Mamm Genome (2011) 22:91–99
DOI 10.1007/s00335-010-9301-7

Tuberculosis as a complex trait: impact of genetic epidemiological study design

Catherine M. Stein · Allison R. Baker

Received: 2 June 2010 / Accepted: 3 November 2010 / Published online: 23 November 2010

Abstract Several studies have suggested a role for human genetic risk factors in the susceptibility to developing tuberculosis (TB). However, results of these studies have been inconsistent, and one potential reason for these inconsistencies is variation in aspects of study design. Specifically, phenotype definitions and population genetic factors have varied dramatically. Since TB is a complex trait, there are many challenges in designing studies to assess appropriately human genetic risk factors for the development of TB as opposed to the acquisition of latent *M. tuberculosis* infection. In this review we summarize these important study design differences, with illustrations from the TB genetics literature. We cite specific examples of studies of the *NRAMP1* (*SLC11A1*) gene and present Fisher's combined *p* values for different stratifications of these studies to further illustrate the impact of study design differences. Finally, we provide suggestions for the design of future genetic epidemiological studies of TB.

C. M. Stein (✉) · A. R. Baker
Department of Epidemiology and Biostatistics, Case Western Reserve University, Wolstein Research Building, Room 1316, 2103 Cornell Rd., Cleveland, OH 44106, USA
e-mail: catherine.stein@case.edu

C. M. Stein
Tuberculosis Research Unit, Case Western Reserve University, Cleveland, OH 44106, USA

Present Address:
A. R. Baker
Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN 37232, USA

Introduction

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, is a significant global public health problem, especially with the rise of the HIV pandemic. Among the one-third of the world infected by *M. tuberculosis*, almost 8 million incident cases of TB are diagnosed annually, and 2 million deaths are attributed to the disease each year.

Support for genetic susceptibility to TB in humans was first provided by twin studies, which have estimated that the concordance rate of TB is higher in monozygotic than in dizygotic twins (Comstock 1978; Kallmann and Reisner 1943). Differences in susceptibility between racial groups have been demonstrated and further support a genetic component; macrophages in blacks are more permissive to tuberculosis infection than those in whites (Crowle and Elkins 1990). A segregation analysis by Shaw et al. (1997) suggested that a general two-locus model was marginally favored over a single-gene codominant model for TB. Several animal models have also provided evidence for a role of genetic factors in TB susceptibility (Blackwell et al. 1995; Flynn et al. 1995; Kramnik et al. 2000; Lurie 1941). Furthermore, candidate gene studies have been conducted in a variety of world populations, though the results have been equivocal (Bellamy 2003).

As discussed in our previous work (Stein et al. 2003), one inherent complication in previous genetic studies of human TB has been the definition of the disease phenotype; as stated by Möller et al. (2010), studies of TB are "exquisitely sensitive to phenotype definition." TB is heterogeneous in its presentation, severity, and duration. TB most often presents with pulmonary disease, but can also affect nearly all organ systems. Even for pulmonary TB, the presentation may differ among people with progressive primary disease and those with reactivation of a

latent form of infection. Since TB is such a heterogeneous disease, it is difficult to define a reliable phenotype for genetic analysis. As we discuss below, diagnostic criteria for TB vary considerably based on available resources. Moreover, TB may develop at any time after infection, and active disease may not present during the data collection phase of the study but rather develop years later. The objective of this review is to examine closely the spectrum of criteria that have been used for clinical characterization of subjects in TB genetics studies and illustrate how these varying criteria may affect study results.

Furthermore, complex traits are characterized by polygenic effects, gene–gene and gene–environment interactions, unclear phenotype definitions, pleiotropy, and genetic heterogeneity. In addition to the aforementioned inconsistency in phenotype definition, many of these genetic epidemiological principles also apply to TB; throughout this review we illustrate examples in the current literature and emphasize the need for ongoing study.

Impact of phenotype definition

TB natural history and disease: factors affecting diagnosis

The pathogenesis of TB can be thought of as a two-stage process (Comstock 1982). In the first stage, exposed individuals can acquire latent *M. tuberculosis* infection (LTBI), in which *M. tuberculosis* establishes a productive infection but does not produce symptoms. LTBI is diagnosed by a positive tuberculin skin test (TST) and/or positive interferon-γ sresponse assay (IGRA) and the absence of clinical signs and symptoms of full-blown disease (ATS/CDC 2000; Nyendak et al. 2008). Approximately 10% of individuals with LTBI progress to develop active TB disease, which is characterized by growth of *M. tuberculosis* in sputum and cultivation in culture or positive acid-fast bacilli (AFB) smear, plus characteristic radiological signs on chest X-ray and hallmark symptoms, including persistent productive cough, fever, and weight loss (ATS/CDC 2000; Garay 2004).

Some studies have provided evidence that these different stages of TB natural history and disease have different genetic influences. Two whole-genome scans (Cobat et al. 2009; Stein et al. 2008) and a candidate gene study (Thye et al. 2009) examined traits related to LTBI and found distinct regions linked to these phenotypes compared to those linked to TB disease. Other studies (Flores-Villanueva et al. 2005; Motsinger-Reif et al. 2010; Stein et al. 2007) have contrasted TB cases to LTBI controls and concluded that observed genetic associations were related to progression from LTBI to active TB disease. However, these studies are in the minority. There are two issues that make the examination of TB genetics literature difficult: (1) characterization of controls and (2) diagnosis of TB. These two issues relate to the two stages of TB pathogenesis, respectively, in which we go into detail below.

