npg

ARTICLE

Polymorphisms in *MC3R* promoter and *CTSZ* 3′UTR are associated with tuberculosis susceptibility

Lindsey A Adams[1], Marlo Möller[1], Almut Nebel[2], Stefan Schreiber[2], Lize van der Merwe[3,4], Paul D van Helden[1] and Eileen G Hoal*,[1]

We have validated the association of two genes on chromosome 20q13.31–33 with tuberculosis susceptibility. A previous genome-wide linkage study performed by Cooke *et al* identified the genes melanocortin-3-receptor (*MC3R*) and cathepsin Z (*CTSZ*) as possible candidates in tuberculosis susceptibility. *MC3R* has been implicated in obesity studies and is known to play a role in many biological systems including the regulation of energy homeostasis and fat metabolism. *CTSZ* has been detected in immune cells, such as macrophages and monocytes, and it is hypothesized that the protein may play a role in the immune response. In our South African population a case–control study confirmed the previously reported association with a single-nucleotide polymorphism (SNP) in *CTSZ* and found an association in *MC3R* with a SNP not previously implicated in tuberculosis susceptibility. Six SNPs in *MC3R* and eight in *CTSZ* were genotyped and haplotypes were inferred. SNP rs6127698 in the promoter region of *MC3R* (cases=498; controls=506) and rs34069356 in the 3′UTR of *CTSZ* (cases=396; controls=298) both showed significant association with tuberculosis susceptibility (*P*=0.0004 and <0.0001, respectively), indicating that pathways involving these proteins, not previously researched in this disease, could yield novel therapies for tuberculosis.
European Journal of Human Genetics (2011) 19, 676–681; doi:10.1038/ejhg.2011.1; published online 2 February 2011

INTRODUCTION

Genome-wide investigations that are free of initial assumptions and are not directly hypothesis driven can identify novel genes involved in complex diseases such as tuberculosis (TB).[1] Most of the genes currently identified by hypothesis-driven studies belong to the innate immune system, the first line of defense against pathogens, but methods based on our current knowledge of disease may overlook genes for which a role has yet to be determined. Among the genes that have been implicated in TB susceptibility, many have well-defined roles in the immune response including HLA class II, *NRAMP1* and *IFNG*.[2] Genome-wide linkage studies are particularly useful because they can identify genes for which a role in infection may not have been suspected. Susceptibility loci for leprosy, a disease related to TB, have been found on chromosomes 10p13 and 20p12.3 in a South Indian population, whereas a Vietnamese study identified chromosome 6q25 as being linked to leprosy susceptibility, which led to the identification of the *PARK 2* and *PACRG* genes.[3–6] The first genome-wide linkage study for TB was conducted in 2000 by Bellamy *et al*,[7] and identified chromosomes 15q and Xq as showing suggestive evidence of linkage to TB susceptibility.

Recently, a multistage strategy was employed by Cooke *et al*[8] to identify a novel locus for TB susceptibility in African populations.[8] An affected sibling pair linkage analysis performed on families recruited from the South African Coloured population in metropolitan Cape Town, and Malawians from the Karonga district indicated one genomic region, 20q13.31–33, as being linked to

TB susceptibility. Forty SNPs within this region were used to screen a large independent Gambian population, and two genes, melanocortin 3 receptor (*MC3R*) and cathepsin Z (*CTSZ*), showed evidence of disease association.[8] Polymorphisms in these genes were further genotyped in populations from Guinea-Bissau and the Republic of Conakry.[8] A polymorphism in the 3′UTR in *CTSZ*, viz CTSZ3P (rs34069356), showed statistically significant disease association (*P*=0.005), with genotype distributions being similar across all three West African populations.[8] Following regression analysis, the initial association seen in the *MC3R* polymorphism MC3R241 (rs3827103) did not remain significant (*P*=0.26), although a trend towards a protective effect of the *MC3R* genotype AA remained.[8]

CTSZ is one of the 11 cysteine proteases of the papain family.[9] In the immune system, cathepsins are involved in antigen processing and maturation of the major histocompatibility complex class II molecules.[10] *CTSZ* is mostly expressed in immune cells, such as macrophages and monocytes, and a role for the protein in the immune response has been hypothesized.[11–13] Cathepsins form a vital component of the lysosomal proteolytic system and are differentially expressed during *Mycobacterium tuberculosis* infection.[14] This expression is specifically associated with macrophages present in the granuloma.[14] Several members of the cathepsins have been implicated in TB.[15] For example, cathepsin L maturation and activity can be impaired by *M. tuberculosis* and *M. avium*, and cathepsin W was identified as a risk factor for the extrapulmonary dissemination of human TB.[15]

