GENOME-WIDE ASSOCIATION STUDIES

Human genetic susceptibility to infectious disease

Stephen J. Chapman and Adrian V. S. Hill

Abstract | Recent genome-wide studies have reported novel associations between common polymorphisms and susceptibility to many major infectious diseases in humans. In parallel, an increasing number of rare mutations underlying susceptibility to specific phenotypes of infectious disease have been described. Together, these developments have highlighted a key role for host genetic variation in determining the susceptibility to infectious disease. They have also provided insights into the genetic architecture of infectious disease susceptibility and identified immune molecules and pathways that are directly relevant to the human host defence.

A characteristic feature of many human infections is that only a proportion of exposed individuals develop clinical disease. Heritable factors were long considered to have a dominant role in explaining this inter-individual variation in susceptibility: the early observation that cases of tuberculosis clustered within households, for example, led to the impression that tuberculosis was an inherited disease. However, the subsequent discovery of bacteria such as *Mycobacterium tuberculosis* as agents of infectious disease — and the experimental demonstration of the communicable nature of infections — focused attention on the pathogen[1], somewhat neglecting the potential importance of host factors.

More recent clues that host genetic variation influences infectious disease susceptibility has come from twin and adoptee studies. Twin studies suggest an important role for host genetic factors in susceptibility to infectious diseases, such tuberculosis, leprosy, poliomyelitis and hepatitis B[1,2–5], although the relative contributions of heredity and environmental factors in tuberculosis susceptibility in twins has been questioned[6]. In the late 1980s, a landmark study reported that adopted children had a markedly increased risk of death from infectious disease if one of their biological parents had died prematurely from infection[7]. Mouse studies have also illustrated the potential importance of host genetic effects, by showing differences between different inbred strains in bacterial loads, cytokine responses and outcomes following bacterial and mycobacterial infection[8]. A considerable role for the host's genetic profile in determining susceptibility to infectious disease is now well established: the classic example is the effect of genetic red-blood-cell disorders on malaria susceptibility[9].

In this Review we focus on two parallel approaches to further our understanding of the genetic basis of infectious disease susceptibility in humans: the use of genome-wide association studies (GWASs) to identify polymorphisms associated with disease, and the study of individuals with rare monogenic defects underlying their susceptibility to a narrow range of pathogens. Such approaches have identified both common and rare human genetic variation underlying susceptibility to infectious disease. These studies implicate novel genetic variants in the immune response to specific pathogens, highlight the role of shared host signalling pathways in the pathogenesis of diverse infectious diseases and provide important insights into the genetic control of immune signalling in humans. Extension of these approaches has the potential to dissect the genetic architecture of infectious disease susceptibility and will serve as a foundation for future studies to evaluate concepts such as host–pathogen genetic interactions and potentially clinical translation.

Approaches for identifying susceptibility genes

Early studies of genetic susceptibility to infectious disease used genome-wide linkage analysis and candidate-gene approaches, and identified only a limited number of independently replicated, strongly associated loci. Linkage approaches have been employed successfully in monogenic disease studies, and were subsequently applied in attempts to define the susceptibility loci underlying common diseases. The most widely used design was the study of affected sibling pairs; this approach had some success in identifying loci linked to some infectious diseases, such as leprosy[3,9,10]. However,

Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. Correspondence to S.J.C. e-mail: schapman@well.ox.ac.uk
doi:10.1038/nrg3114
Published online 7 February 2012

Power
The probability of correctly rejecting the null hypothesis when it is truly false. In genetic association studies, power is determined by the sample size, allele frequency, magnitude of effect and significance threshold, and can be considered as the probability that the study will correctly detect a genuine association.

Population stratification
The presence in a population of several subpopulations that differ in their genetic characteristics, reflecting their different ancestral histories. Such systematic ancestry differences between cases and controls can confound genetic association studies, causing spurious disease associations.

linkage analysis is limited by the difficulty in recruiting numerous multicase families in which two siblings are affected by infectious disease, as well as by a lack of study power[9–11].

Candidate-gene studies comprise the genotyping of polymorphisms in biologically plausible candidate genes, typically in unrelated case and control individuals. The degree of replication between candidate-gene studies is frequently poor. This can be due to various factors, such as small sample sizes (and hence inadequate study power), unidentified population stratification or a failure to correct for multiple testing during statistical analysis. Additional reasons for a failure to replicate associations with infectious diseases may include: differences in the phenotype definition of cases (for example, caused by the use of different pathogen detection assays) and controls (for example, reflecting different ways of assessing the likelihood of exposure to infection); unidentified variation in gene–environment or host–pathogen interactions; or true inter-population genetic heterogeneity[12].

A further limitation of the candidate-gene approach is its reliance on existing and possibly inaccurate biological hypotheses to select genes for study. Despite these limitations, candidate-gene studies have identified a number of robust associations with infectious disease phenotypes. Such studies have been extensively reviewed[3,9–11,13–15]; a systematic discussion of candidate-gene studies is beyond the scope of this Review.

The sequencing of the human genome and the International HapMap project[16,17] — together with developments in microarray-based high-throughput genotyping technology — have led to the possibility of genotyping hundreds of thousands of polymorphisms across the genome, with no assumption about the location of causal variants. The ability of GWASs to identify previously unsuspected genetic associations with common disease has now been demonstrated, although very large sample sizes are required to generate sufficient statistical power to detect true disease associations after correction for multiple testing[18,19]. The current standard for declaring statistical significance at genome-wide level is a combined *P* value (including 'initial discovery' GWAS and replication cohorts) of $<5\times10^{-8}$ (REF. 20). The sample sizes that are required to achieve such significance thresholds may be unrealistic for the study of less-common infectious disease; particular challenges are also posed by identifying genome-wide associations within African populations, who suffer the greatest burden of infectious disease (BOX 1).

GWASs are based on the 'common disease, common variant' hypothesis, and they have been performed largely using SNP arrays that focus only on common genetic polymorphisms (for which the minor allele frequency is >5%)[18,21–23]. Most novel disease-associated variants identified by GWASs are associated with small increments in risk (1.1–1.5-fold); even when taken together, these variants explain only a small fraction of known disease heritability[18,23,24]. Multiple causative rare variants of individually large effect size are considered a likely major source of the 'missing heritability' of common disease, and next-generation sequencing studies aimed at identifying rare variants are widely expected to become a dominant approach in human disease genetics[21–23].

Common polymorphisms identified by GWASs

Various GWASs of infectious diseases have now been reported; markers that show strong evidence of association (combined *P* value $<5\times10^{-8}$) are summarized in TABLE 1, and noteworthy findings are discussed below according to the disease phenotype. These studies have confirmed some previously reported associations from candidate-gene and linkage studies, as well as identifying strong associations between common infectious disease phenotypes and novel genes and pathways.

HIV-1 and AIDS. In common with other major infectious diseases, susceptibility to HIV and the clinical outcome after infection display marked inter-individual variation. Host genetic studies using both candidate-gene and genome-wide strategies have examined various different phenotypes, including susceptibility to HIV-1

Box 1 | Genetic studies of infectious disease in African populations

Most reported large-scale genetic studies have been conducted in populations of European ancestry; the potential clinical benefits resulting from these studies may not therefore be applicable to African populations, who experience the greatest burden of many infectious diseases[18,135,136]. African populations are characterized by greater genetic diversity and shorter regions of linkage disequilibrium (LD) than other major population groups. Therefore, there is a need for genotyping arrays that are denser and specific to ethnic groups in order to provide effective coverage for African populations at the genome-wide level. Indeed, in the Malaria Genomic Epidemiology Network (MalariaGEN) genome-wide association study (GWAS), signals of association at well-established malaria resistance loci were significantly attenuated as a result of only weak LD with genotyped markers[70]. For example, the initial signal near the ungenotyped haemoglobin S (HbS) causal polymorphic variant (rs344) did not reach genome-wide significance because of only weak LD between the genotyped marker SNPs and the causal polymorphism. This limitation was overcome by using population-specific sequencing followed by fine-resolution multipoint imputation; this approach led to a marked increase in the signal of association at the causal HbS locus, from $P=4\times10^{-7}$ to $P=4\times10^{-14}$ (REF. 70). The availability of accurate, population-specific sequence data is essential for such imputation: the tuberculosis GWAS performed in Gambians and Ghanaians reported an imputation error rate of 8% when using the HapMap Yoruban population from Nigeria[72].

Another consequence of the high levels of ethnic diversity in African populations is the risk of false-positive associations resulting from population stratification. The malaria GWAS demonstrated evidence of considerable stratification in the Gambian population, but this was minimized after correction for self-reported ethnicity and principal components analysis[70]. The 1000 Genomes Project will provide further data on sequence variation and haplotype structure in five African populations, and this is anticipated to increase the accuracy of imputation analysis and to facilitate future African GWASs[135,137].

The shorter spans of LD in African populations may in fact confer a considerable methodological advantage, in that they may facilitate the fine mapping of causal variants within a genome-wide association signal. The use of African populations with the phenotype of interest has therefore been proposed as a possible solution to the difficulties in identifying causal variants responsible for GWAS signals in European and Asian populations[135,136].

Genomic studies of developing countries carry practical challenges, such as recruiting large numbers of well-phenotyped individuals in resource-poor settings, as well as particular ethical issues, such as the process of informed consent, safeguarding participants and maintaining anonymity while allowing appropriate data release (reviewed in REF. 69). MalariaGEN provides a model of how such studies may be approached[69,135].

Table 1 | **Markers significantly associated with infectious disease phenotypes in genome-wide studies**

Disease	Phenotype	Population	Sample size*	Most significant marker or markers	SNP location	*P* value*	Odds ratio	Refs
HIV-1 and AIDS	Viral load at set point‡	European	2,554	rs9264942	*HLA-C*	5.9×10^{-32}	NA	33,34
				rs2395029	*HLA-B*, *HCP5*	4.5×10^{-35}	NA	33,34
	Viral load at set point‡	African American	515	rs2523608	*HLA-B*	5.6×10^{-10}	NA	38
	HIV-1 control‡	European	1,712	rs9264942	*HLA-C*	2.8×10^{-35}	2.9	35
				rs4418214	*MICA*	1.4×10^{-34}	4.4	
				rs2395029	*HLA-B*, *HCP5*	9.7×10^{-26}	5.3	
				rs3131018	*PSORS1C3*	4.2×10^{-16}	2.1	
		African American	1,233	rs2523608	*HLA-B*	8.9×10^{-20}	2.6	
				rs2255221	Intergenic	3.5×10^{-14}	2.7	
				rs2523590	*HLA-B*	1.7×10^{-13}	2.4	
				rs9262632	Intergenic	1.0×10^{-8}	3.1	
	Disease progression‡	European	1,071	rs9261174	*ZNRD1*, *RNF39*	1.8×10^{-8}	NA	33,34
	Progression to AIDS 1987‡	European American	755	rs11884476	*PARD3B*	3.4×10^{-9}	NA	41
	Long-term nonprogression‡	European	1,627	rs2395029	*HLA-B*, *HCP5*	6.8×10^{-10}	3.47	42
	Long-term nonprogression‡	European	1,911	rs2234358	*CXCR6*	9.7×10^{-10}	1.85	43
Hepatitis C	Spontaneous clearance	European	1,362	rs8099917	*IL28B*	6.1×10^{-9}	2.31	53
Hepatitis B	Chronic infection	Japanese, Taiwanese	6,387	rs3077	*HLA-DPA1*	2.3×10^{-38}	0.56	60
				rs9277535	*HLA-DPB1*	6.3×10^{-39}	0.57	
Dengue	Dengue shock syndrome	Vietnamese	8,697	rs3132468	*MICB*	4.4×10^{-11}	1.34	65
				rs3765524	*PLCE1*	3.1×10^{-10}	0.80	
Severe malaria	Susceptibility	African (Gambian)	5,900	rs11036238	*HBB*	3.7×10^{-11}	0.63	70
Tuberculosis	Susceptibility	African (Ghana, The Gambia, Malawi)	11,425	rs4334126	18q11.2 (*GATA6*, *CTAGE1*, *RBBP8*, *CABLES1*)	6.8×10^{-9}	1.19	72
Leprosy	Susceptibility	Chinese	11,140	rs3764147	*LACC1*	3.7×10^{-54}	1.68	76
				rs9302752	*NOD2*	3.8×10^{-40}	1.59	
				rs3088362	*CCDC122*	1.4×10^{-31}	1.52	
				rs602875	*HLA-DR-DQ*	5.4×10^{-27}	0.67	
				rs6478108	*TNFSF15*	3.4×10^{-21}	1.37	
				rs42490	*RIPK2*	1.4×10^{-16}	0.76	
Meningococcal disease	Protection	European	7,522	rs1065489	*CFH*	2.2×10^{-11}	0.64	85
				rs426736	*CFHR3*	4.6×10^{-13}	0.63	
Variant Creutzfeldt–Jakob disease	Susceptibility	European, Papua New Guinea	5,183	rs1799990	*PRNP*	2.0×10^{-27}	NA	91

