ORIGINAL INVESTIGATION

S. Malik · C.M.T. Greenwood · T. Eguale · A. Kifle
J. Beyene · A. Habte · A. Tadesse · H. Gebrexabher
S. Britton · E. Schurr

Variants of the *SFTPA1* and *SFTPA2* genes and susceptibility to tuberculosis in Ethiopia

Abstract Lungs are the central organ affected and targeted by *Mycobacterium tuberculosis* and immune processes in the lung are of critical importance in the pathogenesis of tuberculosis. A major lung defense against invading pathogens is provided by surfactant protein A, a multi-chain protein encoded by the *SFTPA1* and *SFTPA2* genes. Here, we investigated polymorphisms in the *SFTPA1* and *SFTPA2* genes for association with tuberculosis in 181 Ethiopian families comprising 226 tuberculosis cases. Four polymorphisms, *SFTPA1* 307A, *SFTPA1* 776T, *SFTPA2* 355C, and *SFTPA2* 751C, were associated with tuberculosis ($P=0.00008$; $P=0.019$, $P=0.029$ and $P=0.042$, respectively). Additional subgroup analysis in male, female and more severely affected patients provided evidence for *SFTPA1*/*2*-covariate interaction. Finally, out of five intragenic haplotypes identified in the *SFTPA1* gene and nine identified in the *SFTPA2* gene, $1A^3$ was most significantly associated with tuberculosis susceptibility ($P=0.026$). These findings suggest that *SFTPA1* and *SFTPA2* modify the risk of tuberculosis susceptibility and that this risk is influenced by additional covariates.

S. Malik · T. Eguale · E. Schurr
McGill Centre for the Study of Host Resistance and Departments of Medicine and Human Genetics, McGill University, Montreal, QC, Canada

C.M.T. Greenwood · J. Beyene
Program in Genetics and Genomic Biology, Hospital for Sick Children, Toronto, ON, Canada

A. Kifle · A. Habte · A. Tadesse · H. Gebrexabher
Armauer Hansen Research Institute, Addis Ababa, Ethiopia

S. Britton
Department of Medicine at Karolinska Hospital, Unit of Infectious Diseases, Huddinge, Stockholm, Sweden

E. Schurr (✉)
Research Institute of the Montreal General Hospital, Rm L11-520, 1650 Cedar Avenue, H3G 1A4 Montreal, QC, Canada
E-mail: erwin.schurr@mcgill.ca

Introduction

Almost one-third of the world population is infected with *Mycobacterium tuberculosis*, the causative agent of tuberculosis resulting in approximately 8 million tuberculosis cases and 1.87 million tuberculosis-related deaths per year (Corbett et al. 2003). Human immunodeficiency virus (HIV) co-infection greatly increases the rate of progression from infection to clinical tuberculosis disease and an increasing number of tuberculosis cases are recorded among AIDS patients (Corbett et al. 2003). Globally, the number of cases of tuberculosis is increasing by 2% per year and tuberculosis cases caused by multidrug resistant strains of the tubercle bacillus are expected to rise even faster (Dye et al. 2002). Nevertheless, in the absence of AIDS, approximately 90% of the persons infected with *M. tuberculosis* never advance to clinical forms of the disease suggesting that effective anti-*M. tuberculosis* immunity exists in the majority of individuals. This natural immunity to tuberculosis is thought to be strongly influenced by host genetic factors; however, the molecular identity and function of such genetic factors remain largely unknown (Casanova and Abel 2002; Malik and Schurr 2002).

The main route of transmission of *M. tuberculosis* is via respiratory droplets that are inhaled into the alveoli of exposed individuals. Inhalation of *M. tuberculosis* triggers a chain of poorly understood events that may result in the invasion of lung alveolar macrophages by tubercle bacilli (El-Etr and Cirillo 2001). Consequently, local innate immune responses in the lungs may be crucial in preventing the infection of alveolar macrophages and also delaying progression from infection to disease. Surfactant protein A (SP-A) is part of the innate immune system in the lung and has been implicated in inflammation and host defense against a wide range of unrelated pathogens (review in Haagsman 2002). SP-A

is most commonly produced by alveolar type-II epithelial cells and non-ciliated (Clara) cells (review in Madsen et al. 2003) and is secreted into the fluid lining on the surface of these cells. SP-A has been shown to enhance the attachment of *M. tuberculosis* to alveolar macrophages (Downing et al. 1995). It is not clear whether SP-A acts as a pro-inflammatory (Blau et al. 1997; Kremlev and Phelps 1994; Kremlev et al. 1997; Song and Phelps 2000; Wang et al. 2000, 2002) or anti-inflammatory mediator (Borron et al. 2000; Hussain et al. 2003; Pasula et al. 1999; Rosseau et al. 1999; Sano et al. 1999). However, a recent report suggested a dichotomous role of SP-A as a pro- and anti-inflammatory agent. SP-A exerts anti-inflammatory effects on macrophages when unbound to pathogen but exerts pro-inflammatory effects when the globular head of SP-A becomes pathogen bound (Gardai et al. 2003).

