Published in final edited form as:

J Med Genet. 2011 April ; 48(4): 217–218. doi:10.1136/jmg.2010.087361.

What did we learn from the genome-wide association study for tuberculosis susceptibility?

Hui-Qi Qu¹, Quan Li², Joseph B McCormick¹, and Susan P Fisher-Hoch¹

¹University of Texas Health Science Center at Houston, School of Public Health, Brownsville, Texas, USA

²The McGill University Health Center (Montreal Children's Hospital), Montréal, Québec, Canada

The genetics literature is replete with results of candidate-gene association studies published mostly before the genome-wide association study (GWAS) became the new standard, but also continuing to this date. Their track record for replication is generally poor and discrepancies are typically attributed to differences in populations or environments. Are these results of any value, or should they be summarily discarded? Recent results from a GWAS for tuberculosis (TB) provide data that might put this question in perspective[1].

TB is a serious health issue in the developing world, due to infection by *Mycobacterium tuberculosis* (MTB). Currently, about one third of the world's population is infected by MTB and 1/10 will develop active TB (<http://www.who.int/tb/en/>). Improved understanding of mechanisms of pathogenesis and host resistance is essential for improved control of TB. Human genetics is an indispensable tool for enhancing the understanding of the molecular basis of many common diseases. There is evidence that genetic factors may be involved in susceptibility to TB infection and activation of TB. This can be demonstrated by the fact of obvious ethnic differences. Blacks have about 1.5–2 fold greater risk than Whites as shown by the tuberculin skin test[2] and concordance for TB is 2.5 fold higher among monozygotic than dizygotic twin pairs[3]. A genome-wide association study (GWAS) allows unbiased examination of hundreds of thousands of genetic markers across the human genome for disease association. Because the GWAS is hypothesis-free, it is proving to be a great success in understanding unknown and unexpected molecular aspects of many human common diseases[4].

Thye et al. combined the GWAS results of two large African cohorts, i.e. one cohort from Ghana comprising 1740 controls and 921 cases, and another from The Gambia comprising 1377 controls and 1309 cases[1]. A novel association tagged by a single-nucleotide polymorphism (SNP) rs4331426 at 18q11.2 was identified. This is a gene-poor region, and therefore the functional mechanism of this association is still a puzzle. Besides the 18q11.2 locus, this study identified another 16 loci with interesting *P*-values in the two GWAS cohorts, but falling short of statistical significance in the replication cohorts (Supplementary Table 1 of Ref.1). Interestingly, none of these 17 loci harbours any genes with known function that may participate in host resistance to MTB infection or TB activation. On the

*Correspondence should be addressed to: Hui-Qi Qu, Ph.D., Assistant Professor, University of Texas Health Science Center at Houston, School of Public Health, Brownsville campus, 80 Fort Brown, SPH Bldg., Brownsville, TX 78520, Ph 956 882 7006, Fax 956 882 5152, Huiqi.Qu@uth.tmc.edu.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in Journal of Medical Genetics and any other BMJ PGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence (<http://group.bmj.com/products/journals/instructions-for-authors/licence-forms>).

Conflict of Interest statement: None declared.

other hand, Thye et al. examined the association of 22 TB candidate genes in the GWAS cohorts (Supplementary Table 5 of Ref.1). These 22 genes have experimentally suggested functions in MTB infection or TB activation, and were associated with TB by previous low-throughput association studies. Four of these candidate genes, i.e. the interferon-gamma gene (*IFNG*), the interferon-gamma receptor-1 gene (*IFNGR1*), the interleukin-12 subunit p40 gene (*IL12B*), and the interleukin-12 receptor beta-1 chain *IL12RB1*), have been identified as bearing mutations that can cause Mendelian susceptibility to mycobacterial diseases (MSMD), which suggests essential roles of these genes in MTB infection [5]. In addition, several other genes are known to be important: The *HLA* class II genes *HLA-DQA1* and *HLA-DRB1* have critical roles in Th1 immunity against MTB infection by presenting MTB antigens on the cell surfaces of macrophages that activate naïve CD4⁺ T cells, and induce Th1 response in conjunction with IL-12 [6]; TB activation induced by anti-tumor necrosis factor α therapy suggests the role of the tumor necrosis factor α gene (*TNF*) in TB [7, 8]; the Vitamin D receptor gene (*VDR*) can mediate the innate immunity responses against MTB infection [9]; the nitric oxide synthase 2A gene (*NOS2A*) catalyzes the production of nitric oxide (NO) in macrophages to destroy MTB [10].

However, in spite of the functional importance of the candidate genes, no association signal with statistical significance was identified in the GWAS cohorts. Nominally significant *P*-values were seen, but as each gene region contains multiple SNPs, no *P*-value can clear the multiple testing correction in each locus. There were 11 regions with the minimum *P*-value < 0.05. The smallest *P*-value among all these regions is 0.002. To examine the family-wise significance of association at these loci, we performed a permutation-based simulation experiment, based on The Gambia GWAS data acquired from the Wellcome Trust Case-Control Consortium (WTCCC). This permutation simulation aimed to develop the empirical distribution of statistical significance metrics under the null hypothesis by randomly assigning the phenotypic data to each individual. In our experiment, the genotype data of SNPs in each candidate gene region were extracted. The number of SNPs in each gene region is the same as described in Ref.1. Subsequently, the phenotype of each individual in the GWAS cohort was randomly permuted, and then genetic association was examined in the simulation data. The minimum *P*-value in each gene region was investigated. We performed 200 iterations of the permutation-based simulation. As shown by the simulation results, 65 out of the 200 simulations (32.5%) have ≥ 11 regions with the minimum *P*-value < 0.05 (Fig.1). Therefore, 11 out of the 22 candidate genes with *P*-value < 0.05 have the random probability of 0.325, which is not a small probability event. In addition, among the 200 simulations, there are 90 simulations (45.0%) with the smallest *P*-value < 0.002. Therefore, the probability of having the smallest *P*-value ≤ 0.002 has the random probability of 0.450 which is not a small probability event either. Therefore, our simulation experiment suggests that there is no over-representation of nominal significant results in the 22 TB candidate genes. True TB susceptibility loci among these 22 genes cannot be assumed.

