Published in final edited form as:

Nat Immunol. 2015 June ; 16(6): 577–583. doi:10.1038/ni.3147.

The impact of host genetic variation on infection with HIV-1

Paul J McLaren¹ and Mary Carrington^{2,3}

¹School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

²Cancer and Inflammation Program, Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA. ³The Ragon Institute of MGH, MIT and Harvard, Cambridge, Massachusetts, USA.

Abstract

The outcome after infection with the human immunodeficiency virus type 1 (HIV-1) is a complex phenotype determined by interactions among the pathogen, the human host and the surrounding environment. An impact of host genetic variation on HIV-1 susceptibility was identified early in the pandemic, with a major role attributed to the genes encoding class I human leukocyte antigens (HLA) and the chemokine receptor CCR5. Studies using genome-wide data sets have underscored the strength of these associations relative to variants located throughout the rest of the genome. However, the extent to which additional polymorphisms influence HIV-1 disease progression, and how much of the variability in outcome can be attributed to host genetics, remain largely unclear. Here we discuss findings concerning the functional impact of associated variants, outline methods for quantifying the host genetic component and examine how available genome-wide data sets may be leveraged to discover gene variants that affect the outcome of HIV-1 infection.

Three decades of research has resulted in a massive amount of information about the pathogenesis and treatment of HIV-1, much of which is applicable to the general understanding of immunology, virology, host genetics and related disciplines. The development of combined antiretroviral therapy (cART) has been astonishingly effective in extending the lives of people who are infected and curbing the spread of HIV-1. Nevertheless, these drugs do not completely eliminate the virus from its cellular reservoir, and development of an effective preventive or therapeutic vaccine remains a major hurdle. The goal of the HIV-1 genetics community is to identify host genetic variation that has an impact on pathogenesis in order to direct the design of therapeutics and vaccines. To date, the main focus has been on identifying phenotypes of differing susceptibility to infection (including in the general population and in high-risk groups that have been exposed but not infected) or markers of disease severity, including viral load (HIV-1 RNA copies per milliliter of plasma), rate of CD4⁺ T cell decline and time to development of AIDS (Fig. 1).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

Correspondence should be addressed to P.J.M. (paul.mclaren@epfl.ch) or M.C. (carringm@mail.nih.gov).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

It is broadly accepted that a minority of people are resistant to HIV-1 infection¹. However, homozygosity for the *CCR5* deletion mutation *CCR5Δ32* is the only genotype that has been consistently identified to protect against HIV-1 infection. Several other variants have been proposed to confer similar protection², but the results have not been replicated in other cohorts, nor have these (or any other variants) been identified in genome-wide association studies (GWAS) of protection against HIV-1 infection^{3,4}. These data indicate that any other genetic effects that protect against infection are of low penetrance or frequency or involve a more complex interaction between two or more genetic variants.

Severity of disease after infection is similarly variable. This variability has often been quantified in nonascertained, population-level samples by measurement of viral load or rate of CD4⁺ T cell decline or through observation of individuals with extreme progression phenotypes⁵. The majority of knowledge about the impact of host genetic variation on HIV-1 outcome has been gained through studies of historical cohorts collected before the development of cART. Among the most valuable cohorts are those including patients for whom seroconversion dates can be determined to within several months or less. Studies of these cohorts were the first to identify specific *HLA* alleles as the primary determinants of rate of progression to AIDS^{6–9}. Seroprevalent cohorts have identified similar *HLA* associations¹⁰, though some differences across cohorts may be due to distinct, albeit related, outcomes.

Many other variants throughout the genome have been claimed to have an association with outcome after HIV-1 infection, but only variation within or near the *HLA* genes and, to a lesser extent, the *CCR5* locus have been replicated in large GWAS^{11,12}. Though these studies confirm the primary impact of variation in *HLA* genes on HIV-1 outcome, they do not address the role of variation in complex genetic regions such as those encoding T cell receptors (TCRs) and immunoglobulins, owing to variation in copy number and/or somatic recombination¹³. For example, variation in the genes encoding killer immunoglobulin-like receptor (KIR) proteins, particularly *KIR3DL1* and *KIR3DS1*, has been associated with HIV-1 outcomes in many genetic and functional studies¹⁰, but these have not been identifiable by GWAS, almost certainly because of the extreme inter- and intragenic diversification of the *KIR* haplotypes¹⁴. Given that GWAS provides an ‘agnostic’ approach to analysis of single-nucleotide polymorphisms (SNPs), they do not assess pairs of functionally related variants (linked or unlinked) that may act in a dependent manner. Again, the *KIR* locus serves as an example, as variation in *KIR* genes is functionally relevant only in the presence of alleles encoding their specific (unlinked) *HLA* ligands; the absence of these *HLA* alleles seriously dilutes any effect of the corresponding *KIR* loci.

Studies have begun to investigate the functional impact of associated variants, tease out more complex associations between HIV-1 outcome and genetic variation and quantify the overall contribution of host genetics to variability in disease progression. Here we will concentrate on these approaches and outline areas of interest for future investigations.