[1]DST/NRF Centre of Excellence for Biomedical Tuberculosis Research/MRC Centre of Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics, Faculty of Health Sciences, Stellenbosch University, Tygerberg, South Africa; [2]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany; [3]Biostatistics Unit, Medical Research Council, Tygerberg, South Africa; [4]Department of Statistics, University of the Western Cape, Bellville, South Africa
*Correspondence: Professor EG Hoal, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research/MRC Centre of Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics, Faculty of Health Sciences, Stellenbosch University, PO Box 19063, Tygerberg 7505, South Africa. Tel: +27 21 938 9412; Fax: +27 21 938 9863; E-mail: egvh@sun.ac.za
Received 5 May 2010; revised 29 December 2010; accepted 7 January 2011; published online 2 February 2011

MC3R belongs to a family of seven-transmembrane G-protein-coupled receptors that transmit their signals through the activation of adenylate cyclase.[16] This receptor is abundantly expressed in brain regions and in a variety of peripheral tissues and has been shown to play a role in many biological systems including the regulation of energy homeostasis and fat metabolism, as well as inflammation.[17,18] Polymorphisms in *MC3R* have been associated with severe obesity and defects in this gene can result in decreased total expression, intracellular retention and defective receptor activation.[19–25] Chen et al[26] showed that inactivating mutations of *MC3R* led to an increase in fat mass with a corresponding decrease in body mass without any change in eating behaviour or metabolic rate.

The initial stage of the original linkage study was conducted by Cooke et al[8] using Malawian and South African Coloured sibling pairs to identify the genomic region. In the second stage, fine mapping was carried out in various West African populations. To validate these findings, we conducted case–control studies for *MC3R* and *CTSZ* in unrelated South African Coloured individuals to determine if polymorphisms in these genes show evidence of disease association.

METHODS

Study population

The population is located in the metropolitan area of Cape Town in the Western Cape Province in South Africa. This area was selected owing to the high incidence of TB in the area as well as the uniform ethnicity, socioeconomic status and low prevalence of human immunodeficiency virus (HIV).[27] We did a population-based case–control association study using unrelated individuals from the South African Coloured population (Table 1). TB patients were identified through bacteriological confirmation (smear positive and/or culture positive). Controls were selected from the same community living under the same conditions including socioeconomic status and availability of health facilities. 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 *M. tuberculosis*.[28] The majority of the control population is therefore TST positive, and with the average age of the controls in this study being 27 years (Table 1), 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. Additional sample characteristics are given in Table 1. There was no overlap between the samples used in this study and the previous linkage study carried out by Cooke et al.[8] Approval from the Ethics Committee of the Faculty of Health Sciences, Stellenbosch University (project number 95/072) was obtained before blood samples were collected with informed consent, and known HIV-positive individuals were excluded from the study. DNA was purified using standard extraction protocols.

Despite having received genetic input from Khoisan, Bantu-speaking, European and Asian antecedents, the South African Coloured currently represents a relatively homogenous population.[27] A previous study genotyped

351 cases and 360 controls for a panel of 25 independent SNPs markers that were not in linkage disequilibrium (LD), and were randomly distributed along the genome and polymorphic among the major contributing ethnic groups.[27] This study showed no significant population stratification.[27] Of the 1186 samples genotyped in our study, 505 overlapped with the previous study that investigated population stratification.[27] We can however not explicitly exclude the possibility of stratification in the South African Coloured population.

Genotyping

Given the size and lack of introns of the *MC3R* gene, it was possible to sequence the full gene to gain maximum information about the polymorphisms in and surrounding it. The single exon of the gene as well as 1000 bp upstream was sequenced in 10 controls and 10 TB patients. In total, six SNPs were detected, including the polymorphism reported by Cooke et al,[8] rs3827103 (Table 2) and all were selected for further analysis. Two of the polymorphisms were located within the single exon of the gene and four were located up to 413 bp upstream of the start codon. All six SNPs were within 654 bp of one another and this fragment was directly sequenced following polymerase chain reaction (PCR) amplification. To sequence the *MC3R* polymorphisms, PCR was used to amplify a 995 bp fragment. PCR reactions were carried out in a total volume of 25 μl. Each reaction contained 100 ng of genomic DNA, 2.5 μl of 10× Reaction Buffer containing 15 mM MgCl$_2$ (JMR Holdings, Kent, UK), 1 μl of 2.5 mM dNTPs (Bioline, London, UK), 0.5 μl of 10 μM forward and reverse primers (5′-AGAATCTCAGGGCCAGGTA-3′ and 5-′GTCCTCGAAGGT-CAGGTAGTC-3′, respectively) (Integrated DNA Technologies, Glasgow, UK) and 0.05 μl of Super-Therm Gold DNA polymerase (JMR Holdings). An Eppendorf Mastercycler PCR System was used for the following cycling programme: 10 min of denaturation at 95 °C followed by 30 cycles of 1 min at 94°C 1 min at 65 °C and 1 min at 72 °C. This reaction was ended by incubation at 72 °C for 10 min, and then a hold at 4 °C control for contamination was carried out by the inclusion of master mix blanks in every batch of samples amplified. Amplicons were sequenced in both forward and reverse direction.

