Review

Host genetics influences on HIV type-1 disease

Jacques Fellay*

1Center for Human Genome Variation, Institute for Genome Sciences & Policy, Duke University, Durham, NC, USA

*Corresponding author: e-mail: jcf20@duke.edu

HIV host genetic studies seek to describe as comprehensively as possible the effect of human genetic variation on the individual response to HIV type-1 (HIV-1) infection. Many associations between specific gene variants and HIV-1 disease outcomes have been reported over the past 15 years. Although most of them have yet to be confirmed or have been proven false-positives, the identification of several definitive genotype–phenotype associations has shed new light on HIV-1 pathogenesis. This review discusses these results in the context of the new genome-wide approaches that now make it possible to globally assess the influence of the host genome on HIV-1-related outcomes.

Introduction

Differences between individuals exposed to HIV type-1 (HIV-1) virus have been observed since the early days of the current pandemic. Susceptibility to HIV-1 infection and natural history of the disease are highly variable, owing to the complex interplay between the virus, its human host and the environment. Over the past 20 years, research in HIV-1 host genetics has unravelled a series of human gene variants that modulate the response to retroviral exposure [1–3], thereby partially explaining why some individuals remain uninfected even when repeatedly exposed to HIV-1, or why a subset of infected patients are able to maintain a normal level of immunity after years of infection.

Most of the discoveries to date resulted from candidate gene studies, in which allelic variants have been analysed in genes that were known or suspected to play a role in HIV-1 pathogenesis and immune response. As a consequence, the genetic markers relevant to HIV-1 disease that have been identified so far are related to genes that can broadly be classified into one of two categories: host genes that are implicated in HIV-1 life cycle, from entry into the target cells to the different intracellular steps that are required for viral replication and propagation; and immune-related genes, coding for canonical innate and adaptive immune response factors, as well as for proteins involved in immune regulation and in specific antiretroviral defence mechanisms (Table 1).

This review will summarize the known host genetic associations with HIV-1 outcomes and put this knowledge in the context of the current genomic era. Indeed, the availability of genome-wide approaches represent a change in paradigm for complex trait genetic studies: global host influences on HIV-1 can now be assessed in single experiments, and individual contributions of numerous genetic variants can be ranked and compared to get a comprehensive view of the effect of human genomic variation on HIV-1 infection [4].

HIV-1 life cycle

Entry: chemokine receptors

The only genetic variants that have been consistently associated with protection against HIV-1 acquisition are those affecting the CC chemokine receptor 5 (CCR5) gene, which encodes the main coreceptor for macrophage-tropic (or R5) strains of HIV-1, normally expressed on CD4+ T-cells. A number of polymorphisms located in the coding and the promoter regions of the CCR5 gene have been found to be associated with HIV-1-related outcomes. Most notably, a 32 base pair deletion (Δ32) in the CCR5 coding region results in the production of a truncated protein that is not expressed on the cell membrane. The CCR5 Δ32 variant affords protection against infection by R5 viruses in individuals who are homozygous for this allele and slows down disease progression in heterozygotes [5–8]. Several other uncommon genetic variants of the CCR5 coding region have been described in individuals from various human populations [9–12], but their effect on HIV-1 coreceptor function has not been completely established, except for the rare CCR5 303T>A change (also referred to as...
m303 or C101X) that introduces a premature stop codon and thus prevents functional coreceptor expression. This mutation was able to entirely block entry of R5 HIV-1 into its target cell \textit{in vitro} [11]; the combination of m303 heterozygosity and the \(\Delta 32\) deletion on the other allele confers resistance to HIV-1 infection [13]. Both \(\Delta 32\) and m303 are found exclusively in patients of European ancestry, with minor allele frequencies of 10% (corresponding to a 1% frequency of homozygous individuals) and 0.2%, respectively.

Promoter variants of the \(CCR5\) gene that influence the pace of progression to AIDS have been reported. By modulating the expression of \(CCR5\), they can have an effect on cellular susceptibility [14]. The so-called HHE haplotype, which carries the previously reported P1 variant, is associated with increased \(CCR5\) expression and faster disease progression [15–17].

