human Genetics*. 2010; 128(2):145–153. doi: [10.1007/s00439-010-0836-1](https://doi.org/10.1007/s00439-010-0836-1) PMID: [20490549](https://pubmed.ncbi.nlm.nih.gov/20490549/)
62. Chimusa ER, Zaitlen N, Daya M, Möller M, van Helden PD, Mulder NJ, et al. Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. *Human Molecular Genetics*. 2013; 23(3):796–809. doi: [10.1093/hmg/ddt462](https://doi.org/10.1093/hmg/ddt462) PMID: [24057671](https://pubmed.ncbi.nlm.nih.gov/24057671/)
63. Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nature Genetics*. 2010; 42(9):739–741. doi: [10.1038/ng.639](https://doi.org/10.1038/ng.639) PMID: [20694014](https://pubmed.ncbi.nlm.nih.gov/20694014/)
64. Fox J. Effect displays in R for generalised linear models. *Journal of Statistical Software*. 2003; 8(15):1–27.
65. Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic Epidemiology*. 2010; 34(6):591–602. doi: [10.1002/gepi.20516](https://doi.org/10.1002/gepi.20516) PMID: [20718045](https://pubmed.ncbi.nlm.nih.gov/20718045/)
66. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006; 38(8):904–909. doi: [10.1038/ng1847](https://doi.org/10.1038/ng1847) PMID: [16862161](https://pubmed.ncbi.nlm.nih.gov/16862161/)
67. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *The American Journal of Human Genetics*. 2004; 74(4):765–769. doi: [10.1086/383251](https://doi.org/10.1086/383251)
68. Gao X, Becker LC, Becker DM, Starmer JD, Province MA. Avoiding the high Bonferroni penalty in genome-wide association studies. *Genetic Epidemiology*. 2010; 34(1):100–105. doi: [10.1002/gepi.20430](https://doi.org/10.1002/gepi.20430) PMID: [19434714](https://pubmed.ncbi.nlm.nih.gov/19434714/)
69. Bžková P, Lumley T, Rice K. Permutation and parametric bootstrap tests for gene-gene and gene-environment interactions. *Annals of Human Genetics*. 2011; 75(1):36–45. doi: [10.1111/j.1469-1809.2010.00572.x](https://doi.org/10.1111/j.1469-1809.2010.00572.x) PMID: [20384625](https://pubmed.ncbi.nlm.nih.gov/20384625/)
70. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*. 2005 Sep; 95(3):221–227. doi: [10.1038/sj.hdy.6800717](https://doi.org/10.1038/sj.hdy.6800717) PMID: [16077740](https://pubmed.ncbi.nlm.nih.gov/16077740/)
71. Kerr KF. Comments on the analysis of unbalanced microarray data. *Bioinformatics*. 2009; 25(16):2035. doi: [10.1093/bioinformatics/btp363](https://doi.org/10.1093/bioinformatics/btp363) PMID: [19528084](https://pubmed.ncbi.nlm.nih.gov/19528084/)
72. Shriner D, Adeyemo A, Gerry NP, Herbert A, Chen G, Doumatey A, et al. Transferability and fine-mapping of genome-wide associated loci for adult height across human populations. *PLoS ONE*. 2009; 4(12):e8398. doi: [10.1371/journal.pone.0008398](https://doi.org/10.1371/journal.pone.0008398) PMID: [20027299](https://pubmed.ncbi.nlm.nih.gov/20027299/)
73. Ramos E, Chen G, Shriner D, Doumatey A, Gerry NP, Herbert A, et al. Replication of genome-wide association studies (GWAS) loci for fasting plasma glucose in African-Americans. *Diabetologia*. 2011; 54(4):783–788. doi: [10.1007/s00125-010-2002-7](https://doi.org/10.1007/s00125-010-2002-7) PMID: [21188353](https://pubmed.ncbi.nlm.nih.gov/21188353/)
74. Alexander D, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. 2009; 19(9):1655–1664. doi: [10.1101/gr.094052.109](https://doi.org/10.1101/gr.094052.109) PMID: [19648217](https://pubmed.ncbi.nlm.nih.gov/19648217/)
75. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2013. ISBN 3-900051-07-0. Available from: <http://www.R-project.org/>.
76. Warnes G, with contributions from Gregor Gorjanc, Leisch F, Man M. genetics: Population Genetics; 2012. R package version 1.3.8. Available from: <http://CRAN.R-project.org/package=genetics>.