CTSZ SNPs that showed an allele frequency above 5% and were located either in the exons or the 3′UTR region of the gene were selected for genotyping from previous publications[8] or online databases such as dbSNP. Eight polymorphisms were selected in total, four of which were located in the 3′UTR region (including rs34069356, the associated SNP reported by Cooke et al[8]) and four in the coding region of the gene. Seven of the *CTSZ* SNPs were genotyped using the SNPlex Genotyping System (Applied Biosystems, Darmstadt, Germany) (Table 2) on an automated platform and data were managed by laboratory information system as described previously.[29] The SNPs were submitted online at the myScience Environment of Applied Biosystems website (http://myscience.appliedbiosystems.com) for assay design. The assay was carried out according to the manufacturer's instructions, CEPH controls were

Table 1 Characteristics of the samples used in this study

All samples (n=1186)	Number	Female (%)	Age, mean± SD
TB cases	628	327 (52%)	29.2 ± 12.2
Controls	558	411 (74%)	33.2 ± 14.3
Samples genotyped for MC3R (n=1020)			
TB cases	504	256 (51%)	33.7 ± 14.1
Controls	516	383 (74%)	29.8 ± 12.0
Samples genotyped for CTSZ (n=857)			
TB cases	481	256 (53%)	32.5 ± 14.7
Controls	376	288 (77%)	26.6 ± 12.3

Table 2 SNPs genotyped in *MC3R* and *CTSZ*

Gene	rs number	Position	Genotyping method
MC3R	rs72650656	Promoter	Sequencing
	rs6127698	Promoter	Sequencing
	rs11575886	Promoter	Sequencing
	rs72650657	Promoter	Sequencing
	rs72650658	Promoter	Sequencing
	rs3827103	Exon 1, missense	Sequencing
CTSZ	rs448943	3′UTR[a]	SNPlex
	rs10369	3′UTR	SNPlex
	rs13720	3′UTR	SNPlex
	rs34069356	3′UTR	Taqman
	rs6064734	Exon 6, synonymous	SNPlex
	rs3787492	Exon 6/intron 5, boundary	SNPlex
	rs163785	Exon 4, synonymous	SNPlex
	rs11540881	Exon 4, synonymous	SNPlex

[a]Untranslated region.

included and assay output files were evaluated with the GeneMapper Analysis Software v.3.5.1 (Applied Biosystems). Alleles were called automatically. The results were verified by inspection of the cluster plots. To ensure genotyping of high quality, the logarithm of the intensity of the fluorescence for a sample had to be greater than 3. SNP rs34069356 was genotyped using a predesigned Custom Taqman SNP Genotyping Assay kit (Applied Biosystems) and fluorescent data was read using the ABI 9700.

Statistical analysis

With the samples successfully genotyped in *MC3R* (504 cases and 516 controls) and *CTSZ* (481 cases and 376 controls) and an expected allele frequency of at least 5% for SNPs genotyped, and assuming a type I error of 5%, the power was 97 and 94%, respectively, to detect an odds ratio of 2.5 in an allelic test. Of the 1186 samples genotyped, 290 TB cases and 246 controls were genotyped for all the *MC3R* and *CTSZ* SNPs. Hardy–Weinberg equilibrium (HWE) was assessed for all SNPs in the TB and control groups. LD patterns based on D' values were summarized with LD heatmaps.

Logistic regression was used to compare the TB and the control group to facilitate adjustment for confounders. As the age and gender both differed significantly between TB cases and controls, all analyses were adjusted for age and gender by including them in the logistic regression models as covariates. We modelled each genotype as the number of minor alleles (additive term), and a dominance term, which is only non-zero for heterozygotes.[30] This model is equivalent to labelling the genotypes. The dominance term was discarded if not significantly different from zero. We inferred haplotypes, of all possible sizes, for both genes, together with the probabilities of the haplotypes being harboured by each individual, and analysed them (adjusting for age and gender) using the methods of Schaid *et al*.[31] We summarized and discussed those models showing significant results (global *P*-values below 0.05). We confirmed all the *P*-values, by running 20 000 simulations (permutations). Results corresponding to *P*-values below 0.05 are described as statistically significant.