Nonclassical effects of classical HLA-encoding genes

A wealth of functional and genetic data has implicated HIV-1 peptide presentation by HLA class I as a main determinant of disease progression. The first HIV-1 GWAS¹⁵ confirmed a major effect of *HLA-B*57* (tagged by rs2395029) on lowering viral load and slowing CD4⁺ T cell decline. This effect had been identified in a genetic study in 1996 (ref. 6), and overwhelming evidence in support of it was reported 4 years later¹⁶. Although a strong protective effect of *HLA-B*57* has been observed repeatedly in HLA-targeted studies and in GWAS, it should be noted that the majority of HIV-infected individuals carrying *HLA-B*57* actually progress to disease at a rate similar to that of those without *HLA-B*57*. No single genetic variant yet identified uniformly confers elite control of HIV-1, and it is likely that additive or synergistic effects of multiple genetic variants are required for HIV-1 control, as seems to be the case for *HLA-B*57*.

A later GWAS focusing on HIV-1 controllers (individuals who maintain extremely low viral load in the absence of therapy) again implied that diversity at the *HLA-B* gene is the primary host genetic influence and identified key amino acid positions in the peptide-binding groove of HLA-B that were strongly associated with HIV-1 control¹². This approach was facilitated through the development of computational methods to accurately infer common classical HLA types and variant amino acid positions by imputation from GWAS data¹⁷. The ability to obtain information on functional variation in this key region of the genome on the basis of genotypes from commercial arrays has broad utility that extends beyond HIV-1 and has been successfully applied to several other traits with a strong HLA component^{18–21}.

In addition to *HLA-B*, a variant located 35 kb upstream of the *HLA-C* gene (rs9264942) was shown to contribute independently to set-point viral load and to HIV-1 control^{12,15}. Importantly, this SNP is also associated with differential expression levels of *HLA-C* mRNA and HLA-C on the cell surface^{22,23} (i.e., it is an expression quantitative trait locus), and genotypes of rs9264942 that correlated with higher *HLA-C* expression were also associated with better HIV-1 control²³. These data suggested that the level of expression represented a mechanism by which variation at loci encoding class I HLA molecules may influence human disease. Although this idea was novel for an *HLA* locus, the level of expression of major histocompatibility complex (MHC) class I had been shown to affect risk of Marek's disease in chickens²⁴—low expression of MHC class I molecules, which are 'promiscuous' in peptide binding, associates with resistance to disease²⁵.

Determining a causal effect of *HLA-C* expression on HIV-1 control was not without controversy, largely owing to the strong linkage disequilibrium between the *HLA-B* and *HLA-C* genes. Because of this, it was proposed that the associations between rs9264942 and HIV-1 outcomes were due primarily to *HLA-B* alleles that are also in linkage disequilibrium with rs9264942 rather than to *HLA-C* expression levels²⁶. To address this concern, follow-up approaches using population genetics, molecular genetics and functional assays were required to determine whether *HLA-C* expression has a direct effect on HIV-1 control (Fig. 2).

Differences in linkage disequilibrium patterns across population backgrounds have been proposed as a powerful mechanism to fine map causal variants from disease-association studies²⁷. In line with this, although the association between rs9264942 and HIV-1 control did not hold up in populations of African ancestry, the link between *HLA-C* cell surface abundance and HIV-1 control was consistent in HIV-infected individuals of both European and African descent²⁸. These population-based analyses supported the model that levels of *HLA-C* expression have a direct effect on HIV-1 outcome, whereas rs9264942 has no direct effect on *HLA-C* expression or HIV-1 disease. Subsequently, it was shown that a polymorphic microRNA-148a (miR-148a) binding site in the 3' untranslated region (UTR) of *HLA-C* (in strong linkage disequilibrium with rs9264942 in people of European ancestry) was partly responsible for differential expression of *HLA-C* alleles²⁹. Some *HLA-C* alleles carry an intact binding site and are inhibited by miR-148a, whereas others escape its binding (and inhibitory effect) owing to variation in the seed region to which miR-148a binds²⁹.

Variation in the *MIR148A* gene (located on chromosome 7, unlinked to the *HLA* genes) also associates with levels of miR-148a expression, affecting the degree of downregulation of *HLA-C* alleles that have an intact miR-148a binding site in the 3' UTR³⁰. Higher expression of miR-148a resulted in lower expression of *HLA-C* and poorer HIV-1 control, but only in individuals carrying alleles with an intact miR-148a binding site. There was no effect of the *MIR148A* variant on *HLA-C* expression or HIV-1 control among those carrying *HLA-C* alleles that escape miR-148a regulation. As the escape variant is fixed at the *HLA-B* locus (i.e., no *HLA-B* allele is regulated by miR-148a), these molecular-genetic data strongly support a direct effect of *HLA-C* expression on HIV-1 that cannot be attributed to alleles of *HLA-B*.

Another approach to distinguishing the causal effect of *HLA-C* expression from that of closely linked loci was the analysis of host-viral sequence associations and cytotoxic T lymphocyte (CTL) functional data²⁸. Viral sequence data from nearly 2,000 individuals showed that the relative frequency of escape mutations (in predicted *HLA-C*-restricted epitopes) associated with increasing *HLA-C* expression, suggesting that high-expressing *HLA-C* alleles mediate greater selection pressure on the virus. These data substantiated a study in Han Chinese people that found that viral mutations in *HLA-C*-associated epitopes occurred more commonly among individuals homozygous for high-*HLA-C* expression alleles as marked by rs9264942 (ref. 31). Further functional data were generated from a data set of CTL responses to overlapping peptides across the HIV-1 proteome in 1,010 HIV-1-infected South Africans^{32,33}. These data showed a strong positive correlation between the relative frequency of responses to specific *HLA-C*-restricted peptides and the level of expression of the allotype²⁸. Such functional data lend further support to a model in which levels of *HLA-C* expression have a direct effect on the control of HIV-1. Interestingly, expression level of *HLA* genes (not just *HLA-C*) has been linked, directly and indirectly, to other diseases, including Crohn's disease²⁸, graft-versus-host disease after unrelated hematopoietic cell transplantation³⁴, hepatitis B virus infection³⁵, spontaneous clearance of hepatitis C virus³⁶ and Parkinson's disease³⁷. Higher *HLA* expression initiates CTL cytotoxicity³⁸ and modulates CTL cytokine secretion³⁹ more efficiently. Given these observations, further understanding of the impact of differential expression of alleles at all *HLA* loci should be pursued.

