GENETICS OF SUSCEPTIBILITY TO INFECTIOUS DISEASES: TUBERCULOSIS AND LEPROSY AS EXAMPLES

SANDRINE MARQUET AND ERWIN SCHURR

McGill Centre for the Study of Host Resistance, McGill University Health Centre Research Institute, Montreal General Hospital, Montreal, Québec, Canada

This paper is available online at http://dmd.aspetjournals.org

In the majority of infectious diseases only a proportion of individuals exposed to a pathogen become infected and develop clinically evident disease. At least in part, this interindividual variability is determined by the combined effect of host proteins encoded by a series of genes that control the quantity and quality of host-parasite interaction and host immune responses. Identification of the most important host susceptibility/resistance genes will allow a better understanding of infectious disease pathogenesis and likely facilitate the development of new therapeutic strategies.

Several approaches can be used to map and identify a host infectious disease susceptibility gene. Three of the most widely used strategies, i.e., mouse models, candidate gene approach, and genome scanning, are briefly presented. To date, at least 11 genes have been implicated in susceptibility/resistance to mycobacterial infection and a short discussion of the experiments implicating individual genes in infectious disease susceptibility is given.

Mouse Models

One approach to identify human disease resistance and susceptibility genes is to identify murine resistance/susceptibility genes. In this strategy, it is assumed that the basic pathology of the infectious disease is similar in the animal model and the human host. Consequently, orthologous genes in mouse and humans are assumed to be important for variable susceptibility/resistance to infection with the same pathogen. Experimental models present several advantages, including the ease of control of the environment, the ready access to strains of defined and genetically homogenous backgrounds, the availability of genetically engineered animals, and the breeding at will of appropriately chosen progenitor strains. A well known example for a susceptibility gene that has been identified in the mouse is the “natural resistance associated macrophage protein 1” (Nramp1). Natural resistance to infection with several intracellular pathogens belonging to the genera Mycobacterium, Leishmania, and Salmonella has been shown to be under control of a single Gl69D amino acid substitution in the Nramp1 protein (Vidal et al., 1995, 1996). A potential problem of the cross-species homology approach is that allelic variants of orthologues can be highly divergent among two species. Hence, the most powerful use of mouse models is the identification of genes and their respective biochemical pathways that are involved in disease susceptibility rather than the identification of specific susceptibility/resistance gene variants.

Candidate Gene Approach

Candidate genes are generally selected on the basis of their known or speculated relevance to disease pathogenesis and the presence of intragenic polymorphisms of possible biological significance. Candidate genes can also be derived based on experiments in mouse models of infectious diseases thereby exploiting the identification of murine resistance/susceptibility loci. Variants within a candidate gene can be analyzed in linkage studies (family studies) and/or in association studies (case-control studies), but in most cases, association studies are used to study the possible biological relevance of polymorphisms in specific candidate genes. With a growing number of gene polymorphisms appearing in public databases each month, the candidate gene strategy has gained tremendously in popularity. Nevertheless, problems remain because it is unlikely that all genes important for susceptibility can be found a priori, and genes with major effects but unknown function can easily be missed. The interpretation of positive results on genetic associations with infectious diseases is frequently complicated by the lack of appropriate corrections for multiple comparisons. Moreover, undetected population admixture or poor choice of control populations probably explain part of the difficulties in reproducing significant marker disease associations.

An alternative approach that has been recently proposed is to scan the whole human genome with a large number of single nucleotide polymorphisms for whole genome association studies. The power to detect marker-phenotype associations following this strategy is presently a matter of controversy. However, it is clear that several million genotypes will need to be generated for such studies. Present genotyping techniques cannot be used for such a task, and until new technologies become available, family studies offer a more efficient and feasible strategy.

