

SCIENCE COMMENTARY

Modern genetics and leprosy susceptibility

JENEFER M. BLACKWELL

Glaxo Professor for Molecular Parasitology and Director, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2XY, UK

With the advent of the genome projects, both human and pathogen, our whole approach to studying the role of host/pathogen interactions in determining infectious disease susceptibility has undergone a remarkable transformation.

The first impact that modern genomics had on analysing genetic susceptibility to infection was the ability to use physical mapping data and 'big DNA' (yeast and bacterial artificial chromosomes, cosmids) clone contigs to pinpoint a gene and identify it on the basis of its genetic map location. The first elegant example of this was the positional cloning¹ of the murine natural resistance associated macrophage protein 1 (*Nramp1*) gene, variously known in earlier studies as *Ity*, *Lsh* and *Bcg* for its role in controlling innate resistance and susceptibility to *Salmonella typhimurium*, *Leishmania donovani* or *Mycobacterium bovis* BCG infection, respectively. In this case, ability to clone the gene on the basis of its map location was vitally dependent on the ability to 'genotype' individual mice according to their phenotypic response to infection. For all three infections (e.g. reference²), segregation of disease phenotypes in the F2 and backcross progeny made between resistant and susceptible inbred mouse strains followed perfect Mendelian inheritance, with resistance almost completely dominant and individual mice falling clearly into non-overlapping resistant or susceptible phenotypes on the basis of liver or spleen pathogen counts. The identification of *Ity/Lsh/Bcg* gene was highly relevant to human disease susceptibility, since we had all been anxious to know whether a human homologue would also control susceptibility to leprosy and/or tuberculosis. In humans, NRAMP1 is globally linked or associated with tuberculosis^{3–8} and is also linked to leprosy susceptibility⁹ and ability to mount a Mitsuda type skin test reaction to leprosy antigens¹⁰ in Vietnam. These studies provided strong support for the concept that murine models of disease can identify disease susceptibility genes that will be important in humans.

The Vidal study¹ was a model of perfection for its time. It involved extensive studies using exon trapping and analysis of tissue expression profiles of candidate exons/genes within the narrowed down interval known to contain the gene in order finally to identify the macrophage-expressed *Nramp1* protein, now renamed solute carrier family 11a member 1 (*Slc11a1*) on the basis of its function as a divalent cation/proton antiporter.¹¹ Now that the first draft of the human sequence is published,^{12,13} we should be able directly to identify the chromosome linked to susceptibility to infection in humans and simply inspect the

Correspondence to: J. H. Blackwell (e-mail: jmb37@cus.cam.ac.uk, Phone: +44-1223-336947, Fax: +44-1223-331206)

region under the linkage curve to identify the disease-associated gene. Or is it really that simple?

The first example of direct identification of a human infectious disease susceptibility gene using modern genomics was another very elegant study in which egg output for schistosomiasis was measured in large multicase pedigrees in Brazil and shown to segregate as a Mendelian single gene trait.¹⁴ DNAs from a subset of 11 informative complex multicase pedigrees was used in a 10–15 centimorgan (cM) microsatellite genome scan, i.e. the DNA for 142 individuals in these pedigrees were genotyped for ~300 microsatellite markers across the genome, and multipoint linkage analyses performed to locate the single major gene that had been predicted. Luckily, only a single major linkage peak was identified with a peak parametric lod score (log likelihood for linkage) of 4.74 (i.e. >10,000:1 odds in favour of linkage). However, although the peak lod score was found near the locus containing the gene (CSF1R) encoding the receptor for colony stimulating factor 1, the linked region contained a host of other potential candidate genes including those encoding interleukin (IL) 4, IL5, IL9, IL13, and interferon regulatory factor 1 (IRF1), for all of which a plausible case could be built for their involvement in determining susceptibility to schistosome infection. Unfortunately, the ability to fine map the region and pinpoint the real susceptibility gene relies on much larger samples sizes and linkage disequilibrium mapping, and such studies are even now still in progress.

The second published attempt at mapping infectious disease susceptibility genes via a genome scan was for tuberculosis, this time using non-parametric linkage analysis of affected sibling pairs.¹⁵ Initially, 92 affected sibling pairs plus parents collected from The Gambia and South Africa were genotyped for 299 microsatellite markers. Seven regions of the genome showed maximum lod scores (MLS) >1 ($P=0.05$). Twenty-two markers from these regions were typed in a second set of 82 affected sibling pairs plus parents, with only chromosomes 15q (MLS = 2) and Xq (MLS = 1.77) retaining lod scores suggestive of linkage but this time at odds $\leq 100:1$. This does not achieve the levels of significance recommended for a genome-wide scan,¹⁶ probably reflecting the fact that this sample size is not sufficient to detect linkage for genes contributing to the complex disease phenotype of pulmonary tuberculosis. It is likely that the search for susceptibility genes for tuberculosis will require much larger sample sizes, and much more effort, to identify the multiple genes that are likely to be contributing to overall disease susceptibility.