Heritability of disease progression and tolerance

Host genetic background (specifically in the *HLA* and *CCR5* regions) strongly influences outcome of HIV-1 infection. Understanding what and how additional host genetic, viral and environmental factors contribute to the variation in disease progression can help drive intervention strategies. To date, *HLA* and *CCR5* loci have been estimated to explain ~13% of variation in viral-load set point¹¹. However, an essential question in understanding all complex phenotypes (including HIV-1 progression) is what amount of the observed trait variation can be attributed to human genetic background overall (i.e., heritability). For HIV-1 infection in particular, it is unclear whether disease progression is affected by common genetic variants outside the *HLA* and *CCR5* genes. Estimating the total heritability of disease progression and understanding how it may be distributed across different classes of genetic variants (such as common SNPs, rare variants and copy-number variants) can inform on what types of study may have the highest impact. This concept has been elegantly applied in the context of many complex-trait GWAS, often showing that loci present on genotyping arrays (or robustly inferred through genotype imputation) falling below the strict statistical thresholds can collectively explain much of the trait variability not attributable to loci of genome-wide significance^{40,41}. This method measures the correlation between phenotypic similarity and genotypic similarity among unrelated individuals across a large number of common SNPs to quantify narrow-sense heritability⁴¹ (i.e., the heritability explained by additive genetic effects) and can help provide understanding of how regulatory variation functions in common diseases. For example, in a study of 11 common diseases, SNPs in DNase I-hypersensitive sites explained, on average, 79% of the heritability measured by GWAS⁴².

In addition to host genetic effects, viral sequence variation (which itself is influenced by host genetic background) is also known to affect HIV-1 infection outcome. A study of 137 heterosexual transmission pairs from Zambia showed that the viral quasi-species transmitted was selected to have amino acid sequences predicted to confer greater viral fitness *in vivo*⁴³. Several studies have also investigated the contribution of viral sequence variation to viral load⁴⁴. Though these studies vary in methodology, combining evidence across them results in an estimate that 33% (with a 95% confidence interval of 20–46%) of the variation in viral-load set point can be attributed to the viral sequence⁴⁴. This degree of heritability was proposed to be the result of HIV-1 evolution to maximize long-term transmission by optimizing the trade-off between high virulence (short life span of host but high viral transmission) and low virulence (long host life span, low viral transmission)⁴⁵. Importantly, however, models of the rates of within- and between-host viral evolution suggest that immune selection pressure within a given host (for example, escape from HLA recognition) may be the dominant driver of viral sequence variation and can also explain the heritability in viral load measured through the viral sequence⁴⁶. Similarly, the use of HIV-1 sequence variation as the phenotype in a GWAS including paired virus and host genetic data for 1,071 affected individuals showed a markedly stronger influence of *HLA* genes on viral sequence than on viral load⁴⁷. Thus, the higher heritability linked to viral sequence than to the host component (33% and 13%, respectively) suggests that current studies may underestimate the

host influence on disease outcome. Identifying the source of this missing heritability of viral load is of great interest.

Even if the contribution of host and viral genetics were fully accounted for, it is highly unlikely that 100% of the variation in viral load would be explained. There is no doubt that environmental and behavioral factors have a large role in determining disease outcome. In a study of 510 Africans infected with HIV-1, a combination of host genetic and behavioral factors explained up to 44% of the variation in HIV-1 progression⁴⁸, with *HLA* genes making the largest individual contribution. Replication of these results in larger studies and determination of which environmental factors are major contributors to disease progression will be of great interest.

As discussed above, genes encoding class I HLAs exert significant selection pressure on HIV-1. This pressure is due to the presentation of viral epitopes to the cellular immune system, resulting in escape mutations that often lead to a fitness cost^{49–52}. Several studies have focused on the consequences of this selection pressure for viral sequence diversity over time at the population level. A study of 2,800 HIV-1–infected individuals across five continents showed a correlation between the frequency of HLA-induced escape mutations in the viral sequence and the population frequency of the allele driving escape⁵³. This relationship was maintained when individuals carrying the escape-driving allele were excluded, showing an absence of reversion mutations after HIV-1 transmission. The importance of this effect was underscored by the observation that the increase in frequency of an escape mutation away from isoleucine at position 135 in HIV-1 reverse transcriptase has resulted in the loss of the protective effect conferred by *HLA-B*51* in the Japanese population (where that allele is at high frequency)⁵³.