We did not use the Bonferroni correction for multiple testing, as such a correction is considered over-conservative when several genetic associations are tested in the same group of individuals,[32] risking the rejection of important findings. Bonferroni correction might also be inappropriate in a situation such as this where there is *a priori* evidence that the genes are associated with TB,[33] whereas Bayesian methods for correction rely on the knowledge of previous probability of involvement, which is currently unknown for most genetic variants.[34]

The freely available (from www.r-project.org) programming environment, R and R packages were used for all statistics. The R package, genetics, was used to estimate genotype and allele frequencies and HWE probabilities.[35] Haplotype frequencies were inferred and analysed using the haplo.stats package[36] and LD heatmaps were drawn with LDheatmap using methods described in Shin *et al*.[37]

RESULTS

Single-point analysis of SNPs in *MC3R*

All SNPs were found to be in HWE in the control group. In the TB cases, two SNPs (rs11575886 and rs3827103) were not in HWE, which might indicate an association with TB (Table 3). SNPs rs6127698, which is located 373 bp upstream of the start codon, showed a statistically highly significant allelic association with TB susceptibility (*P*-value=0.0004). The minor allele, T, was found less frequently in TB cases than in controls. Specifically, the odds of TB is multiplied by 0.69/reduced by more than 30% for each additional T allele, compared with the GG homozygote (OR=0.69; 95% CI: 0.56–0.85), after adjusting for age and gender.

SNP rs11575886 showed a statistically significant association with TB susceptibility (*P*-value=0.0423). The minor homozygote, CC, was found in three cases and not in controls, which resulted in the significant effect we detected. When we combined it with the CT heterozygote, effectively creating a dominant model, the genetic effect was no longer significant (*P*-value=0.5490; OR=0.86; 95% CI: 0.52–1.42) for CC and CT *versus* TT, after adjusting for age and gender.

Single-point analysis of SNPs in *CTSZ*

Of the eight SNPs that were successfully genotyped, three (rs6064734, rs163785 and rs11540881) were monomorphic and excluded from further analyses. Each of the five remaining SNPs was in HWE in controls, and all except rs34069356 was in HWE in TB cases. The same SNP reported by Cooke *et al*[8] to be associated in West Africa (rs34069356) also showed statistically significant evidence of disease association in the South African Coloured population (*P*-value <0.0001). No minor TT genotypes were found, and the TC heterozygote was over-represented in TB patients compared with CC (OR=3.45; 95% CI: 2.10–5.86).

SNP rs13720 showed a significant additive allelic effect (*P*-value=0.0487), with each G allele reducing the odds of TB by 12% (OR=0.78; 95% CI: 0.600–0.999). Fitting a dominant model for G (grouping GG and AG) provided a slightly better fit (*P*-value=0.0412), with AA and AG reducing the odds compared with the wild-type GG (OR=0.73; 95% CI: 0.53–0.99).

Haplotype analysis

The LD heatmaps, based on D', do not show obvious haplotype blocks inside the genes (Supplementary Figure 1). Each SNP is tightly linked with at least one other SNP in the same gene, not necessarily the closest neighbour. Two of the SNPs in *MC3R*, viz rs72650656 and rs72650658, were not included in haplotype analysis as their minor allele frequencies were too low. We examined TB–haplotype association with a progressively larger sliding window. Supplementary Table 1 gives global *P*-values, adjusted for age and gender. We confirmed the *P*-values by doing 20 000 permutations and the resulting *P*-values differed from the logistic regression *P*-values on at most the third decimal (Supplementary Table 2). All those that were significant remained so, and *vice versa*. It is clear that all the haplotypes showing significant associations in *MC3R* contain rs6127698, which had a significant effect individually. Three haplotypes were inferred for the 2-SNP *MC3R* rs6127698–rs11575886 haplotype, G–C (frequency=0.04 in both cases and controls), T–T (frequency=0.29 in controls and 0.22 in cases) and G–T (frequency=0.67 in controls and 0.75 in cases); T–C was not observed at all (Supplementary Table 3). The significant odds ratio for TB was with T–T *versus* G–T (OR=0.69; 95% CI: 0.56–0.84). Longer haplotypes in *MC3R* showed similar effects (Supplementary Table 3). Haplotype rs6127698–rs11575886–rs72650657 showed a significant OR of 0.68 (95% CI: 0.55–0.83) for each T–T–C haplotype (frequency=0.29 in controls and 0.22 in TB cases) compared with the reference G–T–C haplotype (frequency=0.64 in controls and 0.73 in cases). No other haplotype occurred at a frequency of more than 0.05. Four possible haplotypes were not observed in our study group.