A valine to isoleucine change in the \(CCR2\) gene (CCR2-64I variant) has been shown to delay progression to AIDS in a dominant way [18]. The frequency of CCR2-64I does not vary much between populations, with an average frequency of 10–20%. As CCR2 is only a minor coreceptor that HIV-1 does not directly use for host cell entry \textit{in vivo}, and because the V64I substitution is conservative, it is not entirely clear how the 64I allele exerts its protective effect. One proposed explanation is that the 64I allele acts on CCR2 splicing variants that result in down-modulation of \(CCR5\) expression [19], but other studies failed to show an association between CCR2 variation and \(CCR5\) expression levels [20,21].

Variants of the chemokine receptors CX3CR1, CXCR1 (IL8RA) and CXCR6 have been reported to be associated with various HIV-1 outcomes [22–26], but none of these associations could yet be convincingly replicated. More recently, a report suggested that a variant in the Duffy antigen/receptor for chemokines (DARC) modulates HIV susceptibility [27]. DARC is a non-specific chemokine receptor that binds many inflammatory chemokines, but lacks the ability to signal upon ligand binding; it is also the receptor for \(Plasmodium vivax\) malaria. Homozygosity for the null allele \((-46C)\) was therefore strongly enriched in regions affected by \(P. vivax\)

### Table 1. Genetic variants that have been reported to be associated with differences in susceptibility to HIV-1 infection, viral control and/or disease progression

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene function in HIV-1 disease</th>
<th>Variant</th>
<th>dbSNP ID</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CCR5)</td>
<td>Chemokine receptor</td>
<td>(\Delta 32)</td>
<td>rs333</td>
<td>[5–8]</td>
</tr>
<tr>
<td>P1 variant, including in HHE haplotype</td>
<td>rs1800560</td>
<td>[11,13]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(C101X)</td>
<td>rs1799897</td>
<td>[15–17]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(V64I)</td>
<td>rs1799864</td>
<td>[18]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T280M)</td>
<td>rs3732378</td>
<td>[22–24]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(92T)</td>
<td>rs16858811</td>
<td>[26]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Promoter) -46T-C</td>
<td>rs2814778</td>
<td>[27]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Haplotypes)</td>
<td>rs2107538, rs2280788, rs2280789</td>
<td>[30–34]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CCR3)</td>
<td>rs450C&gt;T</td>
<td>rs1719134</td>
<td>[32,35]</td>
<td></td>
</tr>
<tr>
<td>(CCR3L1)</td>
<td>Copy number</td>
<td>NA</td>
<td>[36,37]</td>
<td></td>
</tr>
<tr>
<td>(CCR2)</td>
<td>rs1024610, rs2857657</td>
<td>[38–41]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(SDF-1)</td>
<td>rs1801157</td>
<td>[16,42–46]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Promoter) haplotypes</td>
<td>rs2292179, rs1395319</td>
<td>[44,50]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1604C&gt;G, 1650A&gt;G)</td>
<td>rs8177826, rs6850</td>
<td>[44,53,54]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(H34Y) (127C&gt;T), (R116Q) (407G&gt;A)</td>
<td>rs3740996, rs10838525</td>
<td>[55–58]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(H188R) (557A&gt;G)</td>
<td>rs8177832</td>
<td>[59–61]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(HLA-B)</td>
<td>HLA-B*57</td>
<td>NA</td>
<td>[63–65]</td>
<td></td>
</tr>
<tr>
<td>(HLA-B)</td>
<td>HLA-B*27</td>
<td>NA</td>
<td>[68,69]</td>
<td></td>
</tr>
<tr>
<td>(HLA-B)</td>
<td>HLA-B35Px</td>
<td>NA</td>
<td>[70,71]</td>
<td></td>
</tr>
<tr>
<td>35 kb T&gt;C</td>
<td>rs59264942</td>
<td>[65]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-52G&gt;A</td>
<td>rs1799946</td>
<td>[71,77]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(HLA) class I</td>
<td>HLA homozygosity</td>
<td>NA</td>
<td>[88–90]</td>
<td></td>
</tr>
<tr>
<td>(MBL2)</td>
<td>(Promoter) -52G&gt;A</td>
<td>rs8030737, rs1800450, rs235G&gt;A</td>
<td>[91–95]</td>
<td></td>
</tr>
<tr>
<td>(CD209) (DC-SIGN)</td>
<td>Multiple</td>
<td>rs1800451</td>
<td>[96–98]</td>
<td></td>
</tr>
</tbody>
</table>