Several other studies have convincingly shown an increase in the frequency of HLA-induced mutations over time in population samples, although the rate at which this is occurring and its estimated impact on the epidemic have differed across studies. Comparison of HIV-1 sequence data from 358 men who have sex with men infected with HIV-1 early in the epidemic to a sample of 382 more recently infected individuals with various transmission risk factors showed a significantly higher sequence diversity in the more recent sample, with approximately twofold higher frequency of mutations in HLA epitopes than in non–HLA-selected sites in individuals not carrying the cognate *HLA* allele⁵⁴. This increase in frequency was most pronounced in escape variants restricted by protective HLA allotypes, such as *HLA-B*57:01* and *HLA-B*27:05*. Similarly, a population from Botswana was found to carry more HLA escape variants than did one from South Africa⁵⁵. This accumulation in the population from Botswana was interpreted to be a result of the longer duration of the local HIV-1 epidemic and resulted in an ablation of the protective effects of *HLA-B*57* and *HLA-B*58:01*. Interestingly, a similar loss of protection was not observed for the ‘second-tier’ protective alleles (*HLA-B*39:10*, *HLA-B*81:01* and *HLA-B*42:01*)⁵⁵, nor was an increase in frequency of escape mutations from these alleles. And there was no difference in impact between populations for alleles that associate with higher viral load. Whether this discrepancy is due to different dynamics of viral escape and compensation for these alleles or some other mechanism is unclear. A lower viral replication capacity in isolates from Botswana suggested a lower overall HIV-1 virulence, as HIV-1 has evolved to preserve the

longevity of its host. However, the observation of decreased virulence due to longer duration of disease in Botswana is in contrast to other studies in general. A summary of published reports investigating trends in viral load and CD4⁺ T cell decline identified 32 such studies and observed a relatively equal division among those reporting increased virulence (11 studies), decreased virulence (9) and stable virulence (12)⁵⁶. A detailed investigation of the subset of 20 studies that included sufficient information for meta-analysis resulted in a summary effect of decreasing baseline CD4⁺ T cell counts and increasing viral load, consistent with an increasing virulence⁵⁶. Data on the rate of disease progression across distinct populations is the ultimate acid test for determining the impact of differential viral evolution on the epidemic. However, such data are difficult to acquire given the widespread use of ART and would be difficult to interpret owing to confounding factors such as environmental effects. Understanding the effects of host genetic pressure on HIV-1 diversity over time and how this may influence the broader pandemic will attract much attention.

The idea of pathogen tolerance, in which the host remains healthy despite a relatively high pathogen burden (in contrast to resistance), has been quantified in a population of HIV-1–infected patients from Switzerland⁵⁷. Tolerance was defined as a reduction in CD4⁺ T cell decline despite high viral loads. Though this study did not show a relationship between tolerance and protective *HLA* alleles, in contrast to the data from Botswana, the authors observed variation in tolerance when *HLA-B* alleles were considered in combination. For example, individuals homozygous for *B*35:01* had the lowest tolerance, whereas those heterozygous for *HLA-B*07:02* and *HLA-B*39:01* had the highest. This suggests that in addition to its impact on viral load, host genetic background has an effect on rate of disease progression that is at least partially independent of viral load.

Leveraging public data to understand HIV-1 biology

In addition to large cohort studies of disease traits, researchers in human genetics have undertaken large-scale population-sequencing studies that give an increasingly comprehensive picture of genetic variation in both coding and noncoding regions of the genome. Studies such as the 1000 Genomes Project⁵⁸, the US National Heart, Lung, and Blood Institute Grand Opportunity Exome Sequencing Project (ESP) (<https://esp.gs.washington.edu/drupal/>) and the Exome Aggregation Consortium (ExAC) (<http://exac.broadinstitute.org>) have compiled variation information for >60,000 individuals. Though these data do not directly inform on HIV-1 disease progression, they provide a rich resource for hypothesis generation, in particular when combined with screens that directly interrogate genes required for HIV-1 replication, such as host and virus protein-protein interaction⁵⁹ or screens for HIV-1 dependency factors⁶⁰.

Although the systematic scans outlined above have identified hundreds of human proteins that either directly interact with HIV-1 or are required for viral replication, variation in their encoding genes does not affect HIV-1 disease outcomes by genome-wide studies. This may reflect the fact that these proteins do not harbor common variants (i.e., they are more conserved) and there is insufficient statistical power to identify them by GWAS without samples sizes much larger than those currently available. This hypothesis could be investigated through identification of individuals harboring protein-changing variants in

these genes and testing of their cellular susceptibility to infection *in vitro*. This strategy might be most fruitful for identifying HIV-1 dependency factors that can be partially or fully deleted in human populations without being lethal and whose loss of function in humans results in reduced HIV-1 susceptibility *in vitro* (Fig. 3). *CCR5* fits both criteria; like *CCR5*, such genes would have immediate utility as drug targets.

Alternatively, variation in host genes that interact with HIV-1 may encode functionally redundant molecules whose modification (or ablation) is easily compensated for by products encoded by other genes. Compensatory variation in related genes may have arisen evolutionarily, such that a deleterious variant in one gene may show significant association with an unlinked compensatory variant in a functionally related gene⁶¹. Given the large number of host proteins required for HIV-1 replication and the increasing understanding of how their encoding genes vary in human populations, association strategies aimed at dissecting the effects of functional polymorphisms in these genes in large patient samples would be of great interest.