In *CTSZ*, rs34063956 with and without rs3787492, which were both significant in the single SNP analyses, were included in the statistically significant TB–association haplotypes. Supplementary Table 3 summarizes the model for haplotype rs13720–rs34063956–rs3787492. The A–C–G haplotype (frequency=0.20 in controls and 0.17 in cases) appears to protect (OR=0.73; 95% CI: 0.53–0.99) against TB compared with the G–T–A haplotype (frequency=0.20 in controls and 0.23 in cases). Four possible haplotypes were not observed in our study group.

To assess the combined effect of the *MC3R* rs6127698 and *CTSZ* rs34069656 SNPs, we calculated odds ratios and *P*-values for the allele combinations and adjusted for age and gender (Supplementary Table 4). The global *P*-value was <0.0001 and two statistically significant effects, one protective (T–C combination, frequency=0.29 in controls and 0.20 in cases; OR=0.71; 95% CI: 0.56–0.89) and the

Table 3 Single SNP statistical analysis of *MC3R* and *CTSZ*

	Controls		HWE[a]	TB Cases		HWE	Association
	Count[b]	Freq[c]	P-value	Count	Freq	P-value	P-value[d]
MC3R rs72650656							
Typed	496		1.0000	492		1.0000	
C/C	489	0.99		484	0.98		0.7854
C/G	7	0.01		8	0.02		
C	985	0.99		976	0.99		0.7854
G	7	0.01		8	0.01		
MC3R rs6127698							
Typed	506		0.3334	498		0.1134	
G/G	258	0.51		311	0.62		0.0016
G/T	200	0.40		157	0.32		
T/T	48	0.09		30	0.06		
G	716	0.71		779	0.78		0.0004
T	296	0.29		217	0.22		
MC3R rs11575886							
Typed	506		1.0000	496		0.0216	
T/T	469	0.93		463	0.93		0.0423
T/C	37	0.07		30	0.06		
C/C	0	0.00		3	0.01		
T	975	0.96		956	0.96		0.8854
C	37	0.04		36	0.04		
MC3R rs72650657							
Typed	507		1.0000	495		1.0000	
C/C	480	0.95		474	0.96		0.4066
C/G	27	0.05		21	0.04		
C	987	0.97		969	0.98		0.4066
G	27	0.03		21	0.02		
MC3R rs72650658							
Typed	497		1.0000	490		1.0000	
C/C	493	0.99		480	0.98		0.1205
C/G	4	0.01		10	0.02		
C	990	1.00		970	0.99		0.1205
G	4	0.00		10	0.01		
MC3R rs3827103							
Typed	495		0.1159	488		0.0377	
G/G	242	0.49		214	0.44		0.1939
G/A	197	0.40		202	0.41		
A/A	56	0.11		72	0.15		
G	681	0.69		630	0.65		0.0730
A	309	0.31		346	0.35		
CTSZ rs448943							
Typed	371		1.0000	467		0.0784	
A/A	107	0.29		136	0.29		0.3383
A/G	185	0.50		214	0.46		
G/G	79	0.21		117	0.25		
A	399	0.54		486	0.52		0.4399
G	343	0.46		448	0.48		
CTSZ rs10369							
Typed	375		1.0000	478		0.7984	
C/C	241	0.64		282	0.59		0.3201
C/T	119	0.32		169	0.35		
T/T	15	0.04		27	0.06		
C	601	0.80		733	0.77		0.1673
T	149	0.20		223	0.23		

Table 3 (Continued)

	Controls		HWE[a]	TB Cases		HWE	Association
	Count[b]	Freq[c]	P-value	Count	Freq	P-value	P-value[d]
CTSZ rs13720							
Typed	376		0.1965	477		0.0748	
A/A	245	0.65		333	0.70		0.1221
A/G	112	0.30		124	0.26		
G/G	19	0.05		20	0.04		
A	602	0.80		790	0.83		0.0487
G	150	0.20		164	0.17		
CTSZ rs34069356							
Typed	298		1.0000	396		0.0031	
C/C	276	0.93		303	0.77		<0.0001
C/T	22	0.07		93	0.23		
C	574	0.96		699	0.88		<0.0001
T	22	0.04		93	0.12		
CTSZ rs3787492							
Typed	356		0.2206	451		0.4479	
C/C	188	0.53		253	0.56		0.6111
C/T	148	0.42		174	0.39		
T/T	20	0.06		24	0.05		
C	524	0.74		680	0.75		0.3919
T	188	0.26		222	0.25		