Total Genome Scanning

In this approach, a large number of microsatellites (around 300) evenly spaced across the whole genome are used for linkage analysis employing families with multiple sibs affected by the studied disease. Two types of linkage analysis can be performed: parametric and nonparametric analysis. Parametric linkage analysis by the lod score method requires a defined model specifying the relationship between the phenotype and the factors (environmental and genetic), which have an effect on phenotype expression. For example, such a model can be provided by complex segregation analysis. Genetic linkage analysis tests whether the marker segregates with the disease in pedigrees with multiple affected according to a Mendelian mode of inheritance. The test is formulated as logarithm of the ratio L(θ)/L(θ = 0.5) or lod score, i.e., the likelihood of observing the experimentally determined segregation pattern at a given recombination
frequency $\theta$ compared with the likelihood of the same segregation pattern in the absence of linkage. The objective of parametric linkage analysis is to estimate the recombination frequency ($\theta$) and to test whether $\theta$ is less than 0.5, which is the case when two loci are genetically linked. The nonparametric approach evaluates the statistical significance of excess allele sharing for specific markers among affected sibs and does not require information about the mode of disease inheritance. The genome wide search has the advantage that no a priori knowledge of the structure or function of susceptibility genes is required. Hence, this approach provides the possibility of identifying genes that modulate susceptibility to infectious diseases that had previously not been suspected of playing such a biological role. To date, no final results of a completed genome scan in human mycobacterial diseases have been published. However, several such genome scans are in progress.

**Leprosy (HLA, NRAMP1, TNFA, VDR).** For centuries, leprosy has been associated with an enormous social stigma partly rooted in the claim that leprosy “runs” in certain families. The social suffering of affected and unaffected members of “lepromatous” families is a powerful reminder of the occasional treacherous area in which genetic research can lead. Today we know that environmental, pathogen, and host factors are important for the spread of an infectious disease through an exposed population. For leprosy, four genes have been implicated in modulating susceptibility. To identify these genes, association and linkage studies have been used.

**HLA.** There is strong epidemiological evidence that genetic factors influence susceptibility to leprosy per se and to leprosy type. The role of major histocompatibility complex (MHC) polymorphisms in the determination of leprosy type, i.e., lepromatous or tuberculoid leprosy, has been found in several linkage and association studies, but in most studies only weak linkage or association was detected. This lack of power in association studies is due at least in part to the relatively small sample sizes used to detect allelic association and the high diversity of HLA alleles, which requires multiple allele testing resulting in correction for multiple comparisons and concomitant loss of power. However, a number of case-control studies of both polar types of leprosy and HLA have shown a consistent HLA-DR2 association particularly in Asian populations (Table 1) (van Eden et al., 1980; Todd et al., 1990; Rani et al., 1993). This association has been reconfirmed more recently in an Indian population comprised of 121 patients with lepromatous leprosy, 107 patients with tuberculoid leprosy, and 160 control subjects (Roy et al., 1997) and in a Brazilian population (Visentainer et al., 1997). Further molecular analysis identified specific mutations in pocket 4 of the DRB1-encoded class II molecule that are associated with increased susceptibility to tuberculoid leprosy (Zerva et al., 1996). However, these MHC effects seem not sufficient to explain the entire host genetic susceptibility to leprosy, and several studies were conducted on non-class I or -class II MHC genes such as tumor necrosis factor (TNFA), natural resistance associated macrophage protein 1 (NRAMP1), and vitamin D receptor (VDR).

**TNFA.** In the same Indian population used to show the association between HLA-DR2 and susceptibility to both poles of leprosy severity, there is also a significant higher TNF2 allele frequency in the lepromatous group but not in the tuberculoid group (Table 1) (Roy et al., 1997). The HLA-DR2 and TNF2 alleles are not in strong linkage disequilibrium suggesting that association of lepromatous leprosy with DR2 serotypes and TNF2 is independent despite the proximity of the two loci.

**NRAMP1.** The possible role of NRAMP1 as a genetic factor in leprosy has been extensively studied. However, the first direct evi-
dence for a role of NRAMP1 in susceptibility to leprosy was obtained only recently in a large familial study in South Vietnam. Six NRAMP1 polymorphisms and four closely linked microsatellites were genotyped in 168 patients of 20 multiplex leprosy families, and significant evidence for linkage was obtained (P < 0.005–0.02; Abel et al., 1998) (Table 1). Segregation analysis performed on 285 Vietnamese and 117 Chinese families detected evidence for a major codominant gene only in the Vietnamese population suggesting that the control of susceptibility to leprosy could be genetically heterogeneous according to the ethnic origin of the families. It is possible that genetic heterogeneity will partly explain the lack of linkage between NRAMP1 and leprosy susceptibility in previous reports (Shaw et al., 1993; Roger et al., 1997).