\(DARC\), Duffy antigen/receptor for chemokines; \(dbSNP\) ID, single nucleotide polymorphism database identification hosted at the National Center for Biotechnology Information; \(DEFB1\), \(\beta\)-defensin 1; HIV-1, HIV type-1; \(HLA\), human leukocyte antigen; \(MBL2\), mannose-binding lectin 2; \(MIP\), macrophage inflammatory protein; \(NA\), not applicable; \(PPIA\), peptidyl propyl isomerase A; \(RANTES\), regulated upon activation, normal T-cell expressed and secreted protein; \(SDF-1\), stroma-derived factor 1; \(TSG101\), tumour susceptibility gene 101.
vivax: this allele is nearly fixed in sub-Saharan African populations, whereas the functional allele (-46T) is fixed in European populations. He et al. [27] reported that HIV-1 infected African-Americans have a frequency of the null homozygous genotype of 70%, whereas HIV-1 negative individuals have a frequency of 60%. Therefore, they argued that DARC -46C/C increases susceptibility to infection. Interestingly, in the same study, the null allele (-46C) was associated with slower disease progression if infection occurred.

Entry: chemokines
Chemokines are natural ligands for the same receptors that HIV-1 uses to enter the cells. They can therefore have an effect on HIV-1 entry into host target cells by two means: they compete with the virus for coreceptor binding and/or they reduce the expression of the coreceptors on the cell surface by inducing their internalization. Regulated upon activation, normal T-cell expressed and secreted protein (RANTES), macrophage inflammatory proteins (MIP1α and MIP1β), and stroma-derived factor (SDF-1) are examples of human chemokines that have polymorphisms in their coding genes that have been reported to play a role in differential susceptibility to HIV-1. The CC chemokine, RANTES, encoded by the CCL5 gene, potently inhibits HIV-1 replication in vitro [28]. In vivo, decreased expression of RANTES was shown to accelerate progression, whereas the opposite effect was observed with RANTES up-regulation [29]. Several polymorphisms located in regulatory regions of CCL5 have been grouped in haplotypes that appear to modulate gene expression and, as a consequence, associated with differences in HIV-1 susceptibility and disease progression [30–34].

Several other CC chemokine genes have been scrutinized in many HIV-1 host genetic studies. Polymorphisms in the coding and non-coding regions of the CCL3 gene (MIP1α) have been shown to be associated with both resistance and progression [32,35]. More recently, a detrimental effect was associated with low copy numbers of the CCL3L1 gene (MIP1αP) [36,37], presumably because of modulation of CCL3L1 expression. Analyses of the CCL2-CCL7-CCL11 gene cluster revealed variants that lead to differences in susceptibility to infection [38–40]. Here again, an increase in chemokine expression is proposed as the causal mechanism. Notably, a CCL2 promoter polymorphism, -2578G, that has been associated with a reduced risk of acquiring HIV-1 was shown to increase CCL2 protein expression caused by differential binding of the transcription factor interferon regulatory factor 1 to the polymorphic region [41]. SDF-1 is the natural ligand for CXCR4, the coreceptor for HIV-1 X4 strains. Many publications have looked at variation of the SDF-1 gene (CXCL12) and its relevance to HIV-1 pathogenesis. In particular, a variant located in the 3′-untranslated region of the gene (SDF-1 3′A) was reported to have various effects on progression [16,42–45]. However, no convincing evidence could finally emerge from a large meta-analysis [46] and the importance of SDF-1 variation is still controversial.