The primary structure of SP-A can be divided into four domains, an *N*-terminal domain involved in intermolecular disulfide bonding, a flexible collagen region a hydrophobic neck region containing an amphipathic helix directing protein trimerization, and a C-terminal globular domain involved in lipid binding and Ca^{2+} dependent carbohydrate binding (Garcia-Verdugo et al. 2002; Palaniyar et al. 2001). It is thought that native SP-A is an octadecamer of six trimers, each trimer consisting of SP-A1 and SP-A2 molecules in a 2:1 ratio (Voss et al. 1988, 1991). In humans, the *SP-A* locus has been mapped to chromosome region 10q22-q23 and consists of two functional genes, *SFTPA1* and *SFTPA1* encoding SP-A1 and SP-A2, respectively (Hoover and Floros 1998). Several synonymous and non-synonymous polymorphisms have been identified in the *SFTPA1* and *SFTPA2* genes (DiAngelo et al. 1999) and *SFTPA1/2* alleles have been shown to be associated with infectious and non-infectious pulmonary events including respiratory distress syndrome (RDS) (Floros et al. 2001a, 2001b; Haataja et al. 2001, 2000; Kala et al. 1998; Marttila et al. 2003; Ramet et al. 2000) respiratory syncytial virus infection in infants (Lofgren et al. 2002), allergic bronchopulmonary aspergillosis (ABPB) (Saxena et al. 2003) and pulmonary tuberculosis (Floros et al. 2000; Madan et al. 2002). To further test the hypothesis that *SFTPA1/2* polymorphisms may represent global tuberculosis risk factors, we investigated the association of pulmonary tuberculosis with *SFTPA1/2* in patients from Southeastern Ethiopia. Here, we report the systematic analysis of *SFTPA1* and *SFTPA2* polymorphisms and their association with tuberculosis.

Patients and methods

Study population

A total of 181 nuclear families were enrolled at Hossana hospital in Ethiopia from 1997 to 2001. Hosanna is a rural administrative and agricultural center located approx. 250 km southwest of Addis Ababa. Annually, approximately 3,000 new TB cases are registered by the Hosanna hospital and its affiliated primary health care clinics. At the time of enrolment, the prevalence of HIV positive individuals among TB patients was estimated at approximately 3–5%. The low incidence of HIV positive individuals was the main reason for selecting Hosanna as study area.

Ethnic group was self-reported at the time of enrolment. The majority (> 90%) of families are members of the Hadiya ethnic group. A small number of families belong to the Gurage, Gurage-Silti, Kembata, and Amhara ethnic groups. In addition, a low proportion of families represent inter-group marriages. A major advantage of family-based designs for association studies is that ethnicity can be excluded as a confounding factor.

The phenotype analyzed in the present study was adult pulmonary tuberculosis. Consequently, inclusion criteria for patients were detection of acid fast bacilli in sputum and/or successful cultivation of *M. tuberculosis* from sputum samples. All subjects were interviewed about the duration of common symptoms of tuberculosis. Severity of tuberculosis disease was assessed by recording weight loss over the last 2 months. All subjects were enrolled in the study after written informed consent was obtained. The study was approved by the National Ethics Commission of Ethiopia and the Ethics Committee at the Research Institute of the McGill University Health Centre.

Genotyping analysis

Genomic DNA was extracted from whole blood employing the Nucleon DNA extraction kit (Amersham Biosciences). *SFTPA1* and *SFTPA2* genotyping was based on methods described by DiAngelo et al. (1999). Briefly, a 3.3 kb fragment was amplified from the *SFTPA1* or *SFTPA2* gene. The 3.3 kb fragment served as a template for subsequent genotyping (PCR-RFLP) of five single nucleotide polymorphisms (SNPs) in the *SFTPA1* gene (codons 19, 60, 62, 133, 219) and four SNPs in the *SFTPA2* gene (codons 9, 91, 140, 223). Digested products were run on 12% PAGE to resolve banding patterns.

Statistical analysis

Allele–phenotype associations were investigated employing a family-based association design. Family-based studies avoid confounding of gene–phenotype associations due to inappropriately chosen controls or population substructures. The general principle of this study is to search for a distortion of the transmission of alleles from parents to affected offspring, a strategy that has been termed "Transmission Disequilibrium Test" (TDT). Data were analyzed by the family-based method implemented in the FBAT program which allows tests of

association in instances of incomplete parental data (Rabinowitz and Laird 2000). Furthermore, FBAT allows the use of an empirical variance–covariance estimator for the statistic which is consistent when sibling marker genotypes are correlated, i.e., when families with multiple affected children are being analyzed, and can be used to study haplotype–phenotype associations. In subgroup analyses (e.g., male / female), affected children in multiplex families may belong to different subgroups. Haplotype frequencies and haplotype–specific associations were obtained using the "hbat" command implemented in FBAT (version 1.5.5). Alleles showing some evidence for association were also analyzed by means of conditional logistic regression employing the GASSOC program (Schaid 1996) and SAS (Statistical Analysis System, Cary NC, USA). This analysis allowed us to provide odds-ratio estimates and confidence intervals for tested markers, and to investigate whether the associations varied with age, sex or weight (Schaid 1996). For these conditional models, only families with two genotyped parents were included. Furthermore, one affected child per family was randomly selected when there were multiple affected siblings. Finally, estimates of linkage disequilibrium parameters were obtained with HAPLOVIEW (version 3.2).

