Clinical Medicine 2012, Vol 12, No 6: s12–s16

Infectious disease: bad luck or bad genes?

Stephen J Chapman

Introduction

Infectious disease has been a leading cause of death throughout human history. Indeed, at the time of Thomas Linacre's practice in the sixteenth century, England was subject to a series of devastating epidemics of infectious disease. One of the most characteristic and intriguing features of most human infections is that only a proportion of exposed individuals go on to develop clinical disease. This was noted during Linacre's period; for example, accounts of the Black Death and the English sweating sickness describe how some individuals inexplicably survived while many in their communities perished. The basis of such inter-individual variation in susceptibility to major infectious disease remains incompletely understood and is not fully accounted for by variability in acquired host factors and pathogen virulence. There is increasing evidence that host genetic variation might influence susceptibility to infectious disease in humans.[1]

Early studies examined the concordance rates of common infectious diseases in twins and reported higher rates in monozygotic (identical) twins than in dizygotic (non-identical) twins, suggesting a genetic component to disease susceptibility. Studies of adoptees further suggested that the likelihood of surviving severe infectious disease has a significant inherited component, and indeed this genetic component appeared to exert a greater influence on survival after infectious disease than following cardiovascular disease and cancer.[2] Taken together, the available evidence supports a role for host genetics in influencing susceptibility to, and outcome of, infectious disease in humans — but how do we go about identifying the specific genes responsible?

Human genetic variation and disease

The study of human disease genetics aims to understand the relationship between human genetic variation and the development of disease. The most common type of genetic variation in the human genome is variation in a single nucleotide base. If such variants exist at frequencies greater than 1%, they are called single nucleotide polymorphisms (SNPs); rarer variants are, by convention, termed mutations. Each variant is termed an allele, and there are two alleles at each position (locus): one allele on each of a pair of chromosomes (one allele inherited from the father and one from the mother). Alleles at a given position that are the same comprise a homozygote genotype, whereas different alleles form a heterozygote genotype. The examination of SNP frequencies forms the basis of much study of common disease genetics: association studies look for statistically significant differences in SNP frequencies between individuals with the disease phenotype of interest and healthy controls. Polymorphisms that are located in or around genes might have functional consequences for the protein products of those genes; for example, they might cause an amino-acid substitution that alters a protein's structure or they might affect the transcription levels of a particular protein. In turn, these changes could have physiological consequences and may affect disease susceptibility.

The genetic basis of human disease has traditionally been divided into monogenic and complex disease. In monogenic disease, rare mutations cause disease by exerting large effects on the function of a single gene (eg by a complete gene knockout that leads to the absence of a particular protein); cystic fibrosis is a typical example. Complex disease, on the other hand, has traditionally been considered to result from the combined effects of multiple common polymorphisms, each of which exerts only a modest effect on the individual risk of disease. These polymorphisms might interact with each other and with environmental factors to result in disease. Most common human diseases are widely considered to have such a 'complex' basis, although there is little evidence to support this 'common disease, common variant' hypothesis. It is increasingly recognised that there is considerable overlap between these two extremes: for example, common polymorphisms might act to modify the effect of familial single-gene mutations that result in diseases such as cystic fibrosis, in which the correlation between the genotype (in this case, mutations within the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene) and clinical disease severity is imprecise. Furthermore, it is now clear from very large studies of conditions such as type II diabetes and Crohn's disease that common polymorphisms account for only a small proportion of disease heritability. The source of the remaining 'missing heritability' remains unknown, although one possible explanation comprises multiple uncommon polymorphisms or individually rare mutations of large effect size.

A widely used approach to identify susceptibility genes for human infectious disease is the study of candidate genes. This involves selecting genes for study on the basis of existing evidence that supports a role for that gene in pathogenesis. Candidate gene approaches for infectious disease will be illustrated using the phenotype of invasive pneumococcal disease as an example.