In case–control studies, controls should be similar to cases in every way, only nondiseased. This is why case–control studies can be used to establish a relationship between a (genetic) risk factor and disease. For studies of TB, this means that controls should have had the opportunity to become cases and were therefore exposed to other infectious TB cases. The difficulty with many published reports is that there is no characterization of exposure in controls. Some studies assume that in TB-endemic communities, all individuals are exposed (Hoal et al. 2004; Taype et al. 2006), thereby assuming that all individuals who do not have clinical TB disease are latently infected (LTBI). However, other research has demonstrated that individuals may be persistently exposed to *M. tuberculosis* but never develop LTBI (ATS/CDC 2000; Stein et al. 2008). Some studies characterize exposure in controls by conducting TST on unaffected individuals (e.g., Dubaniewicz et al. 2005; Flores-Villanueva et al. 2005; Motsinger-Reif et al. 2010). Other studies utilize unaffected household members as controls (Cervino et al. 2000; El Baghdadi et al. 2003; Stein et al. 2007; Velez et al. 2009a). However, many other studies have used population controls, but this design runs the danger of possible misclassification bias (Edwards et al. 2005).

These issues have an impact on the interpretation of results, depending on study design. If a study used population controls and it is not known whether these individuals have LTBI, a positive association between a polymorphism and TB could mean that the polymorphism is associated with LTBI (since all TB cases must first develop LTBI) or progression to active TB disease. A second disadvantage is that individuals who are never exposed to infectious TB cases may actually carry the risk polymorphism, but because they are unexposed, they never develop TB. In this case, misclassification bias results in a bias toward the null hypothesis, and this could be problematic when the disease is common (McCarthy et al. 2008).

Second is the issue of TB diagnosis, which varies widely by study site. The gold standard for TB diagnosis is growth of *M. tuberculosis* in culture (ATS/CDC 2000). Many research sites lack facilities available for culture confirmation of TB and thus base diagnosis of TB on a positive AFB smear. Research has shown that an AFB smear is less sensitive than culture and that the AFB smear grade could reflect differences in disease severity (Garay 2004). Furthermore, smear-negative, culture-positive TB is a problem

in developing countries because of HIV coinfection (Garay 2004). Because diagnostic criteria for TB vary by study, this could reflect underlying differences in disease severity. If host genes influence severity of TB, these differences in diagnostic criteria become very important. Furthermore, because sensitivity and specificity differ between culture and AFB smear, the problem of potential misclassification arises, which could introduce further bias into genetic studies.

Another factor related to differences in TB diagnosis is the different *M. tuberculosis* strains. Researchers have categorized *M. tuberculosis* into six main strain lineages that are associated with particular geographical regions (Gagneux and Small 2007) as well as clinical presentation (Thwaites et al. 2008) and rate of progression to active TB disease (de Jong et al. 2008). Therefore, not only do different diagnostic criteria potentially reflect differences in disease severity, as discussed above, but specific *M. tuberculosis* strains may also influence disease severity. The potential impact of strain lineage on genetic epidemiological studies was demonstrated by a recent study that observed an interaction between host genotype and *M. tuberculosis* genotype whereby the TLR2 genotype is associated with TB caused by the Beijing strain (Caws et al. 2008).

Lastly, another contributing factor to the differential diagnosis, clinical presentation, and pathogenesis of TB is age of onset. Diagnosis of TB in children is complicated by the absence of positive *M. tuberculosis* cultures in the vast majority of cases (Lewinsohn et al. 2004) and is often delayed (Iriso et al. 2005). Furthermore, children may be asymptomatic (Iriso et al. 2005). The immune response is different in young children versus adults (Lewinsohn et al. 2008). An association study of *NRAMP1* in pediatric TB cases suggests that this gene may be associated with primary TB disease as opposed to reactivation disease (Malik et al. 2005). Leung et al. (2007) also found a more significant association between *NRAMP1* and TB in younger individuals, with age cutoff of 65 years and youner. Besides these few studies, the confounding influence of age of onset on TB genetics studies has not been explored, but it may have an impact on the heterogeneity among published studies.

Studies of *NRAMP1* as examples

So far we have discussed how phenotype definition matters in the analysis of genetic susceptibility to TB. Next we illustrate these issues using studies of the *NRAMP1* gene, aka *SLC11A1*, as examples. Table 1 summarizes each publication by population, diagnosis method used, information provided about study controls, and the most significant *p* value observed for any polymorphism within the gene. Note that this table includes only those studies that focused on pulmonary TB in all age groups.

First, we can see how different studies diagnosed TB. Some studies used the gold-standard definition of growth of *M. tuberculosis* in culture, while other studies used only positive AFB smear. Other studies had more heterogeneous criteria; for example, many studies diagnosed TB based on growth in culture *or* smear, whereas others also included individuals with response to TB treatment *or* evidence of TB disease on chest X-ray. Although they did not conduct an association analysis of *NRAMP1*, another approach to classifying TB cases is illustrated by the first genome-wide linkage scan of TB (Bellamy et al. 2000), where patients diagnosed based on positive AFB smear were categorized together with individuals with a history of TB treatment. One problem in using TB treatment as a diagnostic criterion for TB disease is that some research sites may be unable to distinguish TB from disease due to nontuberculous mycobacteria, but standard TB therapy cures both diseases. It is also possible that some research sites "overtreat" cases to make sure that patients recover from whatever ails them. Furthermore, some studies included both pulmonary and extrapulmonary TB and analyzed them together in the analysis (Ma et al. 2003; Motsinger-Reif et al. 2010; Rossouw et al. 2003; Taype et al. 2006; Velez et al. 2009a). Thus, the classification of TB cases is not comparable across studies.