VDR. The binding of the active form of vitamin D (1,25-dihydroxy vitamin D) to the vitamin D receptor present on monocytes, macrophages, and activated lymphocytes plays an immunoregulatory role. A single base polymorphism in codon 352 of the VDR gene can be detected as TaqI restriction fragment polymorphism with the two alleles designated “T” and “t”, respectively. The less common allele “t” has been associated with higher levels of VDR mRNA expression in transient transfection assays (Morrison et al., 1992). In a case-control study in Calcutta, the codon 352 TaqI VDR polymorphism was found associated with lepromatous leprosy (P = 0.04) and tuberculoid leprosy (P = 0.004; Roy et al., 1999) (Table 1). The frequency of the tt genotype was significantly increased in tubercuroid patients (21.5%) as compared with controls (7.8%). In contrast, the TT genotype frequency was higher in the lepromatous leprosy group (52.4%) compared with the controls (39.8%). However, in this population no significant association was found between NRAMP1 polymorphisms and leprosy susceptibility.

Tuberculosis (HLA, NRAMP, VDR, IL-1Ra, IL-1b). Numerous studies have analyzed a possible contribution of genetic factors to tuberculosis susceptibility. A consistent conclusion from twin, family, and adoption studies was that the genetic background of the host is an important control element for susceptibility to tuberculosis. Employing candidate gene studies a number of gene variants have been identified that contribute to tuberculosis risk.

HLA. A large number of associations of HLA type with tuberculosis have been reported for different populations. However, a substantial proportion of these tuberculosis associations could not be reproduced in independent studies. One of the most consistent findings that has been reproduced in several populations is that susceptibility to pulmonary tuberculosis appears associated with the HLA-DR2 serotype (Table 1) (Bothamley et al., 1989; Brahmajothi et al., 1991). However, a two-stage case-control study has shown the importance of the HLA-DQβ1*0503 allele in tuberculosis progression (P = 0.005) in a Cambodian population while no significant effect of HLA DR2 alleles was detected (Goldfeld et al., 1998). A recent association study on 126 patients with pulmonary tuberculosis and 87 endemic controls from India indicated that HLA-DRB1*1501 (P = 0.013) and HLA-DQβ1*0601 (P = 0.008) were associated with pulmonary tuberculosis (Ravikumar et al., 1999). Interestingly, no connection has been found between TNFA polymorphisms and tuberculosis risk (Shaw et al., 1997). These results suggest that the variability in the major histocompatibility complex and its relationship to tuberculosis susceptibility deserves further clarification, preferably using HLA analysis on the level of the nucleotide.

NRAMP1. Recently a large case-control study conducted in The Gambia (West Africa) has shown that four alleles of NRAMP1 were significantly associated with susceptibility to tuberculosis (Table 1) (Bellamy et al., 1998). This genetic analysis was performed on 410 smear-test-positive tuberculosis patients and 417 ethnically matched healthy controls. The polymorphic variants of the NRAMP1 gene analyzed correspond to a dinucleotide CA repeat in the 5′ region, a single nucleotide polymorphism in intron 4 (469 + 14G/C), a non-conservative single-base change at codon 543 (D543N), and a 4-base pair deletion in the 3′ region. Combined analysis of the polymorphisms in intron 4 and in the 3′ region detected a strong association with tuberculosis (P < 0.001; Bellamy et al., 1998). Subjects heterozygous for these two variants were four times overrepresented among patients with tuberculosis as compared with those bearing the most common NRAMP1 genotype (odds ratio: 4.07, 95% CI: 1.86–9.12). In a more recent genetic study of a tuberculosis outbreak in a Canadian Aboriginal Community, NRAMP1 was strongly linked to tuberculosis susceptibility (lod score 4.2) (Greenwood et al., 2000). These reports confirm that NRAMP1 is important in modulating susceptibility to tuberculosis and demonstrate that mouse models of infectious disease can identify relevant candidate genes for human disease.