CABLES1, CDK5 and ABL enzyme substrate 1; *CCDC122*, coiled-coil domain-containing 122; *CFH*, complement factor H; *CFHR3*, CFH-related 3; *CTAGE1*, cutaneous T cell lymphoma-associated antigen 1; *CXCR6*, chemokine (C-X-C motif) receptor 6; *GATA6*, GATA-binding protein 6; *HBB*, haemoglobin beta; *HCP5*, HLA complex P5 (non-protein-coding); *HLA*, human leukocyte antigen; *IL28B*, interleukin-28B; *LACC1*, laccase (multicopper oxidoreductase) domain-containing 1; *MIC*, MHC class I polypeptide-related sequence; NA, not applicable or not provided in publication; *NOD2*, nucleotide-binding oligomerization domain-containing 2; *PARD3B*, par-3 partitioning-defective 3 homologue B; *PLCE1*, phospholipase C, epsilon 1; *PRMT6*, protein arginine methyltransferase 6; *PRNP*, prion protein; *PSORS1C3*, psoriasis susceptibility 1 candidate 3 (non-protein coding); *RBBP8*, retinoblastoma-binding protein 8; *RIPK2*, receptor-interacting serine–threonine kinase 2; *RNF39*, ring finger protein 39; *TNFSF15*, tumour necrosis factor [ligand] superfamily member 15; *ZNRD1*, zinc ribbon domain-containing 1. *Results are those reported by the study authors. ‡Phenotype definition is provided in the study reference.

acquisition, viral load following infection, and disease progression[13,25]. Initial candidate-gene approaches identified robust associations between HIV-1 phenotypes and polymorphisms at the chemokine receptor 5–chemokine receptor 2 (*CCR5–CCR2*) locus[26–29], the human leukocyte antigen (HLA) class I region[30,31] and killer immunoglobulin-like receptor (KIR)[25,32] loci. However, many other reported associations have not been confirmed by independent replication (reviewed in REF. 13). Subsequent GWASs of HIV-1 phenotypes have now been reported in different human populations[25,33–44].

HIV-1 enters host cells through attachment to the CD4 receptor and a coreceptor, most commonly either CCR5 or the CXC chemokine receptor CXCR4. Candidate-gene studies showed that individuals who are heterozygous for a 32 base-pair deletion in the cytoplasmic tail of CCR5 (known as CCR5Δ32) progress more slowly to AIDS[26,45]. Furthermore, individuals homozygous for CCR5Δ32, who comprise ~1% of the European population, are highly resistant to acquiring HIV-1 infection, even after repeated exposure[26,46,47]. Conversely, a promoter variant in CCR5 has been associated with increased CCR5 expression, a higher viral-load set point and faster disease progression[27,33,48]. Interestingly, the association with CCR5Δ32 was not observed in an early GWAS[34]; this was probably as a consequence of the rarity of the homozygous state, as well as poor tagging by markers in older genotyping platforms. However, recent, large GWASs were able to demonstrate this association[33,35]. Polymorphism within other chemokine receptor genes has also been implicated in HIV-1 phenotypes. For example, a non-synonymous polymorphism in the related HIV-1 minor coreceptor CCR2 (the CCR2-64I variant at the *CCR5–CCR2* locus) has been reported to associate with improved viral control[28,49].

Both candidate and genome-wide approaches have confirmed a central role for the HLA class I region in HIV-1 infection and progression. HLA class I molecules present viral peptides on the surface of virus-infected cells to $CD8^+$ T cells, and the initiation of this cytotoxic T cell response is essential for the immune control of HIV. The first GWAS of an infectious disease identified HLA SNPs that were highly correlated with viral load during the asymptomatic period of HIV-1 infection in individuals of European ancestry[34] (TABLE 1). The most significant polymorphisms were rs2395029 and rs9264942, two SNPs that exerted independent effects on viral-load set point[34]. rs2395029 is located in the gene HLA complex P5 (non-protein-coding) (*HCP5*) but is in strong linkage disequilibrium (LD) with the HLA allele *B*5701*, and rs9264942 is located in the 5′ region of *HLA-C*. This study also examined disease progression and identified associations with polymorphisms surrounding the genes ring finger protein 39 (*RNF39*) and zinc ribbon domain-containing 1 (*ZNRD1*), which lie close to the *HLA-A* locus in the major histocompatibility complex (MHC)[34]. A recent extension of this study to 2,554 patients confirmed these associated loci, in addition to demonstrating further independent effects within the MHC; however, this study failed to replicate previous candidate gene associations with the exception of the *CCR5–CCR2* locus[33].

Other GWASs examining different phenotypes of HIV-1 infection (for example, extreme HIV-1 nonprogression or very rapid progression) have also highlighted the role of HLA-B. Additionally, they have identified novel loci that require independent replication, such as prospero homeobox 1 (*PROX1*), a negative regulator of interferon-γ (IFNγ) signalling, and the chemokine receptor gene *CXCR6* (REFS 40,43). A recent GWAS examined untreated HIV-1 controllers and progressors in a large, multi-ethnic cohort and identified 313 SNPs with $P < 5 \times 10^{-8}$, all of which were located in the MHC (the most significant, independent associations are presented in TABLE 1)[35]. The study replicated the involvement of the *CCR5–CCR2* locus and extended previous HLA associations through imputation of HLA type, identifying HLA types B*57:01, B*27:05, and B*14 as protective, and C*57 as associated with progression to AIDS. It also identified the specific amino acid positions underlying these association signals. The most significant was located at position 97, which is in the floor of the peptide-binding groove of HLA-B. However, the precise mechanism by which this residue influences the interaction between the HLA and viral peptides to protect against AIDS remains unknown. One possibility is that by encountering fewer self-peptides during thymic development, protective HLA molecules generate a naive $CD8^+$ T cell repertoire with a greater fraction of T cells that recognize viral peptides[50]. The mechanism behind the independent effect of *HLA-C* on HIV-1 infection is also unclear, although it may reflect, in part, variation in *HLA-C* expression[35,51].

Taken together, GWASs have confirmed the major roles of polymorphism in HLA class I loci and the *CCR5–CCR2* locus in the pathogenesis of HIV infection but have not identified additional major susceptibility loci for the phenotypes studied. Of note, the combined effects of HLA class I and *CCR5* explain only approximately 23% of the observed variability in HIV-1 control[35]. The source of the remaining inter-individual variability is currently unknown, but may reflect viral or environmental factors, or the combined effect of multiple host genetic variants that are individually rare but have large effects.

Hepatitis C virus spontaneous clearance and treatment response. Hepatitis C virus (HCV) infects an estimated 180 million individuals worldwide and is a leading cause of cirrhosis and hepatocellular carcinoma. Initial GWASs reported strong associations between SNPs upstream of interleukin-28B (*IL28B*) and the response to IFNα treatment of HCV[52–56]. Subsequent candidate-gene studies and GWASs demonstrated that the *IL28B* genotype also associates with the spontaneous clearance of HCV in individuals of European and African ancestry[53,57]. The biological mechanisms underlying the association between *IL28B* polymorphism and HCV clearance and treatment response remain unknown. Expression of the protein product, IL-28B (also known as IFNλ3) is induced by viral infection and activates the same Janus kinase (JAK)–signal transducer and activator of transcription (STAT) antiviral pathway as does IFNα by binding to a distinct receptor complex[58].

Multiple testing
An analysis in which multiple independent hypotheses are tested within a single data set. Performing multiple independent tests increases the likelihood of a type I error and requires adjustment of the threshold significance level (*P* value); for example, by using the Bonferroni correction.

Principal components analysis
A statistical method that is used to simplify a complex data set by transforming a series of correlated variables into a smaller number of uncorrelated variables called principal components. It is commonly used to correct for stratification in genome-wide association studies.

'Common disease, common variant' hypothesis
This hypothesis states that the genetic variants that underlie complex disease are themselves common.

Heritability
The proportion of phenotypic variation in a given characteristic or state that is due to underlying genetic variation.

Human leukocyte antigens
(HLAs). Glycoproteins encoded by the major histocompatibility complex locus that are located on the surface of antigen-presenting cells and that present antigenic peptides to T cells.

Killer immunoglobulin-like receptors
Receptors expressed at the surface of natural killer (NK) cells that regulate NK cell cytotoxic activity through specific interactions with HLA class I molecules.

Viral-load set point
The HIV RNA level in the blood during the asymptomatic period of HIV-1 infection. It fluctuates around a steady set-point value, although it is variable both within patients over time and among patients. High viral-load set points correlate with rapid disease progression.

Furthermore, there is evidence that IFNα and IFNλ isoforms may exert anti-HCV activity *in vitro* by distinct but complementary mechanisms[59]. Taken together, these findings highlight the key role of IFNλ signalling in the host control of HCV infection, and raise the possibility that the combined use of IFNα and IFNλ may have therapeutic application.

Chronic hepatitis B infection. Hepatitis B virus (HBV) is a major global threat to health that affects >400 million individuals worldwide and is a leading cause of hepatocellular carcinoma. Clinical outcomes following HBV exposure in adulthood are highly variable, with only a minority of individuals developing chronic infection. A GWAS identified strong associations between variants in *HLA-DPA1* and *HLA-DPB1* in the HLA class II region and protection against chronic HBV infection in an Asian population[60]. Independent studies in cohorts from East Asia have confirmed these findings and further suggested that the associations reflect a specific effect on HBV clearance[61–63]. Class II HLA molecules form heterodimers on the surface of antigen-presenting cells, where they bind and present antigen to $CD4^+$ T-helper lymphocytes. HLA-DP molecules display considerable polymorphism, and the observed associations may reflect structural variation in the antigen-binding sites that influence the presentation of viral peptides and the subsequent immune response to HBV. Variants might also affect HLA mRNA expression levels, and lower expression of *HLA-DPA1* and *HLA-DPB1* has been associated with an increased risk of chronic HBV[64].

Dengue. Dengue is a mosquito-borne viral infection that is responsible for 100 million infections each year worldwide. The clinical consequences of infection vary from asymptomatic infection to rapidly fatal disease; the most frequent severe complication is dengue shock syndrome (DSS), which is characterized by an increase in vascular permeability leading to hypovolaemic shock and often death. A recent GWAS of Vietnamese children with DSS identified susceptibility loci at MHC class I polypeptide-related sequence B (*MICB*), which is within the MHC region on chromosome 6, and at phospholipase C, epsilon 1 (*PLCE1*) on chromosome 10 (REF. 65) (TABLE 1). Conditional analysis suggested that the *MICB* SNP accounted for the majority of the MHC association signal, although residual associations were also noted with neighbouring genes, including *HLA-B* and *HLA-C*. Unequivocal identification of the causative gene was therefore not possible but *MICB* remains the most likely candidate[65]. *MICB* encodes an inducible activating ligand that promotes antiviral immunity by natural killer cells and $CD8^+$ T cells after binding their natural killer group 2, member D (NKG2D) receptor[66]; in addition, *MICB* expression is upregulated during human dengue infection[67]. The identification of *PLCE1* is also likely to advance understanding of the pathogenesis of DSS: mutations in this gene have been described as a cause of nephrotic syndrome, a renal disease that is characterized by an impaired integrity of the glomerular basement membrane and that shares some clinical features with DSS[68].

Malaria. The malaria parasite *Plasmodium falciparum* kills one million African children each year, and host genetic factors are known to have an important role in the likelihood of developing severe malaria following infection[69]. The Malaria Genomic Epidemiology Network (MalariaGEN) consortium recently reported their initial results of a GWAS of severe malaria in a large sample size from The Gambia[70] (see also BOX 1). The strongest signal of association was located close to the haemoglobin beta (*HBB*) gene, which contains the classic sickle haemoglobin variant haemoglobin S (HbS) polymorphism. The causal SNP for HbS (rs334) results in the non-synonymous replacement of glutamic acid with valine at amino acid residue 6 of the β-globin chain. Despite rs334 homozygotes experiencing life-threatening sickle cell disease, rs334 heterozygotes have a tenfold reduced risk of severe malaria[10,71]. The GWAS also identified various novel candidate loci that require further study and replication, including variants in *SCO1*, which encodes a protein involved in cytochrome oxidase function, and *DDC*, which encodes dopa decarboxylase, an enzyme involved in the synthesis of dopamine and serotonin[70].