Results

We enrolled a total of 181 families with at least one offspring affected by pulmonary tuberculosis (Table 1). The majority of families belonged to the Hadiya ethnic group, 14 families were of mixed ethnic background, and 41 families belonged to five ethnic groups other than Hadiya. Of the 226 tuberculosis patients, 119 (53%) were male and 107 (47%) were female, 69% of patients belonged to two parent families whereas 31% belonged to one parent families (Table 1). All one parent families included a minimum of two genotyped children. Of the 38 multiplex families, 15 comprised one tuberculosis case in the parental generation. Tuberculosis patients with a weight loss of more than 10 kg during the last 2 months were considered as severely affected, and 17.7% (40/226) of all patients confirmed to this definition (Table 1). There was no significant heterogeneity of patient gender, severe weight loss or age across ethnic groups or family structure (data not shown).

The SFTPA1 and SFTPA2 genes are separated by 50.5 kb of genomic sequence with opposite transcriptional orientation (Fig. 1). There are no additional functional genes located between the two SFTPA genes. A series of single nucleotide polymorphisms (SNP) are found in the coding regions of both genes. Here, we genotyped five SNPs located in the coding region of SFTPA1 and four SNPs located in the coding region of SFTPA2. Non-synonymous polymorphisms occur at amino acid positions A19V (177C/T), V50L (269G/C), and R219W (776G/A) in SFTPA1 and at amino acid positions N9T (110A/C), P91A (355C/G), and K223Q (751A/C) in SFTPA2. Synonymous polymorphisms were also examined and identified at amino acid positions 62P (307G/A) and 133T (520G/A) in SFTPA1 and amino acid position 140S (504T/C) in SFTPA2 (Fig. 1). It is a common practice for SFTPA1 and SFTPA2 to refer intragenic haplotypes, rather than individual SNP variants, as "alleles." However, in this study, we will adhere to the standard nomenclature and refer distinct SNP variants as alleles. The intragenic haplotypes ("alleles") found in the Hosanna study families with minor frequencies > 1% are indicated in Fig. 1.

The subjects enrolled in the study were genotyped for all SNPs. All genotyped markers were in HWE and segregation was consistent with Mendelian rules. Among the unrelated parents of all families, SNPs located within each gene were mostly in significant linkage disequilibrium (LD). The strongest LD was observed between markers SFTPA1-177 and SFTPA1-520 ($r^2 = 0.778$) as well as SFTPA2-504 and SFTPA2-751 ($r^2 = 0.781$). The strongest evidence for LD between markers in the two genes was observed for markers SFTPA1-269 and SFTPA2-110 ($r^2 = 0.406$). For all other intergenic marker combinations evidence for LD was weak ($r^2 < 0.2$). To test for their contribution to risk of tuberculosis, we tested the SFTPA1 and SFTPA2 SNPs individually for association with tuberculosis. Across all markers, SFTPA1 alleles 307A ($P = 0.00008$) and 776T ($P = 0.019$) as well as SFTPA2 alleles 355C ($P = 0.029$) and 751C ($P = 0.042$) provided the strongest evidence for association (Table 2). If applying the Bonferoni correction for multiple comparisons, only allele SFTPA1 307A ($P_{Bonf} = 0.007$) remained significantly associated with tuberculosis although this correction is certainly overly conservative due to the existing linkage disequilibrium among the tested SNPs within each gene.

| Table 1 Characteristics of 181 families comprising 566 individuals enrolled in the present study	Family structure (number of families)	Gender	Number of tuberculosis patients (number of families)			
			Two parent families (125)	Weight loss > 10 kg	One parent families (56)	Weight loss > 10 kg
	Simplex (143)	Male	51	9	24	3
		Female	47	9	21	5
	Multiplex (38)	Male	31	6	13	2
		Female	27	6	12	0
	Total number of tuberculosis patients		156	30	70	10
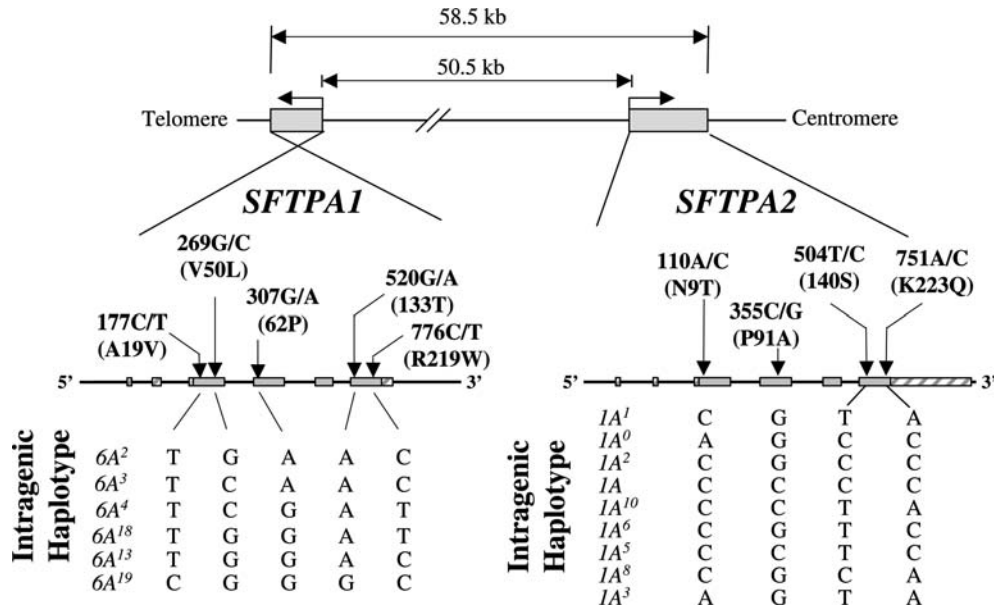
Fig. 1 Schematic representation of *SFTPA1* and *SFTPA2* gene polymorphism on chromosome 10q22. The genomic organization, chromosomal distances in kilo base pairs (kb), and the opposite transcriptional directions of both genes are depicted in the top part of the figure. The exon–intron structure for both genes is shown below with coding exons being represented as *solid boxes*, and non-coding exons as *hatched boxes*. The locations of polymorphisms are indicated by *arrows*. The nucleotide position (accession number NM_005411) for each single nucleotide polymorphism (SNPs) is given on top of each arrow with the corresponding amino acid changes given in parenthesis. For each gene, combinations of SNPs are represented by intragenic haplotypes indicated as $6A^n$ for *SFTPA1* and $1A^n$ for *SFTPA2* based on previously adopted nomenclature (Di Angelo et al. 1999). All intragenic haplotypes given were observed at frequencies >1% in our sample, except $6A^4$ which has been added for discussion purposes