Prospects for the future

There are several routes to discovery of novel gene variants that affect HIV-1 susceptibility and/or disease severity. To date, even the largest published GWAS (>2,500 samples)¹¹ has been underpowered to detect the types of modest-effect-size common variants that have been convincingly associated with other complex traits (Fig. 1). To address this, an ongoing effort by the HIV-1 genomics community seeks to combine available genome-wide SNP data from multiple populations in the hopes of uncovering new associated regions and providing a more detailed understanding of potential causal variants in known regions⁴. Additionally, most genome-wide studies completed have focused on populations of European ancestry. Investigations in other populations, in particular those of African ancestry, have the potential to increase power to detect associated variants (when combined with data from European populations) and to uncover population-specific variants not observed in European samples. Such studies will be facilitated by newer, comprehensive genotyping arrays that are more representative of global diversity. Along these lines, most GWAS data have been collected in men, even though men and women differ in terms of viral-load levels that lead to disease progression^{62,63} and there is evidence to suggest differential immune responses to HIV-1 between the sexes⁶⁴. The genetic basis for these differences is not known and should be investigated.

Although genetic studies using genotyping arrays provide a comprehensive, genome-wide survey of common SNPs, their success relies on the principle that typed variants tag (through linkage disequilibrium) underlying causal variants. This is not always the case. Interestingly, the first GWAS of HIV-1 control observed that there were no individual SNPs that strongly tagged *HLA-B*27:05*, which is well known to reduce patient viral load. Many HLA variants are poorly tagged by individual SNPs¹³. Similarly, the *CCR5* $\Delta 32$ polymorphism is not captured on commercial arrays and, due to the underlying haplotype structure, cannot be accurately imputed⁴. In addition to increasing power through larger samples, studies focusing on genetic effects not observable through GWAS should be a priority. These include structural variants (such as insertion or deletion polymorphisms and

gene duplications), somatic rearrangement (such as TCR recombination), epigenetic modifications and rare variants (<1% frequency). In the near term, exome sequencing studies (i.e., targeted sequencing of the ~1.5% of the genome that is protein coding) should be pursued in population cohorts with disease progression measurements and extreme phenotypes. These studies have the potential to uncover low-frequency functional variants that affect gene function individually or in combination within a single gene or across functionally related genes⁶⁵ Though discovery of genetic associations through this method will require very large sample sizes⁶⁶, such studies may uncover functional polymorphisms in known HIV-1–interacting genes whose impact could be further investigated in laboratory studies.

Additionally, investigating the role of interactions between variants in two or more genes (such as the one described above between *HLA-C* and *MIR148A*) in affecting disease may provide insight. Though the multiple testing burden incurred through exhaustive pairwise genome-wide interaction screens makes discovery of subtle effects unlikely, targeted strategies starting with known regions may alleviate this to some degree. However, it will be crucial for these studies to appropriately control for confounders such as cryptic relatedness and population structure (which are easily corrected for by GWAS) to avoid reporting the types of spurious associations contributed by many candidate-gene studies.

Consideration of additional phenotypes, either in existing cohorts or through generation of new resources, that affect other aspects of HIV-1 infection or act as intermediate phenotypes to susceptibility or disease progression is warranted. These could include additional markers of disease progression or immune parameters, such as levels of immune activation, that may influence susceptibility and can be tested in a hypothesis-driven manner involving specific assays. Although this approach has been underutilized, one example includes the identification of a variant in the gene encoding IRF7, a key regulator of interferon- α (IFN- α), that associates with the amount of IFN- α production by plasmacytoid dendritic cells in response to HIV-1 *in vitro*⁶⁷.

Given the current focus on HIV-1 cure research, another potential avenue for investigation is the genetic influence on the dynamics of the latent reservoir and reactivation. Interestingly, two studies that identified post-treatment controllers (i.e., individuals in whom no viral rebound is detected after cessation of therapy) did not observe an over-representation of protective *HLA* alleles (*HLA-B*57:01* and *HLA-B*27:05*) in these individuals^{68,69}. Though these patients are rare, this observation suggests that host genetic factors influencing reactivation may, if they exist, lie outside of regions known to influence viral replication in untreated patients.

Finally, nesting genetic studies in clinical trials of novel therapies or prevention strategies could contribute to their overall success, as genetic variants can have a large impact on treatment outcome. For example, variants within the *IL28B* gene have been associated with a twofold increase in response to therapy for hepatitis C infection⁷⁰. Similarly, given that the goal of many ongoing vaccine trials is to elicit effective cellular immune responses, understanding the background *HLA* allele frequencies (as was done for the RV144 vaccine trial)⁷¹ could help tailor vaccines to populations or even individuals. Realization of the long-

term goal of completely understanding host genetic factors that affect HIV-1 pathogenesis will require novel methods and lines of investigation. The role of well-phenotyped patient cohorts with genetic consent in achieving this goal cannot be overstated.

ACKNOWLEDGMENTS

We thank A. Bashirova and I. Bartha for helpful suggestions. This project was funded in whole or in part with federal funds from the Frederick National Laboratory for Cancer Research, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government. This research was supported in part by the Intramural Research Program of the National Institutes of Health, Frederick National Laboratory, Center for Cancer Research.

