



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

# Infectogenomics: Insights from the Host Genome into Infectious Diseases

Paul Kellam<sup>1</sup> and Robin A. Weiss<sup>1,\*</sup>

<sup>1</sup>MRC/UCL Centre for Medical Molecular Virology, Division of Infection & Immunity, University College London, London W1T 4JF, UK

\*Contact: r.weiss@ucl.ac.uk

DOI 10.1016/j.cell.2006.02.003

Five years into the human postgenomic era, we are gaining considerable knowledge about host-pathogen interactions through host genomes. This “infectogenomics” approach should yield further insights into both diagnostic and therapeutic advances, as well as normal cellular function.

As with most biological phenomena, disease is the outcome of both nature and nurture. In the case of infectious disease, however, the interaction of two natures, that is, two genomes (host and pathogen), is at play. Environmental and social factors, or nurture, may affect the risk of acquiring infection, and also the risk of becoming ill. How we behave and what environment we live in will determine the number of exposure events (Weiss and McMichael, 2004). The dose of infection and the fitness of the host and pathogen will determine sickness (van Opijnen and Berkhout, 2005). After all, unfit pathogens can make excellent live attenuated vaccines.

The inherent virulence of the pathogen should always be considered in the setting of the host. This interplay is particularly evident in cross-species infections. Herpes B virus elicits little more than cold sores in its natural host, the macaque, whereas in humans it causes a life-threatening encephalitis. Similarly, *Escherichia coli* O157 rarely affects cattle adversely but gives humans severe diarrhea, and H5N1 avian influenza virus is more virulent in geese, chickens, and humans than in ducks. Many host-microbe combinations coevolve with an apparent drive toward decreased pathogenicity of the microbe in its host reservoir, a constraint that can be relieved upon transfer of the microbe to another species.

In studying the severity of infectious diseases, it is not always clear

how much variation to attribute to the virulence of the pathogen and how much to the susceptibility of the host. In some situations, however, the pathogen can be regarded as a constant, thereby revealing the contribution of the host. For instance, the SARS coronavirus that spread from human to human during the 2003 outbreak came from a point source and was essentially an invariant clone. Yet we can discern at least three human host phenotypes of infection, namely, death, recovery, and superspreaders. Moreover, being a superspreader was not highly correlated with severity of disease, and luckily for the human population, superspreaders represented a small proportion of those who became infected. We do not know to what extent human genetic variation determined death, recovery, and infectiousness to others, but it should be studied. Doubtless, correlations will be found in the immunogenetics of the major histocompatibility complex (MHC). However, other factors that are difficult to reduce to one type of gene—possibly specific differences in the coordinated multicell, multipathway response to SARS coronavirus—may have been involved.

Similarly, the contribution of pathogen variation to pathogenesis is complex. Recent clues suggest a functional consequence of RNA virus variation that results from error-prone viral RNA polymerases producing genetically distinct but related viral “quasispecies.” For poliovirus, the

quasispecies itself, rather than the presence of individual adaptive mutations, resulted in increased pathogenesis, supporting the theory that viral colonization of an ecosystem (in this case the host) may be a result of the complex interplay between different members of the quasispecies (Vignuzzi et al., 2006).

## How Pathogens Have Shaped Our Genotype

It is generally agreed that common genetic polymorphisms, such as those in the ABO blood groups and MHC antigens, are maintained in human populations on account of their value in the immune response to infectious diseases. However, clear examples of balanced polymorphisms are few, and we have to beware of altruistic interpretations. Perhaps the best evidence is where there is a tradeoff between resistance to infection and genetic disease. This is seen in the various hemoglobinopathies ( $\alpha$ - and  $\beta$ -thalassemia, and sickle cell anemia) where the resistance to infection with the malaria parasite of those who are heterozygous for the gene mutation counterbalances the lethal effects of recessive disease in homozygotes. The relatively high frequency of cystic fibrosis may also be explained if those heterozygous for mutations in the chloride channel (CTFR) survive severe enteric infections such as typhoid because of reduced dehydration; in this case, heterozygotes may be considered the functional equivalent of oral rehydration therapy.