To obtain estimates of the strength of the genetic effects we also employed conditional logistic regression on the subset of two parent families. Since the analysis could be done only on a family subset there was a drop in the significance of *P* values. Nevertheless, the trend of association between *SFTPA* markers and tuberculosis was the same with *SFTPA1* alleles 307A and 776T as well as *SFTPA2* allele 751C being the strongest risk factors (Table 2).

We decided to further explore the observed association of both *SFTPA* genes with tuberculosis separately in male and female patients, in patients with severe forms versus less severe forms of tuberculosis, and by patient age (≥20 vs. <20 years). The hypothesis behind these analyses is that differences in incidence rates (male vs. female) or disease manifestations are the result of different pathways of pathogenesis, and, hence, genetic control. The measure of severity used was excessive weight loss over the last 2 months. There was evidence for a difference in transmission ratios for *SFTPA2* allele 751C in patients with excessive weight loss ($P = 0.001$; $P_{Bonf} = 0.036$) as well as evidence for gender-specific effects of *SFTPA2* alleles 355C and 751C ($P = 0.0048$; $P_{Bonf} = 0.17$ and $P = 0.0018$; $P_{Bonf} = 0.065$, respectively).

Table 2 Allele frequencies and univariate association analysis of polymorphisms in *SFTPA1* and *SFTPA2* with tuberculosis

Gene	Amino acid polymorphism	Marker	Allele	Allele frequency	Number of families	FBAT* P values	GASSOC P values	GASSOC** odds ratio (95% CI)
SFTPA1	A19V	177C/T	C	0.07	28	NS	NS	ND
	V50L	269G/C	G	0.69	38	NS	NS	ND
	P62P	307G/A	A	0.77	18	0.00008	0.015	1.78 (1.11–2.85)
	T133T	520G/A	G	0.05	23	NS	NS	ND
	R219W	776C/T	T	0.11	42	0.019	0.0043	2.64 (1.25–5.59)
SFTPA2	N9T	110A/C	A	0.17	51	NS	NS	ND
	P91A	355C/G	C	0.22	48	0.029	NS	ND
	S140S	504T/C	T	0.51	50	0.07	0.073	1.94 (0.99–3.79)
	K223Q	751A/C	C	0.49	49	0.042	0.016	2.28 (1.16–4.48)

NS non-significant at $P > 0.1$

*For FBAT results, a dominant model was used

**For the GASSOC results, the following models were used: the *P* values and odds ratios for P62P were estimated under an additive model, i.e., risk increasing with each copy of allele A; for R219W, S140S, and K223Q, recessive models for alleles C were used. *P* values are based on score tests; confidence intervals are based on Wald tests

Table 3 FBAT assessment of the sensitivity of the associations due to covariates

Gene	Allele	P values (number of informative families)					
		Gender		Weight loss		Age	
		Male	Female	Weight loss > 10 kg	Weight loss < 10 kg	Age > 20	Age < 20
SFTPA1	177C	0.63 (18)	0.083 (12)	0.26 (7)	0.65 (19)	0.41 (13)	NS
	269G	0.27 (24)	0.38 (14)	0.56 (14)	0.15 (20)	0.07 (30)	NS
	307A	0.033(12)	ND (7)	0.16 (6)	0.011 (10)	0.020 (12)	NS
	520G	0.48 (13)	0.043 (11)	0.81 (5)	0.32 (15)	NS (10)	NS
	776T	0.053 (26)	0.050 (19)	0.16 (8)	0.042 (26)	0.00088 (18)	NS
SFTPA2	110A	0.67 (23)	0.50 (28)	0.88 (12)	0.19 (31)	NS (19)	NS
	355C	0.23 (18)	0.0048 (32)	0.89 (13)	0.20 (30)	ND (7)	NS
	504T	0.044 (28)	0.50 (22)	0.023 (8)	0.24 (33)	NS (19)	NS
	751C	0.0018 (29)	0.81 (20)	0.0010 (8)	0.47 (31)	0.012 (18)	NS