References

1. Horton RE, McLaren PJ, Fowke K, Kimani J & Ball TB Cohorts for the study of HIV-1-exposed but uninfected individuals: benefits and limitations. *J. Infect. Dis* 202 (suppl. 3), S377–S381 (2010). [PubMed: 20887228]
2. An P & Winkler CA Host genes associated with HIV/AIDS: advances in gene discovery. *Trends Genet* 26, 119–131 (2010). [PubMed: 20149939]
3. Lane J et al. A genome-wide association study of resistance to HIV infection in highly exposed uninfected individuals with hemophilia A. *Hum. Mol. Genet* 22, 1903–1910 (2013). [PubMed: 23372042]
4. McLaren PJ et al. Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. *PLoS Pathog* 9, e1003515 (2013). [PubMed: 23935489]
5. Gurdasani D et al. A systematic review of definitions of extreme phenotypes of HIV control and progression. *AIDS* 28, 149–162 (2014). [PubMed: 24149086]
6. Kaslow RA et al. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat. Med* 2, 405–411 (1996). [PubMed: 8597949]
7. Carrington M et al. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 283, 1748–1752 (1999). [PubMed: 10073943]
8. Gao X et al. Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N. Engl. J. Med* 344, 1668–1675 (2001). [PubMed: 11386265]
9. Keet IP et al. Consistent associations of HLA class I and II and transporter gene products with progression of human immunodeficiency virus type 1 infection in homosexual men. *J. Infect. Dis* 180, 299–309 (1999). [PubMed: 10395843]
10. Martin MP & Carrington M Immunogenetics of HIV disease. *Immunol. Rev* 254, 245–264 (2013). [PubMed: 23772624]
11. Fellay J et al. Common genetic variation and the control of HIV-1 in humans. *PLoS Genet* 5, e1000791 (2009). [PubMed: 20041166]
12. Pereyra F et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 330, 1551–1557 (2010). [PubMed: 21051598]
13. Carrington M, Bashirova AA & McLaren PJ On stand by: host genetics of HIV control. *AIDS* 27, 2831–2839 (2013). [PubMed: 25119688]
14. Pyo CW et al. Recombinant structures expand and contract inter and intragenic diversification at the KIR locus. *BMC Genomics* 14, 89 (2013). [PubMed: 23394822]
15. Fellay J et al. A whole-genome association study of major determinants for host control of HIV-1. *Science* 317, 944–947 (2007). [PubMed: 17641165] This is the first genome-wide association study on an infectious disease trait. This study demonstrates that variation in the MHC region is the primary host genetic influence on HIV outcome in an unbiased, genome-wide screen.
16. Migueles SA et al. *HLA B*5701* is highly associated with restriction of virus replication in a subgroup of HIV-infected long-term nonprogressors. *Proc. Natl. Acad. Sci. USA* 97, 2709–2714 (2000). [PubMed: 10694578]

17. Jia X et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* 8, e64683 (2013). [PubMed: 23762245]
18. Dunstan SJ et al. Variation at *HLA-DRB1* is associated with resistance to enteric fever. *Nat. Genet* 46, 1333–1336 (2014). [PubMed: 25383971]
19. Patsopoulos NA et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet* 9, e1003926 (2013). [PubMed: 24278027]
20. Okada Y et al. Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. *Am. J. Hum. Genet* 95, 162–172 (2014). [PubMed: 25087609]
21. Raychaudhuri S et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet* 44, 291–296 (2012). [PubMed: 22286218]
22. Stranger BE et al. Population genomics of human gene expression. *Nat. Genet* 39, 1217–1224 (2007). [PubMed: 17873874]
23. Thomas R et al. HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nat. Genet* 41, 1290–1294 (2009). [PubMed: 19935663]
24. Kaufman J & Salomonsen J The “minimal essential MHC” revisited: both peptide-binding and cell surface expression level of MHC molecules are polymorphisms selected by pathogens in chickens. *Hereditas* 127, 67–73 (1997). [PubMed: 9420472]
25. Koch M et al. Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. *Immunity* 27, 885–899 (2007). [PubMed: 18083574]
26. Corrah TW et al. Reappraisal of the relationship between the HIV-1-protective single-nucleotide polymorphism 35 kilobases upstream of the HLA-C gene and surface HLA-C expression. *J. Virol* 85, 3367–3374 (2011). [PubMed: 21248048]
27. Pulit SL, Voight BF & de Bakker PI Multiethnic genetic association studies improve power for locus discovery. *PLoS ONE* 5, e12600 (2010). [PubMed: 20838612]
28. Apps R et al. Influence of HLA-C expression level on HIV control. *Science* 340, 87–91 (2013). [PubMed: 23559252] The authors determine the cell surface expression dynamics of all common classical HLA-C allotypes. Accounting for linkage disequilibrium with known allelic effects at HLA-A and HLA-B, they demonstrate an independent impact of HLA-C expression level on HIV control.
29. Kulkarni S et al. Differential microRNA regulation of HLA-C expression and its association with HIV control. *Nature* 472, 495–498 (2011). [PubMed: 21499264]
30. Kulkarni S et al. Genetic interplay between HLA-C and MIR148A in HIV control and Crohn disease. *Proc. Natl. Acad. Sci. USA* 110, 20705–20710 (2013). [PubMed: 24248364]
31. Blais ME et al. High frequency of HIV mutations associated with HLA-C suggests enhanced HLA-C-restricted CTL selective pressure associated with an AIDS-protective polymorphism. *J. Immunol* 188, 4663–4670 (2012). [PubMed: 22474021]
32. Kiepiela P et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* 432, 769–775 (2004). [PubMed: 15592417]
33. Kiepiela P et al. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat. Med* 13, 46–53 (2007). [PubMed: 17173051]
34. Petersdorf EW et al. HLA-C expression levels define permissible mismatches in hematopoietic cell transplantation. *Blood* 124, 3996–4003 (2014). [PubMed: 25323824]
35. Thomas R et al. A novel variant marking HLA-DP expression levels predicts recovery from hepatitis B virus infection. *J. Virol* 86, 6979–6985 (2012). [PubMed: 22496224]
36. Duggal P et al. Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts. *Ann. Intern. Med* 158, 235–245 (2013). [PubMed: 23420232]
37. Wisemann WT et al. Association of Parkinson disease with structural and regulatory variants in the HLA region. *Am. J. Hum. Genet* 93, 984–993 (2013). [PubMed: 24183452]
38. Reits EA et al. Radiation modulates the peptide repertoire, enhances MHC class I expression, and induces successful antitumor immunotherapy. *J. Exp. Med* 203, 1259–1271 (2006). [PubMed: 16636135]

