Tuberculosis xxx (2014) 1–8

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: http://intl.elsevierhealth.com/journals/tube

HOST GENETICS OF SUSCEPTIBILITY

The role of ancestry in TB susceptibility of an admixed South African population☆

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

Molecular Biology and Human Genetics, MRC Centre for Molecular and Cellular Biology, The DST/NRF Centre of Excellence for Biomedical TB Research, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg 7505, South Africa

ARTICLE INFO

Article history:
Received 27 December 2013
Accepted 30 March 2014

Keywords:
Host genetics
Tuberculosis
Admixture
Ancestry informative markers
Association

SUMMARY

Genetic susceptibility to tuberculosis (TB) has been well established and this, taken together with variation in susceptibility observed between different geographic and ethnic populations, implies that susceptibility to TB may in part be affected by ethnicity. In a previous genome-wide TB case–control study (642 cases and 91 controls) of the admixed South African Coloured (SAC) population, we found a positive correlation between African San ancestry and TB susceptibility, and negative correlations with European and Asian ancestries. Since genome-wide data was available for only a small number of controls in the previous study, we endeavored to validate this finding by genotyping a panel of ancestry informative markers (AIMs) in additional individuals, yielding a data set of 918 cases and 507 controls. Ancestry proportions were estimated using the AIMs for each of the source populations of the SAC (African San, African non-San, European, South Asian and East Asian). Using logistic regression models to test for association between TB and ancestry, we confirmed the substantial effect of ancestry on TB susceptibility. We also investigated the effect of adjusting for ancestry in candidate gene TB association studies of the SAC. We report a polymorphism that is no longer significantly associated with TB after adjustment for ancestry, a polymorphism that is significantly associated with TB only after adjustment for ancestry, and a polymorphism where the association significance remains unchanged. By comparing the allele frequencies of these polymorphisms in the source populations of the SAC, we demonstrate that association results are likely to be affected by adjustment for ancestry if allele frequencies differ markedly in the source populations of the SAC.

1. Introduction

According to the World Health Organization (WHO), the highest burden of TB is carried by Asia and Africa [1] and in the USA, there is a marked contrast in incidence between ethnicities [2]. A large proportion of ethnic disparity in TB susceptibility can be attributed to socioeconomic factors [3–5] and the human immunodeficiency virus (HIV) epidemic [6]. The remaining difference, albeit small, could possibly be explained by genetic differences in susceptibility between population groups, since many investigations have shown that genetic factors are involved in the disease [7]. It is thought that certain population groups are more or less susceptible to TB infections, based on the history of their exposure to the disease and the development of resistance due to natural selection. Based on a large study of 165 racially integrated nursing homes in Arkansas (U.S.A.), Stead et al. showed that Europeans are less susceptible to TB infection compared to individuals of African ancestry [8]. This study was however limited due to its inability to control for all behavioral differences between the two groups. Nevertheless, the apparent higher resistance of Europeans to TB could possibly be explained by many centuries of exposure to the disease in densely populated European settlements [9].

The predominant population group in the Western Cape, South Africa, is the five-way admixed group (African San, African non-San, European, South Asian and East Asian) known as the South African Coloured (SAC) [10,11]. Our SAC study participants are ideally suited to test if an association exists between ancestry and TB susceptibility, as they received genetic contributions from both African and European populations, who differ in TB rates, and come from the same high-TB communities with the same socioeconomic status (SES). In a genome-wide TB case–control study of the group (642

☆ 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.

* Corresponding author.
E-mail address: egvh@sun.ac.za (E.G. Hoal).

http://dx.doi.org/10.1016/j.tube.2014.03.012
1472-9792/

cases and 91 controls), Chimusa et al. [12] found a positive correlation between the proportion of African San ancestry and TB susceptibility, and negative correlations with European, South Asian and East Asian ancestries. Due to the small number of controls in the Chimusa et al. study, we endeavored to validate this finding by genotyping a panel of ancestry informative markers (AIMs), described by Daya et al. [13], in additional cases and controls. The selected AIMs were tailored to the SAC, as other panels of AIMs described in the literature did not adequately incorporate African San ancestry, one of the main ancestral components of the SAC. The complex five-way admixture that occurred in the SAC, with dissimilar genetic distances between source populations, was also not adequately modeled in other panels. Daya et al. evaluated various AIM selection strategies, and by comparing ancestry proportions estimated using different panels of AIMs to the gold standard of genome-wide estimated proportions, ensured that the selected panel of AIMs is best suited to ancestry inference in this population.

A previous TB susceptibility study of the SAC stated that no significant population stratification was found in the cohort [14]. This was based on the comparable allele frequency distributions between cases and controls of 25 randomly selected and uncorrelated SNP markers, implying that adjustment for ancestry is not necessary. However, as a result of the complex admixture that occurred in the SAC, a larger number of markers is ideally required to distinguish allele frequency differences that are due to ancestry. In addition, AIMs have greater power to discern ancestries compared to randomly selected markers [15]. Using the panel of AIMs, we therefore also investigate the effect of adjusting for ancestry in candidate gene association studies of susceptibility to TB in the SAC.