NS Non-significant at $P > 0.1$; *ND* Less than 10 families; not considered

There was also a strong evidence for preferential transmission ratio distortion among patients equal or older than 20 years for *SFTPA1* allele 776T ($P = 0.00088$; $P_{Bonf} = 0.032$; Table 3). To obtain estimates of the subgroup-specific strength of the identified *SFTPA1/2* polymorphisms, we conducted a subgroup analysis using the conditional logistic regression model. A summary of the significant results is presented in Table 4, and leads to conclusions similar to Table 3. The strongest risk factors are *SFTPA2* 504T in severe cases (OR = 6.35 for genotypes TT and CT vs. CC in the subgroup of patients with weight loss > 10 kg), *SFTPA2* 751C in the age group 20 years or older (OR = 4.69 for genotypes CC vs. CA and AA), and *SFTPA1* 776T among patients 20 or older (OR = 4.95 for genotypes TT vs. CT and CC). However, for *SFTPA2* 504T the confidence interval is wide since only 17 families contribute to the estimate. For *SFTPA2* 355 the direction of association is inverted for males and females indicating the possibility of gender-specific disease mechanisms. Finally, we conducted tests to assess the impact of age, gender and body weight on *SFTPA1/2* allele transmission and found significant evidence of heterogeneity of the association due to age or weight at several markers (data not shown). However, age, gender and weight are correlated: males tended to have higher weights than females, and older cases tended to weigh more than younger children (data not shown). Therefore, these data do not permit a conclusive interpretation of the separate effects of gender, body weight and age in affecting *SFTPA* genes transmission ratios to tuberculosis patients.

Derived from SNP combinations, five intragenic haplotypes ("alleles") in *SP-A1* ($6A^2$, $6A^3$, $6A^{18}$, $6A^{13}$, and $6A^{19}$) and ten intragenic haplotypes ("alleles") in *SP-A2* ($1A^1$, $1A^2$, $1A^0$, $1A$, $1A^{10}$, $1A^6$, $1A^5$, $1A^3$, $1A^8$, and $1A^9$) were observed in the present study at frequencies > 1%. The $1A^3$ haplotype was significantly overrepresented in patients ($P = 0.026$) but this estimate was based on a very small family count. None of the *SP-A1* intragenic haplotypes were significantly associated with tuberculosis susceptibility (Table 5). Once corrected for multiple testing none of the intragenic haplotypes were significantly associated with tuberculosis. Extended haplotypes spanning the *SFTPA1* and *SFTPA2* genes were distributed equally among patients (data not shown). To probe for a possible interaction among single sites significantly associated with tuberculosis (Table 2), we analyzed haplotypes of these sites only, but failed to detect significant evidence of association (data not shown).

Table 4 Subgroup estimates of odds ratios

Allele	Number of families	Model (Allele)	Subgroup	Odds ratio (95%CI)
SFTPA1-307G/A	70	Additive (A)	WL < 10 kg	2.13 (1.16–3.94)
	56	Recessive (A)	Male	1.88 (0.95–3.73)
SFTPA1-269G/C	45	Dominant (C)	Age≥20	3.10 (1.25–7.68)
SFTPA1-776C/T	41	Recessive (T)	Age≥20	4.95 (1.38–17.74)
SFTPA2-355C/G	43	Recessive (G)	Male	N/A ($P = 0.008$)
	51	Recessive (C)	Female	2.43 (1.04–5.72)
SFTPA2-504T/C	17	Dominant (T)	WL > 10 kg	6.35 (1.15–35.06)
SFTPA2-751A/C*	88	Recessive (C)	All	2.28 (1.16–4.48)
	47	Recessive (C)	Male	2.95 (1.19–7.34)
	19	Recessive (C)	WL > 10 kg	3.25 (1.06–9.97)
	39	Recessive (C)	Age≥20 years	4.69 (1.23–17.81)

WL Weight loss *N/A* Wald tests gave estimates of infinity for the odds ratio since too few individuals were homozygous for allele "C"*Conditional logistic models with covariates age, weight at enrolment, or weight loss showed marginal significance for heterogeneity at 751A/C (age: $P = 0.051$; weight at enrolment: $P = 0.033$; weight loss > 10 kg: $P = 0.064$). Three other markers also demonstrated some heterogeneity with weight at enrolment (110A/C: $P = 0.036$, 307G/A: $P = 0.032$; 177C/T: $P = 0.067$)

Table 5 Allele frequencies and association analysis of intragenic haplotypes in the *SFTPA1* and *SFTPA2* genes with tuberculosis

Gene	Intragenic haplotypes	Haplotype frequency	Number of families	P value*
SFTPA1	6A^2	0.49	56	NS
	6A^3	0.28	55	NS
	6A^{18}	0.09	29	NS
	6A^{13}	0.07	17	NS
	6A^{19}	0.04	16	NS
SFTPA2	1A^1	0.37	56	NS
	1A^0	0.17	36	NS
	1A^2	0.17	38	NS
	1A	0.11	24	NS
	1A^{10}	0.09	17	NS
	1A^6	0.02	5	NS
	1A^5	0.02	8	0.04
	1A^8	0.02	5	NS
	1A^3	0.01	5	0.026