39. Faroudi M et al. Lytic versus stimulatory synapse in cytotoxic T lymphocyte/target cell interaction: manifestation of a dual activation threshold. *Proc. Natl. Acad. Sci. USA* 100, 14145–14150 (2003). [PubMed: 14610278]
40. Stahl EA et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet* 44, 483–489 (2012). [PubMed: 22446960]
41. Yang J et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet* 42, 565–569 (2010). [PubMed: 20562875] Missing heritability is the proportion of trait variability known to be attributable to genetic effects that has not been explained by GWAS. Using human height as a model, the authors describe a method for quantifying the heritability explained by all common SNPs and show that this accounts for a large fraction of the missing heritability.
42. Gusev A et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet* 95, 535–552 (2014). [PubMed: 25439723]
43. Carlson JM et al. HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science* 345, 1254031 doi:10.1126/science.1254031 (2014). [PubMed: 25013080]
44. Fraser C et al. Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. *Science* 343, 1243727 doi:10.1126/science.1243727 (2014). [PubMed: 24653038]
45. Fraser C, Hollingsworth TD, Chapman R, de Wolf F & Hanage WP Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis. *Proc. Natl. Acad. Sci. USA* 104, 17441–17446 (2007). [PubMed: 17954909]
46. van Dorp CH, van Boven M & de Boer RJ Immuno-epidemiological modeling of HIV-1 predicts high heritability of the set-point virus load, while selection for CTL escape dominates virulence evolution. *PLoS Comput. Biol* 10, e1003899 (2014). [PubMed: 25522184]
47. Bartha I et al. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *Elife* 2, e01123 (2013). [PubMed: 24171102] In a combined analysis of host and pathogen genetics, the authors use HIV sequence diversity as phenotype in multiple GWAS. The authors demonstrated that using sites of viral escape as an intermediate phenotype was more powerful than using clinical markers of disease progression to detect HLA associations.
48. Mackelprang RD et al. Host genetic and viral determinants of HIV-1 RNA set-point among HIV-1 seroconverters from sub-Saharan Africa. *J. Virol* 89, 2104–2111 (2015). [PubMed: 25473042]
49. Frater AJ et al. Effective T-cell responses select human immunodeficiency virus mutants and slow disease progression. *J. Virol* 81, 6742–6751 (2007). [PubMed: 17409157]
50. Leslie AJ et al. HIV evolution: CTL escape mutation and reversion after transmission. *Nat. Med* 10, 282–289 (2004). [PubMed: 14770175]
51. Martinez-Picado J et al. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *J. Virol* 80, 3617–3623 (2006). [PubMed: 16537629]
52. Schneidewind A et al. Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. *J. Virol* 81, 12382–12393 (2007). [PubMed: 17804494]
53. Kawashima Y et al. Adaptation of HIV-1 to human leukocyte antigen class I. *Nature* 458, 641–645 (2009). [PubMed: 19242411] In a study of HIV diversity in multiple geographic regions, the authors show a correlation between frequency of viral escape mutations and the restricting HLA alleles. This adaptation resulted in the loss of protective effect of *HLA-B*51* in a Japanese sample over time.
54. Cotton LA et al. Genotypic and functional impact of HIV-1 adaptation to its host population during the North American epidemic. *PLoS Genet* 10, e1004295 (2014). [PubMed: 24762668]
55. Payne R et al. Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. *Proc. Natl. Acad. Sci. USA* 111, E5393–E5400 (2014). [PubMed: 25453107]
56. Herbeck JT et al. Is the virulence of HIV changing? A meta-analysis of trends in prognostic markers of HIV disease progression and transmission. *AIDS* 26, 193–205 (2012). [PubMed: 22089381]
57. Regoes RR et al. Disentangling human tolerance and resistance against HIV. *PLoS Biol* 12, e1001951 (2014). [PubMed: 25226169]

