Minireview

Genome-wide association studies are coming for human infectious diseases

Sonia Davila and Martin L Hibberd

Address: Genome Institute of Singapore, Infectious Diseases, 60 Biopolis Street, 138672 Singapore.

Correspondence: Martin L Hibberd. Email: hibberdml@gis.a-star.edu.sg

Published: 10 February 2009

Genome Medicine 2009, **1**:19 (doi:10.1186/gm19)

The electronic version of this article is the complete one and can be found online at http://genomemedicine.com/content/1/2/19

Abstract

A genetic contribution to infectious disease in human populations has long been suspected and is now supported by more than 50 years of epidemiological evidence showing, for example, infection rates to be much higher than disease rates. In successful family studies of high-penetrance effects, single gene mutations have been identified that reveal a molecular mechanism leading to increased risk of a specific infectious disease. However, in population-based studies, genetic variants conferring host susceptibility to various infectious diseases have been difficult to uncover. Although mutations such as that in the *CCR5* gene, which confers protection against HIV infection, have been reliably discovered, polymorphisms affecting larger proportions of a population have been hard to prove definitively. The recent arrival of the genome-wide association study format, currently being applied to Kawasaki disease, tuberculosis, malaria, HIV, dengue and others, gives us hope that these challenges can finally be met, with implications for population-based treatment and prognosis strategies.

Infectious agents continue to have a major influence on human evolution as a result of the widespread nature of these agents in the environment and the predominance of infectious disease in children, resulting in a clear impact on allele transmission depending on the outcome of this early encounter. For example, in areas with endemic dengue disease, such as Vietnam, seroprevalence studies show that infection has occurred in over 88% of children by the age of 15 and yet disease that results in hospitalization occurs in less than 0.1% [1]. Is this infectious disease distribution predominantly due to widespread general resistance over multiple genes or is susceptibility due to rare mutations in a few specific genes? This question on the nature of the genetic change, rare or common, has remained, despite numerous population studies of infectious disease.

Candidate gene studies

Although family-specific mutations have been shown to have indisputable consequences in infectious disease [2-5], their effects in the general population are thought to be small [6]. There are also some very well established population-based mutations with clear effects on disease, from sickle cell trait protection against malaria [7] to *CCR5*, a known receptor of HIV, defective alleles in which have been associated with resistance to HIV infection [8]. Indeed, rare mutations leading to increased susceptibility to infectious disease can be identified in the general population, when looked for carefully [9]. Interestingly, the evidence from this field of work is currently pointing towards the concept of 'one gene, one infection' [10]. The strong evidence that a mutation in a single gene frequently leads to increased susceptibility to

only one infection is also supported by animal studies, which have used forward genetics to show that, although many genes may be involved in recognizing a single pathogen, mutations in only one or a few lead to increased susceptibility to disease [11].

Despite the clear evidence identifying these so called 'major genes', it is unlikely that mutations in single genes are the only causes of genetic susceptibility in a population. So far, the work done to extend this approach to a population basis has used candidate gene studies and, in general, when they harmonize with the major genes findings, significant insights can be gained from them. For example, in meningococcal disease the 'major genes' lie in the complement pathway, with mutations in many of these genes (each restricted to just a few individuals) leading to greatly increased risk of disease [12]. Using one of these genes (that encoding mannose binding lectin) as a candidate gene in a population-based study has supported and extended these findings to the general population, revealing a polymorphism that can alter the initiation of this pathway, and thus the susceptibility to disease, in a large number of individuals [13]. However, candidate gene approaches are ultimately restricted to those few genes that can be strongly implicated by current knowledge, and there is little understanding of their overall contribution to disease in relation to other genes.

Moving towards genome-wide studies

Two of the first efforts to apply a genome-wide approach to multi-factorial infectious diseases were done in schistosomiasis and tuberculosis, carried out in Brazilian and African families, respectively [14,15]. The tools available at the time only allowed linkage studies using a few hundred genetic markers spread across the whole genome. Nowadays, data generated after the sequencing of the human genome and the HapMap project have identified millions of single nucleotide polymorphisms facilitating the implementation of genome-wide association studies (GWASs). The GWAS approach, although new, is already well described and has significant methodological improvements compared with previous approaches; GWAS analysis has already enabled the successful identification and replication of novel genes for several complex disorders such as Crohn's disease, rheumatoid arthritis, type 1 and type 2 diabetes, macular degeneration and prostate cancer [16-18].

