Susceptibility to HIV Infection—Disentangling Host Genetics and Host Behavior

Amalio Telenti[1] and John P. A. Ioannidis[2]

[1]Institute of Microbiology and University Hospital, University of Lausanne, Lausanne, Switzerland; [2]Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece

(See the article by Shrestha et al., on pages 16–26.)

In 1874, Francis Galton, a cousin of Charles Darwin, coined the idiom "nature and nurture" to cover all of the influences that determine an adult's constitution—the sum total of their particular physical and behavioral characteristics. In this issue of the *Journal*, Shrestha et al. [1] apply state-of-the-art genetics along with explicit modeling of an "environmental" component (high risk exposure) to disentangle the nature and nurture of an individual's susceptibility to HIV-1. The study is an example of the rapidly evolving field of host genetics in infectious diseases.

Nature offers 2 experiments on HIV-1 susceptibility: the uniqueness of long-term nonprogressors and the highly exposed uninfected individual. Both represent situations that hide a large amount of heterogeneity in mechanisms and biology. However, the condition of being "exposed uninfected" appears to be particularly intractable, because it grafts a multifactorial trait onto the poor efficiency of transmission of HIV-1 and the rarity of cohorts of individuals that are characterized as highly exposed and persistently seronegative. Particular studies have examined heterosexual couples with discordant HIV serostatus, highly exposed sex workers, and highly exposed men having sex with men. The mechanisms identified or invoked to modulate susceptibility to infection in the various reports emphasize the relevance of differences in acquired immunity, through the role of protective cytotoxic T lymphocyte responses and NK cell activity in the context of specific HLA class I alleles (reviewed in [2]), as well as differences in humoral responses at mucosal surfaces [3]. The other area of research has included assessment of expression of chemokine receptors and their ligands, as well as the role of genetic variants of those molecules (reviewed in [2]). In addition, CD4 T cells from healthy blood donors differ markedly in their susceptibility to HIV-1 infection [4], and a relative resistance to HIV-1 infection of CD4 T cells from exposed uninfected individuals has been described [5].

The study by Shrestha et al. [1] aimed at freeing itself from the constraint of the low numbers of the above unique populations—and, thus, the constraint of limited statistical power—by using a standardized cumulative risk exposure measurement on a larger number of individuals ($n = 789$). They applied this approach to the evaluation of genetic variants of 9 genes involved in HIV-1 entry and replication, with a focus on the biology of CCR5-mediated pathways. Not included was *CCL3L1*, coding for the CCR5 ligand macrophage inflammatory protein 1 (MIP1)–α-P and present in the human genome at variable copy numbers, which define different levels of risk for infection [6].

Most genetic epidemiology studies to date have been unable to deal efficiently with environmental factors [7]. Although taking genetic "measurements" (i.e., genotyping) is straightforward, measuring environmental exposures can be difficult and imprecise. Usually, we have to rely on traditional epidemiological methods (e.g., questionnaires), with all of their shortcomings. Shrestha et al. use an innovative approach to incorporate a multivariate propensity score for cumulative high risk exposure. Propensity methods were originally developed to estimate which parameters are associated with the "risk" of receiving a specific treatment in nonrandomized databases [8]. Patients may have had different reasons for being given one or another treatment, and propensity models offered a way to balance the comparison of these treatments for such uneven treatment preferences. In the genetic set-

Received 20 July 2005; accepted 20 July 2005; electronically published 22 November 2005.

Financial support: Swiss National Science Foundation (to A.T.); General Secretariat for Research and Technology, Greece (to J.P.A.I.).

Potential conflicts of interest: none reported.

Reprints or correspondence: Dr. Amalio Telenti, Institute of Microbiology, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland (amalio.telenti@chuv.ch); or Dr. John P. A. Ioannidis (jioannid@cc.uoi.gr).

The Journal of Infectious Diseases 2006;193:4–6

0022-1899/2006/19301-0002$15.00

ting, genotypes are the equivalent of "treatments." Typically, we rely on Mendelian randomization [9] and assume that high risk exposure and confounding factors would also be randomly assigned to different genotypes. This means that we accept that exposure would be similar regardless of genotype. One could go even a step further and examine whether there is any interaction between risk exposure and genetic profile. However, demonstration of interaction effects would require even larger sample sizes.

There are more challenges to be overcome in modeling environmental factors. Multivariate models have shortcomings, and predictive models also need validation, both internally (through bootstrapping, for example) and externally (in different populations). There is increasing evidence of the limitations of appraising prognostic factors [10]. Moreover, the discriminatory ability of predictive models may still be limited, even if several risks are considered. For example, even in Shrestha et al.'s state-of-the-art study, the high-risk quintile, compared with the low-risk quintile, had only a 2-fold difference in risk (49% vs. 24%).