77. Sinnwell J, Schaid D. haplo.stats: Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous; 2013. R package version 1.6.8. Available from: <http://CRAN.R-project.org/package=haplo.stats>.
78. Wickham H. ggplot2: elegant graphics for data analysis. Springer New York; 2009. Available from: <http://had.co.nz/ggplot2/book>.
79. Kanakry CG, Li Z, Nakai Y, Sei Y, Weinberger DR. Neuregulin-1 regulates cell adhesion via an ErbB2/phosphoinositide-3 kinase/Akt-dependent pathway: potential implications for schizophrenia and cancer. *PLoS ONE*. 2007; 2(12):e1369. doi: [10.1371/journal.pone.0001369](https://doi.org/10.1371/journal.pone.0001369) PMID: [18159252](https://pubmed.ncbi.nlm.nih.gov/18159252/)



80. Marballi K, Quinones MP, Jimenez F, Escamilla MA, Raventós H, Soto-Bernardini MC, et al. In vivo and in vitro genetic evidence of involvement of neuregulin 1 in immune system dysregulation. *Journal of Molecular Medicine*. 2010; 88(11):1133–1141. doi: [10.1007/s00109-010-0653-y](https://doi.org/10.1007/s00109-010-0653-y) PMID: [20625696](https://pubmed.ncbi.nlm.nih.gov/20625696/)
81. Benzel I, Bansal A, Browning BL, Galwey NW, Maycox PR, McGinnis R, et al. Interactions among genes in the ErbB-Neuregulin signalling network are associated with increased susceptibility to schizophrenia. *Behavioral and Brain Functions*. 2007;3:31. doi: [10.1186/1744-9081-3-31](https://doi.org/10.1186/1744-9081-3-31) PMID: [17598910](https://pubmed.ncbi.nlm.nih.gov/17598910/)
82. Hayes MG, Pluzhnikov A, Miyake K, Sun Y, Ng MC, Roe CA, et al. Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. *Diabetes*. 2007; 56(12):3033–3044. doi: [10.2337/db07-0482](https://doi.org/10.2337/db07-0482) PMID: [17846124](https://pubmed.ncbi.nlm.nih.gov/17846124/)
83. Jeon CY, Murray MB. Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies. *PLoS Medicine*. 2008; 5(7):e152. doi: [10.1371/journal.pmed.0050152](https://doi.org/10.1371/journal.pmed.0050152) PMID: [18630984](https://pubmed.ncbi.nlm.nih.gov/18630984/)
84. Khader SA, Cooper AM. IL-23 and IL-17 in tuberculosis. *Cytokine*. 2008; 41(2):79–83. doi: [10.1016/j.cyto.2007.11.022](https://doi.org/10.1016/j.cyto.2007.11.022) PMID: [18218322](https://pubmed.ncbi.nlm.nih.gov/18218322/)
85. Songane M, Kleinnijenhuis J, Alisjahbana B, Sahiratmadja E, Parwati I, Oosting M, et al. Polymorphisms in autophagy genes and susceptibility to tuberculosis. *PloS ONE*. 2012; 7(8):e41618. doi: [10.1371/journal.pone.0041618](https://doi.org/10.1371/journal.pone.0041618) PMID: [22879892](https://pubmed.ncbi.nlm.nih.gov/22879892/)
86. Wenning AS, Neblung K, Straub B, Wolfs MJ, Sappok A, Hoth M, et al. TRP expression pattern and the functional importance of TRPC3 in primary human T-cells. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2011; 1813(3):412–423. doi: [10.1016/j.bbamcr.2010.12.022](https://doi.org/10.1016/j.bbamcr.2010.12.022)
87. McLeod IX, He YW. Editorial: TRPV1: how thymocytes sense stress and respond with autophagy. *Journal of Leukocyte Biology*. 2012; 92(3):409–411. doi: [10.1189/jlb.0612269](https://doi.org/10.1189/jlb.0612269) PMID: [22936836](https://pubmed.ncbi.nlm.nih.gov/22936836/)
88. Fukami T, Satoh H, Williams YN, Masuda M, Fukuhara H, Maruyama T, et al. Isolation of the mouse Tsl1 and Tsl2 genes, orthologues of the human TSLC1-like genes 1 and 2 (TSL1 and TSL2). *Gene*. 2003; 323:11–18. doi: [10.1016/j.gene.2003.09.018](https://doi.org/10.1016/j.gene.2003.09.018) PMID: [14659875](https://pubmed.ncbi.nlm.nih.gov/14659875/)
