Published in final edited form as:

J Infect Dis. 2008 January 1; 197(1): 4–6. doi:10.1086/524118.

Genes Influencing Susceptibility to Infection

William A. Petri Jr.¹, Beth D. Kirkpatrick², Rashidul Haque³, and Priya Duggal⁴

¹University of Virginia, Charlottesville ²University of Vermont, Burlington ³International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh ⁴National Human Genome Research Institute, National Institutes of Health, Baltimore, Maryland

In this issue of the *Journal*, an article [1] reports the results of a human genome scan for quantitative trait loci (QTLs) influencing susceptibility to *Ascaris lumbricoides*. The work is important both for the insight that it provides into how the body resists heavy infection with this important soil-transmitted intestinal helminth and for what it promises in this new era of investigation and discovery concerning which human genes influence susceptibility to infection.

Every infectious-diseases physician has puzzled over the differing sensitivities of patients to infection or to the medications used to treat infection. Why do some HIV-infected patients progress slowly to disease? How do persons with positive tuberculin test results not develop active tuberculosis? What predisposes patients to adverse effects from antibiotics? The discipline of genetic epidemiology now has both the technology and statistical rigor to answer many of these questions. Some of the most notable answers to date include the resistance to severe *Plasmodium falciparum* malaria provided by the sickle cell trait [2] and to *Plasmodium vivax* malaria due to Duffy blood group negativity [3], the 32-bp deletion in the CCR5 chemokine receptor that renders homozygous individuals resistant to infection with CCR5-tropic HIV [4], and the association of the variant in the peptide-binding groove of hsp70 with abacavir hypersensitivity [5].

Infections provide a remarkable window into the extensive genetic diversity of humans and, through evolution, are likely responsible for a substantial fraction of that diversity. Humans not only differ by ~12 million single-nucleotide polymorphisms (SNPs) but also have extensive variation in the numbers of copies of individual genes [6]. Common genetic polymorphisms influence susceptibility to disease, whether it be as a result of infection, inflammation, cancer, or degenerative diseases. For example, the HLA class I and class II alleles, responsible for antigen presentation to CD8⁺ and CD4⁺ T cells, respectively, are the most genetically diverse region in the human genome. Heterozygosity at HLA alleles in general correlates with resistance to infection, likely because heterozygosity provides additional variants of class I and class II molecules to present microbial antigens to the acquired immune system. There are, in addition, associations between HLA haplotypes and resistance to specific pathogens, including *Mycobacterium tuberculosis*, HIV, *Mycobacterium leprae* [7], and, in our own research, *Entamoeba histolytica* [8]. Importantly, there are also numerous and common (>10% frequency in the population) polymorphisms in human immune response genes that are associated with, or likely will be found to be associated with, infection and inflammation [7, 9].

© 2007 by the Infectious Diseases Society of America. All rights reserved.

Reprints or correspondence: Dr. William A. Petri, Jr., Div. of Infectious Diseases and International Health, University of Virginia, PO Box 801340, Charlottesville, VA 22908-1340 (wap3g@virginia.edu).

Potential conflicts of interest: none reported.

This study by Williams-Blangero et al. is remarkable for being one of the first genomewide linkage scans for an infectious disease. *A. lumbricoides* joins leishmaniasis [10, 11], schistosomiasis [12], leprosy [13, 14], and tuberculosis [15], diseases for which families have been used to identify regions in the human genome that may control susceptibility to disease. Individuals infected with *A. lumbricoides* vary in the number of parasites infecting the intestine, with a minority having a high worm burden [16]. Those with a large number of worms in their intestine are more likely to suffer from malnutrition and intestinal obstruction [16]. In Williams-Blangero et al.'s study, parasite load was measured as the number of eggs per gram of feces in the Jirel population in Nepal. The Jirel population is small, geographically isolated, and stable, with a low migration rate. The relatedness of all individuals was delineated with a single pedigree. This extended pedigree allowed the study of related individuals living in different households and unrelated individuals living in the same household, permitting the delineation of environmental from genetic effects on the phenotype (worm burden). This allowed estimation that between 30% and 50% of the worm burden was due to human genetics [17]. Nine Jirel villages participated in the study, with sampling conducted in 42% of the ~3000 villagers.

A 10-cM whole-genome scan of 1258 members of the Jirel population is reported in this article, refining with greater statistical power an earlier study identifying genetic regions associated with the intensity of *Ascaris* infection [18]. A variance component linkage analysis was used, which measures the covariance for a given genetic marker among individuals in a family for the QTL (in this case, worm burden). Three QTLs exhibited genomewide significance: on chromosome 13 near a major candidate gene, *TNFSF13B*, which is involved in the regulation of B cell activation and immunoglobulin secretion; chromosome 8 in a 33-Mb region encompassing 330 known or predicted genes; and chromosome 11 in a region with 54 known genes. The major task that lies ahead for these investigators will be to identify the specific genes at these 3 loci that influence the intensity of *Ascaris* infection. The benefit from identifying these genes promises to be 2-fold: first, a better understanding of the pathogenic interaction of the parasite and host that could provide new approaches to the treatment or prevention of infection; and second, a more general elucidation of the regulation and function of the innate and acquired immune responses of the intestine.