Tuberculosis. *M. tuberculosis* infects an estimated one-third of the world's population and accounts for two million deaths per year. Only 10% of individuals infected with *M. tuberculosis* go on to develop tuberculosis disease, and the factors that determine this inter-individual variation in susceptibility remain only partly understood. Multiple human loci have been implicated in host susceptibility to tuberculosis on the basis of linkage analysis and candidate-gene studies[10], although the degree of replication between these reports is low. The only reported GWAS of tuberculosis recently described a susceptibility locus within a gene-poor but highly conserved region on chromosome 18q11.2, suggesting a possible regulatory effect on a currently unknown gene that is located elsewhere on the chromosome[72]. However, additional research is required to pinpoint the causative variant and to explore its functional significance.

Leprosy. Leprosy is a chronic infectious disease that is caused by the intracellular pathogen *Mycobacterium leprae*. This pathogen exhibits a tropism for peripheral nerve Schwann cells and macrophages, and affects the skin and peripheral nervous system. Initial linkage and candidate-gene approaches implicated various gene variants in leprosy susceptibility. In particular these were the *HLA-DR* region, the parkin and *PARK2* co-regulated (*PARK2–PACRG*) locus, and lymphotoxin alpha (*LTA*)[3,73–75]. More recently, the first GWAS of leprosy susceptibility reported convincing associations with markers in six genetic loci: *HLA-DR-DQ*, receptor-interacting serine–threonine kinase 2 (*RIPK2*), tumour necrosis factor [ligand] superfamily member 15 (*TNFSF15*), laccase (multicopper oxidoreductase) domain-containing 1 (*LACC1*; previously known as *C13orf31*), coiled-coil domain-containing 122 (*CCDC122*) and nucleotide-binding oligomerization domain-containing 2 (*NOD2*) (REF. 76) (TABLE 1). A weak association was also observed with leucine-rich repeat

Linkage disequilibrium
(LD). The occurrence of alleles together at frequencies greater than are expected by chance alone.

Major histocompatibility complex
(MHC). A large complex of tightly linked genes on human chromosome 6, many of which are involved in the immune response. The HLA genes are located within the MHC.

HIV-1 nonprogression
HIV-1 disease progression may be assessed by a number of measures, which can be used to define subgroups of individuals with long-term nonprogression. It can be defined, for example, as asymptomatic HIV-1 infection and a stable $CD4^+$ T cell count for ten or more years in the absence of treatment.

HIV-1 controllers
A small subgroup of individuals infected with HIV-1 who maintain very low or undetectable plasma HIV viral loads in the absence of antiviral therapy.

Imputation
A statistical method that infers the genotype of untyped variants based on the known genotypes of a smaller number of nearby markers and the known linkage disequilibrium relationships in fully genotyped individuals (provided by HapMap reference data).

Hypovolaemic shock
A life-threatening condition in which fluid loss results in a reduction in the intravascular blood volume and inadequate tissue perfusion.

Conditional analysis
A form of statistical testing that considers the independent contributions of variants in a region of association containing multiple variants in linkage disequilibrium, in order to identify the most likely causal variant or variants.

kinase 2 (*LRRK2*). The earlier associations with *PARK2–PACRG* and *LTA* did not replicate in this study, although not all of the associated markers were represented on the GWAS array[3,76,77]. Associations between leprosy and the *HLA-DR-DQ* region, *LACC1* and *CCDC122* were subsequently replicated in an Indian population[78,79], and a further association was observed with the I602S functional SNP in the Toll-like receptor 1 (*TLR1*) gene[78]. The association between *NOD2* and leprosy was not replicated in the Indian study but has been reported in a Nepalese population[78,80]. The lack of complete replication between these studies may reflect: different sample sizes and hence study power; the use of different genotyping arrays; differences in allele frequency and haplotype structure between populations; and/or true population-specific variation in susceptibility (for example, resulting from differing environmental exposures that are necessary for disease). The *TLR1* I602S variant, for example, is very rare in Chinese populations (1.7%).

The genetic association between the *HLA-DR-DQ* region and susceptibility is already well-established on the basis of linkage and candidate-gene analyses[3,75]. HLA-DR molecules on the surface of phagocytes present *M. leprae* antigens to $CD4^+$ T cells, leading to the generation of a T-helper 1 response. This triggers the production of IFNγ and IL-2, macrophage maturation, granuloma formation and mycobacterial containment[3]. Rare mutations leading to impaired IFNγ signalling have been described in association with other phenotypes of mycobacterial disease (BOX 2).

Although the biological functions of *LACC1* and *CCDC122* are poorly understood, a striking feature of the remaining associations with *RIPK2*, *TNFSF15* and *NOD2*, as well as with *TLR1* is that they each encode proteins involved in innate immunity (FIG. 1). Key questions remain regarding the mechanisms by which associated loci influence the development of disease; for example, it is unclear how reduced inflammatory signalling through the hypofunctional TLR1 I602S variant confers protection against leprosy[78]. A further surprising finding is the overlap between susceptibility loci for leprosy and the inflammatory bowel condition Crohn's disease: *NOD2*, *TNFSF15*, *LRRK2* and the 13q14 locus containing *LACC1* and *CCDC122* have each been reported as susceptibility loci for Crohn's disease[81,82]. This suggests a shared immunological basis for these two diseases and provides further evidence to support a role for impaired recognition and processing of mycobacteria in the pathogenesis of Crohn's disease, although this remains controversial[77,83].

Meningococcal disease. Meningococcal disease is caused by the Gram-negative bacterium *Neisseria meningitidis* and is a leading cause of severe sepsis and meningitis in young adults. As with other major bacterial causes of sepsis, asymptomatic carriage of the meningococcus is common in the general population yet life-threatening, invasive disease develops in only a minority of individuals. Previous candidate-gene studies of meningococcal disease were largely characterized by small sample sizes and low levels of replication (reviewed in REF. 14), although a possible association with a promoter region polymorphism in the complement factor H (*CFH*) gene was reported[84]. A GWAS identified multiple SNPs in *CFH* and the adjacent CFH-related 3 (*CFHR3*) gene that robustly associated with meningococcal disease[85] (TABLE 1). These polymorphisms were in complete LD with one another, and further research is required to pinpoint the functional variant or variants. The pattern of associated SNPs differed from those previously associated with age-related macular degeneration at this locus[86]. A central role for complement in meningococcal host defence is already well-established: *N. meningitidis* is known to bind host CFH in order to evade complement-mediated killing, and very rare, highly penetrant mutations in complement components underlie some cases of susceptibility to recurrent meningococcal disease[87] (TABLE 2). Nevertheless, the demonstration that common variation at the *CFH–CFHR3* locus robustly associates with susceptibility should further focus research on this pathway in meningococcal disease.

Variant Creutzfeldt–Jakob disease. Prion diseases are transmissible, fatal neurodegenerative conditions that are characterized by post-translational modifications of host-encoded prion proteins into a pathological isoforms that have a tendency to aggregate and form amyloid fibrils. Transmission of bovine spongiform encephalopathy (BSE) prions to humans occurred in the 1990s in the United Kingdom, leading to approximately 200 cases of variant Creutzfeldt–Jakob disease (vCJD). Other forms of human prion disease include sporadic CJD, iatrogenic CJD and kuru. A polymorphism at codon 129 (which encodes methionine or valine) of the human prion protein (*PRNP*) gene is an established susceptibility factor for all types of prion disease[88,89]; familial forms of CJD are well-described and are linked to rare mutations in *PRNP*[90].

The involvement of the *PRNP* polymorphism at codon 129 has been confirmed in a recent GWAS involving 119 patients with vCJD in the discovery phase, followed by replication studies in patients with other prion diseases[91]. Consistent with previous studies, every case of vCJD genotyped was found to be homozygous for the variant encoding methionine at codon 129 (TABLE 1), which is the genotype of approximately 40% of healthy Europeans. This represents the strongest association between a common polymorphism and infectious disease reported to date. The study also suggested other candidate susceptibility loci. These included an independent SNP downstream of *PRNP* and SNPs upstream of the retinoic acid receptor beta (*RARB*) and stathmin 2 (*STMN2*) genes[91]. These candidates are biologically plausible: retinoic acid regulates prion protein expression in cell cultures, and STMN2 has a role in neuronal microtubule stability[91]. However, the sample sizes are small, which reflects the rarity of these conditions, and confirmation in larger studies is needed. Exposure to BSE was probably widespread in the UK population, and it remains uncertain whether further cases with longer incubation periods may subsequently emerge, perhaps in association with different codon 129 genotypes; a recent case report of probable vCJD in a *PRNP* codon 129 heterozygous individual supports this possibility[92].

Innate immunity
An immediate, non-specific and evolutionarily conserved ancient form of immune response to foreign infectious agents. It constitutes the first line of defence against invading organisms.

Crohn's disease
A type of inflammatory bowel disease characterized by granulomatous inflammation that may occur in any region of the gastrointestinal tract.

Complement
A complex protein cascade that is involved in both innate and adaptive immunity. Complement activation results in pathogen opsonization and cell lysis, and occurs through three different pathways: alternate, classical and lectin.

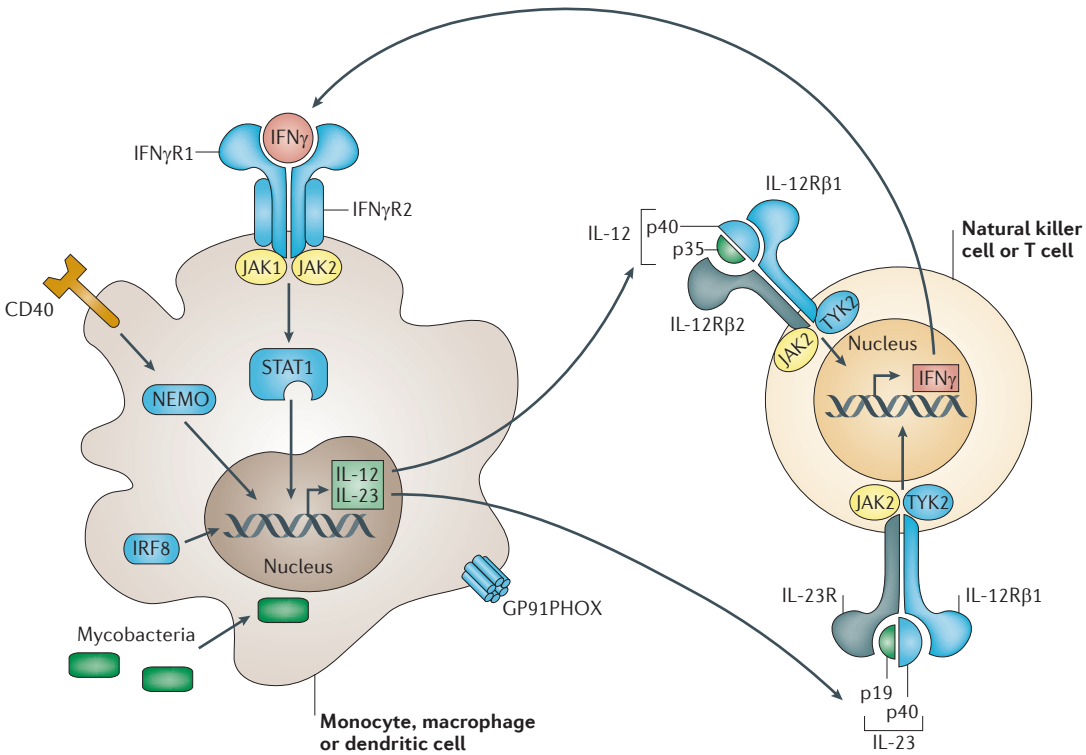
Respiratory burst
The rapid generation and release of reactive oxygen species that follows exposure of phagocytes (primarily neutrophils and macrophages) to microorganisms or inflammatory mediators.

Box 2 | IL-12–IL-23–IFNγ pathway defects and Mendelian susceptibility to mycobacterial disease

Patients with Mendelian susceptibility to mycobacterial disease (MSMD) are vulnerable to disease caused by weakly pathogenic mycobacteria, in particular non-tuberculous mycobacteria and the bacille Calmette–Guérin (BCG) vaccine, as well as displaying increased susceptibility to invasive salmonellosis, tuberculosis, and severe viral infection in some cases[95,138]. Genetic defects in members of the interleukin-12 (IL-12)–IL-23–interferon-γ (IFNγ) signalling pathway have been identified in association with MSMD (see the figure; gene defects underlying MSMD are shown in blue). Following infection with intracellular pathogens, monocytes, macrophages and dendritic cells secrete IL-12 and IL-23 (the cell on the left in the figure). These cytokines drive the development and differentiation of naive T cells into T-helper 1 cells, leading to IFNγ production (the cell on the right in the figure). IFNγ signalling is mediated through the Janus kinase (JAK)–signal transducer and activator of transcription (STAT) pathway, involving the phosphorylation of JAK1 and JAK2 and the subsequent activation of STAT1, which mediates transcription of IFNγ-inducible genes.