Host genetic susceptibility to invasive pneumococcal disease

Infection by *Streptococcus pneumoniae* (the pneumococcus) is the leading cause of community-acquired pneumonia in the UK

Stephen J Chapman, senior lecturer and consultant in respiratory medicine, Oxford Biomedical Research Centre and Oxford University Hospitals NHS Trust, UK

and a major cause of death worldwide. Invasive pneumococcal disease (IPD) is defined by the isolation of *Streptococcus pneumoniae* from a normally sterile site, most commonly blood or cerebrospinal fluid. As with many other major bacterial causes of sepsis, asymptomatic carriage of the pneumococcus is common in the general population, yet invasive disease develops in only a minority of infected people. Host genetic factors are increasingly recognised as contributing to IPD susceptibility. Indeed, several very rare monogenic primary immunodeficiency (PID) conditions have been described as a cause of IPD in children.[3] Mutations in the genes *NEMO* (NF-κB essential modulator), *IRAK4* (interleukin-1 receptor-associated kinase 4), *MyD88* (myeloid differentiation primary response gene 88) and *NFKBIA* (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha) have been identified as causes of this phenotype.[1,3] These genes encode proteins that are involved in the Toll-like receptor (TLR)-nuclear factor-κB (NF-κB) signalling pathway (Fig 1), which is involved in the early recognition of invading pathogens such as the pneumococcus and the generation of a pro-inflammatory host response. Toll-like receptors (TLR) are pattern-recognition receptors that recognise conserved components on the surface of pathogens.[4] TLR2, for example, forms heterodimers with TLR1 or TLR6 and recognises the lipoteichoic acid and peptidoglycan cell wall components of Gram-positive bacteria such as *Streptococcus pneumoniae*. It also recognises other agonists such as components of mycobacteria and the malaria parasite. TLR4 recognises lipopolysaccharides in Gram-negative bacterial cell walls and the

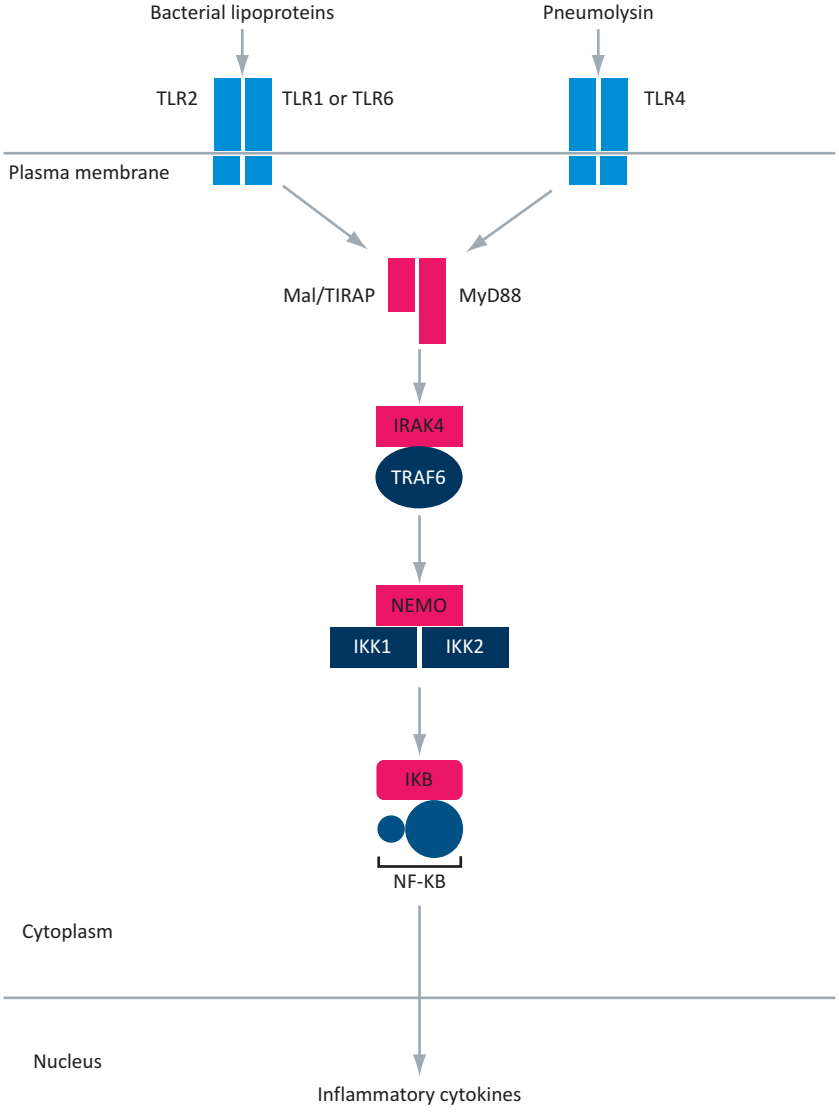

Fig 1. Toll-like receptor (TLR)-nuclear factor-κB (NF-κB) signalling and susceptibility to pneumococcal disease. Genes in which common polymorphisms and/or rare mutations are associated with susceptibility to invasive pneumococcal disease are highlighted in pink. The pathway is complex and not all of its components are shown. IκB = inhibitor of NF-κB; IKK = IκB kinase; IRAK = interleukin 1 receptor-associated kinase; Mal = MyD88-adaptor like; MyD88 = myeloid differentiation primary response protein 88; NEMO = NF-κB essential modulator; TIRAP = Toll–interleukin 1 receptor domain-containing adaptor protein; TRAF = tumour necrosis factor receptor-associated factor.