[a]P-values for exact test of Hardy–Weinberg equilibrium (HWE) stratified by TB susceptibility status.
[b]Genotype and allelic count.
[c]Genotype and allelic frequencies.
[d]P-value for association between polymorphism and TB susceptibility, adjusted for age and gender.

other risk (G–T combination, frequency=0.03 in controls and 0.10 in cases; OR=3.16; 95% CI: 1.75–5.73), compared with the reference G–C (frequency=0.68 in both groups), were detected.

DISCUSSION

We have validated the association between TB and both *CTSZ* and *MC3R*, which was first identified in a genome-wide linkage study. A number of linkage analyses have been conducted in an attempt to identify novel loci involved in susceptibility to TB. In 2000, Bellamy *et al*[7] found evidence that chromosomes 15q and Xq may have linkage to TB, whereas Greenwood *et al*[38] found significant linkage with chromosome region 2q35. A study of the Brazilian population implicated chromosomes 10, 11 and 20, and in a Moroccan population, Baghdadi *et al*[39] showed significant linkage between TB and chromosome 8q12–q13.[40]

A genome-wide linkage study by Cooke *et al*[8] identified a locus on chromosome 20q13.31–33 containing *MC3R* and *CTSZ*, which showed linkage with TB susceptibility. In 2008, Stein *et al*[41] performed a genome linkage study in a large population from Uganda and replicated this finding (*P*=0.002), identifying a 25 cM long region containing both *MC3R* and *CTSZ*.

We have now conducted an independent, unrelated case–control study and found that the same SNP implicated in *CTSZ* by Cooke *et al*,[8] rs34069356, showed evidence of disease association in the South African Coloured population (*P*<0.0001, adjusted for age and gender). Cooke *et al*[8] determined that TT homozygous individuals were more susceptible to TB, but we found no individuals with the TT genotype. TC heterozygotes were however significantly over-represented in TB patients (23% in cases *versus* 7% in controls).

This polymorphism results in a non-conservative amino-acid change of a non-polar alanine to a polar, uncharged threonine. Although it does not appear that the amino-acid substitution occurs in an active site on *CTSZ*, the introduction of a hydroxyl side chain has many possible implications. Threonine has an uncharged, polar side chain, making the amino-acid hydrophilic. Unlike alanine (a hydrophobic amino acid typically located on the interior of a protein), threonine is typically located on the exterior of a protein where the hydroxyl side chain is free to interact with surrounding water molecules. Threonine, but not alanine, is also subject to a number of post-translational modifications (PTMs), including phosphorylation by threonine kinases, O-linked glycosylation and acetylation.[42] Phosphorylation is known to regulate the activity of proteins, and as the amino-acid substitution introduced by SNP rs34069356 occurs close to the *N*-terminal of the protein, it is likely that this threonine is available for phosphorylation. The introduction of a hydroxyl group and a number of PTMs is likely to affect protein folding, intracellular localization and protein activity.[42]

A SNP located 373 bp upstream of the *MC3R* gene (rs6127698) was significantly associated with TB (*P*=0.0004, adjusted for age and gender). This SNP is predicted by Genomatix to create an alternative transcription factor binding site (http://www.genomatix.de/). The SNP in the single exon of *MC3R* associated in the study of Cooke *et al*[8] (rs3827103) was not significantly associated with TB susceptibility in the South African Coloured population. SNP rs6127698 was not genotyped by Cooke *et al*[8] and it should be noted that the r^2 value between rs3827103 and rs6127698 is quite low (in controls: r^2=0.16, *D′*=0.99, data not shown), indicating that these two alleles are not completely predictive of each other in our population. This might explain why we did not find rs3827103 to be associated with TB in our

study. The effects of the creation of an alternative transcription factor binding site are difficult to predict as a polymorphism in this region may result in either an increase or decrease in the transcription of the *MC3R* gene. Further studies must be performed to obtain a better understanding of the effect of such a polymorphism.