Technical advances in the facility and cost of human genetic analysis are catalyzing what will be an explosive growth in whole-genome association, linkage, and candidate gene analyses. Sequencing of the human genome, cataloguing of SNPs between individuals, and the HapMap project, which provided information on the patterns of inheritance of SNPs (haplotypes) in 4 different populations, have enabled the application of human genetics to infectious diseases. New technologies that now allow the relatively low-cost genotyping of 500,000 to 1 million SNPs per individual and, as important, the design of software programs to analyze the data and relate it to phenotypes have removed the last obstacle from genetic analysis of infection.

The last several months have seen a flurry of publications of whole-genome association studies for chronic diseases, including type I and II diabetes, Crohn disease, hypertension, and coronary artery disease [19–22]. Insights into pathogenesis from these studies give a glimpse of the power of whole-genome scans in the study of infectious diseases. For example, genetic risk factors for type I diabetes are involved in immune regulation, supporting a role for an autoimmune reaction against the pancreatic β cell in causation. In Crohn disease, a theme of dysregulated immune response to the enteric bacterial flora is emerging, with the discovery of associations with *CARD15* (also known as *NOD2*, involved in intracellular recognition of bacterial products), interleukin-23 receptor, *IRGM* (immunity-related guanosine triphosphatase, a guanosine triphosphate-binding protein also involved in

the elimination of intracellular bacteria), *MST1* (macrophage-stimulating 1, which regulates peritoneal macrophage activity), and *PTPN2* (protein tyrosine phosphatase, nonreceptor type 2, which negatively regulates inflammatory responses directed by T cells) [15].

The role of the infectious-diseases investigator in genetic studies is central. Determining what the important questions are, identifying the populations to be studied, and defining the infectious phenotype are paramount. The most sophisticated of genetic and epidemiological tools are useless when applied to a poorly defined phenotype. The infection measure could be qualitative (infection vs. no infection or localized vs. disseminated infection) or quantitative (for example, the “load” of the microorganism, whether it be viral load with hepatitis C virus or egg count with *Ascaris*). Because large populations are needed to detect small effects from multiple genes, identification of the patient populations to be studied is also a critical role for the infectious-diseases investigator.

Limitations of current approaches center on the fact that susceptibility to infectious diseases is polygenic; the contribution of many genes means that the contribution of each to susceptibility will be modest (odds ratios of <2) and that large sample sizes (a minimum of 1000 case patients and 1000 control subjects for a genomewide association) or many pedigrees for linkage studies will be required. Moreover, although the expense of genotyping is dropping, for the most part the cost still outstrips the budget of a National Institutes of Health (NIH) R01 grant, requiring special funding mechanisms, such as those available through the Center for Inherited Disease Research of the NIH or the Genetic Association Information Network. Candidate gene analysis, which can be less costly, requires preconceived assumptions about which gene is important, thereby negating the power of genetics to use experiments in nature (genetic polymorphisms) to teach us which genes are important. Another problem is that SNPs associated with a phenotype are most likely not to result in a stop codon or even an amino acid substitution but are to be found in noncoding DNA outside of exons. Thus, the SNP will be of unknown biological significance until its potential effects on gene expression are understood. Some regions of the human genome are in large areas of linkage disequilibrium. In the absence of frequent recombination within an area, it may not be possible with genetics alone to identify the responsible gene or polymorphism. Finally, validation of studies may be difficult because of small sample sizes, potential environmental confounders, and even region-specific polymorphisms in the microbe.

Validation or replication of a study is essential to understand its significance and to sort true-positive from false-positive associations [23]. Identification of the true causal SNP is foremost (i.e., which nearby SNPs are also in linkage disequilibrium with the phenotype and which SNP shows the greatest association with the phenotype). The ability to replicate a study in a different population starts with a rigorous study design and a clear-cut definition of the phenotype, with a careful matching of case patients to control subjects. Quality-control methods used for genotyping (bar coding to limit sample misidentification, similar DNA quality and concentration), testing for Mendelian consistency if related individuals are studied, measurement of deviations from Hardy-Weinberg equilibrium in case patients and control subjects that could reflect technical issues in genotyping or population stratification, and validation of most critical results with an independent genotyping method are all important. Statistical analyses need to take into account correction of tests of statistical significance for multiple testing, with very low *P* values ($P < 5 \times 10^{-7}$) being the strongest evidence for association in genomewide studies. The biological plausibility of the hypothesis for why the gene could be involved and, ideally, animal model data for its involvement are supportive. Of course, not all studies are validated, and it is important that scientific journals also be willing to publish articles reporting failures to replicate genetic-association studies.