pneumococcal toxin pneumolysin. Recognition of microbial components leads to the recruitment of adaptor proteins (such as MyD88 and Mal/TIRAP [toll-interleukin 1 receptor (TIR) domain containing adaptor protein]) and to the activation of a complex series of kinases (including IRAK4 and NEMO) (Fig 1).[4] This leads to the phosphorylation and degradation of inhibitors of NF-κB (IκBs), resulting in the release of NF-κB which translocates into the nucleus and activates the transcription of more than 100 genes that are involved in the inflammatory host response.

Common polymorphisms within genes in this pathway also appear to be associated with susceptibility to IPD and other infectious disease phenotypes. For example, polymorphisms in the gene *NFKBIA*, as well as in the related genes *NFKBIZ* and *NFKBIL2*, associate with susceptibility to IPD in adults.[5–7] These genes each encode IκB proteins that act as inhibitors of NF-κB. Taken together, these findings suggest that both very rare and common genetic variation in the control of NF-κB appears to be important in determining IPD susceptibility. A polymorphism in the gene that encodes the adaptor protein Mal/TIRAP has also been described in association with not only IPD but also with bacteraemia, malaria and tuberculosis in different human populations.[8] Mal/TIRAP functions as an adaptor protein for signalling via TLR2 and TLR4; its observed association with susceptibility to such a broad range of infectious pathogens reflects the convergence of signalling networks that involve these receptors and the central role of TLR signalling. The disease-associated SNP encodes an amino acid substitution of serine (S) to leucine (L) at position 180 of the Mal/TIRAP protein. Interestingly, protection against the diseases was associated with the heterozygous genotype. Functional studies suggested that S180L heterozygotes generated increased pro-inflammatory cytokine responses compared to SS homozygotes, and that 180L homozygotes produced the highest cytokine responses.[9] These findings imply that the moderately increased inflammatory response in S180L heterozygotes is beneficial (probably reflecting enhanced early clearance of pathogen), whereas the excessive inflammatory response of 180L homozygotes might result in a predisposition to invasive or severe disease.

A further approach to the selection of candidate genes for IPD came from insights into autoimmune disease. A common SNP in the gene *PTPN22*, which encodes the lymphoid protein tyrosine phosphatase Lyp, has emerged as a major susceptibility locus for human autoimmunity. Lyp acts to inhibit immune cell signalling, in particular by suppressing the activation of T cells, although the functional consequences of the disease-associated SNP remain controversial. This functional *PTPN22* SNP was found to also associate with susceptibility to IPD and, interestingly, the direction of association was the same as that observed with autoimmune disease.[10] The major selective pressure exerted by infectious disease is widely considered to have encouraged the emergence of autoimmune susceptibility loci. This finding suggests, on the other hand, that in some cases bacterial disease and autoimmunity may have a shared genetic basis.

Genome-wide association studies of human infectious disease

The sequencing of the human genome and the development of the International HapMap project, combined with advances in high-throughput genotyping technology, have made it possible to genotype many hundreds of thousands of SNPs spread at intervals across the human genome. Genome-wide association studies (GWAS) utilise this approach to compare allele frequencies in disease cases and healthy controls across most of the genome. Stringent thresholds for declaring statistical significance are required in GWAS, reflecting the large number of independent tests of association that are performed. As a result, very large sample sizes are required. Nevertheless, GWAS have now been carried out for a number of major infectious disease phenotypes (Table 1, reviewed by Chapman and Hill).[1] These studies have identified the human leukocyte antigen (HLA) as a major susceptibility region for many, but not all, of the infectious diseases studied. A significant advantage of GWAS over the candidate gene approach is the potential to identify previously unsuspected genetic associations using an unbiased, genome-wide approach. For example, a recent GWAS of tuberculosis reported an association within a gene desert, a gene-poor but highly conserved region of chromosome 18q11.2, suggesting that this region has a possible regulatory effect on an unknown gene elsewhere.[11] The genes nearest to the associated SNP are *GATA6* (GATA binding protein 6), *CTAGE1* (cutaneous T-cell lymphoma-associated antigen 1), *RBBP8* (retinoblastoma binding protein 8) and *CABLES1* (Cdk5 and Abl enzyme substrate 1), of which the most interesting candidate is *GATA6*, a transcription factor that regulates arachidonate 15-lipoxygenase (*ALOX15*). *ALOX15* has been reported to regulate cytokine release in lung epithelial cells and the related lipoxygenase gene *ALOX5* has been implicated in tuberculosis susceptibility in a human candidate gene study and in mouse models. Further research is needed to identify the causative variant and to characterise its functional significance. Such studies might also reveal a role for this locus in protection against other major causes of mortality, as the relatively high frequency of the tuberculosis susceptibility allele in African populations raises the possibility that it might be subject to a counterbalancing selective pressure.