Mutations underlying MSMD have been identified in the genes encoding IFNγ receptor 1 (*IFNGR1*), IFNγ receptor 2 (*IFNGR2*), STAT1 (*STAT1*), the p40 subunit of IL-12 and IL-23 (*IL12B*), IL-12 receptor β1 (*IL12RB1*) and nuclear factor-κB (NF-κB) essential modulator (*NEMO*)[95,133,138]. The specific NEMO mutations seem to confer susceptibility to mycobacterial disease by partially impairing CD40 signalling, leading to defective IL-12 production[138]. Homozygous mutations in the JAK family member tyrosine kinase 2 (*TYK2*) have been also described in association with susceptibility to mycobacterial disease, albeit in the setting of a broad-spectrum immunodeficiency[130]. Genetic defects underlying MSMD lead to an impaired ability to produce or respond to IFNγ, highlighting the essential role of this cytokine in protective immunity to mycobacteria, and to a lesser extent other intracellular pathogens, such as *Salmonella* spp. However, patients with MSMD are usually otherwise healthy and seem to be resistant to most other pathogens. The clinical penetrance of some disorders (for example, *IL12RB1* deficiency) is low, perhaps reflecting variation in pathogen exposure or the effect of additional, currently unknown host genetic loci. Furthermore, the clinical status of IL-12 p40 and IL-12 receptor β1 (IL-12Rβ1) deficiency frequently improves with age. The possible role of common genetic variation in the IL-12–IL-23–IFNγ axis has also been studied, and candidate-gene approaches have implicated common polymorphism within *IFNG* (encoding IFNγ) and *IFNGR1* in susceptibility to tuberculosis, although replication studies have reported variable results[72,140–143].

Very recently, gene defects in *CYBB* (encoding the GP91PHOX subunit of the phagocyte NADPH oxidase complex)[144] and interferon regulatory factor 8 (*IRF8*)[145] have been described as novel causes of MSMD. The MSMD-associated *CYBB* mutations seem to impair the respiratory burst only in macrophages, in contrast to other *CYBB* mutations that impair microbial killing by all phagocytes, resulting in chronic granulomatous disease[144]. The *IRF8* mutation was associated with selective depletion of IL-12-producing $CD11c^+CD1c^+$ myeloid dendritic cells, probably accounting for the enhanced mycobacterial susceptibility. Interestingly, a different *IRF8* mutation resulted in a complete lack of mononuclear phagocyte subgroups and a more severe, broader-spectrum immunodeficiency[145]. The figure is modified, with permission, from REF. 139 © (2007) Macmillan Publishers Ltd. All rights reserved.



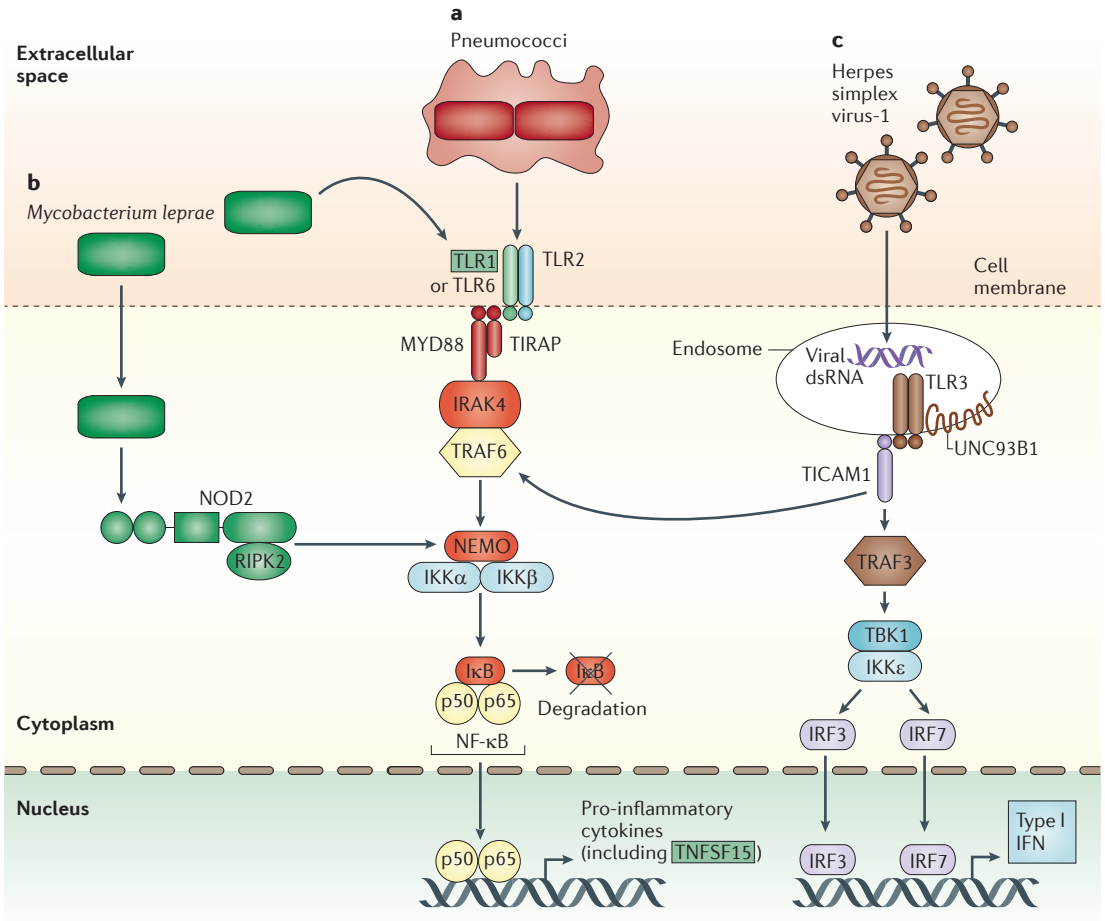


Figure 1 | **Genetic variation in TLR signalling and susceptibility to infectious disease.** Toll-like receptors (TLRs) are pattern-recognition receptors that initiate intracellular signalling after the recognition of distinct pathogen-associated molecular patterns[96]. Examples are presented of genetic variants reported in association with susceptibility to three specific infectious disease phenotypes: invasive disease due to *Streptococcus pneumoniae* infection (genetic associations are shown in orange) (**a**), leprosy (caused by *Mycobacterium leprae*; genetic associations are shown in green (**b**) and herpes simplex encephalitis (genetic associations are shown in brown) (**c**). **a** | TLR2 forms heterodimers with TLR1 or TLR6 and recognizes, among other agonists, the lipoteichoic acid and peptidoglycan cell wall components of Gram-positive bacteria such as *S. pneumoniae*. Recognition of microbial products leads to the recruitment of the protein adaptors myeloid differentiation primary response protein 88 (MYD88) and Toll–interleukin 1 receptor domain-containing adaptor protein (TIRAP), followed by the activation of a kinase cascade. This cascade includes the IL-1 receptor-associated kinases (IRAKs) and the IκB kinase (IKK) complex, comprising IKKα, IKKβ and NF-κB essential modulator (NEMO; also known as IKKγ). The cascade results in the degradation of inhibitors of NF-κB (IκB)[96]. This releases the transcription factor nuclear factor-κB (NF-κB) to translocate into the nucleus and control the expression of various pro-inflammatory genes, including tumour necrosis factor [ligand] superfamily member 15 (*TNFSF15*). **b** | TLR1–TLR2 heterodimers also lead to NF-κB activation after recognition of *M. leprae*[3]. The cytosolic pattern recognition receptor, nucleotide-binding oligomerization domain-containing 2 (NOD2) recognizes muramyl dipeptide in the mycobacterial cell wall and then interacts with receptor-interacting serine–threonine kinase 2 (RIPK2) to activate NF-κB through the recruitment of NEMO[3,81,166]. **c** | TLR3 recognizes viral dsRNA and uses the adaptor TIR domain-containing adapter molecule 1 (TICAM1) to activate the transcription factors interferon regulatory factor 3 (IRF3) and IRF7 through interactions with TNF receptor-associated factor 3 (TRAF3). IFN, interferon.

Single-gene mutations

The current interest in the role of individually rare, large-effect variants in susceptibility to common disease is particularly relevant to infectious disease, as it parallels the description of an increasing number of single-gene defects underlying specific infectious-disease phenotypes. Indeed, the identification of mutations in individual immune-related genes that underlie susceptibility to only a narrow range of different pathogens in otherwise healthy children and adults has led to an evolving concept of the nature of genetically determined immunodeficiency[93]. Furthermore, such 'selective' primary immunodeficiency (PID) conditions may also display incomplete penetrance, in marked contrast to the traditional concept of PID, which was restricted to highly penetrant, single-gene mutations in multicase families leading to recurrent and diverse childhood infections[93]. A well-studied example of a selective PID is Mendelian

Penetrance
The proportion of individuals with a given genotype who display a particular phenotype.

Table 2 | **Single-gene variants associated with increased susceptibility or resistance to specific infectious disease phenotypes**

Pathogen	Phenotype	Probable mechanism	Gene or genes	Refs
Streptococcus pneumoniae	Susceptibility to invasive disease	Impaired TLR–IL-1R signalling	*IRAK4*	94,100,103
			MYD88	94,97
Neisseria meningitidis	Susceptibility to invasive disease	Membrane attack complex deficiency	*C5-C9*	15,87, 167-169
		Properdin deficiency	*CFP*	
		Factor D deficiency	*CFD*	
Encapsulated bacteria (for example, *S. pneumoniae*, *N. meningitidis* and *Haemophilus influenzae*)	Susceptibility to invasive disease	Classical complement pathway deficiency	*C2*, *C1Q*, *C1R*, *C1S*, *C4*	15,87, 169,170
Mycobacteria	Susceptibility (MSMD)	Impaired IFNγ response	*IFNGR1*, *IFNGR2*, *STAT1*	95,138
		Impaired IFNγ production	*IL12B*, *IL12RB1*, *NEMO*	
		Impaired macrophage respiratory burst	*CYBB*	144
		Impaired differentiation of dendritic cell subgroups	*IRF8*	145
HIV-1	Resistance	Absence of coreceptor for pathogen	*CCR5*	26,46,47
HSV-1	Susceptibility to HSV encephalitis	Impaired production of IFNα, IFNβ and/or IFNλ	*UNC93B1*	118
			TLR3	119
			TRAF3	120
Human herpesvirus-8	Classic Kaposi's sarcoma	T cell deficiency	*STIM1*	171
Human papillomaviruses	Epidermodysplasia verruciformis	Unknown	*TMC6*, *TMC8*	172
Norovirus	Resistance	Absence of receptor for pathogen	*FUT2*	173,174
Plasmodium vivax	Resistance	Absence of coreceptor for pathogen	*DARC*	175
Candida albicans	Chronic mucocutaneous candidiasis	Impaired IL-17 immunity	*CARD9*	176
			IL17RA	177
			IL17F	177
			STAT1	126,127
Bovine spongiform encephalopathy prion	Susceptibility to variant Creutzfeldt–Jakob disease	Post-translational modification of host cellular prion protein	*PRNP*	88,91

C1, *C2*, *C4*, *C5–C9*, complement components; *CARD9*, caspase recruitment domain family member 9; *CCR5*, chemokine (C-C motif) receptor 5; *CFD*, complement factor D; *CFP*, complement factor properdin; *CYBB*, cytochrome *b*-245, beta polypeptide (encodes GP91PHOX); *DARC*, Duffy blood group chemokine receptor; *FUT2*, fucosyltransferase 2; HSV, herpes simplex virus; IFNγ, interferon-γ; *IFNGR*, IFNγ receptor; *IL*, interleukin; *IL12RB1*, IL-12 receptor β1; *IRAK4*, interleukin-1 receptor-associated kinase 4; *IRF8*, interferon regulatory factor 8; MSMD, Mendelian susceptibility to mycobacterial disease; *MYD88*, myeloid differentiation primary response gene 88; *NEMO*, nuclear factor-κB essential modulator; *PRNP*, prion protein; *STAT1*, signal transducer and activator of transcription 1; *STIM1*, stromal-interaction molecule 1; TLR, Toll-like receptor; *TMC*, transmembrane channel-like (also known as *EVER* genes); *TRAF3*, TNF receptor-associated factor 3; *UNC93B1*, *unc-93* homologue B1.

susceptibility to mycobacterial disease (MSMD) (BOX 2). More recently, the identification of genetic defects in TLR signalling has led to considerable insights into human immunity to invasive pneumococcal disease and herpes simplex viral encephalitis (discussed below).