Why should we do this work? Infection is the result of the interaction between host, parasite, and environment, superimposed on the continual evolution and adaptation of the microbial and human genomes to each other. Already the study of genes associated with susceptibility to infection has given us new drugs for HIV infection, identified patients at risk of hypersensitivity reactions to an antiretroviral drug, and provided insight into the interaction of the malaria parasite with the red blood cell. We are entering an era in which human genetic epidemiology will help us to care not only for the Jirel villager east of Kathmandu with *Ascaris* infection but also for the individual living with HIV/AIDS in Charlottesville.

Acknowledgments

Financial support: National Institutes of Health (grant AI 43596).

References

1. Williams-Blangero S, VandeBerg JL, Subedi J, Jha B, Corrêa-Oliveira R, Blangero J. Localization of multiple quantitative trait loci influencing susceptibility to infection with *Ascaris lumbricoides*. *J Infect Dis*. 2008; 197:66–71. (in this issue). [PubMed: 18171287]
2. Allison AC. Protection afforded by sickle-cell trait against subtertian malarial infection. *Br Med J*. 1954; 1:290–294. [PubMed: 13115700]
3. Tournamille C, Colin Y, Catron JP, et al. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet*. 1995; 10:224–228. [PubMed: 7663520]
4. Lederman MM, Penn-Nicholson A, Cho M, Mosier D. Biology of CCR5 and its role in HIV infection and treatment. *JAMA*. 2006; 296:815–826. [PubMed: 16905787]
5. Martin AM, Nolan D, Gaudieri S, et al. Predisposition to abacavir hypersensitivity conferred by HLA-B*5701 and a haplotypic Hsp70-Hom variant. *Proc Natl Acad Sci USA*. 2004; 101:4180–4185. [PubMed: 15024131]
6. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007; 315:848–853. [PubMed: 17289997]
7. Hill AVS. Aspects of genetic susceptibility to human infectious diseases. *Annu Rev Genet*. 2006; 40:469–486. [PubMed: 17094741]
8. Duggal P, Haque R, Roy S, et al. Influence of human leukocyte antigen class II alleles on susceptibility to *Entamoeba histolytica* infection in Bangladeshi children. *J Infect Dis*. 2004; 189:520–526. [PubMed: 14745711]
9. Burgner D, Jamieson SE, Blackwell JM. Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? *Lancet Infect Dis*. 2006; 6:653–663. [PubMed: 17008174]
10. Jamieson SE, Miller EN, Peacock CS, et al. Genome-wide scan for visceral leishmaniasis susceptibility genes in Brazil. *Genes Immun*. 2007; 8:84–90. [PubMed: 17122780]
11. Miller EN, Fadl M, Mohamed HS, et al. Y chromosome lineage- and village-specific genes on chromosomes 1p22 and 6q27 control visceral leishmaniasis in Sudan. *PLoS Genet*. 2007; 3:e71. [PubMed: 17500593]
12. Marquet S, Abel L, Hillaire D, et al. Genetic localization of a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31–q33. *Nat Genet*. 1996; 14:181–184. [PubMed: 8841190]
13. Mira MT, Alcais A, Nguyen VT, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature*. 2004; 427:636–640. [PubMed: 14737177]
14. Siddiqui RM, Meisner S, Tosh K, et al. A major susceptibility locus for leprosy in India maps to chromosome 10p13. *Nat Genet*. 2001; 27:439–441. [PubMed: 11279529]
15. Baghdadi JE, Orlova M, Alter A, et al. An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. *J Exp Med*. 2006; 203:1679–1684. [PubMed: 16801399]
16. Crompton DWT. *Ascaris* and ascariasis. *Adv Parasitol*. 2001; 48:285–373. [PubMed: 11013758]

17. Williams-Blangero S, Subedi J, Upadhayay PR, et al. Genetic analysis of susceptibility to infection with *Ascaris lumbricoides*. Am J Trop Med Hyg. 1999; 60:921–926. [PubMed: 10403321]
18. Williams-Blangero S, VandeBerg JL, Subedi J, et al. Genes on chromosomes 1 and 13 have significant effects on *Ascaris* infection. Proc Natl Acad Sci USA. 2002; 99:5533–5538. [PubMed: 11960011]
19. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]
20. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 2007; 316:1336–1341. [PubMed: 17463249]
21. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007; 316:1331–1336. [PubMed: 17463246]
22. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007; 316:1341–1345. [PubMed: 17463248]
23. NCI-NHGRI Working Group. Replicating genotype-phenotype associations. Nature. 2007; 447:655–660. [PubMed: 17554299]