Second, we can see the wide variety of designs used to categorize individuals as controls. Some studies utilized unaffected individuals within households who were clearly exposed to TB and therefore had the opportunity to develop disease but did not. Other studies conducted TST in unaffected individuals. Not only does a positive TST indicate past exposure to a TB case, it also allows the authors to deduce that the genetic association observed is with progression to active disease. However, the clinical status of other controls is unclear. For example, many studies utilized blood donors, so nothing is known about the past exposure or LTBI status of these individuals. Other studies recruited individuals presenting at clinics for diseases other than TB, so again, both past exposure and LTBI classification are unknown.

Notice that of the 12 studies that demonstrated an association between *NRAMP1* and TB, only four used culture positivity as their diagnosis method. Also note that only one of the *NRAMP1* associations has been observed in studies in which exposure has been characterized in some way, by either utilizing household members as controls or evaluating TST in nondiseased individuals. To investigate differences in a more quantitative fashion, we used Fisher's method for combining *p* values (Fisher 1950). When we combined *p* values, we looked for the most significant *p* value for *any* marker within the gene. This answers the

Table 1 Summary of TB association genetic studies of *NRAMP1/SLC11A1*[a]

Population (Ref.)	TB diagnostic criteria[b]	Characterization of controls	*M. tuberculosis* lineage group[c]	Association *p* value
Gambia (Bellamy et al. 1998)	Smear +	Healthy blood donors	6, 4	<0.001 unadjusted, 0.004 adjusted
Gambia (Awomoyi et al. 2002)	Smear +	Healthy blood donors	6, 4	0.024
Malawi (Fitness et al. 2004)	Smear + OR culture + OR histology	Unrelated with no history of infectious disease	4, 2, 1	0.014 (HIV− cases)0.046 (HIV+ cases)
Morrocco (El Baghdadi et al. 2003)	Culture +	Healthy family members	4	0.33
Tanzania (Søborg et al. 2007)	Culture +	Blood donors	1, 3	0.010
Guinea (Cervino et al. 2000)	Microscopy (smear +? Culture +?)	Unaffected relatives	6	0.036
South Africa (Hoal et al. 2004)	Smear + OR culture +	Unrelated healthy	4, 2	0.002
Caucasian and African American (Velez et al. 2009a)	Culture + OR past diagnosis	Household members in close contact	4	0.03
Caucasian (Ma et al. 2002)	Culture + OR response to TB treatment	Clinic patients without infectious disease	4	<0.010
Caucasian, African-American, and Asian (Motsinger-Reif et al. 2010)	Culture +	Tuberculin skin test positive	4	Not reported
Cambodia (Delgado et al. 2002)	Smear +	Hospital/clinic patients	2, 4	0.009
China (Liu et al. 2004)	Smear + OR culture + OR symptoms and radiological evidence; Males only	Unrelated healthy males	2	0.030
Japan (Abe et al. 2003)	Smear + OR culture +	No history of TB disease	2	Not reported
Japan (Gao et al. 2000)	Smear +	Random clinic patients	2	< 0.0001
Taiwanese (Liaw et al. 2002)	Culture +	Clinic patients without pulmonary disease	4	0.114
Japan (Akahoshi et al. 2004)	Smear +	Healthy blood donors without history of pulmonary or inflammatory disease	2	0.144
Thai (Vejbaesya et al. 2007)	Culture +	Healthy blood bank donors	2, 1	0.54
China (Leung et al. 2007)	Culture +	Hospital patients and healthy blood donors	2	0.0163
Korea (Ryu et al. 2000)	Culture + (unclear)	No history of TB disease	2	0.020
Japan (Kusuhara et al. 2007)	Smear + OR culture +	Unrelated healthy	2	0.096
Poland (Dubaniewicz et al. 2005)	Culture +	TST negative	4	0.63

[a] This table is limited to studies published in English so that case and control definitions could be determined. It is also limited to studies of pulmonary TB in all age groups

[b] Smear + refers to AFB smear positive. Culture + could include more stringent definitions such as culture positive, smear positive, and radiological evidence consistent with TB

[c] *M. tuberculosis* strain lineage based on Gagneux et al. (2006). 1 = Indo-Oceanic, 2 = East-Asian, 3 = East-African-Indian, 4 = Euro-American, 5 = West-African lineage 1 (not detected in these above populations), 6 = West-African lineage 2

question of whether the gene is associated with TB at all; as we point out later, with the variation seen in polymorphisms analyzed in this gene, this may actually be the most appropriate question. Although not nearly as rigorous as a full meta-analysis, this method provides a quick look at these different groupings of studies. We do not present a proper meta-analysis here because of the various stratifications we are interested in; we refer the interested reader to published meta-analysis of this gene (Li et al. 2006) for additional insight.