The first GWAS in infectious disease was done in 2007, when Fellay *et al.* [19] determined genetic components influencing the viral load of HIV-positive patients during the asymptomatic phase of the disease. However, this study did not investigate host genetic susceptibility. The first GWAS of infectious disease susceptibility was carried out in a group of patients affected with Kawasaki disease [20], a self-limiting acute vasculitis mostly affecting children below 5 years of age [21]. This study started by genotyping a small group of European origin patients and controls with an Affymetrix 250K Nsp chip; a follow-up group comprising affected children and parents was then used to confirm the initial findings; and finally a third stage, including fine-mapping of eight associated genes, further validated the genetic associations. Interestingly, five of the eight fine-mapped genes formed a connected network that also showed significant differences in transcript levels between acute and convalescent stages of the disease, suggesting that multiple genes in a pathway may be involved in susceptibility to disease, something that may turn out to be important for other infectious diseases. However, in comparison with disorders such as inflammatory bowel disease or macular degeneration, for which the small number of genetic variants detected were shown to confer a highly significant disease risk to carriers [17,22], the study on Kawasaki disease [20] has not found such a profound effect, hinting that infectious disease may not yield easily to the GWAS approach. It is also interesting to see that the *HLA* locus, such a clear candidate region for infectious disease because its role in discriminating self from non-self was thought to have evolved to protect from infection, did not cause the dominant association seen in many autoimmune diseases [23,24], and this may also turn out to be true for many infectious disease studies.

Facing new challenges

There are several challenges for the application of GWAS to determine genetic variants that confer susceptibility to infectious diseases. First, the current design of commercially available single nucleotide polymorphism chips is skewed towards common variants, with at least 1% presence in the population, following the 'common variant, common disease' hypothesis [25,26]. However, if susceptibility to an infectious disease is caused by one or a combination of rare variants, as previously reported for other disorders such as autism [27], we would not be able to detect them using the current technology. In that scenario, sequencing of the candidate genes would be the most direct way to tackle the problem. Second, structural variants, such as copy number variants that were barely analyzed in the past, are currently being investigated in more detail and have been successfully linked to susceptibility to schizophrenia [28,29]. Recent development of new technologies as well as analysis tools will facilitate the study of copy number variants, which might potentially be involved in host susceptibility to infectious diseases. Third, a landmark paper by the Wellcome Trust Case Control Consortium [16] introduced the concept of using the same set of population controls, without detailed phenotype information for different diseases, in case-control studies. Although there is the possibility of misclassification, meaning that a proportion of controls might have the disease or might develop it in the future, this could be a very useful approach for the many infectious diseases for which the incidence rates in the general

population are low (less than 1%), such as meningococcal disease, severe dengue and others. It is worth noting, however, that for infectious disease the environmental agent (the pathogen) that triggers the disease is usually known and past exposure to the agent is often easily measured (using serology), criteria that are frequently not available in diseases such as cancer or autoimmunity. This means that it is possible in infectious disease to select controls with known (antibody detected) exposure, but without disease. Finally, another important aspect to take into account is the role of the pathogen's genetic variability and its interaction with host genetics. An interesting study by Caws *et al.* [30] has hinted at a relationship between host and pathogen genotypes in the development of tuberculosis.

When these challenges are successfully addressed, a GWAS will offer substantial insights into the role of common variation in an infectious disease and, with its comprehensive nature, will hopefully enable us to elucidate the involvement of multiple genes or previously unsuspected pathways in disease.

Clinical significance

Moving from genetic discovery to direct clinical relevance has not been as difficult for infectious disease as it has been for many other diseases, perhaps because there is a relatively detailed understanding of the immune system. For instance, patients with Mendelian diseases characterized by diminished production of interferon or interferon pathway activation could be treated with recombinant forms of interferon, helping them to mount an immune response against various pathogens. Although there is still no treatment using the *CCR5* gene, it has been a tempting target for therapies in the HIV field and may yet prove to be effective. In another example, a polymorphism associated with increased serum concentrations of complement Factor H was revealed to be a susceptibility factor for meningococcal disease, as Factor H bound to and protected the causative bacteria from complement attack [31]. This is a finding with implications for the vaccine community, which showed that antibodies that prevent this binding enable complement-mediated killing of the bacteria and, when incorporated into a vaccine, has been the first successful strategy to prevent group B meningococcal disease [32].

With the arrival of the GWAS approach to infectious disease, we can expect many more genetic loci to be revealed, molecular pathways to disease described and thus therapeutic targets identified, with the hope that these can be quickly translated into treatments.

In conclusion, the first GWASs of infectious diseases have been published [19,20], and more will come in the near future, with important implications for our understanding of these diseases. Although initial hints suggest that big hits will not lead the way, the unraveling of disease mechanisms through networks of genes should reveal novel ways of tackling infectious diseases in the clinic.

Abbreviations

GWAS, genome-wide association study.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

SD and MLH are funded by the Singapore, Agency for Science, Technology and Research (A*STAR).