An overview of single-gene variants reported in association with marked susceptibility or resistance to specific infectious disease phenotypes is presented in TABLE 2. It should be noted that although these immunodeficiency states are widely considered to be selective, the specificity of pathogen susceptibility is not always absolute. As examples, invasive staphylococcal and Gram-negative bacterial disease occur in a minority of patients with deficiency of either IL-1 receptor-associated kinase 4 (*IRAK4*) or myeloid differentiation primary response gene 88 (*MYD88*)[94], and invasive salmonellosis is common in some forms of MSMD[95]. Nevertheless, the range of pathogen susceptibility is typically much narrower in humans than that observed in the corresponding mouse knockout models[93].

The TLR–NF-κB pathway and invasive pneumococcal disease. The TLR–nuclear factor-κB (NF-κB) signalling pathway has a central role in the early recognition of invading pathogens and the initiation of an inflammatory host response[96]. A specific role for TLR signalling and NF-κB activation in the human immune response to infection with *Streptococcus pneumoniae* (the pneumococcus) has been firmly established following the investigation of individuals with rare PIDs that resulted in invasive pneumococcal disease (IPD). These studies identified causative mutations in four genes in the TLR–NF-κB pathway that resulted in impaired NF-κB activation: NF-κB essential modulator (*NEMO*; also known as *IKBKG*, this encodes the I-κB kinase regulatory subunit IKKγ), *NFKBIA* (which encodes the IκBα inhibitor of

Box 3 | **Expanding the boundaries of infectious disease genetic studies**

Most studies have assessed variation solely in either the host or pathogen genome, whereas it is likely that the outcome of exposure to an infectious agent reflects the interaction between specific human and pathogen genotypes. The nature of this interaction may drive the co-evolution of the host and pathogen, and adaptation of pathogens to host polymorphism in specific human populations may in part explain the geographical distribution of pathogen strains. For example, human genetic variation in the expression of fucosylated blood group antigens in the gastrointestinal tract seems to affect the susceptibility to infection with *Helicobacter pylori* and norovirus in a strain-specific manner[146–148]. Additionally, *Mycobacterium tuberculosis* displays considerable genomic variation across different geographic regions[149], suggesting that it has adapted to specific human populations.

Consistent with the existence of a range of specific host–pathogen interactions, a genome-wide screen using small interfering RNAs (siRNAs) to systematically inhibit the expression of individual host genes revealed considerable variability in the host determinants of intracellular pathogen load across a panel of *M. tuberculosis* field isolates. This study found that the genotypic and phenotypic characteristics of the pathogen influenced the types of host molecules identified[150]. Animal studies have further demonstrated substantial interactive effects between host genetic backgrounds and mycobacterial strains on susceptibilities to infections[151]. Furthermore, a specific association between Toll-like receptor 2 (*TLR2*) polymorphism and susceptibility to the East-Asian (Beijing) strain of *M. tuberculosis* has been reported in humans[152].

The degree of pathogen diversity may directly influence the success of infectious disease genome-wide association studies (GWASs). For example, host genetic polymorphism might exert a greater influence on inter-individual variation in susceptibility in the setting of the unusually low levels of *Mycobacterium leprae* genomic diversity[77,153]. This might explain the greater number of loci identified in GWASs of leprosy when compared to studies of more genetically diverse pathogens such as *M. tuberculosis*[72,76,77]. Factors such as the dose and route of infection — as well as the possibility of co-infection with multiple pathogen strains — may also be involved in determining the outcome of the host–pathogen interaction. Such factors are difficult to account for in genetic association studies, although even a crude analysis of exposure intensity seemed to increase the power to detect genetic associations with tuberculosis in human family-based studies[154].

Gene–gene and gene–environment interactions are even more challenging to model. For example, epistatic effects have been demonstrated in susceptibility to infection in *Drosophila melanogaster* Toll signalling pathways[155]. However, the possibility of analogous epistatic interactions existing between infectious-disease-associated variants in human TLR–nuclear factor-κB (NF-κB) signalling (FIG. 1) has not been explored. The effects of host genetic polymorphism on phenotype may also be dependent on specific environmental factors. For example, preliminary data suggest an interactive effect between polymorphism in iron-regulatory genes and host iron status on HIV infection mortality[156].

The relative contributions to clinical phenotypes of host and pathogen genetic variability and environmental influences, as well as their potential interactive, joint effects, remain largely unknown. Integrative analysis of host, pathogen and environmental variation represents a major challenge, and it is dependent on detailed clinical and microbiological phenotyping, as well as extremely large sample numbers.

NF-κB), *IRAK4* and *MYD88* (REFS 93,94,97–102) (FIG. 1). The *NEMO* and *NFKBIA* mutations interrupt multiple innate and adaptive pathways that signal to NF-κB, including the TLR pathway. The subsequent immunodeficiency states are typically severe and encompass susceptibility to a broad range of pathogens, including encapsulated bacteria, atypical mycobacteria, fungi and viruses[98,99,101,102]. By contrast, mutations in *IRAK4* and *MYD88* disrupt only TLR and IL-1 receptor signalling and associate with a narrower range of infectious pathogens in childhood, primarily pyogenic encapsulated bacteria and, in particular, recurrent IPD[94,97,100,103]. Children with *IRAK4* or *MYD88* mutations are typically resistant to other bacteria, mycobacteria, viruses and fungi, and classically their immunodeficiency improves with age[94].

Although these mutations seem to be extremely rare, candidate-gene studies have reported common polymorphism within TLR–NF-κB pathway genes in association with related infectious disease phenotypes[104]. For example, common polymorphisms in *NFKBIA* and the related genes *NFKBIZ* and *NFKBIL2*, which all encode inhibitors of NF-κB, have been described in association with susceptibility to IPD in adults[105–107]. A functional polymorphism in the gene encoding the Toll–interleukin 1 receptor domain-containing adaptor protein (*TIRAP*) has been reported in different human populations in association with susceptibility not only to IPD, but also to other Gram-positive and Gram-negative bacterial infections, malaria and tuberculosis[108]. Independent studies of tuberculosis and sepsis have reported variable results, including a lack of association, association with different *TIRAP* polymorphisms, and clear replication[109–116]. The initial direction of effect was of heterozygote protection, with an overall odds ratio for infectious disease in heterozygotes of ~0.5. Interestingly, the heterozygous genotype resulted in intermediate levels of inflammatory cytokine responses following intravenous lipopolysaccharide challenge, suggesting that extremes of TLR–NF-κB signalling may be detrimental[117].

TLR3–IFN pathways and HSV encephalitis. Herpes simplex virus-1 (HSV-1) infection is widespread in children, and the life-threatening complication herpes simplex encephalitis (HSE) develops in only a small minority of individuals. Mutations in *unc-93* homologue B1 (*UNC93B1*)[118], *TLR3* (REF. 119) and TNF receptor-associated factor 3 (*TRAF3*)[120] have been identified as rare causes of isolated susceptibility to HSE in otherwise healthy children who seem to be resistant to other forms of HSV disease and indeed to other viruses.

Odds ratio
A comparison of the odds of exposure to a susceptible genetic variant in cases versus controls.

Lipopolysaccharide
A Gram-negative bacterial cell wall component that is recognized by Toll-like receptor 4 and triggers a pro-inflammatory host response.

Box 4 | **Opportunities for clinical translation of infectious disease host genetics**

A major challenge is to develop strategies for translating insights from the genetic basis of infectious disease into improved patient outcomes. The greatest potential for rapid clinical translation probably comes from the field of pharmacogenomics, in which one of best examples is of hypersensitivity reactions to HIV therapy. Severe hypersensitivity to the antiretroviral drug abacavir is strongly associated with *HLA-B*5701*. Clinical trials have shown that genetic testing is of benefit in preventing abacavir hypersensitivity, and testing before the prescription of abacavir is now recommended in clinical guidelines and is widely practiced[157–161].

Another well-studied example is genetic variation in response to hepatitis C virus (HCV) combination therapy with pegylated interferon-α (IFNα) and the antiviral drug ribavirin. Multiple genome-wide association studies (GWASs) in different populations have reported strong associations between treatment response and interleukin-28B (*IL28B*) polymorphism[52–56]. The size of the genotypic effect on sustained viral response was approximately twofold in European populations, threefold in African populations and significantly larger in Asian populations. However, the *IL28B* genotype alone is not sufficiently discriminatory to inform clinical decision-making[162]. In addition to genetic predictors of treatment response, GWASs have also investigated the haematological side effects of IFN–ribavirin therapy and identified significant associations with functional variants in the inosine triphosphatase (*ITPA*) gene[163–165]. Future studies should integrate genotypic information with other host and viral factors into combined predictive models to prospectively evaluate treatment responses and durations, as well as adverse event profiles.

Insights from pharmacogenomics will probably also influence the treatment of other infectious diseases. One possibility is that the considerable genetic variation in Toll-like receptor–nuclear factor-κB (TLR–NF-κB) signalling (FIG. 1) may influence therapeutic strategies aimed at manipulating this pathway, such as in the treatment of sepsis. Clinical trials that fail to take into account human genetic variation may miss relevant effects on subgroups of individuals, such as those with extremes of inflammatory signalling. Indeed, such an effect may partly account for the disappointing outcomes of clinical trials of anti-inflammatory agents for the treatment of sepsis. There is a need to identify the functional genetic variants controlling inter-individual variation in TLR–NF-κB signalling and to stratify clinical trials of immunomodulatory agents by host genotype.

The use of genetic information to predict infectious disease risk in individuals is unlikely to alter clinical practice in the near future, and the predictive value of genetically based risk assessment remains poor, even in more-extensively studied, non-infectious disease phenotypes[20]. Clinical translation is more likely to result from the characterization of the molecular pathways involved in disease and the identification of novel targets for immunomodulatory drugs or vaccines. For example, the identification of specific gene defects underlying herpes simplex encephalitis raises the possibility that the use of IFNα treatment to supplement antiviral agents may be of therapeutic benefit in this condition[118]. The application of systems biology to integrate genome-wide studies, including genome, transcriptome and proteome analyses, as well as small interfering RNA (siRNA) screens, may be a particularly powerful approach for identifying novel therapeutic targets.

TLR3-dependent induction of IFNα, IFNβ and IFNλ was found to be impaired in these children, highlighting a crucial role for this signalling pathway in immunity against HSV-1 in the central nervous system[118–120] (FIG. 1). However, the *TLR3* and *UNC93B1* mutations exhibit incomplete clinical penetrance, perhaps reflecting the effects of additional host genetic factors, unknown environmental factors, or viral factors such as HSV-1 genetic variation[119]. Furthermore, only a small number of such patients have been identified and studied to date; therefore, the precise nature and range of pathogen susceptibilities resulting from these immunodeficiencies remain to be fully elucidated.

Conclusions and future directions

Recent studies have clearly implicated innate and acquired immune loci in human host defences against infectious disease. These approaches have confirmed a major role for HLA in the susceptibility to many, but not all, infectious diseases, and there is a need for further rigorous examination of the HLA region, which may also be relevant to non-infectious disease. This is likely to require very large sample sizes, specific sequence-based genotyping of HLA alleles, and novel analytical approaches (reviewed in REF. 121). The GWAS approach has also revealed many novel infectious disease susceptibility loci, although considerable further work is required in many cases to identify the causative alleles, their functional consequences and the biological mechanisms by which they influence disease pathogenesis.

Identification of novel 'selective' immunodeficiencies has also led to major insights into the pathways that are directly relevant for human host defences against specific pathogens. These studies have revealed that underlying gene defects cluster, unsurprisingly, on the same or related immune pathways (BOX 2; FIG. 1). Similar results have been reported in other diseases; for example, the clustering of rare mutations in neurosignalling and neurodevelopmental pathways in association with autism and schizophrenia[22]. 'Pathway-based' sequencing strategies may therefore represent a potentially cost-effective method to uncover novel rare variants that are associated with disease[18,22]. Furthermore, the description of monogenic predisposition to a narrow range of pathogens in otherwise apparently healthy individuals has led to the intriguing hypothesis that multiple, diverse, individually rare mutations in key immune genes and pathways may substantially affect susceptibility to severe phenotypes of infectious disease in the general population[93]. Such a possibility has also been suggested, with supporting data, for rare variants in *TLR4* and susceptibility to meningococcal disease[122]. Indeed, the enormous selective pressure exerted by infectious disease suggests that most major susceptibility variants will be rare. The use of whole-exome

'Pathway-based' sequencing strategies
Joint analysis of variants within a group of related genes in the same biological pathway, as opposed to analysis of individual markers.