The third published example of genome scanning to identify susceptibility genes for infectious disease is the recently reported study on leprosy.¹⁷ In this case, the authors were more successful in ultimately identifying a region on chromosome 10p13 with a non-parametric MLS of 4.09 ($P=0.000007$) (Figure 1). To obtain this, 388 markers were initially typed on 103 affected sibling pairs plus parents. Twenty-eight regions were identified with MLS >1.0. Thirty-seven markers from these regions were typed in an additional 142 affected sibling pair families. In the overall analysis, only chromosome 10p13 showed a multipoint MLS >3.0. A further eight markers typed in this region yielded the final MLS of 4.09. Examination of public domain databases provides a single obvious candidate, the gene encoding mannose receptor C type 1 (MRC1), that maps under the 10p13 interval¹⁸ positive for linkage to leprosy susceptibility. Recent studies¹⁹ provide new structural insights into the molecular deciphering of mycobacterial lipoglycan binding to C-type lectins such as the macrophage mannose receptor, providing a rational basis to target MRC1 as the most appropriate candidate within the linkage interval on 10p13. Polymorphisms within the MRC1 gene are now being analysed in allelic association studies.

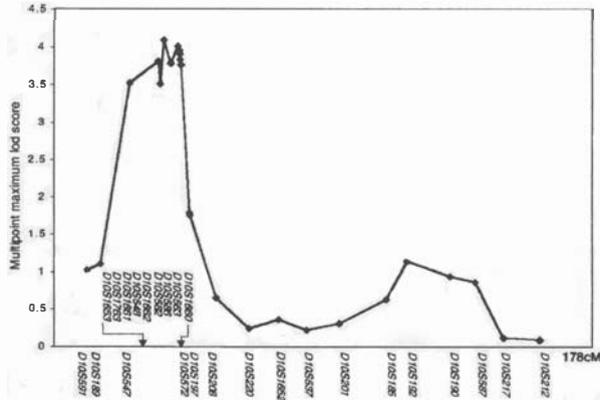


Figure 1. Maximum likelihood multipoint map for 25 microsatellite markers on chromosome 10, including 8 markers flanking *D10S548*. The maximum multipoint MLS for the region is 4.09 (P value = 0.000007) and corresponded to *D10S166*. Reproduced with permission from Siddiqui *et al.*¹⁷

The results of this Indian study are interesting in relation to previous studies suggesting a role for HLA in determining leprosy susceptibility, including many studies based in India. In the genome scan, markers within HLA were not positive for linkage. Hence, the authors conclude that HLA is not a major susceptibility gene in India.¹⁷ In our own studies of leprosy in Brazil, we have found a peak parametric lod score for linkage to HLA of 5.78 ($P=0.00000025$) in 73 two and three generation extended multicase pedigrees analysed using combined segregation and linkage analysis,²⁰ similar to that used in the schistosomiasis study. Affected sib pair analysis of the same families for 63 nuclear families with 138 affected sibling pairs (equivalent to 91 affected sib pairs plus parents after weighting for multiple affected sib pairs in a family) provides an MLS of only 2.12 ($P=0.002$), demonstrating the increased power provided by the larger two and three generation pedigrees. Nevertheless, the locus-specific risk for siblings of patients (λ_s) for the peak of linkage at HLA (calculated from the ratio of the expected proportion of affected sibling pairs sharing zero alleles identical by descent (0.25) and the observed proportion²¹) based on this set of affected sibling pairs is 1.79, compared to a value of 1.66 for the peak of linkage on chromosome 10p13 in the Indian study (245 independent affected sibling pairs¹⁷). Hence, we may conclude that HLA makes an equivalent if not greater contribution to the total genetic component of susceptibility to leprosy *per se* in Brazil as the chromosome 10p13 gene does in India, highlighting the fact that different genes may contribute to disease susceptibility in different geographic regions. Alternatively, or in addition, the Indian study took place in an area where the incidence of paucibacillary leprosy was high and none of the families had only multibacillary affected siblings, whereas the Brazilian study took place in an area with high incidence of multibacillary leprosy. Interestingly, detailed investigation of HLA in a similar sample of 98 multicase pedigrees with pulmonary tuberculosis from the same region of Brazil failed to show any evidence for linkage or allelic association (I. Davidson, J. M. Blackwell & M.-A. Shaw, unpublished observation). For leprosy, allelic association was highly significant in both the HLA DR/DQ ($P=0.001-0.024$) and TNFA ($P=3.6 \times 10^{-5}$) gene regions.²⁰ Hence, these two important mycobacterial infections of humans show quite distinct genetic regulation, particularly with respect to the role of the major histocompatibility complex in this region of Brazil.