There are also examples of host genetic resistance to infections in which the homozygous state is not lethal. The delta 32 deletion in the gene encoding the CCR5 chemokine receptor is frequent in the Caucasian population. Homozygotes appear to be healthy and are resistant to infection by the majority of strains of HIV-1 that require this receptor to enter their target cells. Heterozygotes also benefit from relative resistance to infection and, when infected, progress to AIDS more slowly. But like cholera and CTFR, HIV is too recent an arrival in humans to account for the selection of delta 32 CCR5. Resistance to smallpox or the plague has been postulated as the original selective force, but compelling evidence has not been forthcoming.

#### Host Variation in Infectious Disease

The occurrence of these classical Mendelian traits that have large discrete effects suggests that searching the human genome will reveal polymorphisms that affect susceptibility to specific infectious diseases. To date, however, scanning the genome for single-nucleotide polymorphisms (SNPs) or other markers has not been particularly fruitful in revealing genetic loci associated with a trait for infectious diseases. This is due to the fact that much of the genetic variation underlying infectious disease susceptibility is complex, involving the combination of many loci, and is further confounded by SNP variations between different populations and ethnic groups. Such surveys, therefore, require careful design, and it may be that identifying loci in animal models is a more straightforward way of highlighting loci and genes in humans. Correspondingly, it has been more fruitful thus far to identify polymorphisms in candidate human genes for infectious disease.

#### Multigenic Variation: HIV and AIDS

Both HIV infection and progression to AIDS are influenced by the host genotype, and in turn, the host imposes selection on the virus. Obvious candidate genes have been investigated

in HIV pathogenesis. As expected, both class I and class II genes in the MHC region affect HIV susceptibility and disease progression (Carrington and O'Brien, 2003), but it is the combination of "good" and "bad" alleles that determines the phenotype of the disease. There is even a suggestion that rare MHC alleles offer stronger cellular immune responses to HIV because the viral peptide sequences have mutated to evade the more common MHC genotypes (Scherer et al., 2004).

An excellent example of multigenic variation in AIDS is the complexity of host receptor-ligand interactions. Most of the transmissible strains of HIV-1 require the CCR5 chemokine receptor as a fusion receptor after the virus particles bind to the CD4 receptor on host T cells. There are widespread polymorphisms in the promoter region of the CCR5 gene that affect the density of receptor expression on the surface of CD4<sup>+</sup> T lymphocytes. There is also polymorphism in the number of gene copies of its main ligand, the CCL3L1 form of the cytokine MIP-1 $\alpha$ . Careful analysis of susceptibility to HIV infection and progression to AIDS in large human cohorts reveals that it is the combination of CCR5 and CCL3L1 polymorphisms that has the most telling effect on disease (Gonzalez et al., 2005). High chemokine expression combined with low receptor expression delays the development of AIDS in the host. This makes sense from our understanding of HIV entry into cells because CCL3L1 competes with HIV for a limiting number of receptors.

Intracellular host restriction factors, such as APOBEC3 and Trim5 $\alpha$ , also affect HIV infection. Polymorphisms in these proteins have been analyzed principally by comparison of host species. The virus had to develop mutations in the proteins that interact with these host factors to adapt to its new human host. However, recent data indicate that polymorphisms within human populations may also play a role in susceptibility to HIV (An et al., 2004; Sawyer et al., 2006) and routes of infection (Shrestha et al., 2006).

#### Host Genetics of Common Infections

Regarding common infections that are seldom pathogenic, we can envisage an infectogenomics approach analogous to pharmacogenetics. The appearance of a disease or symptom following exposure to an infectious agent can be regarded as an unusual "side effect" just like an adverse reaction to a drug.