*NS non-significant at $P > 0.1$

Discussion

We examined nine exonic SNPs in *SFTPA1* and *SFTPA2* and found significant association between *SFTPA1* alleles 307A ($P = 0.00008$) and 776T ($P = 0.019$), and *SFTPA2* alleles 751C ($P = 0.042$) and 355C ($P = 0.029$) with tuberculosis. Since there is a limited LD between polymorphisms in the *SFTPA1/2* genes these results indicate independent association of both *SFTPA* genes with tuberculosis. These data replicate previous findings in a Mexican tuberculosis cohort (Floros et al. 2000), and, hence, implicate *SFTPA1/2* alleles as risk factors of tuberculosis in two ethnically and geographically distinct populations. Specifically, in the Mexican population the *6A^4* and *1A^3* intragenic haplotypes had been found to be risk factors of tuberculosis. In our study, only the 1A^3 *SFTPA2* intragenic haplotype reached borderline significance. In both the Mexican and the Ethiopian populations the *1A^3* haplotype is rare (~1%) and the evidence for association is weak. Nevertheless, detection of association in two distinct studies and populations is supportive of a true effect. However, the evidence for association with tuberculosis of individual polymorphisms was stronger among Ethiopian patients suggesting allelic heterogeneity of the *SFTPA1/2* effects on tuberculosis risk between the two populations.

There is circumstantial evidence to support the suggestions that the tested variants are linked to altered biological function. For example, genotypic variants in the *SFTPA1/2* genes have been shown to mediate differential *SFTPA1/2* mRNA expression, *SFTPA1/SFTPA2* mRNA ratios, and alternative splicing (Karinch et al. 1997; Wang et al. 2003). Furthermore, alleles in the *SFTPA1* and *SFTPA2* genes influence the ability of SP-A to self-aggregate and to induce LPS aggregation (Garcia-Verdugo et al. 2002), and, consistent with an inflammatory role of SP-A, *SFTPA1* and *SFTPA2* polymorphisms have been shown to influence the pro-

duction of TNF-α and IL-8 in THP-1 cells (Wang et al. 2000, 2002). Among the *SFTPA1/2* polymorphisms associated with tuberculosis in the present study, the *SFTPA2* 751A > C polymorphism results in a lysine to glutamine change. Glutamine is a neutral amino acid with a polar amide group, whereas lysine is a positively charged amino acid with a basic side chain. Given the differences in physiochemical properties of the two allelic amino acids and their location in the carbohydrate recognition domain (CRD) of SP-A, (Garcia-Verdugo et al. 2002) the *SFTPA2*-751A/C polymorphism may influence the attachment of *M. tuberculosis* to alveolar macrophages. It is not known how the synonymous P62P polymorphism, *SFTPA1* 307G/A, might affect SP-A function . However, this polymorphism is in close proximity to the exon–intron splice junction and may impact on mRNA maturation. It is well documented that silent mutations that affect splicing regulation can result in serious disorders such as spinal muscular atrophy, frontotemporal dementia and parkinsonism (Cartegni et al. 2002). The *SFTPA2* 355C > G polymorphism, identified in a previous small study (Madan et al. 2002) and by subgroup analysis in the Ethiopian patients, results in a proline to alanine change. It is known that proline normally stabilizes collagen triple helices due to conformational restrictions of the pyrrolidine ring and the presence of tertiary amides while alanine substitutions tend to destabilize the triple helix (Kersteen and Raines 2001). Likewise, it seems possible that a nonconserved arginine to tryptophane substitution at amino acid position 219 located in the CRD of SP-A1 (*SFTPA1* 776C > T) might impact on protein function but direct experimental evidence is lacking.

The data obtained by subgroup analysis and conditional logistic models suggest that the impact of genetic *SFTPA1/2* risk factors is modulated by additional covariates. For example, when stratifying the present study by gender, 355C (proline) was significantly overrepresented in female patients ($P = 0.005$), and logistic regression analysis suggested an inverse effect of this polymorphism in males and females (Table 4). The gender specificity of risk factors for tuberculosis was confirmed by logistic regression analysis for additional *SFTPA1/2* markers. Since it has been reported that the synthesis of surfactant is stimulated by estrogen (Chu and Rooney 1985), it is conceivable that the interplay between estrogen and *SFTPA1/2* polymorphisms may result in gender-specific SP-A activity in tuberculosis pathogenesis. Our data suggest that two *SFTPA2* polymorphisms, 504T and 751C, have their strongest effect among patients with more severe forms of tuberculosis. A recent study demonstrated that SP-A is potent modulator of inflammation in the lung, and that presence of SP-A is critical for protection of lung segments not involved in tuberculosis pathogenesis from inflammatory damage (Gold et al. 2004). Hence, any polymorphism resulting in reduced lung SP-A concentrations or activity can significantly add to tuberculosis pathogenicity.

In the present study, we investigated a number of common exonic polymorphisms for association with tuberculosis. Several of the amino acid polymorphisms had previously been found associated with increased disease risk and possibly altered protein function. In our study, there was a pronounced drop in strength of association from individual SNPs to intra- or intergenic haplotypes. Together with the a priori biological evidence, we take this observation as indication that the tuberculosis-associated single sites are functional. Hence, our data can be taken as a demonstration for the case that direct haplotype association analyses are less powerful than direct SNP association studies, contrary to indirect association studies where the inverse is expected to occur. On the other hand, the overall SNP density in our study was low and we have not established a complete LD pattern of the *SFTPA1* and *SFTPA2* genomic region. To reach a more definite conclusion with respect to the power of haplotype and marker association studies a much higher SNP density will be required. Given that the studied amino acid variants have a genetic effect typical for complex trait risk factors (OR 2–4), the *SFTPA1/2* genes and tuberculosis could serve as an experimental model to study the power and usefulness of common SNP variants, their haplotypes and less common functional variants and their haplotypes to detect association with tuberculosis.