So what can we learn from these new genome-based family studies of susceptibility to infectious disease? Firstly, two of the three studies performed to date have identified major genes controlling disease susceptibility that conform to the genome-wide criteria for statistically significant linkage.¹⁶ In both cases, obvious candidate genes have been identified under the region of linkage. For schistosomiasis, a candidate rich region was identified, which has made it difficult to identify the specific gene regulating egg output. For leprosy, there appears at present to be just one obvious candidate (MRC1) under the linked region in India, but confirmation of its role has yet to be published and its global relevance to leprosy susceptibility requires further investigation. For tuberculosis, genetic regulation appears more complex, and will require further work to uncover the multiple genes and mechanisms likely to regulate susceptibility. As for schistosomiasis, this might involve careful stratification of the disease phenotype, with better recording of clinical criteria across the spectrum of disease. The same is true for leprosy, where the studies in both India and Brazil have focussed on areas differing in incidence of paucibacillary versus multibacillary leprosy. Nevertheless, these ground-breaking studies applying modern whole-genome approaches to infectious disease susceptibility promise to yield rich rewards in furthering our understanding of the genes and mechanisms that regulate infectious disease susceptibility.

References

- ¹ Vidal SM, Malo D, Vogan K *et al.* Natural resistance to infection with intracellular parasites: isolation of a candidate for *Bcg*. *Cell*, 1993; **73**: 469–485.
- ² Bradley DJ. Genetic control of natural resistance to *Leishmania donovani*. *Nature*, 1974; **250**: 353–354.
- ³ Shaw M-A, Collins A, Peacock CS *et al.* 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*, 1997; **78**: 35–45.
- ⁴ Bellamy R, Ruwende C, Corrah T *et al.* Variation in the NRAMP1 gene is associated with susceptibility to tuberculosis in West Africans. *N Engl J Med*, 1998; **338**: 640–644.
- ⁵ Ryu S, Park YK, Bai GH *et al.* 3'UTR polymorphisms in the NRAMP1 gene are associated with susceptibility to tuberculosis in Koreans. *Int J Tuberc Lung Dis*, 2000; **4**: 577–580.
- ⁶ Greenwood CM, Fujiwara TM, Boothroyd LJ *et al.* Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. *Am J Hum Genet*, 2000; **67**: 405–416.
- ⁷ Gao PS, Fujishima S, Mao XQ *et al.* Genetic variants of NRAMP1 and active tuberculosis in Japanese populations. International Tuberculosis Genetics Team. *Clin Genet*, 2000; **58**: 74–76.
- ⁸ Cervino AC, Lakiss S, Sow O, Hill AV. Allelic association between the NRAMP1 gene and susceptibility to tuberculosis in Guinea-Conakry. *Ann Hum Genet*, 2000; **64**: 507–512.
- ⁹ Abel L, Sanchez FO, Oberti J *et al.* Susceptibility to leprosy is linked to the human NRAMP1 gene. *J Infect Dis*, 1998; **177**: 133–145.
- ¹⁰ Alcais A, Sanchez FO, Thuc NV *et al.* Granulomatous reaction to intradermal injection of lepromin (Mitsuda reaction) is linked to the human NRAMP1 gene in Vietnamese leprosy sibships. *J Infect Dis*, 2000; **181**: 302–308.
- ¹¹ Goswami T, Bhattacharjee A, Babal P *et al.* Natural resistance-associated macrophage protein 1 is an H+/bivalent cation antiporter. *Biochem J*, 2001; **354**: 511–519.
- ¹² Lander ES, Linton LM, Birren B *et al.* Initial sequencing and analysis of the human genome. *Nature*, 2001; **409**: 860–921.
- ¹³ Venter JC, Adams MD, Myers EW *et al.* The sequence of the human genome. *Science*, 2001; **291**: 1304–1351.
- ¹⁴ Marquet S, Laurent A, 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.
- ¹⁵ Bellamy R, Beyers N, McAdam KP *et al.* Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. *Proc Natl Acad Sci USA*, 2000; **97**: 8005–8009.
- ¹⁶ Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet*, 1995; **11**: 241–247.
- ¹⁷ Siddiqui MR, Meisner S, Tosh K *et al.* A major susceptibility locus for leprosy in India maps to chromosome 10p13. *Nat Genet*, 2001; **27**: 439–441.
- ¹⁸ Eichbaum Q, Clerc P, Bruns G *et al.* Assignment of the human macrophage mannose receptor gene (MRC1) to 10p13 by in situ hybridization and PCR-based somatic cell hybrid mapping. *Genomics*, 1994; **22**: 656–658.

- ¹⁹ Nigou J, Vercellone A, Puzo G. New structural insights into the molecular deciphering of mycobacterial lipoglycan binding to C-type lectins: lipoarabinomannan glycoform characterization and quantification by capillary electrophoresis at the subnanomole level. *J Mol Biol*, 2000; **299**: 1353–1362.
- ²⁰ Shaw M-A, Donaldson IJ, Collins A *et al.* Association and linkage of leprosy phenotypes with HLA class II and tumour necrosis factor genes. *Genes Immunity*, 2001 (in press).
- ²¹ Risch N. Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet*, 1987; **40**: 1–14.