A combined analysis of the individually significant *CTSZ* and *MC3R* SNPs revealed two statistically significant effects on TB susceptibility, one protective and the other risk. In the study carried out by Cooke *et al*,[8] the strongest evidence for linkage was within the region 20q13.31–33 with a single point LOD score of 3.1, $P=10^{-4}$ and a maximum-likelihood score MLS of 2.8, $P=0.00008$. In our study, the G–T allele combination of the associated SNPs had an odds ratio of 3.16, 95% CI: 1.75–5.73, $P=0.0001$. The linkage LOD score of Cooke *et al*[8] and our association OR are high and, although these are different measures, it is tempting to speculate that the associated alleles together gave rise to the original observed linkage signal that led us to focus on this region. However, it is also possible that these SNPs are in LD with SNPs located in other adjacent genes and are therefore essentially tag SNPs of a non-neighbouring group of highly correlated SNPs ('bin') as was the case for the *PARK2/PACRG* gene in leprosy.[43] We did not detect any bins for the associated SNPs using the LDselect algorithm as implemented by the Genome Variation Server (http://gvs.gs.washington.edu/GVS/index.jsp).[44]

This study has validated the findings implicating *MC3R* and *CTSZ* in TB susceptibility and provides convincing evidence to motivate further investigation into the mechanisms of action of their respective pathways in TB progression.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