With the wrong genotype, such adverse reactions can be severe indeed. For example, 85% of the global human population, or approximately 5 billion people, are persistently infected with Epstein-Barr virus (EBV). Most of us become infected in infancy without diagnosed illness, although infection in adolescence causes infectious mononucleosis. But if an infant is homozygous for the Duncan's syndrome allele, primary infection leads to uncontrollable, lethal mononucleosis. Curiously, this gene appears to affect the response to EBV exclusively; it is not related to a more general immune deficiency.

Other infections may exhibit mild symptoms, except in rare cases of severity, such as those caused by cytomegalovirus (CMV) or human herpes virus type 6 (also see the Essay by S. Falkow, page 699 of this issue). The throat infection, *Neisseria meningococcus*, circulates among asymptomatic carriers but causes meningitis on rare occasions. Perhaps we should not only seek microbial markers of virulence but also spend more effort on identifying the human genetic factors that predispose to invasion by pathogens. Moreover, common infections with low virulence may represent the unknown environmental trigger for diseases that are not clearly infectious—such as, multiple sclerosis, asthma, and acute lymphocytic leukemia—when they infect a genotype predisposed to the disease.

#### Diagnostic and Prognostic Signatures in the Host Transcriptome

The host genotype ultimately manifests its function through differences in gene transcription or functional

differences in proteins. Advances in analysis of gene expression profiling via microarrays, and also in proteomics, not only show how pathogens remodel the host's gene expression patterns (Jenner and Young, 2005) but also provide a wealth of candidate genes for genetic susceptibility studies. Some of these changes are common to large groups of infection; for instance, many viruses stimulate interferon response pathways. Changes in gene expression profiles can also betray the type of pathogen present with the potential of identifying the pathogen through host gene expression (Huang et al., 2001). Thus, gene expression patterns in blood could serve as a window into the pathogenesis and diagnosis of infectious diseases. Whole blood and purified peripheral blood mononuclear cells exhibit interpatient variation, reflected in the global gene expression patterns of the samples but nevertheless can be interpreted through knowledge of other clinical parameters such as differential blood counts, temperature, and gender. When these are accounted for, the remaining gene expression changes indicate both general infection and disease-specific responses (Griffiths et al., 2005; Whitney et al., 2003).

Transcriptional profiles may also be expected to aid prognosis and responses to treatment as is now the case in oncology (Bild et al., 2006). For example, some HIV-infected patients on antiretroviral therapy do not respond with falling viral load even when the virus itself shows no markers of drug resistance, and it may be that host factors determine nonresponsiveness in these cases. Similarly, only a proportion of patients with hepatitis C infection respond to interferon treatment, and transcriptional profiling may indicate why this is so. This is an area where infectogenomics and pharmacogenomics merge.

### Combating the Pathogen through Host Functional Genomics

Transcriptional profiles specific for infected cells also provide insights into known pathways of gene regula-

tion and reveal new ways of treating or managing infections (Kellam, 2006). For example, a study of gene expression among B cell lymphomas delineated a distinct interaction of Kaposi sarcoma associated virus (KSHV) from that of EBV. Because KSHV-infected lymphoma cells show upregulation of the vitamin D receptor pathway, known inhibitors of that pathway may be explored therapeutically (Jenner et al., 2003). Similarly, cyclo-oxygenase 2 (COX-2) inhibitors were successfully tested for reduction of CMV titer after detection of upregulation of COX-2 expression in infected cells. In addition, the HIV-1 Nef protein increases cholesterol biosynthesis, suggesting that cholesterol-lowering drugs such as statins may have an antiviral effect in vivo (del Real et al., 2004).

Thus new roles for known drugs are emerging from studies of infectogenomics. Such insights could provide an increased pharmaceutical inventory based on drugs that have already undergone extensive toxicity and efficacy screening as treatment for other diseases (Kellam, 2006).