Whole-exome
The collection of protein-coding regions (exons) in the genome. It is widely presumed that most genetic variants that exert a large effect on disease susceptibility will be located in coding regions. Because these regions comprise only 1% of the human genome, whole-exome sequencing represents a potentially efficient strategy for the identification of rare, large-effect variants underlying disease.

Severe combined immunodeficiency
A range of primary immunodeficiencies due to severe defects in T cell (and often B cell) production and function and defective cytokine signalling that result in early-onset severe infections by a range of pathogens.

Chronic mucocutaneous candidiasis
Persistent or recurrent superficial infections of the skin, nails or mucous membranes by *Candida* species.

Hyper-IgE syndrome
A primary immunodeficiency characterized by eczema, greatly elevated serum levels of immunoglobulin E (IgE) and recurrent skin and pulmonary infections.

or whole-genome sequencing in individuals with extreme phenotypes of infectious disease is now a realistic possibility and has enormous potential to identify novel immunodeficiency states.

It is increasingly clear that both rare and common variants in the same molecular pathway may underlie related disease phenotypes; for example, rare mutations in complement components and common variation at the *CFH–CFHR3* locus affect susceptibility to meningococcal disease[85,87], and both rare and common TLR–NF-κB pathway variants influence IPD susceptibility[93,94,97–102,105–108]. A similar pattern is observed with broader-spectrum immunodeficiencies, such as for the JAK–STAT cytokine signalling pathway, in which mutations in pathway members underlie both monogenic PIDs and common infectious disease. JAK–STAT variants in monogenic PIDs include JAK3 deficiency in severe combined immunodeficiency[123,124]; different *STAT1* mutations in association with mycobacterial infection, severe viral disease or chronic mucocutaneous candidiasis[125–127]; *STAT3* mutations in hyper-IgE syndrome[128,129]; and tyrosine kinase 2 (TYK2) deficiency in 'variant' hyper-IgE syndrome[130]. An example of a JAK–STAT pathway variant in common infectious disease is a common polymorphism in cytokine inducible SH2-containing protein (CISH) in association with multiple infectious diseases[131]; CISH is a 'suppressor of cytokine signalling' (SOCS) protein that controls STAT5. The nature and extent of interactions between rare and common variants is currently unknown, but common variants may act as 'modifier' genes that affect the clinical penetrance of rarer variants, analogously to the genetic model described in monogenic conditions such as cystic fibrosis[132–134]. Application of next-generation sequencing technology has the potential to define the relative contributions of common and rare variants to disease susceptibility at both the individual and population levels. Such an approach will be reliant on large, well-phenotyped sample collections, and will need to address potential interactions with pathogen genomic variation and environmental factors, as well as the possible contributions of epistatic and epigenetic effects (BOX 3).

Despite great breakthroughs in human genomics, disappointingly few improvements in clinical care have resulted to date. Studies of host genetics have suggested potential translational approaches for infectious disease (BOX 4). Implementation is likely to be highly challenging and it might be argued that research into the acquired causes of infection may elucidate results that are more directly modifiable and hence may have a more direct clinical impact, particularly in low-income nations. In this context, it is essential that advances in the understanding of human infectious disease genetics are viewed not as a final research end point but as a foundation for innovative approaches to improve clinical care. The major challenge for today's researchers is to justify the investment in human genomics by translating insights from genetic susceptibility to infection into improved preventative strategies and novel therapies for infectious disease.

1. Kaufman, S. H. E. Robert Koch, the Nobel prize, and the ongoing threat of tuberculosis. *N. Engl. J. Med.* **353**, 2423–2426 (2005).
2. Comstock, G. W. Tuberculosis in twins: a re-analysis of the Prophit survey. *Am. Rev. Respir. Dis.* **117**, 621–624 (1978).
3. Misch, E., Berrington, W. R., Vary, J. C. & Hawn, T. R. Leprosy and the human genome. *Microbiol. Mol. Biol. Rev.* **74**, 589–620 (2010).
4. Herndon, C. N. & Jennings, R. G. A twin family study on susceptibility to poliomyelitis. *Am. J. Hum. Genet.* **3**, 17–46 (1951).
5. Lin, T. M. *et al.* Hepatitis B virus markers in Chinese twins. *Anticancer Res.* **9**, 737–741 (1989).
6. van der Eijk, E. A., van de Vosse, E., Vandenbroucke, J. P. & van Dissel, J. T. Heredity versus environment in tuberculosis in twins. *Am. J. Respir. Crit. Care Med.* **176**, 1281–1288 (2007).
7. Sorensen, T. I. A., Nielsen, G. G., Andersen, P. K. & Teasdale, T. W. Genetic and environmental influences on premature death in adult adoptees. *New Engl. J. Med.* **318**, 727–732 (1988).
A landmark study of adoptees that suggests that a host genetic component contributes significantly to the risk of death from infectious disease.
8. Gingles, N. A. *et al.* Role of genetic resistance in invasive pneumococcal infection: identification and study of susceptibility and resistance in inbred mouse strains. *Infect. Immun.* **69**, 426–434 (2001).
9. Cooke, G. C. & Hill, A. V. S. Genetics of susceptibility to human infectious disease. *Nature Rev. Genet.* **2**, 967–977 (2001).
10. Hill, A. V. Aspects of genetic susceptibility to human infectious diseases. *Annu. Rev. Genet.* **40**, 469–486 (2006).
11. Burgner, D., Jamieson, S. E. & Blackwell, J. M. Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? *Lancet Infect. Dis.* **6**, 653–663 (2006).
12. Stein, C. M. Genetic epidemiology of tuberculosis susceptibility: impact of study design. *PLoS Pathog.* **7**, e1001189 (2011).
13. An, P. & Winkler, C. A. Host genes associated with HIV/AIDS: advances in gene discovery. *Trends Genet.* **26**, 119–131 (2010).
14. Brouwer, M. C., Read, R. C. & van de Beek, D. Host genetics and outcome in meningococcal disease: a systematic review and meta-analysis. *Lancet Infect. Dis.* **10**, 262–274 (2010).
15. Brouwer, M. C. *et al.* Host genetic susceptibility to pneumococcal and meningococcal disease: a systematic review and meta-analysis. *Lancet Infect. Dis.* **9**, 31–44 (2009).
16. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
17. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
18. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Rev. Genet.* **9**, 356–369 (2008).
19. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
20. Manolio, T. Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* **363**, 166–176 (2010).
21. Gloyn, A. *et al.* Variation across the allele frequency spectrum. *Nature Genet.* **42**, 648–650 (2010).
22. McClellan, J. & King, M.-C. Genetic heterogeneity in human disease. *Cell* **141**, 210–217 (2010).
23. Cirulli, E. & Goldstein, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Rev. Genet.* **11**, 415–425 (2010).
24. Kraft, P. & Hunter, D. J. Genetic risk prediction – are we there yet? *N. Engl. J. Med.* **360**, 1701–1703 (2009).
25. Fellay, J., Shianna, K. V., Telenti, A. & Goldstein, D. B. Host genetics and HIV-1: the final phase? *PLoS Pathog.* **6**, 10 (2010).
26. Dean, M. *et al.* Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CKR5* structural gene. *Science* **273**, 1856–1862 (1996).
27. Martin, M. P. *et al.* Genetic acceleration of AIDS progression by a promoter variant of *CCR5*. *Science* **282**, 1907–1911 (1998).
28. Smith, M. W. *et al.* Contrasting genetic influence of *CCR2* and *CCR5* variants on HIV-1 infection and disease progression. *Science* **277**, 959–965 (1997).
29. Huang, Y. *et al.* The role of a mutant *CCR5* allele in HIV-1 transmission and disease progression. *Nature Med.* **2**, 1240–1243 (1996).
30. Hendel, H. *et al.* New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *J. Immunol.* **162**, 6942–6946 (1999).
31. Flores-Villanueva, P. O. *et al.* Associations of MHC ancestral haplotypes with resistance/susceptibility to AIDS disease development. *J. Immunol.* **170**, 1925–1929 (2003).
32. Carrington, M., Martin, M. P. & van Bergen, J. KIR-HLA intercourse in HIV disease. *Trends Microbiol.* **16**, 620–627 (2008).
33. Fellay, J. *et al.* Common genetic variation and the control of HIV-1 in humans. *PLoS Genet.* **5**, e1000791 (2009).
34. Fellay, J. *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science* **317**, 944–947 (2007).
The first GWAS of an infectious disease.
35. The International HIV Controllers Study. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* **330**, 1551–1557 (2010).
A large GWAS of HIV-1 controllers and progressors that extends our understanding of previous HLA associations with HIV-1 control and identifies the specific amino acid positions underlying these signals.
36. Petrovski, S. *et al.* Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. *AIDS* **25**, 513–518 (2011).
37. Joubert, B. R. *et al.* A whole genome association study of mother-to-child transmission of HIV in Malawi. *Genome Med.* **2**, 17 (2010).
38. Pelak, K. *et al.* Host determinants of HIV-1 control in African Americans. *J. Infect. Dis.* **201**, 1141–1149 (2010).