First, to assess the impact of the TB diagnosis method used, we grouped studies with culture positivity as the only

diagnostic criterion versus all other studies. There was a significant association with TB in both categories ($p = 0.0008$ and $p < 10^{-5}$, respectively), so it does not appear that the TB diagnosis method has a significant impact on the overall conclusion of association between TB and *NRAMP1*.

Second, we analyzed the impact of characterization of controls on association results. Here, we grouped together studies in which controls had been clinically characterized, in terms of either exposure to an infectious TB case or TST (Cervino et al. 2000; Dubaniewicz et al. 2005; El Baghdadi et al. 2003; Motsinger-Reif et al. 2010; Velez et al. 2009a). When p values across these studies were combined, the result was still statistically significant ($p = 0.012$), although it must be interpreted with caution. Because a p value was not reported by Motsinger-Reif et al. (2010), we conservatively inferred a nonsignificant p value of 0.055. However, if we assumed a higher p value of, say, 0.3, the pooled p value was less significant ($p = 0.04$). If the studies were restricted to those with family-based controls, the result is significant, but not overwhelmingly so ($p = 0.014$). When we consider that these gene-centric analyses utilized the smallest p value across the polymorphisms analyzed for *NRAMP1*, one might argue that the more appropriate significant cutoff would account for multiple comparisons. In this case, if we consider that most studies analyzed four polymorphisms within *NRAMP1*, a conservative Bonferroni-like significance threshold would be $\alpha^* = 0.05/4 = 0.0125$; then, the latter analysis that restricted studies to those with family-based controls would not show a statistically significant association between TB and *NRAMP1*. By contrast, the combined p value in the remaining studies was highly significant ($p < 10^{-10}$). Does this contrast suggest an influence of misclassification of controls because of undocumented exposure? There are not enough studies to firmly conclude this. In addition, there are not enough studies with documented TST to conclude whether associations with *NRAMP1* are with progression from LTBI to active TB disease or simply susceptibility to LTBI.

Third, we examined the potential impact of the *M. tuberculosis* strain. Using the global phylogeography of *M. tuberculosis* strain lineage presented by Gagneux et al. (2006), we grouped studies by the *M. tuberculosis* lineage found in the study population (Table 1). Malawi was not studied by Gagneux et al., so we conservatively assumed that the Indo-Oceanic (present in Tanzania), East African (present in Tanzania and South Africa), and Euro-American (in South Africa) lineages were prevalent there. Note that there was a great ecological simplification made here; we assumed that all of the cases analyzed in articles from in a given geographic region were caused by the same strain(s). However, it presents an interesting hypothesis that genetic association studies of *NRAMP1* may differ based on the endemic *M. tuberculosis* strain. It is not trivial to analyze these data because many populations were infected by more than one lineage, and populations that have unique combinations (e.g., Tanzania has both Indo-Oceanic and East-African-Indian lineages) cannot be pooled with others. Studies conducted in populations with the East-Asian strain were highly statistically significant ($p < 10^{-5}$). When populations infected by Euro-American and West-African 2 strains were analyzed together, there also was a statistically significant association ($p < 10^{-4}$). When populations with only the Euro-American strain were analyzed, the result was less significant ($p = 0.033$), although strong conclusions cannot be drawn about this. Again, if we apply the adjustment for multiple comparisons described above ($\alpha^* = 0.0125$), then this analysis of populations with Euro-American lineage were not statistically significant. Thus, we do not have evidence that there is a strain lineage by *NRAMP1* interaction, although specific studies investigating this hypothesis are warranted. These results must be interpreted with caution because our assumption that the results of Gagneux et al. (2006) adequately capture the variation within these study populations is likely an oversimplification. Indeed, recent research has suggested that the distribution of the bacterial population structure has changed significantly over time (van der Spuy et al. 2009), so it is possible that strain diversity in study populations has changed.

Other complicating factors

Population genetics

Another explanation for differing results by population includes genetic heterogeneity or inestimable polygenic effects (Deng 2001; Möller et al. 2010). Polygenic effects may be caused by the combined effects of several rare variants (Bodmer and Bonilla 2008). Rare variants in TB have been much understudied. One study has done extensive resequencing of Toll-like receptor (TLR) genes and found association with TB (Ma et al. 2007). We have also conducted full-exon resequencing of TLR genes and identified novel polymorphisms in Ugandan and South African populations (Baker et al. 2009). Copy number variants, which have been associated with HIV acquisition and progression as well as autoimmune diseases (McCarroll and Altshuler 2007), may also prove to be important to TB susceptibility. Because these variants have low frequency, they must be treated as rare genetic variants in analysis, and thus they require either large sample sizes or novel analytical approaches to detect trait associations. Analysis of large extended pedigrees would be ideal for the

detection and analysis of rare variants (Manolio et al. 2009).

Linkage disequilibrium (LD) is another population genetic factor that may explain differences between studies. LD differs widely by population (Jakobsson et al. 2008), and since trait-associated polymorphisms may actually be untyped and in LD with genotyped markers, variations in LD may confound the ability to detect and replicate marker–trait associations. This is illustrated by the *NRAMP1* association study by Velez et al. (2009a, 2009b); this study did not find association between TB and the four polymorphisms in *NRAMP1* that were examined in previous studies, but it did find association between TB and other single nucleotide polymorphisms (SNPs) in the gene. If these additional SNPs had not been analyzed, the association between TB and *NRAMP1* in that population likely would have been missed. In terms of understanding how specific genes convey risk for developing disease, it may be more important to first understand which gene is associated with disease risk and then identify the specific polymorphism(s). Since TB is a complex trait, there is likely genetic heterogeneity, implying that different variants influence disease risk in different populations. Thus, it is important to conduct thorough genotyping when conducting genetic epidemiological studies rather than focus on a few promoter or nonsynonymous SNPs.