Shared pathways underlying infectious disease

An emerging theme from genetic studies of infectious disease susceptibility is that, in many cases, the underlying variants appear to cluster on shared biological pathways.[1] In addition to the associations described above between IPD susceptibility and both common and rare genetic variants in TLR-NF-κB pathway signalling (Fig 1), other examples include: rare genetic defects in TLR3 signalling and susceptibility to herpes simplex encephalitis in childhood; both rare mutations in complement components and common polymorphism in *CFH* (encoding complement factor H) that are associated with susceptibility to meningococcal

Table 1. Examples of genome-wide association studies of human infectious disease.

Disease	Phenotype	Gene(s)
HIV-1/AIDS	Viral load, disease progression	*HLA-B, HLA-C*
Hepatitis C	Chronic infection, response to interferon-α treatment	*IL28B*
Hepatitis B	Chronic infection	*HLA-DP*
Severe malaria	Protection	*HBB*
Meningococcal disease	Protection	*CFH, CFHR3*
Tuberculosis	Susceptibility	18q11.2 region
Leprosy	Susceptibility	*LACC1, NOD2, CCDC122, HLA-DR-DQ, TNFSF15, RIPK2, TLR1*

CCDC122 = coiled-coil domain containing 122; *CFH* = complement factor H; *CFHR3* = CFH-related protein 3; *HBB* = haemoglobin beta; *HLA* = human leukocyte antigen; *IL28B* = interleukin-28B; *LACC1* = laccase (multicopper oxidoreductase) domain containing 1 (also referred to as *C13orf31*); *NOD2* = nucleotide-binding oligomerisation domain containing 2; *RIPK2* = receptor-interacting serine–threonine kinase 2; *TLR1* = Toll-like receptor 1; *TNFSF15* = tumour necrosis factor [ligand] superfamily member 15.

disease; and both rare and common genetic variants in the Janus kinase (JAK)-signal transducer and activator of transcription (STAT) cytokine signalling pathway that underlie both primary immunodeficiency and common infectious disease.[1,12] Interestingly, there is also increasing evidence to suggest that common immune-related genes and pathways are shared among infectious, autoimmune and chronic inflammatory conditions; the example of *PTPN22* in bacterial disease and autoimmunity is described above.[10] Another example is the striking overlap between susceptibility genes that underlie Crohn's disease and leprosy. *NOD2* (nucleotide-binding oligomerisation domain containing 2), *TNFSF15* (tumour necrosis factor [ligand] superfamily member 15), *LRRK2* (leucine-rich repeat kinase 2), the 13q14 locus containing *LACC1* (laccase [multicopper oxidoreductase] domain containing 1) and *CCDC122* (coiled-coil domain containing 122) have been reported as susceptibility loci for both of these conditions, which are each characterised by granulomatous inflammation. These findings suggest that focusing future research, such as sequencing strategies and functional studies, on key immune pathways might reveal major insights into apparently diverse human diseases.

Rare gene variants and susceptibility to common infectious disease?