58. Abecasis GR et al. A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073 (2010). [PubMed: 20981092]
59. Jäger S et al. Global landscape of HIV–human protein complexes. *Nature* 481, 365–370 (2012).
60. Zhu J et al. Comprehensive identification of host modulators of HIV-1 replication using multiple orthologous RNAi reagents. *Cell Reports* 9, 752–766 (2014). [PubMed: 25373910]
61. Akashi H, Osada N & Ohta T Weak selection and protein evolution. *Genetics* 192, 15–31 (2012). [PubMed: 22964835]
62. Farzadegan H et al. Sex differences in HIV-1 viral load and progression to AIDS. *Lancet* 352, 1510–1514 (1998). [PubMed: 9820299]
63. Sterling TR et al. Sex differences in longitudinal human immunodeficiency virus type 1 RNA levels among seroconverters. *J. Infect. Dis* 180, 666–672 (1999). [PubMed: 10438353]
64. Kontinen YT, Hanninen A & Fuellen G Plasmacytoid dendritic cells, Janus-faced sentinels: progesterone, guilty or innocent? *Immunotherapy* 1, 929–931 (2009). [PubMed: 20635911]
65. Kiezun A et al. Exome sequencing and the genetic basis of complex traits. *Nat. Genet* 44, 623–630 (2012). [PubMed: 22641211]
66. MacArthur DG et al. Guidelines for investigating causality of sequence variants in human disease. *Nature* 508, 469–476 (2014). [PubMed: 24759409]
67. Chang J et al. Polymorphisms in interferon regulatory factor 7 reduce interferon-alpha responses of plasmacytoid dendritic cells to HIV-1. *AIDS* 25, 715–717 (2011). [PubMed: 21297422]
68. Lodi S et al. Immunovirologic control 24 months after interruption of antiretroviral therapy initiated close to HIV seroconversion. *Arch. Intern. Med* 172, 1252–1255 (2012). [PubMed: 22826124]
69. Sáez-Cirión A et al. Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI Study. *PLoS Pathog* 9, e1003211 (2013). [PubMed: 23516360]
70. Ge D et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 461, 399–401 (2009). [PubMed: 19684573]
71. Prentice HA et al. HLA class I, KIR, and genome-wide SNP diversity in the RV144 Thai phase 3 HIV vaccine clinical trial. *Immunogenetics* 66, 299–310 (2014). [PubMed: 24682434]
72. Deeks SG & Walker BD Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity* 27, 406–416 (2007). [PubMed: 17892849]
73. Bartha I, McLaren PJ, Ciuffi A, Fellay J & Telenti A GuavaH: a compendium of host genomic data in HIV biology and disease. *Retrovirology* 11, 6 (2014). [PubMed: 24428872]

Figure 1.

Opportunities and obstacles for human genetic studies of HIV-1 outcomes. **(a)** Genetic studies of pre-infection phenotypes have focused on susceptibility by comparing samples from HIV-1-infected people to the general population or to cohorts of high-risk exposed-uninfected individuals. Studies of disease progression in infected groups have commonly used set-point viral load, rate of CD4+ T cell decline and time to AIDS or death as markers of progression. Other phenotypes with potential impact on susceptibility to infection or severity of disease—including degree of immune activation, magnitude of acute viremia and the size of the latent reservoir—may also be informative. **(b)** A commonly proposed model of genetic architecture of disease traits in which variant frequency is inversely correlated with variant impact. Common variants (>1% frequency) of moderate to low effect size can be detected in large patient samples by GWAS using genotyping arrays. Rare variants (<1%) can be detected by sequencing studies that provide base-pair-level resolution. **(c)** Curves showing the sample size required for 80% power to detect common (5%, 10% and 30%) and rare (1% and 0.5%) genetic variants across a range of effect sizes (odds ratio) at genome-wide significance ($P < 5 \times 10^{-8}$). Curves were modeled on case-control studies assuming an equally distributed sample (n cases = n controls) and a trait prevalence of 5% (the approximate population proportion of viremic controllers)⁷². The diamonds indicate the properties of the well-described effects of *HLA-B*57:01* (5% population frequency) and *CCR5* Δ32 heterozygosity (10% population frequency) with odds ratios as reported in HIV-1 controllers¹². The dashed line is set at a sample size of 2,500, the largest published GWAS of HIV-1 progression phenotypes to date¹¹. Variants affecting HIV-1 progression with properties falling above the dashed line would not have been detected by current studies.

Figure 2. Approaches to supporting a direct functional impact of HLA-C expression on HIV-1 disease outcome. **(a)** Higher expression of HLA-C associates with protection against HIV-1 in African and European Americans despite distinct *HLA-C* allele frequencies and linkage disequilibrium patterns with *HLA-B* in these two groups. **(b)** *MIR148A* insertion and deletion polymorphism associates with different expression levels, affecting the extent to which *HLA-C* is downregulated and degree of HIV-1 control in individuals carrying an intact miR-148a binding site in the *HLA-C* 3' UTR. The *MIR148A* variant has no effect on HIV-1 control in those carrying *HLA-C* alleles that escape miR-148a regulation. **(c)** Higher expression of HLA-C on HIV-1-infected targets results in a greater frequency of CTL responses to specific HLA-C-restricted peptides.

Figure 3.

HIV-1-dependency genes identified by genome-wide short interfering RNA (siRNA) screens are relatively conserved. **(a)** Selection as measured by the ratio of nonsynonymous substitutions per nonsynonymous site (d_N) to synonymous substitutions per synonymous site (d_S) across primate species for 99 high-confidence HIV-1 dependency factors (HDFs)⁶⁰ (dashed line) and a background set of 16,618 genes (solid line). HDFs are relatively depleted of nonsynonymous variation (Wilcoxon rank sum $P = 0.01$), suggesting conservation across species. **(b)** Example plot summarizing human variation across the HDF PARP9 in a set of ~6,300 individuals from the NHLBI ESP. The protein background (gray box) is plotted with vertical bars indicating amino acid location of protein-modifying variants (blue, missense; red, stop-gain (nonsense); yellow, frameshift). Amino acid locations of stop-gain and frameship variants are given below the protein background. Top, bars indicate the frequency of each plotted variant (green, rare variants; purple, common variants). This variation could be exploited to test the importance of the encoding gene in HIV-1 replication *ex vivo*. Protein plot was generated by the GuavaH web tool⁷³. MAF, minor allele frequency.