89. Fukuhara H, Kuramochi M, Nobukuni T, Fukami T, Saino M, Maruyama T, et al. Isolation of the TSL1 and TSL2 genes, members of the tumor suppressor TSLC1 gene family encoding transmembrane proteins. *Oncogene*. 2001; 20(38):5401–5407. doi: [10.1038/sj.onc.1204696](https://doi.org/10.1038/sj.onc.1204696) PMID: [11536053](https://pubmed.ncbi.nlm.nih.gov/11536053/)
90. Galibert L, Diemer GS, Liu Z, Johnson RS, Smith JL, Walzer T, et al. Nectin-like protein 2 defines a subset of T-cell zone dendritic cells and is a ligand for class-I-restricted T-cell-associated molecule. *Journal of Biological Chemistry*. 2005; 280(23):21955–21964. doi: [10.1074/jbc.M502095200](https://doi.org/10.1074/jbc.M502095200) PMID: [15781451](https://pubmed.ncbi.nlm.nih.gov/15781451/)
91. Luce MJ, Burrows PD. The neuronal EGF-related genes NELL1 and NELL2 are expressed in hemopoietic cells and developmentally regulated in the B lineage. *Gene*. 1999; 231(1):121–126. doi: [10.1016/S0378-1119\(99\)00093-1](https://doi.org/10.1016/S0378-1119(99)00093-1) PMID: [10231576](https://pubmed.ncbi.nlm.nih.gov/10231576/)
92. Franke A, Hampe J, Rosenstiel P, Becker C, Wagner F, Häsler R, et al. Systematic association mapping identifies NELL1 as a novel IBD disease gene. *PLoS ONE*. 2007; 2(8):e691. doi: [10.1371/journal.pone.0000691](https://doi.org/10.1371/journal.pone.0000691) PMID: [17684544](https://pubmed.ncbi.nlm.nih.gov/17684544/)
93. Uhlin-Hansen L, Eskeland T, Kolset SO. Modulation of the expression of chondroitin sulfate proteoglycan in stimulated human monocytes. *Journal of Biological Chemistry*. 1989; 264(25):14916–14922. PMID: [2768247](https://pubmed.ncbi.nlm.nih.gov/2768247/)
94. Fallahi-Sichani M, Kirschner DE, Linderman JJ. NF-κB signaling dynamics play a key role in infection control in tuberculosis. *Frontiers in Physiology*. 2012; 3. doi: [10.3389/fphys.2012.00170](https://doi.org/10.3389/fphys.2012.00170) PMID: [22685435](https://pubmed.ncbi.nlm.nih.gov/22685435/)
95. Kingma PS, Zhang L, Ikegami M, Hartshorn K, McCormack FX, Whitsett JA. Correction of pulmonary abnormalities in Sftpd<sup>-/-</sup> mice requires the collagenous domain of surfactant protein D. *Journal of Biological Chemistry*. 2006; 281(34):24496–24505. doi: [10.1074/jbc.M600651200](https://doi.org/10.1074/jbc.M600651200) PMID: [16787926](https://pubmed.ncbi.nlm.nih.gov/16787926/)
96. Kanazawa N, Okafuji I, Kambe N, Nishikomori R, Nakata-Hizume M, Nagai S, et al. Early-onset sarcoidosis and CARD15 mutations with constitutive nuclear factor-κB activation: common genetic etiology with Blau syndrome. *Blood*. 2005; 105(3):1195–1197. doi: [10.1182/blood-2004-07-2972](https://doi.org/10.1182/blood-2004-07-2972) PMID: [15459013](https://pubmed.ncbi.nlm.nih.gov/15459013/)
97. D’Cunha J, Knight E, Haas AL, Truitt RL, Borden EC. Immunoregulatory properties of ISG15, an interferon-induced cytokine. *Proceedings of the National Academy of Sciences*. 1996; 93(1):211–215. doi: [10.1073/pnas.93.1.211](https://doi.org/10.1073/pnas.93.1.211)
98. Lin Y, Jamison S, Lin W. Interferon-γ activates nuclear factor-κ B in oligodendrocytes through a process mediated by the unfolded protein response. *PloS ONE*. 2012; 7(5):e36408. doi: [10.1371/journal.pone.0036408](https://doi.org/10.1371/journal.pone.0036408) PMID: [22574154](https://pubmed.ncbi.nlm.nih.gov/22574154/)