39. Dalmasso, C. *et al.* Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS genome wide association 01 study. *PLoS ONE* **3**, e3907 (2008).
40. Herbeck, J. T. *et al.* Multistage genomewide association study identifies a locus at 1q41 associated with rate of HIV-1 disease progression to clinical AIDS. *J. Infect. Dis.* **210**, 618–626 (2010).
41. Troyer, J. L. *et al.* Genome-wide association study implicates PARD3B-based AIDS restriction. *J. Infect. Dis.* **203**, 1491–1502 (2011).
42. Limou, S. *et al.* Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS genomewide association study 02). *J. Infect. Dis.* **199**, 419–426 (2009).
43. Limou, S. *et al.* Multiple-cohort genetic association study reveals CXCR6 as a new chemokine receptor involved in long-term nonprogression to AIDS. *J. Infect. Dis.* **202**, 908–915 (2010).
44. Le Clerc, S. *et al.* Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS genomewide association study 03). *J. Infect. Dis.* **200**, 1194–1201 (2009).
45. de Roda Husman, A. M. *et al.* Association between *CCR5* genotype and the clinical course of HIV-1 infection. *Ann. Intern. Med.* **127**, 882–890 (1997).
46. Liu, R. *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367–377 (1996).
47. Samson, M. *et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the *CCR-5* chemokine receptor gene. *Nature* **382**, 722–725 (1996).
48. Salkowitz, J. R. *et al. CCR5* promoter polymorphism determines macrophage CCR5 density and magnitude of HIV-1 propagation *in vitro*. *Clin. Immunol.* **108**, 234–240 (2003).
49. Rizzardi, G. P. *et al. CCR2* polymorphism and HIV disease. Swiss HIV Cohort. *Nature Med.* **4**, 252–253 (1998).
50. Kosmrlj, A. *et al.* Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature* **465**, 350–354 (2010).
51. Thomas, R. *et al.* HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nature Genet.* **41**, 1290–1294 (2009).
52. Ge, D. *et al.* Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).
53. Rauch, A. *et al.* Genetic variation in *IL28B* is associated with chronic hepatitis C and treatment failure: a genome-wide association study. *Gastroenterology* **138**, 1338–1345 (2010).
54. Suppiah, V. *et al. IL28B* is associated with response to chronic hepatitis C interferon-α and ribavirin therapy. *Nature Genet.* **41**, 1100–1104 (2009).
55. Tanaka, Y. *et al.* Genome-wide association of *IL28B* with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C. *Nature Genet.* **41**, 1105–1109 (2009).
56. Ochi, H. *et al. IL-28B* predicts response to chronic hepatitis C therapy – fine-mapping and replication study in Asian populations. *J. Gen. Virol.* **92**, 1071–1081 (2011).
References 52–56 report strong genomic associations with response to IFNα treatment of hepatitis C. Together, these papers provide a rare example of a genetic association that may rapidly translate into clinical benefit.
57. Thomas, D. L. *et al.* Genetic variation in *IL28B* and spontaneous clearance of hepatitis C virus. *Nature* **461**, 798–801 (2009).
58. O'Brien, T. R. Interferon-alfa, interferon-λ and hepatitis C. *Nature Genet.* **41**, 1048–1050 (2009).
59. Marcello, T. *et al.* Interferons α and λ inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics. *Gastroenterology* **131**, 1887–1898 (2006).
60. Kamatani, Y. *et al.* A genome-wide association study identifies variants in the *HLA-DP* locus associated with chronic hepatitis B in Asians. *Nature Genet.* **41**, 591–595 (2009).
61. An, P. *et al.* A common HLA-DPA1 variant is a major determinant of hepatitis B virus clearance in Han Chinese. *J. Infect. Dis.* **203**, 943–947 (2011).
62. Guo, X. *et al.* Strong influence of human leukocyte antigen (HLA)-DP gene variants on development of persistent chronic hepatitis B virus carriers in the Han Chinese population. *Hepatology* **53**, 422–428 (2011).
63. Wang, L. *et al.* Evaluation of genetic susceptibility loci for chronic hepatitis B in Chinese: two independent case-control studies. *PLoS ONE* **6**, e17608 (2011).
64. O'Brien, T. R. *et al.* Risk alleles for chronic hepatitis B are associated with decreased mRNA expression of *HLA-DPA1* and *HLA-DPB1* in normal human liver. *Genes Immun.* **12**, 428–433 (2011).
65. Khor, CC. *et al.* Genome-wide association study identifies susceptibility loci for dengue shock syndrome at *MICB* and *PLCE1*. *Nature Genet.* **43**, 1139–1141 (2011).
The first GWAS for susceptibility to dengue. It identifies robust, novel associations with variants in *MICB* and *PLCE1*.
66. Champsaur, M. & Lanier, L. L. Effect of NKG2D ligand expression on host immune responses. *Immunol. Rev.* **235**, 267–285 (2010).
67. Hoang, L. T. *et al.* The early whole-blood transcriptional signature of dengue virus and features associated with progression to dengue shock syndrome in Vietnamese children and young adults. *J. Virol.* **84**, 12982–12994 (2010).
68. Hinkes, B. *et al.* Positional cloning uncovers mutations in *PLCE1* responsible for a nephrotic syndrome variant that may be reversible. *Nature Genet.* **38**, 1397–1405 (2006).
69. Malaria Genomic Epidemiology Network. A global network for investigating the genomic epidemiology of malaria. *Nature* **456**, 732–737 (2008).
70. Jallow, M. *et al.* Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nature Genet.* **41**, 657–665 (2009).
This study of severe malaria in The Gambia provides valuable insights into the challenges involved in performing effective GWASs in African populations.
71. Allison, A. C. Protection afforded by sickle-cell trait against subtertain malarial infection. *BMJ* **1**, 290–294 (1954).
72. Thye, T. *et al.* Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nature Genet.* **42**, 739–741 (2010).
73. Mira, M. T. *et al.* Susceptibility to leprosy is associated with *PARK2* and *PACRG*. *Nature* **427**, 636–640 (2004).
74. Alcaïs, A. *et al.* Stepwise replication identifies a low-producing lymphotoxin-α allele as a major risk factor for early-onset leprosy. *Nature Genet.* **39**, 517–522 (2007).
75. Geluk, A. & Ottenhoff, T. H. HLA and leprosy in the pre and postgenomic eras. *Hum. Immunol.* **67**, 439–445 (2006).
76. Zhang, F. R. *et al.* Genomewide association study of leprosy. *N. Engl. J. Med.* **361**, 2609–2618 (2009).
The first GWAS of leprosy susceptibility. It identifies multiple strong associations, many of which cluster in innate immune pathways and overlap with susceptibility loci for Crohn's disease.
77. Schurr, E. & Gros, P. A common genetic fingerprint in leprosy and Crohn's disease? *N. Engl. J. Med.* **361**, 2666–2668 (2009).
78. Wong, S. H. *et al.* Leprosy and the adaptation of human Toll-like receptor 1. *PLoS Pathog.* **6**, e1000979 (2010).
79. Wong, S. H., Hill, A. V. & Vannberg, F. O. Genomewide association study of leprosy. *N. Engl. J. Med.* **362**, 1446–1447 (2010).
80. Berrington, W. R. *et al.* Common polymorphisms in the *NOD2* gene region are associated with leprosy and its reactive states. *J. Infect. Dis.* **201**, 1422–1435 (2010).
81. Kanneganti, T.-D., Lamkanfi, M. & Nunez, G. Intracellular NOD-like receptors in host defense and disease. *Immunity* **27**, 549–559 (2007).
82. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature Genet.* **42**, 1118–1125 (2010).
83. Marks, D. J. B., Rahman, F. Z., Sewell, G. W. & Segal, A. W. Crohn's disease: an immune deficiency state. *Clin. Rev. Allergy Immunol.* **38**, 20–31 (2010).
84. Haralambous, E. *et al.* Factor H, a regulator of complement activity, is a major determinant of meningococcal disease susceptibility in UK Caucasian patients. *Scand. J. Infect. Dis.* **38**, 764–771 (2006).
85. Davila, S. *et al.* Genome-wide association study identifies variants in the *CFH* region associated with host susceptibility to meningococcal disease. *Nature Genet.* **42**, 772–776 (2010).
86. Hughes, A. E. *et al.* A common *CFH* haplotype, with deletion of *CFHR1* and *CFHR3*, is associated with lower risk of age-related macular degeneration. *Nature Genet.* **38**, 1173–1177 (2006).
87. Degn, S. E., Jensenius, J. C. & Thiel, S. Disease-causing mutations in genes of the complement system. *Am. J. Hum. Genet.* **88**, 689–705 (2011).
88. Zeidler, M., Stewart, G., Cousens, S. N., Estibeiro, K. & Will, R. G. Codon 129 genotype and new variant CJD. *Lancet* **350**, 668 (1997).
89. Mead, S. *et al.* Balancing selection at the prion protein gene consistent with prehistoric kurulike epidemics. *Science* **300**, 640–643 (2003).
90. Gambetti, P., Kong, Q., Zou, W., Parchi, P. & Chen, S. G. Sporadic and familial CJD: classification and characterisation. *Br. Med. Bull.* **66**, 213–239 (2003).
91. Mead, S. *et al.* Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study. *Lancet Neurol.* **8**, 57–66 (2009).
92. Kaski, D. *et al.* Variant CJD in an individual heterozygous for *PRNP* codon 129. *Lancet* **374**, 2128 (2009).
93. Casanova, J. L. & Abel, L. Primary immunodeficiencies: a field in its infancy. *Science* **317**, 617–619 (2007).
This review introduces a paradigm shift in primary immunodeficiency by describing the concept of 'selective' immunodeficiency.
94. Picard, C. *et al.* Clinical features and outcome of patients with IRAK-4 and MyD88 deficiency. *Medicine* **89**, 403–425 (2010).
95. van de Vosse, E., van Dissel, J. T. & Ottenhoff, T. H. Genetic deficiencies of innate immune signalling in human infectious disease. *Lancet Infect. Dis.* **9**, 688–698 (2009).
96. Kawai, T. & Akira, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nature Immunol.* **11**, 373–384 (2010).
97. von Bernuth, H. *et al.* Pyogenic bacterial infections in humans with MyD88 deficiency. *Science* **321**, 691–696 (2008).
98. Zonana, J. *et al.* A novel X-linked disorder of immune deficiency and hypohidrotic ectodermal dysplasia is allelic to incontinentia pigmenti and due to mutations in IKK-γ (NEMO). *Am. J. Hum. Genet.* **67**, 1555–1562 (2000).
99. Doffinger, R. *et al.* X-linked anhidrotic ectodermal dysplasia with immunodeficiency is caused by impaired NF-κB signaling. *Nature Genet.* **27**, 277–285 (2001).
100. Picard, C. *et al.* Pyogenic bacterial infections in humans with IRAK-4 deficiency. *Science* **299**, 2076–2079 (2003).
This study, together with reference 97, provides clear evidence that single-gene mutations disrupting TLR and IL-1 receptor signalling may result in a narrow spectrum immunodeficiency in otherwise healthy children.
101. Courtois, G. *et al.* A hypermorphic IκBα mutation is associated with autosomal dominant anhidrotic ectodermal dysplasia and T cell immunodeficiency. *J. Clin. Invest.* **112**, 1108–1115 (2003).
102. Janssen, R. *et al.* The same IκBα mutation in two related individuals leads to completely different clinical syndromes. *J. Exp. Med.* **200**, 559–568 (2004).
103. Medvedev, A. E. *et al.* Distinct mutations in *IRAK-4* confer hyporesponsiveness to lipopolysaccharide and interleukin-1 in a patient with recurrent bacterial infections. *J. Exp. Med.* **198**, 521–531 (2003).
104. Misch, E. A. & Hawn, T. R. Toll-like receptor polymorphisms and susceptibility to human disease. *Clin. Sci.* **114**, 347–360 (2008).
105. Chapman, S. J. *et al.* IκB genetic polymorphisms and invasive pneumococcal disease. *Am. J. Respir. Crit. Care Med.* **176**, 181–187 (2007).
106. Chapman, S. J. *et al. NFKBIZ* polymorphisms and susceptibility to pneumococcal disease in European and African populations. *Genes Immun.* **11**, 319–325 (2010).
107. Chapman, S. J. *et al.* Common *NFKBIL2* polymorphisms and susceptibility to pneumococcal disease: a genetic association study. *Crit. Care* **14**, R227 (2010).
108. Khor, C. C. *et al.* A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nature Genet.* **39**, 523–528 (2007).
109. Kumpf, O. *et al.* Influence of genetic variations in TLR4 and TIRAP/Mal on the course of sepsis and pneumonia and cytokine release: an observational study in three cohorts. *Crit. Care* **14**, R103 (2010).
110. Hamann, L. *et al.* Low frequency of the TIRAP S180L polymorphism in Africa, and its potential role in malaria, sepsis, and leprosy. *BMC Med. Genet.* **10**, 65 (2009).
111. Nejentsev, S. *et al.* Analysis of association of the TIRAP (MAL) S180L variant and tuberculosis in three populations. *Nature Genet.* **40**, 261–262; author reply 262–263 (2008).