The genetic architecture of common, severe infectious disease remains largely unknown. The increasing interest in rare variants as a source of the 'missing heritability' of common human disease is particularly relevant to the field of infectious disease for two reasons. First, infection has been associated with significant childhood disease throughout human history and consequently exerts an enormous evolutionary selective pressure. This suggests that most major susceptibility variants are likely to be rare. Exceptions will occur if a variant is subject to opposing selective pressures as is the case, for example, with the haemoglobinopathies and malaria resistance, although such examples are likely to be rare. Second, an intriguing feature of the primary immunodeficiency states that result from IRAK4 and MyD88 deficiencies is that they affect otherwise healthy individuals and confer increased susceptibility to only a narrow spectrum of pathogens, frequently solely IPD.[3] An increasing number of such 'selective' primary immunodeficiencies have now been described, including susceptibility to non-tuberculous mycobacteria in the setting of mutations in interleukin-12/interleukin-23/interferon-γ signalling, and susceptibility to herpes simplex encephalitis in otherwise healthy children who have mutations in TLR3 signalling.[1,3] The identification of these conditions provides proof-of-principle that rare genetic variants can underlie susceptibility to severe infectious diseases in otherwise healthy individuals. The contribution of individually rare mutations to severe infectious disease at the population level remains unknown, although the widespread application of next-generation sequencing technology is likely to define the relative contributions of common polymorphism and rare mutations to disease susceptibility.

Potential for clinical translation

To date, the major advances in human genomics have resulted in disappointingly few improvements in clinical care. A widely considered avenue for clinical translation is the possibility of utilising knowledge of an individual's genetic profile to predict their likelihood of disease development or outcome. At present, however, understanding of the genetic basis of disease development is too incomplete to facilitate outcome prediction for most common diseases. Another opportunity for clinical translation stems from the identification of novel genes and biological pathways, which may lead to new therapeutic targets for drugs and vaccines. An increased understanding of disease biology might also facilitate a more robust sub-classification of disease. The identification of genetically determined molecular sub-phenotypes underlying clinically indistinguishable diseases might allow the therapeutic targeting of patient subgroups and hence deliver a new level of precision to clinical trials. A potential

example of such an approach is the management of sepsis, in which no drug has consistently been shown to be of benefit. As described above, subgroups of patients have particularly weak or excessive inflammatory responses as a result of their genetic background. Such subgroups might benefit from targeted immunomodulatory therapy, and this opportunity could be missed in a clinical trial that uses the current approach of grouping all patients together and testing only for an overall benefit. An increased understanding of the genetic basis of infectious disease susceptibility and outcome at the individual genomic level may allow the stratification of clinical trials on the basis of host genotype and pave the way for personalised medicine.

Financial & competing interests disclosure

The author has no relevant affiliations or financial involvement with any organisation or entity with a financial interest in, or financial conflict with, the subject matter or materials discussed in this manuscript.

Acknowledgements

Supported by Medical Research Council, UK, and NIHR Oxford Biomedical Research Centre.

References

1 Chapman SJ, Hill AV. Human genetic susceptibility to infectious disease. *Nat Rev Genet* 2012;13:175–88.
2 Sorensen TIA, Nielsen GG, Andersen PK, Teasdale TW. Genetic and environmental influences on premature death in adult adoptees. *New Engl J Med* 1988;318:727–32.
3 Casanova JL, Abel L. Primary immunodeficiencies: a field in its infancy. *Science* 2007;317:617–9.
4 Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol* 2010;11:373–84.
5 Chapman SJ, Khor CC, Vannberg FO *et al.* IκB genetic polymorphisms and invasive pneumococcal disease. *Am J Respir Crit Care Med* 2007;**176:**181–7.
6 Chapman SJ, Khor CC, Vannberg FO *et al. NFKBIZ* polymorphisms and susceptibility to pneumococcal disease in European and African populations. *Genes Immun* 2010;11:319–25.
7 Chapman SJ, Khor CC, Vannberg FO *et al.* Common *NFKBIL2* polymorphisms and susceptibility to pneumococcal disease: a genetic association study. *Crit Care* 2010;14:R227.
8 Khor CC, Chapman SJ, Vannberg FO *et al.* A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nat Genet* 2007;39:523–8.
9 Ferwerda B, Alonso S, Banahan K *et al.* Functional and genetic evidence that the Mal/TIRAP allele variant 180L has been selected by providing protection against septic shock. *Proc Natl Acad Sci USA* 2009;106:10272–7.
10 Chapman SJ, Khor CC, Vannberg FO *et al.* PTPN22 and invasive bacterial disease. *Nat Genet* 2006;38:499–500.
11 Thye T, Vannberg FO, Wong SH *et al.* Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat Genet* 2010;42:739–41.
12 Khor CC, Vannberg FO, Chapman SJ *et al. CISH* and susceptibility to infectious diseases. *N Engl J Med* 2010;362:2092–101.

**Address for correspondence: Dr SJ Chapman,
Oxford Centre for Respiratory Medicine, Churchill Hospital,
Oxford University Hospitals, Old Road,
Headington, Oxford OX3 7LE.
Email: schapman@well.ox.ac.uk**