### Conclusions

The functional genomics of the host is of crucial importance in analyzing host-pathogen interactions. Host genetic variation plays a key role in determining the outcome of many potentially pathogenic infections, and the prevalent pathogens have influenced the genetic make-up of human populations. Infectogenomics can be harnessed to identify infectious states, to understand the host response, to predict disease outcomes, to monitor responses to antimicrobial therapies, and to indicate promising new types of treatment. In addition, we should acknowledge that the disease state can inform our understanding of normality. Just as virology led us to oncogenes, tumor suppressor proteins, membrane trafficking pathways, and other aspects of molecular cell biology in the past, so can studies of the perturbation of the transcriptome by infection open new vistas onto "systems" biology today.

### REFERENCES

- An, P., Bleiber, G., Duggal, P., Nelson, G., May, M., Mangeat, B., Alobwede, I., Trono, D., Vlahov, D., Donfield, S., et al. (2004). *J. Virol.* 78, 11070–11076.
- Bild, A.H., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.B., Harpole, D., Lancaster, J.M., Berchuck, A., et al. (2006). *Nature* 439, 353–357.
- Carrington, M., and O'Brien, S.J. (2003). *Annu. Rev. Med.* 54, 535–551.
- del Real, G., Jimenez-Baranda, S., Mira, E., Lacalle, R.A., Lucas, P., Gomez-Mouton, C., Alegret, M., Pena, J.M., Rodriguez-Zapata, M., Alvarez-Mon, M., et al. (2004). *J. Exp. Med.* 200, 541–547.
- Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R.J., Freedman, B.I., Quinones, M.P., Bamshad, M.J., et al. (2005). *Science* 307, 1434–1440.
- Griffiths, M.J., Shafi, M.J., Popper, S.J., Hemingway, C.A., Kortok, M.M., Wathen, A., Rockett, K.A., Mott, R., Levin, M., Newton, C.R., et al. (2005). *J. Infect. Dis.* 191, 1599–1611.
- Huang, Q., Liu, D., Majewski, P., Schulte, L.C., Korn, J.M., Young, R.A., Lander, E.S., and Hacohen, N. (2001). *Science* 294, 870–875.
- Jenner, R.G., and Young, R.A. (2005). *Nat. Rev. Microbiol.* 3, 281–294.
- Jenner, R.G., Maillard, K., Cattini, N., Weiss, R.A., Boshoff, C., Wooster, R., and Kellam, P. (2003). *Proc. Natl. Acad. Sci. USA* 100, 10399–10404.
- Kellam, P. (2006). *Genome Biol.* 7, 201.
- Sawyer, S.L., Wu, L.I., Akey, J.M., Emerman, M., and Malik, H.S. (2006). *Curr. Biol.* 16, 95–100.
- Scherer, A., Frater, J., Oxenius, A., Agudelo, J., Price, D.A., Gunthard, H.F., Barnardo, M., Perrin, L., Hirschel, B., Phillips, R.E., and McLean, A.R. (2004). *Proc. Natl. Acad. Sci. USA* 101, 12266–12270.
- Shrestha, S., Strathdee, S.A., Galai, N., Oleksyk, T., Fallin, M.D., Mehta, S., Schaid, D., Vlahov, D., O'Brien, S.J., and Smith, M.W. (2006). *J. Infect. Dis.* 193, 16–26.
- van Opijnen, T., and Berkhout, B. (2005). *Rev. Med. Virol.* 15, 219–233.
- Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E., and Andino, R. (2006). *Nature* 439, 344–348.
- Weiss, R.A., and McMichael, A.J. (2004). *Nat. Med.* 10, S70–S76.
- Whitney, A.R., Diehn, M., Popper, S.J., Alizadeh, A.A., Boldrick, J.C., Relman, D.A., and Brown, P.O. (2003). *Proc. Natl. Acad. Sci. USA* 100, 1896–1901.