Complex genetic effects

Complex traits are often characterized by gene–gene and gene–environment interactions, but such interaction effects in TB have been understudied. One study suggested interaction between the *NOS2A* gene and *IFNGR1* and *TLR4* (Velez et al. 2009b), though neither *IFNGR1* nor *TLR4* had significant main effects in this analysis. A second study by the same research group found interaction between *NRAMP1* and *TLR2*, but *TLR2* did not have a significant main effect (Velez et al. 2009a). This suggests that many important genes may influence TB in combination with other genes, but these interactions may be overlooked because their individual effects did not meet criteria for statistical significance. Alternative statistical approaches may be used to discover epistatic effects; such an analysis was conducted by Motsinger-Reif et al. (2010) who used multifactor dimensionality reduction to identify a potential gene–gene interaction between the *TLR4* and *TNFα* genes.

Epidemiological factors also may modify the effects of genes. *M. tuberculosis* gene by human gene interaction, as discussed above, is one example of this. Another important effect modifier is HIV seropositivity. The effect of HIV on TB genetics has been understudied because most research studies exclude HIV-infected individuals from both their case and control populations, and HIV is a strong confounding factor of TB immune response. However, our previous work has shown an interaction between HIV serostatus and the TNF receptor 1 gene (Stein et al. 2007). Future studies may suggest that certain genetic effects are important only in HIV-uninfected individuals, while other genes modify TB susceptibility in HIV-infected individuals. On the other hand, studies examining interaction effects require large sample sizes (Velez et al. 2009b).

Summary and conclusions

Although numerous studies examining genetic risk factors for TB have been conducted, we are only beginning to scratch the surface of understanding the role of host genetics in TB susceptibility. Here we have described several principles of genetic epidemiology—phenotype definition, population genetics, and complex genetic effects—and we have illustrated how these factors cloud the current TB genetics literature. These factors should all be considered when synthesizing the literature. Although many of our combined p values attained statistical significance when considering *NRAMP1* studies as examples, we do not wish to imply that this is true of studies of other genes. Many of these p values are highly significant due to the large number of p values that were combined. Moreover, these p values are all significant because we analyzed the most significant published p value. If we had conducted these analyses separately for each polymorphism, we likely would have had different results. For genes that have been studied less extensively (see Möller and Hoal 2010; Möller et al. 2010 for recent reviews), the impact of study design on study results is unknown. Although it is impossible to change these study design components retrospectively, future studies should consider these issues when developing the study's design.

Seven genome-wide linkage scans for TB have been published. Recently, the first genome-wide association study (GWAS) of TB was published (Thye et al. 2010). Interestingly, this GWAS detected statistically significant association between TB and a novel genomic region on chromosome 18 ($p < 10^{-8}$), but this region appears to be a gene desert, so no new candidate genes were immediately proposed by this study. Because of the increased power of association analysis over linkage analysis for detection of common variants with smaller effect sizes (Ardlie et al. 2002; Risch and Merikangas 1996), a GWAS analysis of TB should certainly provide new clues to the genetic underpinnings of TB risk. However, in order to provide sufficient statistical power for a GWAS, thousands of study subjects are needed. In the Thye et al. (2010) GWAS, populations from Ghana, The Gambia, and Malawi were

analyzed using meta-analysis. As we show in this review, merging data across studies should be done with extreme caution because of differences in study design. Other GWAS analyses have merged data across studies, using meta-analysis techniques, and have successfully identified loci associated with Crohn's disease and type II diabetes (Barrett et al. 2008; Zeggini et al. 2008); however, heterogeneity in study design may mask underlying associations when unaccounted for (Heid et al. 2009). However, if we are interested in detecting rare variants underlying TB risk, linkage analysis may be more powerful (Ardlie et al. 2002), so family studies will be advantageous, as stated above.

The ultimate goal in understanding TB genetics is to understand which factors make individuals more susceptible to developing disease in order to facilitate the development of better vaccines and other therapeutics. Another field of study involved in reaching this goal is immunology. Very few studies have examined genetic influences on the TB immune response (Hawn et al. 2007; Shey et al. 2010; Stein et al. 2007, 2008; Wheeler et al. 2006), and more studies are needed. Furthermore, the observation that some genes are associated with more than one phenotype demonstrates pleiotropic effects. For example, studies have shown association with both immunological traits and TB disease outcomes (Shey et al. 2010; Stein et al. 2007). These results could reflect the fact that these traits are on the same pathway (immune response influencing disease outcome) or that the traits themselves are correlated. In addition, some researchers have advocated the study of the "genetics of health" for a different perspective (Nadeau and Topol 2006), which would be useful in vaccine development. The few studies on TB that have focused on resistance to *M. tuberculosis* infection (Stein et al. 2008), or on LTBI and not TB disease (Cobat et al. 2009; Stein et al. 2007; Thye et al. 2009), are beginning to offer insight into the genetics of individuals who infected with *M. tuberculosis* and remain healthy.