Intracellular life cycle
Inside its target cell, HIV-1 interacts with numerous host proteins: some act as antiviral factors (see Intrinsic antiviral factors), but most of them are HIV-1 dependency factors (HDFs), necessary for sustaining viral replication. HIV-1 has the capacity to hijack numerous human proteins in order to successfully complete its life cycle, as demonstrated in several recent large-scale small interfering RNA screens [47–49]. Because human genes that act as HDFs represent ideal candidate genes, many of them have been investigated in host genetic studies. However, only two such genes have variants that were shown to play a role in the modulation of HIV-1 pathogenesis and were confirmed in subsequent studies: tumour susceptibility gene 101 (TSG101) and peptidyl propyl isomerase A (PPIA; encoding the cyclophilin A protein).

The TSG101 protein interacts with the P6 product of HIV-1 Gag and is known to be crucial in the budding process of new viral particles from the plasma membrane of infected cells: haplotypes constructed from two polymorphisms located in the 5′-region of the gene (-183T>C and +181A>C) were shown to influence multiple outcomes, including viral load and disease progression measured by CD4+ T-cell decrease [44,50]. The cyclophilin A protein is incorporated into the HIV-1 virion as a result of an interaction with the viral capsid protein: the mechanism by which cyclophilin A is able to enhance HIV-1 infection is still largely unknown, although it has been suggested that it is involved in the uncoating of the viral core [51] and can also act as a cofactor for the anti-HIV-1 TRIM5 protein [52]. Several variants located in regulatory regions of the PPIA gene have been shown to influence CD4+ T-cell depletion and possibly susceptibility to infection [44,53,54].

Intrinsic antiretroviral factors
In addition to host proteins that are necessary for HIV-1 to efficiently replicate, the human genome encodes molecules that have an antiretroviral function. Most notably, TRIM5α and APOBEC3G have both been shown to be potent inhibitors of retroviral replication.

TRIM5α restricts HIV-1 at a post-entry, pre-integration stage in the viral life cycle by recognizing the incoming retroviral capsid and promoting its premature disassembly. This inhibition process is species-specific.
In fact, the capacity of human TRIM5α to inhibit HIV-1 replication is limited, whereas its Simian orthologue restricts the human virus very efficiently. In vitro experiments have shown that several polymorphisms in the human TRIM5α gene correlate with differences in antiretroviral potency, but most of the in vivo data suggest that common human variants of TRIM5α have no effect or only modest influence on HIV-1 disease outcomes [55–58].

APOBEC3G was first identified as a cellular factor able to restrict replication of HIV-1 viruses lacking the accessory protein Vif [59]. Its cytidine deaminase activity affects viral replicative capacity by introducing G to A hypermutations in the HIV-1 DNA. An H186R coding change observed mainly in African populations was reported to be associated with accelerated progression to AIDS, although the in vitro antiviral activity of the 186R enzyme was not inferior to that of the common H186 variant [60]. A thorough analysis of APOBEC3G polymorphisms in Caucasians did not show any association with HIV-1 control [61]. Here again, larger studies in ethnically-controlled populations are warranted. A recent report sheds further light on the role of the APOBEC3 gene locus in retroviral pathogenesis: a human APOBEC3H allele encoding a protein with potent activity against retroviruses and retrotransposons was identified in African populations [62].

**Immunity**

**Human leukocyte antigens**

The most prominent and consistent associations identified in HIV-1 host genetic studies are those between human leukocyte antigen (HLA) genes and disease outcomes. Three genes (HLA-A, HLA-B and HLA-C) encode the HLA class I proteins. Fundamental to the adaptive immune response, HLA molecules are expressed at the cell surface where they present antigenic epitopes and, notably, viral peptides to CD8+ T-cells, thereby initiating a cytotoxic T-cell response.