99. Davila S, Hibberd ML, Dass RH, Wong HE, Sahiratmadja E, Bonnard C, et al. Genetic association and expression studies indicate a role of toll-like receptor 8 in pulmonary tuberculosis. *PLoS Genetics*. 2008; 4(10):e1000218. doi: [10.1371/journal.pgen.1000218](https://doi.org/10.1371/journal.pgen.1000218) PMID: [18927625](https://pubmed.ncbi.nlm.nih.gov/18927625/)
100. Rosenzweig SD, Holland SM. Defects in the interferon- and interleukin-12 pathways. *Immunological Reviews*. 2005; 203(1):38–47. doi: [10.1111/j.0105-2896.2005.00227.x](https://doi.org/10.1111/j.0105-2896.2005.00227.x) PMID: [15661020](https://pubmed.ncbi.nlm.nih.gov/15661020/)
101. Kuenzel S, Till A, Winkler M, Häslér R, Lipinski S, Jung S, et al. The nucleotide-binding oligomerization domain-like receptor NLRC5 is involved in IFN-dependent antiviral immune responses. *The Journal of Immunology*. 2010; 184(4):1990–2000. doi: [10.4049/jimmunol.0900557](https://doi.org/10.4049/jimmunol.0900557) PMID: [20061403](https://pubmed.ncbi.nlm.nih.gov/20061403/)
102. Seldin M, Pasaniuc B, Price A. New approaches to disease mapping in admixed populations. *Nature Reviews Genetics*. 2011;. doi: [10.1038/nrg3002](https://doi.org/10.1038/nrg3002) PMID: [21709689](https://pubmed.ncbi.nlm.nih.gov/21709689/)
103. Bhangale TR, Rieder MJ, Nickerson DA. Estimating coverage and power for genetic association studies using near-complete variation data. *Nature Genetics*. 2008; 40(7):841–843. doi: [10.1038/ng.180](https://doi.org/10.1038/ng.180) PMID: [18568023](https://pubmed.ncbi.nlm.nih.gov/18568023/)
104. Chimusa ER, Daya M, Möller M, Ramesar R, Henn BM, van Helden PD, et al. Determining Ancestry Proportions in Complex Admixture Scenarios in South Africa Using a Novel Proxy Ancestry Selection Method. *PLoS ONE*. 2013 Sep; 8(9):e73971. doi: [10.1371/journal.pone.0073971](https://doi.org/10.1371/journal.pone.0073971) PMID: [24066090](https://pubmed.ncbi.nlm.nih.gov/24066090/)
105. Daya M, van der Merwe L, Galal U, Möller M, Salie M, Chimusa ER, et al. A Panel of Ancestry Informative Markers for the Complex Five-Way Admixed South African Coloured Population. *PLoS ONE*. 2013; 8(12):e82224. doi: [10.1371/journal.pone.0082224](https://doi.org/10.1371/journal.pone.0082224) PMID: [24376522](https://pubmed.ncbi.nlm.nih.gov/24376522/)
106. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, et al. The genetic structure and history of Africans and African Americans. *Science*. 2009; 324(5930):1035–1044. doi: [10.1126/science.1172257](https://doi.org/10.1126/science.1172257) PMID: [19407144](https://pubmed.ncbi.nlm.nih.gov/19407144/)
107. Patterson N, Petersen DC, Van Der Ross RE, Sudoyo H, Glashoff RH, Marzuki S, et al. Genetic structure of a unique admixed population: implications for medical research. *Human Molecular Genetics*. 2010; 19(3):411–419. doi: [10.1093/hmg/ddp505](https://doi.org/10.1093/hmg/ddp505) PMID: [19892779](https://pubmed.ncbi.nlm.nih.gov/19892779/)
108. Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, et al. Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. *The American Journal of Human Genetics*. 2010; 86(4):611–620. doi: [10.1016/j.ajhg.2010.03.015](https://doi.org/10.1016/j.ajhg.2010.03.015)
109. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*. 2006; 103(8):2869–2873. doi: [10.1073/pnas.0511240103](https://doi.org/10.1073/pnas.0511240103)
110. Abel L, El-Baghdadi J, Bousfiha AA, Casanova JL, Schurr E. Human genetics of tuberculosis: a long and winding road. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2014; 369(1645):20130428. doi: [10.1098/rstb.2013.0428](https://doi.org/10.1098/rstb.2013.0428)