Here, we show that genetic polymorphisms in the *SFTPA1* and *SFTPA2* genes are risk factors for pulmonary tuberculosis in an Ethiopian population. This finding confirms a previous investigation in a Mexican case control population (Floros et al. 2000), and therefore identifies *SFTPA* polymorphisms as global risk factors for tuberculosis. We also provide evidence that the genetic risk conferred by distinct *SFTPA* variants is sensitive to severity of disease, gender and age. This finding points to the possibility of complex gene–environment interactions. A focus for future studies will be to understand the functional consequences of the identified *SFTPA1/2* polymorphisms for pathogenesis of tuberculosis.

References

Blau H, Riklis S, Van Iwaarden JF, McCormack FX, Kalina M (1997) Nitric oxide production by rat alveolar macrophages can be modulated in vitro by surfactant protein A. Am J Physiol 272:L1198–L1204

Borron P, McIntosh JC, Korfhagen TR, Whitsett JA, Taylor J, Wright JR (2000) Surfactant-associated protein A inhibits LPS-induced cytokine and nitric oxide production in vivo. Am J Physiol Lung Cell Mol Physiol 278:L840–L847

Cartegni L, Chew SL, Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285–298

Casanova JL, Abel L (2002) Genetic dissection of immunity to mycobacteria: the human model. Annu Rev Immunol 20:581–620

Chu AJ, Rooney SA (1985) Estrogen stimulation of surfactant synthesis. Pediatr Pulmonol 1:S110–S411

Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, Dye C (2003) The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch Intern Med 163:1009–1021

DiAngelo S, Lin Z, Wang G, Phillips S, Ramet M, Luo J, Floros J (1999) Novel, non-radioactive, simple and multiplex PCR-cRFLP methods for genotyping human SP-A and SP-D marker alleles. Dis Markers 15:269–281

Downing JF, Pasula R, Wright JR, Twigg HL 3rd, Martin WJ 2nd (1995) Surfactant protein A promotes attachment of Mycobacterium tuberculosis to alveolar macrophages during infection with human immunodeficiency virus. Proc Natl Acad Sci USA 92:4848–4852

Dye C, Williams BG, Espinal MA, Raviglione MC (2002) Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. Science 295:2042–2046

El-Etr SH, Cirillo JD (2001) Entry mechanisms of mycobacteria. Front Biosci 6:D737–D747

Floros J, Lin HM, Garcia A, Salazar MA, Guo X, DiAngelo S, Montano M, Luo J, Pardo A, Selman M (2000) Surfactant protein genetic marker alleles identify a subgroup of tuberculosis in a Mexican population. J Infect Dis 182:1473–1478

Floros J, Fan R, Diangelo S, Guo X, Wert J, Luo J (2001a) Surfactant protein (SP) B associations and interactions with SP-A in white and black subjects with respiratory distress syndrome. Pediatr Int 43:567–576

Floros J, Fan R, Matthews A, DiAngelo S, Luo J, Nielsen H, Dunn M, Gewolb IH, Koppe J, van Sonderen L, Farri-Kostopoulos L, Tzaki M, Ramet M, Merrill J (2001b) Family-based transmission disequilibrium test (TDT) and case-control association studies reveal surfactant protein A (SP-A) susceptibility alleles for respiratory distress syndrome (RDS) and possible race differences. Clin Genet 60:178–187

Garcia-Verdugo I, Wang G, Floros J, Casals C (2002) Structural analysis and lipid-binding properties of recombinant human surfactant protein a derived from one or both genes. Biochemistry 41:14041–14053

Gardai SJ, Xiao YQ, Dickinson M, Nick JA, Voelker DR, Greene KE, Henson PM (2003) By binding SIRPalpha or calreticulin/CD91, lung collectins act as dual function surveillance molecules to suppress or enhance inflammation. Cell 115:13–23

Gold JA, Hoshino Y, Tanaka N, Rom WN, Raju B, Condos R, Weiden MD (2004) Surfactant protein A modulates the inflammatory response in macrophages during tuberculosis. Infect Immun 72:645–650

Haagsman HP (2002) Structural and functional aspects of the collectin SP-A. Immunobiology 205:476–489

Haataja R, Ramet M, Marttila R, Hallman M (2000) Surfactant proteins A and B as interactive genetic determinants of neonatal respiratory distress syndrome. Hum Mol Genet 9:2751–2760

Haataja R, Marttila R, Uimari P, Lofgren J, Ramet M, Hallman M (2001) Respiratory distress syndrome: evaluation of genetic susceptibility and protection by transmission disequilibrium test. Hum Genet 109:351–355

Hoover RR, Floros J (1998) Organization of the human SP-A and SP-D loci at 10q22-q23. Physical and radiation hybrid mapping reveal gene order and orientation. Am J Respir Cell Mol Biol 18:353–362