Finally, unlike most genetic epidemiological studies that are conducted in developed countries, studies of infectious diseases like TB must focus on developing countries in Africa and Asia where the disease is endemic. Studies in Africa offer a variety of challenges (Sirugo et al. 2008), including limited resources for diagnosis of TB (Gustafson et al. 2001), geographic variations in *M. tuberculosis* lineage, and population-level variations in LD. Future studies must be mindful of these issues. From a public health perspective, these populations will gain the most from genetic epidemiological studies of TB, though the challenges are great.

Acknowledgments This work was funded in part by National Institutes of Health (NIH) grant R01-HL096811 from NHLBI Tuberculosis Research Unit, grant HHSN266200700022C/N01-AI70022 from NIAID, and grant T32-HL0756 from NHLBI. We also thank Dr. W. Henry Boom for helpful discussions.

References

Abe T, Iinuma Y, Ando M, Yokoyama T, Yamamoto T et al (2003) *NRAMP1* polymorphisms, susceptibility and clinical features of tuberculosis. J Infect 46:215–220

Akahoshi M, Ishihara M, Remus N, Uno K, Miyake K et al (2004) Association between IFNA genotype and the risk of sarcoidosis. Hum Genet 114(5):503–509

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3(4): 299–309

ATS/CDC (2000) Diagnostic Standards and Classification of Tuberculosis in Adults and Children. This official statement of the American Thoracic Society and the Centers for Disease Control and Prevention was adopted by the ATS Board of Directors, July 1999. This statement was endorsed by the Council of the Infectious Disease Society of America, September 1999. Am J Respir Crit Care Med 161(4 Pt 1):1376–1395

Awomoyi A, Marchant A, Howson JM, McAdam KP, Blackwell JM et al (2002) Interleukin-10, polymorphism in *SLC11A1* (formerly *NRAMP1*), and susceptibility to tuberculosis. J Infect Dis 186:1808–1814

Baker AR, Randhawa AK, Shey MS, de Kock M, Kaplan G et al (2009) Comparison of genotype frequencies in Toll-like receptor genes in Ugandans, South Africans, and African HapMap populations. Presented at the American Society of Human Genetics annual meeting, Honolulu, HI 22 October 2009.

Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH et al (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 40(8): 955–962

Bellamy R (2003) Susceptibility to mycobacterial infections: the importance of host genetics. Genes Immun 4:4–11

Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC et al (1998) Variations in the *NRAMP1* gene and susceptibility to tuberculosis in West Africans. N Engl J Med 338:640–644

Bellamy R, Beyers N, McAdam KP, Ruwende C, Gie R et al (2000) Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. Proc Nat Acad Sci USA 97:8005–8009

Blackwell J, Barton CH, White JK, Searle S, Baker AM et al (1995) Genomic organizaton and sequence of the human *NRAMP* gene: identification and mapping of a promotor region polymorphism. Mol Med 1:194–205

Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 40(6):695–701

Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NT et al (2008) The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. PLoS Pathog 4(3):e1000034

Cervino A, Lakiss S, Sow O, Hill AV (2000) Allelic association between the *NRAMP1* gene and susceptibility to tuberculosis in Guinea-Conakry. Ann Hum Genet 64:507–512

Cobat A, Gallant CJ, Simkin L, Black GF, Stanley K et al (2009) Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. J Exp Med 206(12):2583–2591

Comstock G (1978) Tuberculosis in twins: a re-analysis of the Prophit Study. Am Rev Respir Dis 117:621–624

Comstock G (1982) Epidemiology of tuberculosis. Am Rev Respir Dis 125:8–15

Crowle A, Elkins N (1990) Relative permissiveness of macrophages from black and white people for virulent tubercle bacilli. Infect Immun 58:632–638

de Jong BC, Hill PC, Aiken A, Awine T, Antonio M et al (2008) Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in The Gambia. J Infect Dis 198(7):1037–1043

Delgado J, Baena A, Thim S, Goldfeld AE (2002) Ethnic-specific genetic associations with pulmonary tuberculosis. J Infect Dis 186:1463–1468

Deng HW (2001) Population admixture may appear to mask, change or reverse genetic effects of genes underlying complex traits. Genetics 159(3):1319–1323

Dubaniewicz A, Jamieson SE, Dubaniewicz-Wybieralska M, Fakiola M, Nancy Miller E et al (2005) Association between *SLC11A1* (formerly *NRAMP1*) and the risk of sarcoidosis in Poland. Eur J Hum Genet 13(7):829–834

Edwards BJ, Haynes C, Levenstien MA, Finch SJ, Gordon D (2005) Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. BMC Genet 6(1):18

El Baghdadi J, Remus N, Benslimane A, El Annaz H, Chentoufi M et al (2003) Variants of the human *NRAMP1* gene and susceptibility to tuberculosis in Morocco. Int J Tuberc Lung Dis 7:599–602

Fisher R (1950) Statistical methods for research workers, vol. V, 11th edn. Liver & Boyd, Edinburgh, Scotland

Fitness J, Floyd S, Warndorff DK, Sichali L, Malema S et al (2004) Large-scale candidate gene study of tuberculosis susceptibility in the Karonga district of Nothern Malawi. Am J Trop Med Hyg 71:341–349

Flores-Villanueva PO, Ruiz-Morales JA, Song CH, Flores LM, Jo EK et al (2005) A functional promoter polymorphism in monocyte chemoattractant protein-1 is associated with increased susceptibility to pulmonary tuberculosis. J Exp Med 202(12):1649–1658