References

1. Thai KT, Binh TQ, Giao PT, Phuong HL, Hung le Q, Van Nam N, Nga TT, Groen J, Nagelkerke N, de Vries PJ: **Seroprevalence of dengue antibodies, annual incidence and risk factors among children in southern Vietnam.** *Trop Med Int Health* 2005, **10:**379-386.
2. Altare F, Durandy A, Lammas D, Emile JF, Lamhamedi S, Le Deist F, Drysdale P, Jouanguy E, Doffinger R, Bernaudin F, Jeppsson O, Gollob JA, Meinl E, Segal AW, Fischer A, Kumararatne D, Casanova JL: **Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency.** *Science* 1998, **280:**1432-1435.
3. Altare F, Lammas D, Revy P, Jouanguy E, Doffinger R, Lamhamedi S, Drysdale P, Scheel-Toellner D, Girdlestone J, Darbyshire P, Wadhwa M, Dockrell H, Salmon M, Fischer A, Durandy A, Casanova JL, Kumararatne DS: **Inherited interleukin 12 deficiency in a child with bacille Calmette-Guerin and *Salmonella enteritidis* disseminated infection.** *J Clin Invest* 1998, **102:**2035-2040.
4. de Jong R, Altare F, Haagen IA, Elferink DG, Boer T, van Breda Vriesman PJ, Kabel PJ, Draaisma JM, van Dissel JT, Kroon FP, Casanova JL, Ottenhoff TH: **Severe mycobacterial and *Salmonella* infections in interleukin-12 receptor-deficient patients.** *Science* 1998, **280:**1435-1438.
5. Picard C, Puel A, Bonnet M, Ku CL, Bustamante J, Yang K, Soudais C, Dupuis S, Feinberg J, Fieschi C, Elbim C, Hitchcock R, Lammas D, Davies G, Al-Ghonaium A, Al-Rayes H, Al-Jumaah S, Al-Hajjar S, Al-Mohsen IZ, Frayha HH, Rucker R, Hawn TR, Aderem A, Tufenkeji H, Haraguchi S, Day NK, Good RA, Gougerot-Pocidalo MA, Ozinsky A, Casanova JL: **Pyogenic bacterial infections in humans with IRAK-4 deficiency.** *Science* 2003, **299:**2076-2079.
6. Hill AV: **Aspects of genetic susceptibility to human infectious diseases.** *Annu Rev Genet* 2006, **40:**469-486.
7. Allison AC: **Protection afforded by sickle-cell trait against subtertian malareal infection.** *Br Med J* 1954, **1:**290-294.
8. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, Saragosti S, Lapoumeroulie C, Cognaux J, Forceille C, Muyldermans G, Verhofstede C, Burtonboy G, Georges M, Imai T, Rana S, Yi Y, Smyth RJ, Collman RG, Doms RW, Vassart G, Parmentier M: **Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene.** *Nature* 1996, **382:**722-725.
9. Smirnova I, Mann N, Dols A, Derkx HH, Hibberd ML, Levin M, Beutler B: **Assay of locus-specific genetic load implicates rare Toll-like receptor 4 mutations in meningococcal susceptibility.** *Proc Natl Acad Sci USA* 2003, **100:**6075-6080.
10. Casanova JL, Abel L: **Human genetics of infectious diseases: a unified theory.** *EMBO J* 2007, **26:**915-922.
11. Beutler B, Eidenschenk C, Crozat K, Imler JL, Takeuchi O, Hoffmann JA, Akira S: **Genetic analysis of resistance to viral infection.** *Nat Rev Immunol* 2007, **7:**753-766.
12. Schneider MC, Exley RM, Ram S, Sim RB, Tang CM: **Interactions between *Neisseria meningitidis* and the complement system.** *Trends Microbiol* 2007, **15:**233-240.