VDR. Variations in the vitamin D receptor gene were analyzed in the same Gambian population that was enrolled to demonstrate the NRAMP1 association with tuberculosis. Homozygous patients for a polymorphism at codon 352 (genotype tt) were significantly under-represented among those with tuberculosis (P = 0.01; Bellamy et al., 1999) suggesting a role of VDR in the pathogenesis of tuberculosis (Table 1). Recently it has also been shown that serum vitamin D deficiency may contribute to the high occurrence of tuberculosis among Gujarati Asian subjects (P = 0.008) (Wilkinson et al., 2000). Moreover, the VDR genotypes of 91 untreated tuberculosis patients and 116 healthy people who had been sensitized to tuberculosis indicated that VDR polymorphisms are involved in tuberculosis susceptibility (Wilkinson et al., 2000).

MBL (or MBP). Mannose binding lectin (MBL) also called mannose binding protein (MBP) activates the classical complement pathway and phagocytosis leading to neutralization of the pathogen. To investigate the role of MBL gene polymorphisms in tuberculosis susceptibility, a recent study was carried out with Indian patients. Two hundred and two pulmonary tuberculosis patients and 109 controls were genotyped for three MBL (codons 52, 54, and 57) functional variants affecting the structure of MBL and associated with low serum levels. A significantly increased genotype frequency of mutant homozygotes was noted in pulmonary tuberculosis patients compared with healthy controls (P = 0.008) (Table 1) (Selvaraj et al., 1999). On the other hand, a protective effect of the G54D MBL allele on tuberculosis meningitis was noted in a South African population (Hoal-van Helden et al., 1999).

IL-1Ra and IL-1b. The proinflammatory cytokine interleukin-1b (IL-1b) and the interleukin-1 receptor antagonist (IL-1Ra), which is a specific inhibitor of IL-1 activity, have been shown to be induced in vitro by Mycobacterium tuberculosis. Moreover patients with tuberculosis present an elevated serum concentration of IL-1Ra and IL-1Ra was identified as a marker of disease activity in tuberculosis (Juffermans et al., 1998). Within the IL-1b gene, two biallelic polymorphisms have been identified at positions −511 and +3953, respectively. An 86-base pair variable number of tandem repeats polymorphism with five different alleles is found in the 5′ region, although expression of mRNA for IL-1b was slightly higher in patients with the IL-1b(+3953) A1+ allele (P =
0.04). In a case-control study comprised of 114 healthy subjects and 89 patients with tuberculosis, no significant difference was found in IL-1b and IL-1Ra allele or genotype frequencies between the two groups. However, the IL-1Ra A2/-IL-1b (+3951) A1+ genotype combinations are overrepresented in patients with tuberculous pleurisy (92%) compared with M. tuberculosis-sensitized control patients (57%) or patients with other disease forms (56%) (Table 1). The mechanistic basis underlying the above associations is unknown. However, considering the relatively small number of individuals enrolled in the study and the fact that the association with tuberculosis type is of borderline significance if corrected for multiple comparisons, it is important that the results be repeated in distinct populations.

Genetic Analysis of Tuberculosis Mouse Models

To identify new candidate genes for tuberculosis severity, a recent genome wide analysis was performed in a mouse model (Lavebratt et al., 1999). The severity of tuberculosis was approximated with loss of body weight after M. tuberculosis infection. Hence, body weight at 20 days postinfection was used as phenotype for quantitative trait loci (QTL) analysis. Among females, QTLs on chromosomes 9 and 3 were significantly linked to postinfection body weight (Lod score 6.68 and 3.92, respectively), two other suggestive linkages were found on chromosomes 8 and 17. For males, suggestive linkages were found on chromosomes 5 and 10. Identification of syntenic regions in the human genome will provide candidate genome regions for the identification of tuberculosis susceptibility loci.

Hypersusceptibility to Mycobacterial Infection

Genetic studies of patients with severe idiopathic disseminated infections due to weakly pathogenic Mycobacteria revealed the presence of mutations in four different genes: IFNgRI, IFNgR2, IL-12Rb1, and IL-12p40.