2. Materials and methods

2.1. Sample collection and ethical approval

A large study group of South African Coloured individuals was recruited from the Cape Town suburbs of Ravensmead and Uitsig (the same community recruited in the Chimusa et al. study). This district was selected due to its high TB incidence, uniform SES and low prevalence of HIV [14]. TB patients were identified through bacteriological confirmation (smear positive and/or culture positive). Healthy individuals that had no previous history of TB were selected as controls. The controls were living under the same conditions as TB patients, including SES status and availability of health facilities. HIV positive individuals were excluded from the study.

Approval for the study was obtained from 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. The research was conducted according to the principles expressed in the Declaration of Helsinki.

2.2. Genotyping, quality control and ancestry proportion estimation

Our sample bank comprises 955 case and 521 control samples, collected between 1994 and 2007. 969 Samples were genotyped on the Affymetrix GeneChip Human Mapping 500K Array Set in 2008, 888 of which passed quality control criteria (of which a subset of 733 unrelated individuals were used by Chimusa et al. to test for association between TB susceptibility and ancestry) [12]. The sample bank was also used to perform a number of candidate gene studies between 2003 and 2013, summarized in Supplementary Table 1.

120 AIMs were selected to distinguish the source populations of the SAC [13] and genotyped in 918 samples. The genotyping was performed at the Institute for Clinical Molecular Biology at the Christian-Albrechts University in Kiel, Germany, using the Sequenom iPLEX platform (114 SNPs) and TaqMan assays (6 SNPs). 4 SNPs and 27 samples failed our quality control criteria and were removed from the data set. The remaining 116 AIMs were also available in 888 samples genotyped on the Affymetrix platform. 316 Samples overlapped between the Affymetrix and Sequenom data sets, and were used to assess the concordance between the platforms. One of these samples had 73 discordant genotypes between the two platforms, which were most likely due to mislabeling of the physical sample. After excluding this sample, the mean proportion of discordant genotypes was 0.0047. In total, AIMs were available for 1425 samples (918 Sequenom samples + 888 Affymetrix samples − 316 overlapping samples − 27 samples that failed quality control − 38 samples with either missing sex or age information).

Ancestry proportions were estimated for the combined Sequenom and Affymetrix AIM data set, jointly for each of the five source populations of the SAC, using the methodology and source populations described by Daya et al. [13]. As some of the samples were collected from families, the combined data set contained individuals that were related to one another. Since the multinomial model used to estimate ancestry proportions assumes independence, ancestry proportions were estimated in separate batches, each batch comprised of unrelated individuals.

We re-examined a number of previous candidate gene studies performed by our research group (236 SNPs of 78 genes genotyped in 11 published and 18 unpublished studies), now adjusting for ancestry. An unrelated subset of individuals from the combined data set described above was used in each of these studies. We report the genotype model results of rs2243639, rs2569190 and rs34069356 in order to illustrate the possible effects of ancestry adjustment. These polymorphisms were genotyped using a Taqman assay, a PCR-RFLP method [16], and the SNPlex Genotyping System [17], respectively.

2.3. Statistical analysis

Mixed-effects logistic regression models (generalized linear models with a binomial family and logit link) were used to test for association between TB susceptibility and ancestry. A reduced data set that excluded samples used in the Chimusa et al. [12] study ($n = 696$) was modeled, as well as a complete data set of all samples for which AIMs are available ($n = 1425$). Models were fitted for the individual effect of each of the source ancestries of the SAC, as well as a combined model that included all the ancestries. Age and sex were adjusted for by including them in the models as fixed effects. A family identifier was specified as a random effect to adjust for relatedness between groups of individuals (251 of the 1425 individuals could be grouped into 101 families).

Logistic regression models were used to test for association between TB susceptibility and genotype, adjusting for age and sex, and then adjusting for age, sex and ancestry. (Mixed-effects models were not used, since the individuals were unrelated.) The possibility of genotyping errors was assessed by evaluating Hardy–Weinberg equilibrium (HWE) in controls (exact test).

In this study, associations corresponding to a p-value of 0.05 or less were considered significant. The Bonferroni correction for multiple testing was not used as it may be over-conservative when several genetic associations are tested in the same group [18]. Most multiple testing correction methods may be unsuitable when there is a priori evidence that genes are associated with a phenotype [19,20]. In addition, the reported genotypic association tests are

Please cite this article in press as: Daya M, et al., The role of ancestry in TB susceptibility of an admixed South African population, Tuberculosis (2014), http://dx.doi.org/10.1016/j.tube.2014.03.012