Hussain S, Wright JR, Martin WJ (2003) Surfactant protein A decreases nitric oxide production by macrophages in a tumor necrosis factor-alpha-dependent mechanism. Am J Respir Cell Mol Biol 28:520–527

Kala P, Ten Have T, Nielsen H, Dunn M, Floros J (1998) Association of pulmonary surfactant protein A (SP-A) gene and respiratory distress syndrome: interaction with SP-B. Pediatr Res 43:169–177

Karinch AM, deMello DE, Floros J (1997) Effect of genotype on the levels of surfactant protein A mRNA and on the SP-A2 splice variants in adult humans. Biochem J 321:39–47

Kersteen EA, Raines RT (2001) Contribution of tertiary amides to the conformational stability of collagen triple helices. Biopolymers 59:24–28

Kremlev SG, Phelps DS (1994) Surfactant protein A stimulation of inflammatory cytokine and immunoglobulin production. Am J Physiol 267:L712–L719

Kremlev SG, Umstead TM, Phelps DS (1997) Surfactant protein A regulates cytokine production in the monocytic cell line THP-1. Am J Physiol 272:L996–L1004

Lofgren J, Ramet M, Renko M, Marttila R, Hallman M (2002) Association between surfactant protein A gene locus and severe respiratory syncytial virus infection in infants. J Infect Dis 185:283–289

Madan T, Saxena S, Murthy KJ, Muralidhar K, Sarma PU (2002) Association of polymorphisms in the collagen region of human SP-A1 and SP-A2 genes with pulmonary tuberculosis in Indian population. Clin Chem Lab Med 40:1002–1008

Madsen J, Tornoe I, Nielsen O, Koch C, Steinhilber W, Holmskov U (2003) Expression and localization of lung surfactant protein A in human tissues. Am J Respir Cell Mol Biol 29:591–597

Malik S, Schurr E (2002) Genetic susceptibility to tuberculosis. Clin Chem Lab Med 40:863–868

Marttila R, Haataja R, Ramet M, Pokela ML, Tammela O, Hallman M (2003) Surfactant protein A gene locus and respiratory distress syndrome in Finnish premature twin pairs. Ann Med 35:344–352

Palaniyar N, Ikegami M, Korfhagen T, Whitsett J, McCormack FX (2001) Domains of surfactant protein A that affect protein oligomerization, lipid structure and surface tension. Comp Biochem Physiol A Mol Integr Physiol 129:109–127

Pasula R, Wright JR, Kachel DL, Martin WJ 2nd (1999) Surfactant protein A suppresses reactive nitrogen intermediates by alveolar macrophages in response to Mycobacterium tuberculosis. J Clin Invest 103:483–490

Rabinowitz D, Laird N (2000) A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum Hered 50:211–223

Ramet M, Haataja R, Marttila R, Floros J, Hallman M (2000) Association between the surfactant protein A (SP-A) gene locus and respiratory-distress syndrome in the Finnish population. Am J Hum Genet 66:1569–1579

Rosseau S, Hammerl P, Maus U, Gunther A, Seeger W, Grimminger F, Lohmeyer J (1999) Surfactant protein A down-regulates proinflammatory cytokine production evoked by Candida albicans in human alveolar macrophages and monocytes. J Immunol 163:4495–4502

Sano H, Sohma H, Muta T, Nomura S, Voelker DR, Kuroki Y (1999) Pulmonary surfactant protein A modulates the cellular response to smooth and rough lipopolysaccharides by interaction with CD14. J Immunol 163:387–395

Saxena S, Madan T, Shah A, Muralidhar K, Sarma PU (2003) Association of polymorphisms in the collagen region of SP-A2 with increased levels of total IgE antibodies and eosinophilia in patients with allergic bronchopulmonary aspergillosis. J Allergy Clin Immunol 111:1001–1007

Schaid DJ (1996) General score tests for associations of genetic markers with disease using cases and their parents. Genet Epidemiol 13:423–449

Song M, Phelps DS (2000) Interaction of surfactant protein A with lipopolysaccharide and regulation of inflammatory cytokines in the THP-1 monocytic cell line. Infect Immun 68:6611–6617

Voss T, Eistetter H, Schafer KP, Engel J (1988) Macromolecular organization of natural and recombinant lung surfactant protein SP 28–36. Structural homology with the complement factor C1q. J Mol Biol 201:219–227

Voss T, Melchers K, Scheirle G, Schafer KP (1991) Structural comparison of recombinant pulmonary surfactant protein SP-A derived from two human coding sequences: implications for the chain composition of natural human SP-A. Am J Respir Cell Mol Biol 4:88–94

Wang G, Phelps DS, Umstead TM, Floros J (2000) Human SP-A protein variants derived from one or both genes stimulate TNF-alpha production in the THP-1 cell line. Am J Physiol Lung Cell Mol Physiol 278:L946–L954

Wang G, Umstead TM, Phelps DS, Al-Mondhiry H, Floros J (2002) The effect of ozone exposure on the ability of human surfactant protein a variants to stimulate cytokine production. Environ Health Perspect 110:79–84

Wang G, Guo X, Floros J (2003) Human SP-A 3′-UTR variants mediate differential gene expression in basal levels and in response to dexamethasone. Am J Physiol Lung Cell Mol Physiol 284:L738–L748