Even if we ignore the complexity conferred by the interaction of the host genetic background with risk behavior, the sample size required to detect modest genetic effects (i.e., odds ratios in the 1.1–1.8 range) can be quite large. Most genetic epidemiology studies conducted in the past have been underpowered [11]. An advantage of the study by Shrestha et al. is that, given its sample size, modest genetic effects for relatively common alleles would not be missed. However, even with 789 subjects, there is not enough power to pursue polymorphisms with allele frequencies <10%. It is unknown how much such uncommon polymorphisms contribute to the genetic background of HIV-1 susceptibility. For example, the strongest protective factor known to date, homozygosity for the *CCR5-Δ32* allele, occurs in only 1% of white individuals and is practically absent in people of African descent.

Multiple testing can also lead to spurious findings that are not validated by subsequent studies [12]. Of the 50 polymorphisms tested by Shrestha et al., 2 or 3 would be expected to show significant associations ($P < .05$) just by chance. With the advent of discovery-oriented approaches [13], it is possible to screen hundreds of thousands of genetic variants across relatively large data sets. At first pass, a typical whole-genome association study may yield a few thousand "promising" polymorphisms for any complex trait. Evidently, all of these tentative associations need further validation in independent data sets. Thus, proposed genetic associations should be considered to be works in progress [14]. Haplotype analyses, such as those utilized by Shrestha et al., are also becoming the rule. The HapMap project [15] aims to describe most existing genetic variability by typing only a limited number of polymorphisms per gene. However, haplotypes may not always adequately represent the full variability in diverse populations, and comparing selected haplotypes rather than all possible haplotypes without clear justification is a dangerous practice.

Another major threat to the validity of genetic epidemiology studies is the selective publication and selective reporting of the analyses being performed. Shrestha et al. provide an excellent example of how all tested polymorphism associations should be reported, regardless of whether they lead to significant results. The availability of electronic online supplements should obviate problems related to the mass of currently assembled databases, even for the most data-rich analyses. In publications that present the discovery of a few formally statistically significant associations, it is unknown whether the investigators tested many others but are reporting only the most significant ones. Such selective reporting may lead to a distorted literature, and replication of the claimed findings would be uncommon [12]. Formal statistical significance should not be used as a criterion for dissemination of research in genetic epidemiology.

Finally, we still need more empirical evidence on how to link epidemiological evidence with data on the biological and functional significance of postulated gene-disease associations [16]. Of the 4 candidate protective polymorphisms identified by Shrestha et al., one is in a noncoding region, another is synonymous (no amino acid change), and a third is at odds with previous reports. Only 1 polymorphism, *CCR2 64I,* has a biological rationale to support its role [17] and concurrent evidence for protection against disease progression [18]. Nevertheless, this polymorphism has not been associated with risk of HIV-1 transmission via the vertical route [19]. Squaring epidemiology with biology can be difficult. However, we should keep an open mind, since there is much we do not know about the intricacies of the human genome and its function. For example, polymorphisms in noncoding regions may have important indirect biological effects.

Overall, the extensive work of Shrestha et al. yielded a modest output—4 "hits" among 50 genetic variants in 9 genes. Is this a satisfactory outcome for a large and costly experiment? The answer is a sober "yes." Sooner or later, all genes linked to the pathogenesis of a given disease will be genotyped, and all putative genetic markers (associated with a study phenotype) will be followed by using the appropriate genetic tools to identify linked functional variants, by establishing the appropriate biological experiments, and by validation across several populations and cohorts. For susceptibility to HIV-1 infection, the list of candidates is long, because limited work has been done on variants of genes that are needed for viral replication in the cell [20, 21]. The list is longer if we move from acquired immunity to the innate immunity or to genes coding for cellular antiretroviral defense mechanisms. Species-specific single–amino acid differences in apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G (APOBEC3G) [22, 23], associated

with hypermutation of the viral DNA, and in tripartite motif–containing protein 5 (TRIM5)–α [24, 25], an antiretroviral protein acting after HIV-1 infects the cell, are enough to provide adequate protection against HIV-1 in some primates, whereas the human versions are ineffective. Characterization of variants of these genes in humans has just started [26, 27].

Against a background of "hype" and excitement around all things genomic, there is one more lesson to learn from most studies trying to make sense of the complex biology behind interindividual differences. It is true that some "big" discoveries and breakthroughs will happen. For example, the identification of the *CCR5-Δ32* mutation not only resulted in a paradigm shift in our understanding of pathogenesis, but it also helped in the development of a new class of antiretroviral agents. However, much work in genetics and genomics is unavoidably antlike, going through scores of data to, piece by piece, reconstruct the pathogen-host interaction—this time, starting from the host side. It is unlikely that the identification of multigene effects will change the management of HIV-1 disease in the near future. It would even be erroneous to suggest that genetic tests that, in isolation, confer small quanta of information should be integrated into patient care. One has to verify first that they are indeed true and reproducible, possible to standardize for clinical use, amenable to use by information-overwhelmed clinicians, readily interpretable, able to improve outcomes, and cost-effective. None of the validated or postulated genetic markers of HIV-1 disease susceptibility or progression meets all of these criteria for clinical use.