Flynn J, Goldstein MM, Chan J, Triebold KJ, Pfeffer K et al (1995) Tumor necrosis factor-alpha is required in the protective immune response against *Mycobacterium tuberculosis* in mice. Immunity 2:561–572

Gagneux S, Small PM (2007) Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. Lancet Infect Dis 7(5):328–337

Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC et al (2006) Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc Nat Acad Sci USA 103(8):2869–2873

Gao PS, Fujishima S, Mao XQ, Remus N, Kanda M et al (2000) Genetic variants of *NRAMP1* and active tuberculosis in Japanese populations. Clin Genet 58:74–76

Garay S (2004) Pulmonary tuberculosis. In: Rom WN, Garay SM (eds) Tuberculosis. Lippincott Williams & Wilkins, Philadelphia, pp 345–394

Gustafson P, Gomes VF, Vieira CS, Jensen H, Seng R et al (2001) Tuberculosis mortality during a civil war in Guinea-Bissau. JAMA 286(5):599–603

Hawn TR, Misch EA, Dunstan SJ, Thwaites GE, Lan NT et al (2007) A common human TLR1 polymorphism regulates the innate immune response to lipopeptides. Eur J Immunol 37(8):2280–2289

Heid IM, Huth C, Loos RJ, Kronenberg F, Adamkova V et al. (2009) Meta-analysis of the INSIG2 association with obesity including 74,345 individuals: does heterogeneity of estimates relate to study design? PLoS Genet 5(10):e1000694

Hoal EG, Lewis LA, Jamieson SE, Tanzer F, Rossouw M et al (2004) *SLC11A1* (*NRAMP1*) but not *SLC11A2* (*NRAMP2*) polymorphisms are associated with susceptibility to tuberculosis in a high-incidence community in South Africa. Int J Tuberc Lung Dis 8(12):1464–1471

Iriso R, Mudido PM, Karamagi C, Whalen C (2005) The diagnosis of childhood tuberculosis in an HIV-endemic setting and the use of induced sputum. Int J Tuberc Lung Dis 9(7):716–726

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM et al (2008) Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451(7181):998–1003

Kallmann F, Reisner D (1943) Twin studies on the significance of genetic factors in tuberculosis. Am Rev Tuberc 47:549–574

Kramnik I, Dietrich WF, Demant P, Bloom BR (2000) Genetic control of resistance to experimental infection with virulent *Mycobacterium tuberculosis*. Proc Nat Acad Sci USA 97: 8560–8565

Kusuhara K, Yamamoto K, Okada K, Mizuno Y, Hara T (2007) Association of IL12RB1 polymorphisms with susceptibility to and severity of tuberculosis in Japanese: a gene-based association analysis of 21 candidate genes. Int J Immunogenet 34(1):35–44

Leung KH, Yip SP, Wong WS, Yiu LS, Chan KK et al (2007) Sex- and age-dependent association of *SLC11A1* polymorphisms with tuberculosis in Chinese: a case control study. BMC Infect Dis 7:19

Lewinsohn DA, Gennaro ML, Scholvinck L, Lewinsohn DM (2004) Tuberculosis immunology in children: diagnostic and therapeutic challenges and opportunities. Int J Tuberc Lung Dis 8(5): 658–674

Lewinsohn DA, Zalwango S, Stein CM, Mayanja-Kizza H, Okwera A et al (2008) Whole blood interferon-γ responses to *Mycobacterium tuberculosis* antigens in young household contacts of persons with tuberculosis in Uganda. PLoS One 3(10):e3407

Li HT, Zhang TT, Zhou YQ, Huang QH, Huang J (2006) *SLC11A1* (formerly *NRAMP1*) gene polymorphisms and tuberculosis susceptibility: a meta-analysis. Int J Tuberc Lung Dis 10(1):3–12

Liaw YS, Tsai-Wu JJ, Wu CH, Hung CC, Lee CN et al (2002) Variations in the *NRAMP1* gene and susceptibility of tuberculosis in Taiwanese. Int J Tuberc Lung Dis 6:454–460

Liu W, Cao WC, Zhang CY, Tian L, Wu XM et al (2004) *VDR* and *NRAMP1* gene polymorphisms in susceptibility to pulmonary tuberculosis among the Chinese Han population: a case-control study. Int J Tuberc Lung Dis 8:428–434

Lurie M (1941) Heredity, constitution and tuberculosis: an experimental study. Am Rev Tuberc 44(suppl):1–125

Ma X, Dou S, Wright JA, Reich RA, Teeter LD et al (2002) 5′ dinucleotide repeat polymorphism of *NRAMP1* and susceptibility to tuberculosis among Caucasian patients in Houston, Texas. Int J Tuberc Lung Dis 6:818–823

Ma X, Reich RA, Wright JA, Tooker HR, Teeter LD et al (2003) Association between interleukin-8 gene alleles and human susceptbility to tuberculosis disease. J Infect Dis 188:349–355

Ma X, Liu Y, Gowen BB, Graviss EA, Clark AG et al (2007) Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. PLoS One 2(12):e1318

Malik S, Abel L, Tooker H, Poon A, Simkin L et al (2005) Alleles of the *NRAMP1* gene are risk factors for pediatric tuberculosis disease. Proc Nat Acad Sci USA 102(34):12183–12188

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA et al (2009) Finding the missing heritability of complex diseases. Nature 461(7265):747–753