13. Hibberd ML, Sumiya M, Summerfield JA, Booy R, Levin M: **Association of variants of the gene for mannose-binding lectin with susceptibility to meningococcal disease. Meningococcal Research Group.** *Lancet* 1999, **353:**1049-1053.
14. Marquet S, Abel L, Hillaire D, Dessein H, Kalil J, Feingold J, Weissenbach J, Dessein AJ: **Genetic localization of a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31-q33.** *Nat Genet* 1996, **14:**181-184.
15. Bellamy R, Beyers N, McAdam KP, Ruwende C, Gie R, Samaai P, Bester D, Meyer M, Corrah T, Collin M, Camidge DR, Wilkinson D, Hoal-Van Helden E, Whittle HC, Amos W, van Helden P, Hill AV: **Genetic susceptibility to tuberculosis in Africans: a genome-wide scan.** *Proc Natl Acad Sci USA* 2000, **97:**8005-8009.
16. Welcome Trust Case Control Consortium: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447:**661-678.
17. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J: **Complement factor H polymorphism in age-related macular degeneration.** *Science* 2005, **308:**385-389.
18. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, Greenway SC, Stram DO, Le Marchand L, Kolonel LN, Frasco M, Wong D, Pooler LC, Ardlie K, Oakley-Girvan I, Whittemore AS, Cooney KA, John EM, Ingles SA, Altshuler D, Henderson BE, Reich D: **Multiple regions within 8q24 independently affect risk for prostate cancer.** *Nat Genet* 2007, **39:**638-644.
19. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL, McMichael AJ, Haynes BF, Telenti A, Goldstein DB: **A whole-genome association study of major determinants for host control of HIV-1.** *Science* 2007, **317:**944-947.
20. Burgner D, Davila S, Breunis WB, Ng SB, Li Y, Bonnard C, Ling L, Wright VJ, Thalamuthu A, Odam M, Shimizu C, Burns JC, Levin M, Kuijpers TW, Hibberd ML, on behalf of the International Kawasaki Disease Genetics Consortium: **A genome-wide association study identifies novel and functionally related susceptibility loci for Kawasaki disease.** *PLoS Genet* 2009, **5:**e100319.
21. Burns JC, Glode MP: **Kawasaki syndrome.** *Lancet* 2004, **364:**533-544.
22. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH: **A genome-wide association study identifies IL23R as an inflammatory bowel disease gene.** *Science* 2006, **314:**1461-1463.
23. Graham RR, Kozyrev SV, Baechler EC, Reddy MV, Plenge RM, Bauer JW, Ortmann WA, Koeuth T, Gonzalez Escribano MF, Pons-Estel B, Petri M, Daly M, Gregersen PK, Martin J, Altshuler D, Behrens TW, Alarcon-Riquelme ME: **A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus.** *Nat Genet* 2006, **38:**550-555.
24. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, Liew A, Khalili H, Chandrasekaran A, Davies LR, Li W, Tan AK, Bonnard C, Ong RT, Thalamuthu A, Pettersson S, Liu C, Tian C, Chen WV, Carulli JP, Beckman EM, Altshuler D, Alfredsson L, Criswell LA, Amos CI, Seldin MF, Kastner DL, Klareskog L, Gregersen PK: **TRAF1-C5 as a risk locus for rheumatoid arthritis - a genomewide study.** *N Engl J Med* 2007, **357:**1199-1209.
25. Risch N, Merikangas K: **The future of genetic studies of complex human diseases.** *Science* 1996, **273:**1516-1517.
26. Lander ES: **The new genomics: global views of biology.** *Science* 1996, **274:**536-539.
27. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MA, Green T, Platt OS, Ruderfer DM, Walsh CA, Altshuler D, Chakravarti A, Tanzi RE, Stefansson K, Santangelo SL, Gusella JF, Sklar P, Wu BL, Daly MJ: **Association between microdeletion and microduplication at 16p11.2 and autism.** *N Engl J Med* 2008, **358:**667-675.
28. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, Stray SM, Rippey CF, Roccanova P, Makarov V, Lakshmi B, Findling RL, Sikich L, Stromberg T, Merriman B, Gogtay N, Butler P, Eckstrand K, Noory L, Gochman P, Long R, Chen Z, Davis S, Baker C, Eichler EE, Meltzer PS, *et al.*: **Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia.** *Science* 2008, **320:**539-543.
29. Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgeirsson TE, Sigurdsson A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Moller HJ, Hartmann A, Shianna KV, *et al.*: **Large recurrent microdeletions associated with schizophrenia.** *Nature* 2008, **455:**232-236.
30. Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NT, Thuong NT, Stepniewska K, Huyen MN, Bang ND, Loc TH, Gagneux S, van Soolingen D, Kremer K, van der Sande M, Small P, Anh PT, Chinh NT, Quy HT, Duyen NT, Tho DQ, Hieu NT, Torok E, Hien TT, Dung NH, Nhu NT, Duy PM, van Vinh Chau N, Farrar J: **The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*.** *PLoS Pathog* 2008, **4:**e1000034.
31. Haralambous E, Dolly SO, Hibberd ML, Litt DJ, Udalova IA, O'Dwyer C, Langford PR, Simon Kroll J, Levin M: **Factor H, a regulator of complement activity, is a major determinant of meningococcal disease susceptibility in UK Caucasian patients.** *Scand J Infect Dis* 2006, **38:**764-771.
32. Beernink PT, Welsch JA, Bar-Lev M, Koeberling O, Comanducci M, Granoff DM: **Fine antigenic specificity and cooperative bactericidal activity of monoclonal antibodies directed at the meningococcal vaccine candidate factor H-binding protein.** *Infect Immun* 2008, **76:**4232-4240.