1 Clementi M, Di Gianantonio E: Genetic susceptibility to infectious diseases. *Reprod Toxicol* 2006; **21**: 345–349.
2 Möller M, Hoal EG: Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. *Tuberculosis (Edinb)* 2010; **90**: 71–83.
3 Mira MT, Alcaïs A, Nguyen VT *et al*: Susceptibility to leprosy is associated with *PARK2* and *PACRG*. *Nature* 2004; **427**: 636–640.
4 Siddiqui MR, Meisner S, Tosh K *et al*: A major susceptibility locus for leprosy in India maps to chromosome 10p13. *Nat Genet* 2001; **27**: 439–441.
5 Tosh K, Meisner S, Siddiqui MR *et al*: A region of chromosome 20 is linked to leprosy susceptibility in a South Indian population. *J Infect Dis* 2002; **186**: 1190–1193.
6 Mira MT, Alcaïs A, Van Thuc N *et al*: Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. *Nat Genet* 2003; **33**: 412–415.
7 Bellamy R, Beyers N, McAdam KP *et al*: Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. *Proc Natl Acad Sci USA* 2000; **97**: 8005–8009.
8 Cooke GS, Campbell SJ, Bennett S *et al*: Mapping of a novel susceptibility locus suggests a role for *MC3R* and *CTSZ* in human tuberculosis. *Am J Respir Crit Care Med* 2008; **178**: 203–207.
9 Santamaria I, Velasco G, Pendas AM, Fueyo A, Lopez-Otin C: Cathepsin Z, a novel human cysteine proteinase with a short propeptide domain and a unique chromosomal location. *J Biol Chem* 1998; **273**: 16816–16823.
10 Zavasnik-Bergant T, Turk B: Cysteine cathepsins in the immune response. *Tissue Antigens* 2006; **67**: 349–355.
11 Journet A, Chapel A, Kieffer S, Louwagie M, Luche S, Garin J: Towards a human repertoire of monocytic lysosomal proteins. *Electrophoresis* 2000; **21**: 3411–3419.
12 Garin J, Diez R, Kieffer S *et al*: The phagosome proteome: insight into phagosome functions. *J Cell Biol* 2001; **152**: 165–180.
13 Kos J, Sekirnik A, Premzl A *et al*: Carboxypeptidases cathepsins X and B display distinct protein profile in human cells and tissues. *Exp Cell Res* 2005; **306**: 103–113.
14 Stewart JN, Rivera HN, Karls R, Quinn FD, Roman J, Rivera-Marrero CA: Increased pathology in lungs of mice after infection with an alpha-crystallin mutant of *Mycobacterium tuberculosis*: changes in cathepsin proteases and certain cytokines. *Microbiology* 2006; **152**: 233–244.
15 Kim DK, Park GM, Hwang YI: Microarray analysis of gene expression associated with extrapulmonary dissemination of tuberculosis. *Respirology* 2006; **11**: 557–565.
16 Lee EJ, Lee SH, Jung JW *et al*: Differential regulation of cAMP-mediated gene transcription and ligand selectivity by MC3R and MC4R melanocortin receptors. *Eur J Biochem* 2001; **268**: 582–591.
17 Wang SX, Fan ZC, Tao YX: Functions of acidic transmembrane residues in human melanocortin-3 receptor binding and activation. *Biochem Pharmacol* 2008; **76**: 520–530.
18 Getting SJ, Lam CW, Leoni G, Gavins FN, Grieco P, Perretti M: [D-Trp8]-gamma-melanocyte-stimulating hormone exhibits anti-inflammatory efficacy in mice bearing a nonfunctional MC1R (recessive yellow e/e mouse). *Mol Pharmacol* 2006; **70**: 1850–1855.
19 Lee YS, Poh LK, Kek BL, Loke KY: The role of melanocortin 3 receptor gene in childhood obesity. *Diabetes* 2007; **56**: 2622–2630.
20 Tao YX: Functional characterization of novel melanocortin-3 receptor mutations identified from obese subjects. *Biochim Biophys Acta* 2007; **1772**: 1167–1174.
21 Tao YX, Segaloff DL: Functional characterization of melanocortin-3 receptor variants identify a loss-of-function mutation involving an amino acid critical for G protein-coupled receptor activation. *J Clin Endocrinol Metab* 2004; **89**: 3936–3942.
22 Mencarelli M, Walker GE, Maestrini S *et al*: Sporadic mutations in melanocortin receptor 3 in morbid obese individuals. *Eur J Hum Genet* 2008; **16**: 581–586.
23 Rached M, Buronfosse A, Begeot M, Penhoat A: Inactivation and intracellular retention of the human I183N mutated melanocortin 3 receptor associated with obesity. *Biochim Biophys Acta* 2004; **1689**: 229–234.
24 Feng N, Young SF, Aguilera G *et al*: Co-occurrence of two partially inactivating polymorphisms of *MC3R* is associated with pediatric-onset obesity. *Diabetes* 2005; **54**: 2663–2667.
25 Lee YS, Poh LK, Loke KY: A novel melanocortin 3 receptor gene (*MC3R*) mutation associated with severe obesity. *J Clin Endocrinol Metab* 2002; **87**: 1423–1426.
26 Chen AS, Marsh DJ, Trumbauer ME *et al*: Inactivation of the mouse melanocortin-3 receptor results in increased fat mass and reduced lean body mass. *Nat Genet* 2000; **26**: 97–102.
27 Barreiro LB, Neyrolles O, Babb CL *et al*: Promoter variation in the DC-SIGN encoding gene *CD209* is associated with tuberculosis. *PLoS Med* 2006; **3**: e20.
28 Gallant CJ, Cobat A, Simkin L *et al*: The impact of age and sex on anti-mycobacterial immunity of children and adolescents in an area of high tuberculosis incidence. *Int J Tubercul Lung Dis* 2010; **14**: 952–958.
29 Hampe J, Wollstein A, Lu T *et al*: An integrated system for high throughput TaqMan based SNP genotyping. *Bioinformatics (Oxford, England)* 2001; **17**: 654–655.
30 Cordell HJ, Clayton DG: Genetic association studies. *Lancet* 2005; **366**: 1121–1131.
31 Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA: Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002; **70**: 425–434.
32 Nyholt DR: A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004; **74**: 765–769.
33 Perneger TV: What's wrong with Bonferroni adjustments. *BMJ* 1998; **316**: 1236–1238.
34 Campbell H, Rudan I: Interpretation of genetic association studies in complex disease. *Pharmacogenom J* 2002; **2**: 349–360.
35 Warnes G, Gorman D, Leisch F, Man M: Genetics: population genetics. http://CRAN.R-project.org/package=genetics, 2008.
36 Sinwell JP, Schaid DJ: haplo.stats: statistical analysis of haplotypes with traits and covariates when linkage phase is ambiguous. http://CRAN.R-project.org/package=haplo.stats, 2010.
37 Shin J-H, Blay S, McNeney B, Graham J: LDheatmap: an R function for graphical display of pairwise linkage disequilibrium between single nucleotide polymorphisms. *J Statist Software* 2006; **16**: Code Snippet 3.
38 Greenwood CM, Fujiwara TM, Boothroyd LJ *et al*: Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. *Am J Hum Genet* 2000; **67**: 405–416.
39 Baghdadi JE, Orlova M, Alter A *et al*: An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. *J Exp Med* 2006; **203**: 1679–1684.
40 Miller EN, Jamieson SE, Joberty C *et al*: Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians. *Genes Immun* 2004; **5**: 63–67.
41 Stein CM, Zalwango S, Malone LL *et al*: Genome scan of *M. tuberculosis* infection and disease in Ugandans. *PLoS ONE* 2008; **3**: e4094.
42 Cooper GM (2000). *The Cell. A Molecular Approach*, 2nd edn. Sinauer Associates: Sunderland, MA.
43 Alter A, Alcaïs A, Abel L, Schurr E: Leprosy as a genetic model for susceptibility to common infectious diseases. *Hum Genet* 2008; **123**: 227–235.
44 Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA: Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004; **74**: 106–120.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)