How should infectious-diseases specialists react to the growing number of studies on host genetics? Over the past 20 years, they have been asked to learn a great deal of microbiology, then basic immunology, molecular biology, and biostatistics. Now they will need to learn genetics and genetic nomenclature and to build genetic cohorts [28] and collaborations between cohorts working on the same field [29]. Large challenges lie ahead.

References

1. Shrestha S, Strathdee SA, Galai N, et al. Behavioral risk exposure and host genetics of susceptibility to HIV-1 infection. J Infect Dis **2006**; 193:16–26 (in this issue).
2. Kaslow RA, Dorak T, Tang JJ. Influence of host genetic variation on susceptibility to HIV type 1 infection. J Infect Dis **2005**; 191(Suppl 1): S68–77.
3. Mazzoli S, Trabattoni D, Lo CS, et al. HIV-specific mucosal and cellular immunity in HIV-seronegative partners of HIV-seropositive individuals. Nat Med **1997**; 3:1250–7.
4. Ciuffi A, Bleiber G, Munoz M, et al. Entry and transcription as key determinants of differences in CD4 T cell permissiveness to HIV-1 infection. J Virol **2004**; 78:10747–54.
5. Paxton WA, Martin SR, Tse D, et al. Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high-risk sexual exposure. Nat Med **1996**; 2:412–7.
6. Gonzalez E, Kulkarni H, Bolivar H, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science **2005**; 307:1434–40.
7. Botto LD, Khoury MJ. Commentary: facing the challenge of gene-environment interaction: the two-by-four table and beyond. Am J Epidemiol **2001**; 153:1016–20.
8. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med **1997**; 127:757–63.
9. Davey SG, Ebrahim S. What can Mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ **2005**; 330:1076–9.
10. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med **2000**; 19:453–73.
11. Ioannidis JP. Genetic associations: false or true? Trends Mol Med **2003**; 9:135–8.
12. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet **2001**; 29:306–9.
13. Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat Genet **2005**; 37:413–7.
14. Ioannidis JP. Why most published research findings are false. PLoS Med **2005**; 2:e124.
15. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P, International HapMap Consortium. A haplotype map of the human genome. Nature **2005**; 437:1299–320.
16. Rebbeck TR, Spitz M, Wu X. Assessing the function of genetic variants in candidate gene association studies. Nat Rev Genet **2004**; 5: 589–97.
17. Nakayama EE, Tanaka Y, Nagai Y, Iwamoto A, Shioda T. A CCR2-V64I polymorphism affects stability of CCR2A isoform. AIDS **2004**; 18:729–38.
18. Ioannidis JP, Rosenberg PS, Goedert JJ, et al. Effects of *CCR5-Δ32, CCR2-64I,* and *SDF-1 3′A* alleles on HIV-1 disease progression: an international meta-analysis of individual-patient data. Ann Intern Med **2001**; 135:782–95.
19. Contopoulos-Ioannidis DG, O'Brien TR, Goedert JJ, Rosenberg PS, Ioannidis JP. Effect of *CCR5-Δ32* heterozygosity on the risk of perinatal HIV-1 infection: a meta-analysis. J Acquir Immune Defic Syndr **2003**; 32:70–6.
20. Bleiber G, May M, Martinez R, et al. Use of a combined *ex vivo*/*in vivo* population approach for screening of human genes involved in the human immunodeficiency virus type 1 life cycle for variants influencing disease progression. J Virol **2005**; 79:12674–80.
21. Greene WC, Peterlin BM. Charting HIV's remarkable voyage through the cell: basic science as a passport to future therapy. Nat Med **2002**; 8:673–80.
22. Schrofelbauer B, Chen D, Landau NR. A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). Proc Natl Acad Sci USA **2004**; 101: 3927–32.
23. Mangeat B, Turelli P, Liao S, Trono D. A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. J Biol Chem **2004**; 279:14481–3.
24. Yap MW, Nisole S, Stoye JP. A single amino acid change in the SPRY domain of human Trim5alpha leads to HIV-1 restriction. Curr Biol **2005**; 15:73–8.
25. Stremlau M, Perron M, Welikala S, Sodroski J. Species-specific variation in the B30.2(SPRY) domain of TRIM5α determines the potency of human immunodeficiency virus restriction. J Virol **2005**; 79:3139–45.
26. An P, Bleiber G, Duggal P, et al. APOBEC3G genetic variants and their influence on AIDS. J Virol **2004**; 78:11070–6.
27. Do H, Vasilescu A, Diop G, et al. Exhaustive genotyping of the CEM15 (APOBEC3G) gene and absence of association with AIDS progression in a French cohort. J Infect Dis **2005**; 191:159–63.
28. Telenti A. The genetic cohorts: facing the new challenges in infectious diseases: the HIV model. Enferm Infecc Microbiol Clin **2004**; 22:337–41.
29. Ioannidis JP, Bernstein J, Boffetta P, et al. A network of investigator networks in human genome epidemiology. Am J Epidemiol **2005**; 162:302–4.