IFNgR1—complete and partial deficiency. In a Maltese family, four children with recessive susceptibility to atypical mycobacterial infection (M. fortuitum, M. chelonei, and M. avium) presented a homozygous point mutation at nucleotide position 395 in the gene for interferon gamma receptor 1 (IFNgR1) (Table 1). This mutation introduced a stop codon leading to a truncated protein that lacks the transmembrane and cytoplasmic domains of the receptor (Newport et al., 1996). A second study indicated that one child with fatal disseminated BCG infection was homozygous for a frameshift deletion in this gene (Jouanguy et al., 1996) with absence of expression of the receptor at the cell surface. IFNgR1 deficiency has also been identified in a patient with disseminated M. smegmatis infection (Pierraudigier et al., 1997). These studies demonstrated that the IFNgR1 gene was responsible for the children’s immune deficiency and underlined the importance of the IFNγ pathway in immunity to mycobacterial infection. Moreover complete IFNgR1 deficiency appears to be an autosomal immune disorder associated with severe and selective infection by poorly pathogenic Mycobacteria, and a complete absence of mature granuloma formation.

Partial IFNgR1 deficiency due to a homozygous missense mutation has been found in two patients, one patient with disseminated infection after BCG vaccination and the second with clinical tuberculosis without BCG vaccination (Table 1) (Jouanguy et al., 1997). Moreover, partial as opposed to complete IFNgR1 deficiency is associated with mature granulomas and a milder course of mycobacterial infection. These studies suggested that ligation of IFNγ to IFNgR is essential for immunity against BCG and M. tuberculosis and that subtle mutations in these genes may alter disease susceptibility on the population level.

IFNgR2—complete deficiency. A complete genetic deficiency of IFNgR2 (the IFNgR signaling chain) has been found in one patient with severe disseminated infection caused by M. fortuitum and M. avium (Table 1). Molecular analysis indicated the presence of a homozygous recessive frameshift deletion in the IFNgR2 gene, which resulted in a premature stop codon in the region encoding the extracellular domain (Altare et al., 1998c; Dorman and Holland 1998). The absence of additional mutations in the IFNgR1 gene was confirmed.

IL-12Rb1—complete deficiency. Sequence analysis of the IL-12Rb1 gene in several patients suffering from disseminated mycobacterial infections revealed a homozygous recessive mutation, which introduces a premature stop codon in the extracellular domain (Table 1) (Altare et al., 1998a; de Jong et al., 1998). The receptor was not expressed but the consequences of this deficiency are apparently much less severe compared with complete IFNgR deficiency (Newport et al., 1996; Altare et al., 1998b; Ottenhoff et al., 1998). Interestingly, patients with IL-12Rb1 deficiency develop mature BCG granulomas.

IL-12p40—complete deficiency. A large homozygous deletion (373 nucleotides) within the IL-12p40 subunit gene was found in a child with curable BCG and Salmonella enteridis infection (Table 1) (Altare et al., 1998d). Similar to IL-12Rb1 deficiency, IL-12p40 deficiency has no effect on mature BCG granuloma formation. Further studies suggested that susceptibility to mycobacterial infection of patients with genetically impaired IL-12-mediated immunity is due to insufficient IFNγ-mediated immunity. Moreover, the milder clinical phenotypes of IL-12p40- and IL-12Rb1-deficient patients compared with patients with complete IFNgR1 or IFNgR2 deficiency are explained by a residual (IL-12-independent) IFNγ secretion. Recently, Jouanguy et al. (1999) described a hotspot for small deletions in human IFNgR1 that confer dominant susceptibility to infection by poorly virulent mycobacteria. Molecular genetic analysis was performed on 18 patients from several generations of 12 unrelated families and 12 independent mutations were found at a single site.

Taken together, these studies suggest that IFNγ and IL-12 are two important cytokines in human defense against mycobacteria and salmonella. Particularly, the type-1 cytokine pathway appears essential since it seems that type-1 deficiencies in patients are not compensated by any other protective immune mechanism.

Conclusion

To identify susceptibility/resistance genes in infectious diseases different strategies can be used. To date, candidate gene analysis has been the method of choice in most studies. Contrary to genome-scan analysis, this approach allows the possibility of identifying genes that exert a small or moderate effect on susceptibility to infection. However, this approach is limited to known candidate genes, and the results should be confirmed in independent studies. Another challenge of candidate gene analysis is the identification of the causal mutation of the disease. This can be relatively straightforward in the case of major mutations but can be a difficult undertaking in the case of subtle polymorphic variations. The characterization of the susceptibility genes and their underlying causative mutations has important implications not only for a better understanding of disease pathogenesis but also for the control and development of new therapeutic strategies for infectious diseases.

References


GENETICS OF SUSCEPTIBILITY TO INFECTIOUS DISEASES