In summary, the study by Thye et al. provides a proof-of-principle that the GWAS approach can be used to understand the molecular mechanisms of TB, as it has been successfully used in other common human diseases. On the other hand, the majority of the genetic associations reported by candidate gene studies in the pre-GWAS era cannot be replicated by GWAS studies [11]. The GWAS study on TB further supports this observation. Much work remains to be done. TB is epidemic in the developing world. Advanced genomic technologies have not been used sufficiently to study this disease. Furthermore, the genetic effect on TB susceptibility is more complicated than many other common human diseases because of the serious confounding effects of environmental factors involved in MTB exposure, infection, and TB activation. More attention on this field is warranted to clarify the molecular mechanisms of this ancient disease, through the application of current research technologies and improved study design.

Acknowledgments

This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113 and 085475.

References

1. Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, Sirugo G, Sisay-Joof F, Enimil A, Chinbuah MA, Floyd S, Warndorff DK, Sichali L, Malema S, Crampin AC, Ngwira B, Teo YY, Small K, Rockett K, Kwiatkowski D, Fine PE, Hill PC, Newport M, Lienhardt C, Adegbola RA, Corrah T, Ziegler A, Morris AP, Meyer CG, Horstmann RD, Hill AVS. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat Genet.* 2010; 42(9):739–741. [PubMed: 20694014]
2. Stead WW, Senner JW, Reddick WT, Lofgren JP. Racial differences in susceptibility to infection by *Mycobacterium tuberculosis*. *N Engl J Med.* 1990; 322(7):422–427. [PubMed: 2300105]
3. Comstock GW. Tuberculosis in twins: a re-analysis of the Proffit survey. *Am Rev Respir Dis.* 1978; 117(4):621–624. [PubMed: 565607]
4. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest.* 2008; 118(5):1590–1605. [PubMed: 18451988]
5. Casanova JL, Blanche S, Emile JF, Jouanguy E, Lamhamedi S, Altare F, Stephan JL, Bernaudin F, Bordigoni P, Turck D, Lachaux A, Albertini M, Bourrillon A, Dommergues JP, Pocard MA, Le Deist F, Gaillard JL, Griscelli C, Fischer A. Idiopathic disseminated bacillus Calmette-Guerin infection: a French national retrospective study. *Pediatrics.* 1996; 98(4 Pt 1):774–778. [PubMed: 8885960]
6. North RJ, Jung Y-J. Immunity to Tuberculosis. *Annual Review of Immunology.* 2004; 22(1):599–623.
7. Lienhardt C, Bennett S, Del Prete G, Bah-Sow O, Newport M, Gustafson P, Manneh K, Gomes V, Hill A, McAdam K. Investigation of Environmental and Host-related Risk Factors for Tuberculosis in Africa. I. Methodological Aspects of a Combined Design. *Am J Epidemiol.* 2002; 155(11):1066–1073. [PubMed: 12034586]
8. Bennett S, Lienhardt C, Bah-Sow O, Gustafson P, Manneh K, Del Prete G, Gomes V, Newport M, McAdam K, Hill A. Investigation of Environmental and Host-related Risk Factors for Tuberculosis in Africa. II. Investigation of Host Genetic Factors. *Am J Epidemiol.* 2002; 155(11):1074–1079. [PubMed: 12034587]
9. Liu PT, Stenger S, Li H, Wenzel L, Tan BH, Krutzik SR, Ochoa MT, Schaubert J, Wu K, Meinken C, Kamen DL, Wagner M, Bals R, Steinmeyer A, Zugel U, Gallo RL, Eisenberg D, Hewison M, Hollis BW, Adams JS, Bloom BR, Modlin RL. Toll-Like Receptor Triggering of a Vitamin D-Mediated Human Antimicrobial Response. *Science.* 2006; 311(5768):1770–1773. [PubMed: 16497887]
10. Nicholson S, Bonecini-Almeida Mda G, Lapa e Silva JR, Nathan C, Xie QW, Mumford R, Weidner JR, Calaycay J, Geng J, Boechat N, Linhares C, Rom W, Ho JL. Inducible nitric oxide synthase in pulmonary alveolar macrophages from patients with tuberculosis. *J Exp Med.* 1996; 183(5):2293–2302. [PubMed: 8642338]
11. Kavvoura FK, Ioannidis JP. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. *Hum Genet.* 2008; 123(1):1–14. [PubMed: 18026754]

Fig.1.

Number of loci with the minimum $P < 0.05$ in the permutation-based simulation experiment. X-axis represents 200 simulations; Y-axis represents the number of loci with the minimum $P < 0.05$. The horizontal line corresponds to 11 loci with $P < 0.05$. In this experiment, 65 out of the 200 simulations (32.5%) have ≥ 11 regions with the minimum P -value < 0.05 .