McCarroll SA, Altshuler DM (2007) Copy-number variation and association studies of human disease. Nat Genet 39(7 Suppl):S37–S42

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J et al (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9(5):356–369

Möller M, Hoal EG (2010) Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. Tuberculosis (Edinb) 90(2):71–83

Möller M, de Wit E, Hoal EG (2010) Past, present and future directions in human genetic susceptibility to tuberculosis. FEMS Immunol Med Microbiol 58(1):3–26

Motsinger-Reif AA, Antas PR, Oki NO, Levy S, Holland SM et al (2010) Polymorphisms in IL-1beta, vitamin D receptor Fok1, and Toll-like receptor 2 are associated with extrapulmonary tuberculosis. BMC Med Genet 11:37

Nadeau JH, Topol EJ (2006) The genetics of health. Nat Genet 38(10):1095–1098

Nyendak M, Lewinsohn DA, Lewinsohn D (2008) The use of interferon-gamma release assays in clinical practice. In: Davies P, Barnes PF, Gordon SB (eds) Clinical Tuberculosis. Hodder Arnold, London

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273(5281):1516–1517

Rossouw M, Nel HJ, Cooke GS, van Helden PD, Hoal EG (2003) Association between tuberculosis and a polymorphic NFκB binding site in the interferon γ gene. Lancet 361:1871–1872

Ryu S, Park YK, Bai GH, Kim SJ, Park SN et al (2000) 3′UTR polymorphisms in the *NRAMP1* gene are associated with susceptibility to tuberculosis in Koreans. Int J Tuberc Lung Dis 4:577–580

Shaw MA, Collins A, Peacock CS, Miller EN, Black GF et al (1997) Evidence that genetic susceptibility to *Mycobacterium tuberculosis* in a Brazilian population is under oligogenic control: linkage study of the candidate genes *NRAMP1* and *TNFA*. Tuberc Lung Dis 78:35–45

Shey MS, Randhawa AK, Bowmaker M, Smith E, Scriba TJ et al (2010) Single nucleotide polymorphisms in toll-like receptor 6 are associated with altered lipopeptide- and mycobacteria-induced interleukin-6 secretion. Genes Immun 11(7):561–572

Sirugo G, Hennig BJ, Adeyemo AA, Matimba A, Newport MJ et al (2008) Genetic studies of African populations: an overview on disease susceptibility and response to vaccines and therapeutics. Hum Genet 123(6):557–598

Søborg C, Andersen AB, Range N, Malenganisho W, Friis H et al (2007) Influence of candidate susceptibility genes on tuberculosis in a high endemic region. Mol Immunol 44(9):2213–2220

Stein C, Guwatudde D, Nakakeeto M, Peters P, Elston RC et al (2003) Heritability analysis of cytokines as intermediate phenotypes of tuberculosis. J Infect Dis 187:1679–1685

Stein CM, Zalwango S, Chiunda AB, Millard C, Leontiev DV et al (2007) Linkage and association analysis of candidate genes for TB and TNFalpha cytokine expression: evidence for association with IFNGR1, IL-10, and TNF receptor 1 genes. Hum Genet 121(6):663–673

Stein CM, Zalwango S, Malone LL, Won S, Mayanja-Kizza H et al (2008) Genome scan of *M. tuberculosis* infection and disease in Ugandans. PLoS One 3(12):e4094

Taype CA, Castro JC, Accinelli RA, Herrera-Velit P, Shaw MA et al (2006) Association between *SLC11A1* polymorphisms and susceptibility to different clinical forms of tuberculosis in the Peruvian population. Infect Genet Evol 6(5):361–367

Thwaites G, Caws M, Chau TT, D'Sa A, Lan NT et al (2008) The relationship between *Mycobacterium tuberculosis* genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. J Clin Microbiol 46(4):1363–1368

Thye T, Browne EN, Chinbuah MA, Gyapong J, Osei I et al (2009) IL10 haplotype associated with tuberculin skin test response but not with pulmonary TB. PLoS One 4(5):e5420

Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I et al (2010) Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet 42(9):739–741

van der Spuy GD, Kremer K, Ndabambi SL, Beyers N, Dunbar R et al (2009) Changing *Mycobacterium tuberculosis* population highlights clade-specific pathogenic characteristics. Tuberculosis (Edinb) 89(2):120–125

Vejbaesya S, Chierakul N, Luangtrakool P, Sermduangprateep C (2007) *NRAMP1* and TNF-alpha polymorphisms and susceptibility to tuberculosis in Thais. Respirology 12(2):202–206

Velez DR, Hulme WF, Myers JL, Stryjewski ME, Abbate E et al (2009a) Association of *SLC11A1* with tuberculosis and interactions with *NOS2A* and *TLR2* in African-Americans and Caucasians. Int J Tuberc Lung Dis 13(9):1068–1076

Velez DR, Hulme WF, Myers JL, Weinberg JB, Levesque MC et al (2009b) *NOS2A*, *TLR4*, and *IFNGR1* interactions influence pulmonary tuberculosis susceptibility in African-Americans. Hum Genet 126(5):643–653

Wheeler E, Miller EN, Peacock CS, Donaldson IJ, Shaw MA et al (2006) Genome-wide scan for loci influencing quantitative immune response traits in the Belem family study: comparison of methods and summary of results. Ann Hum Genet 70(Pt 1):78–97

Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL et al (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 40(5):638–645