112. Hawn, T. R. *et al.* A polymorphism in Toll-interleukin 1 receptor domain containing adaptor protein is associated with susceptibility to meningeal tuberculosis. *J. Infect. Dis.* **194**, 1127–1134 (2006).
113. Castiblanco, J. *et al.* TIRAP (MAL) S180L polymorphism is a common protective factor against developing tuberculosis and systemic lupus erythematosus. *Infect. Genet. Evol.* **8**, 541–544 (2008).
114. Dissanayeke, S. B. *et al.* Polymorphic variation in TIRAP is not associated with susceptibility to childhood TB but may determine susceptibility to TBM in some ethnic groups. *PLoS ONE* **4**, e6698 (2009).
115. Selvaraj, P. *et al.* Toll-like receptor and *TIRAP* gene polymorphisms in pulmonary tuberculosis patients of South India. *Tuberculosis* **90**, 306–310 (2010).
116. Ladhani, S. N. *et al.* Association between single-nucleotide polymorphisms in *Mal/TIRAP* and interleukin-10 genes and susceptibility to invasive haemophilus influenzae serotype b infection in immunized children. *Clin. Infect. Dis.* **51**, 761–767 (2010).
117. Ferwerda, B. *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* **106**, 10272–10277 (2009).
118. Casrouge, A. *et al.* Herpes simplex virus encephalitis in human UNC-93B deficiency. *Science* **314**, 308–312 (2006).
119. Zhang, S. Y. *et al.* TLR3 deficiency in patients with herpes simplex encephalitis. *Science* **317**, 1522–1527 (2007).
120. Perez de Diego, R. *et al.* Human TRAF3 adaptor molecule deficiency leads to impaired Toll-like receptor 3 response and susceptibility to herpes simplex encephalitis. *Immunity* **33**, 400–411 (2010).
References 118–120 provide the first demonstration that selective primary immunodeficiency underlies susceptibility to HSE.
121. Blackwell, J. M., Jamieson, S. E. & Burgner, D. HLA and infectious diseases. *Clin. Microbiol. Rev.* **22**, 370–385 (2009).
122. Smirnova, I. *et al.* Assay of locus-specific genetic load implicates rare Toll-like receptor 4 mutations in meningococcal susceptibility. *Proc. Natl Acad. Sci. USA* **100**, 6075–6080 (2003).
123. Macchi, P. *et al.* Mutations of Jak-3 gene in patients with autosomal severe combined immune deficiency (SCID). *Nature* **377**, 65–68 (1995).
124. Russell, S. M. *et al.* Mutation of Jak3 in a patient with SCID: essential role of Jak3 in lymphoid development. *Science* **270**, 797–800 (1995).
125. Dupuis, S. *et al.* Impaired response to interferon-α/β and lethal viral disease in human STAT1 deficiency. *Nature Genet.* **33**, 388–391 (2003).
126. Liu, L. *et al.* Gain-of-function human STAT1 mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis. *J. Exp. Med.* **208**, 1635–1648 (2011).
127. van de Veerdonk, F. L. *et al.* STAT1 mutations in autosomal dominant chronic mucocutaneous candidiasis. *N. Engl. J. Med.* **365**, 54–61 (2011).
128. Holland, S. M. *et al. STAT3* mutations in the hyper-IgE syndrome. *N. Engl. J. Med.* **357**, 1608–1619 (2007).
129. Minegishi, Y. *et al.* Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome. *Nature* **448**, 1058–1062 (2007).
130. Minegishi, Y. *et al.* Human tyrosine kinase 2 deficiency reveals its requisite roles in multiple cytokine signals involved in innate and acquired immunity. *Immunity* **25**, 745–755 (2006).
131. Khor, C. C. *et al. CISH* and susceptibility to infectious diseases. *N. Engl. J. Med.* **362**, 2092–2101 (2010).
132. Gu, Y. *et al.* Identification of *IFRD1* as a modifier gene for cystic fibrosis lung disease. *Nature* **458**, 1039–1042 (2009).
133. Alcaïs, A., Abel, L. & Casanova, J. L. Human genetics of infectious diseases: between proof of principle and paradigm. *J. Clin. Invest.* **119**, 2506–2514 (2009).
134. Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. Human genetic variation and its contribution to complex traits. *Nature Rev. Genet.* **10**, 241–251 (2009).
135. Teo, Y. *et al.* Methodological challenges of genome-wide association analysis in Africa. *Nature Rev. Genet.* **11**, 149–160 (2010).
136. Rosenberg, N. *et al.* Genome-wide association studies in diverse populations. *Nature Rev. Genet.* **11**, 356–366 (2010).
137. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
138. Zhang, S.-Y. *et al.* Inborn errors of interferon (IFN)-mediated immunity in humans: insights into the respective roles of IFN-α/β, IFN-γ, and IFN-λ in host defense. *Immunol. Rev.* **226**, 29–40 (2008).
139. Maródi, L. & Notarangelo, L. D. Immunological and genetic bases of new primary immunodeficiencies. *Nature Rev. Immunol.* **7**, 851–861 (2007).
140. Cooke, G. S. *et al.* Polymorphism within the interferon-γ/receptor complex is associated with pulmonary tuberculosis. *Am. J. Respir. Crit. Care Med.* **174**, 339–343 (2006).
141. Vannberg, F. O., Chapman, S. J. & Hill, A. V. Human genetic susceptibility to intracellular pathogens. *Immunol. Rev.* **240**, 105–116 (2011).
142. Pacheco, A. G., Cardoso, C. C. & Moraes, M. O. IFNG +874T/A, IL10 -1082G/A and TNF -308G/A polymorphisms in association with tuberculosis susceptibility: a meta-analysis study. *Hum. Genet.* **123**, 477–484 (2008).
143. Awomoyi, A. A. *et al.* No association between interferon-γ receptor-1 gene polymorphism and pulmonary tuberculosis in a Gambian population sample. *Thorax* **59**, 291–294 (2004).
144. Bustamante, J. *et al.* Germline *CYBB* mutations that selectively affect macrophages in kindreds with X-linked predisposition to tuberculous mycobacterial disease. *Nature Immunol.* **12**, 213–221 (2011).
145. Hambleton, S. *et al. IRF8* mutations and human dendritic-cell immunodeficiency. *N. Engl. J. Med.* **365**, 127–138 (2011).
146. Le Pendu, J., Ruvoën-Clouet, N., Kindberg, E. & Svensson, L. Mendelian resistance to human norovirus infections. *Semin. Immunol.* **18**, 375–386 (2006).
147. Aspholm-Hurtig, M. *et al.* Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin. *Science* **305**, 519–522 (2004).
148. Lindesmith, L. C. *et al.* Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med.* **5**, e31 (2008).
149. Gagneux, S. *et al.* Variable host–pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 2869–2873 (2006).
150. Kumar, D. *et al.* Genome-wide analysis of the host intracellular network that regulates survival of *Mycobacterium tuberculosis*. *Cell* **140**, 731–743 (2010).
151. Di Pietrantonio, T., Correa, J. A., Orlova, M., Behr, M. A. & Schurr, E. Joint effects of host genetic background and mycobacterial pathogen on susceptibility to infection. *Infect. Immun.* **79**, 2372–2378 (2011).
152. Caws, M. *et al.* The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog.* **4**, e1000034 (2008).
153. Monot, M. *et al.* On the origin of leprosy. *Science* **308**, 1040–1042 (2005).
154. Schurr, E. Is susceptibility to tuberculosis acquired or inherited? *J. Intern. Med.* **261**, 106–111 (2007).
155. Lazzaro, B. P., Sceurman, B. K. & Clark, A. G. Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. *Science* **303**, 1873–1876 (2004).
156. McDermid, J. M. *et al.* Mortality in HIV infection is independently predicted by host iron status and SLC11A1 and HP genotypes, with new evidence of a gene-nutrient interaction. *Am. J. Clin. Nutr.* **90**, 225–233 (2009).
157. Mallal, S. *et al.* Association between presence of *HLA-B*5701*, *HLA-DR7*, and *HLA-DQ3* and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* **359**, 727–732 (2002).
158. Martin, A. M. *et al.* Predisposition to abacavir hypersensitivity conferred by *HLA-B*5701* and a haplotypic *Hsp70-Hom* variant. *Proc. Natl Acad. Sci. USA* **101**, 4180–4185 (2004).
159. Young, B. *et al.* First large, multicenter, open-label study utilizing *HLA-B*5701* screening for abacavir hypersensitivity in North America. *AIDS* **22**, 1673–1675 (2008).
160. Colombo, S. *et al.*, The *HCP5* single-nucleotide polymorphism: a simple screening tool for prediction of hypersensitivity reaction to abacavir. *J. Infect. Dis.* **198**, 864–867 (2008).
161. Hetherington, S. *et al.* Genetic variations in *HLA-B* region and hypersensitivity reactions to abacavir. *Lancet* **359**, 1121–1122 (2002).
162. Clark, P. J., Thompson, A. L. & McHutchison, J. G. *IL28B* genomic-based treatment paradigms for patients with chronic hepatitis C infection: the future of personalized HCV therapies. *Am. J. Gastroenterol.* **106**, 38–45 (2011).
163. Fellay, J. *et al. ITPA* gene variants protect against anaemia in patients treated for chronic hepatitis C. *Nature* **464**, 405–408 (2010).
164. Ochi, H. *et al.* Inosine triphosphate pyrophophatase polymorphism affects ribavirin-induced anaemia and outcome of therapy – a genome-wide study of Japanese hepatitis C virus patients. *Gastroenterology* **139**, 1190–1197 (2010).
165. Tanaka, Y. *et al.* Genome-wide association study identified ITPA/DDRGK1 variants reflecting thrombocytopenia in pegylated interferon and ribavirin therapy for chronic hepatitis C. *Hum. Mol. Genet.* **20**, 3507–3516 (2011).
166. Hitotsumatsu, O. *et al.* The ubiquitin-editing enzyme A20 restricts nucleotide-binding oligomerization domain containing 2-triggered signals. *Immunity* **28**, 381–390 (2008).
167. Sprong, T. *et al.* Deficient alternative complement pathway activation due to factor D deficiency by 2 novel mutations in the complement factor D gene in a family with meningococcal infections. *Blood* **107**, 4865–4870 (2006).
168. Fijen, C. A., Kuijper, E. J., te Bulte, M. T., Daha, M. R. & Dankert, J. Assessment of complement deficiency in patients with meningococcal disease in The Netherlands. *Clin. Infect. Dis.* **28**, 98–105 (1999).
169. Skattum, L., van Deuren, M., van der Poll, T. & Truedsson, L. Complement deficiency states and associated infections. *Mol. Immunol.* **48**, 1643–1655 (2011).
170. Jonsson, G. *et al.* Hereditary C2 deficiency in Sweden: frequent occurrence of invasive infection, atherosclerosis, and rheumatic disease. *Medicine* **84**, 23–34 (2005).
171. Byun, M. *et al.* Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *J. Exp. Med.* **207**, 2307–2312 (2010).
172. Ramoz, N. *et al.* Mutations in two adjacent novel genes are associated with epidermodysplasia verruciformis. *Nature Genet.* **32**, 579–581 (2002).
173. Lindesmith, L. *et al.* Human susceptibility and resistance to Norwalk virus infection. *Nature Med.* **9**, 548–553 (2003).
174. Thorven, M. *et al.* A homozygous nonsense mutation (428G-->A) in the human secretor (*FUT2*) gene provides resistance to symptomatic norovirus (GGII) infections. *J. Virol.* **79**, 15351–15355 (2005).
175. Miller, L. H., Mason, S. J., Clyde, D. F. & McGinniss, M. H. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *N. Engl. J. Med.* **295**, 302–304 (1976).
176. Glocker, E. O. *et al.* A homozygous *CARD9* mutation in a family with susceptibility to fungal infections. *N. Engl. J. Med.* **361**, 1727–1735 (2009).
177. Puel, A. *et al.* Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. *Science* **332**, 65–68 (2011).

Acknowledgements
Research in the authors' laboratory is funded by the Wellcome Trust, the UK Medical Research Council and the UK National Institute for Health Research (NIHR) Oxford Biomedical Research Centre.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
MalariaGEN: http://www.malariagen.net
1000 Genomes Project: http://www.1000genomes.org
***Nature Reviews Genetics* series on Genome-wide association studies:** http://www.nature.com/nrg/series/gwas/index.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Autobiographies

Stephen J. Chapman

Stephen J. Chapman is a Senior Lecturer and Consultant Physician in respiratory medicine at the Oxford Biomedical Research Centre and Oxford Radcliffe Hospitals National Health Service (NHS) Trust, UK. His major research interest is the study of human genetic susceptibility to severe bacterial respiratory disease, in particular pneumococcal disease and thoracic empyema. He is a practicing clinician with specialist interests in the management of acute respiratory infection and chronic suppurative lung disease (adult cystic fibrosis and bronchiectasis).

Adrian V. S. Hill

Adrian V. S. Hill is Professor of Human Genetics at the University of Oxford, UK, and has had a long-standing interest in the field of the human genetics of infectious diseases, particularly in major bacterial diseases of developing countries. He also directs the Jenner Institute in Oxford, where his research group works on the development of new vaccination strategies against intracellular pathogens, especially to prevent malaria.

Online summary

- Common and rare human genetic variation influences susceptibility to infectious disease.
- Genome-wide association studies have identified various strong associations between genetic polymorphisms and susceptibility to common infectious disease phenotypes, such as HIV-1, hepatitis B and C viruses, dengue, malaria, tuberculosis, leprosy, meningococcal disease and prion disease.
- Studies have confirmed an important role for human leukocyte antigen (HLA) variation in susceptibility to many, but not all, common infectious diseases.
- Application of genome-wide approaches to African populations is challenging, which reflects in part the high levels of ethnic diversity in Africa and a lack of population-specific genotyping arrays.
- An increasing number of rare, single-gene defects have been described in association with susceptibility to a narrow range of different pathogens in otherwise healthy individuals.
- Examples of mutations associated with such 'selective' immunodeficiencies include: interleukin-12–interleukin-23–interferon-γ (IFNγ) pathway defects and susceptibility to mycobacterial disease; Toll-like receptor (TLR)–nuclear factor-κB pathway defects and invasive pneumococcal disease; and TLR3–IFNα–IFNβ pathway mutations and herpes simplex viral encephalitis.
- Both common and rare variants in the same molecular pathway may underlie related infectious disease phenotypes.
- The relative contributions of common and rare variants to infectious disease susceptibility at the individual and population level are currently unknown. The application of next-generation sequencing technology is likely to define the genetic architecture of infectious disease.

ToC Blurb

SERIES ON GENOME-WIDE ASSOCIATION STUDIES

000 **Human genetic susceptibility to infectious disease**

Stephen J. Chapman and Adrian V. S. Hill

The discovery of common and rare genetic variants underlying infectious disease susceptibility in humans highlights the involvement of shared host signalling pathways in diverse infectious diseases, and provides insights into the genetic control of immune signalling.

Subject categories

Disease genetics, infectious disease