The HLA class I genes present an extreme allelic diversity and are in fact the most polymorphic genes in the human genome. Once a particular individual gets infected by HIV-1, the potency of the elicited immune response depends on the retroviral epitopes that their HLA alleles are able to present to CD8+ T-cells. The HLA allele that is most consistently associated with potent control of HIV-1 is B*57 [63–65], with B*5701 observed almost exclusively in Caucasians and B*5703 mostly seen in individuals of African ancestry. Of note, a single nucleotide polymorphism that is a proxy for HLA-B*5701 showed the strongest association with HIV-1 viral control or long-term non-progression in all three genome-wide association studies published to date in the HIV-1 field [65–67]. There is also clear epidemiological and functional evidence for effective restriction of HIV-1 by HLA-B*27 [68,69]. By contrast, HLA-B*35Px (including B*3502, B*3503, B*3504 and B*3501) is associated with faster progression to AIDS [70,71]. Several haplotypes in the major histocompatibility complex (MHC) and various HLA supertypes have been implicated in HIV-1 control [72–76], but most of the associations are likely to be a result of individual alleles that are included in these groups and to the long-range linkage disequilibrium structure of the MHC region. Thus, HLA molecules have the ability to present different HIV-1 epitopes, which result in variable restriction of the virus by the CD8+ T-cells. Yet, this is not the only way that the MHC region contributes to the interindividual differences observed in HIV-1 pathogenesis: homozygosity for HLA-A, HLA-B and/or HLA-C reduces the repertoire of antigen presentation, limits the number of epitopes recognized by cytotoxic T-lymphocytes and results in a faster disease progression [71,77]. Additionally, although most of the associations with HIV-1 outcomes identified in the MHC region involve the HLA-B gene, there is some evidence that the importance of genetic variation in HLA-C might have been underestimated [78]. A genome-wide association study of determinants of HIV-1 viraemia in seroconverters identified a polymorphism in the 5′ region of HLA-C that associates with both viral control and expression of the gene [65], suggesting that the amount of available HLA class I molecules can also play a role in the efficacy of the immune response.

**Killer cell immunoglobulin-like receptors**

In addition to their importance for acquired immunity processes, HLA molecules are also ligands for the killer cell immunoglobulin-like receptors (KIR). The KIR receptors are preferentially expressed at the surface of natural killer (NK) cells and regulate their activation status through inhibitory or activating signalling. NK cells represent an essential innate immune defence mechanism against viruses, being able not only to kill infected cells but also to produce cytokines.

Certain HLA–KIR combinations have epistatic influences on the outcome of HIV-1 infection [79]: KIR3DL1 and KIR3DS1, which are expressed as allelic variants of the same locus on chromosome 19, have both been shown to protect against disease progression when found in combination with HLA-B molecules that have a Bw4 serological specificity. Various combinations of inhibitory KIR3DL1 alleles and HLA Bw4 molecules have been associated with lower HIV-1 viraemia and slower disease progression [80]. By contrast, the activating allele KIR3DS1 has been associated with lower viraemia and a delayed progression to AIDS when found alone [81] or in combination with HLA Bw4 molecules that have an isolateucine at position 80 (Bw4-80I) [82,83]. Functionally, KIR3DS1...
has been shown to correlate with strong inhibition of HIV-1 replication [84] and with higher NK cell effector functions in early HIV-1 disease [85]. Recent studies also suggest that KIR3DL1 and KIR3DS1 could also play a role in differential susceptibility to HIV-1 infection [86,87].

Other immune-related molecules
All human genes that are related to immunity or inflammation can be suspected to play a role in HIV-1 pathogenesis. Many of them have been the subject of genetic studies. For example, polymorphisms in genes coding for several cytokines and cytokine receptors, which are key regulators of the inflammatory homeostasis, have been associated with both resistance to HIV-1 infection and progression of the disease. Yet, in the absence of conclusive replication, most of the associations published so far remain controversial.

Defensins are small peptides produced mainly by epithelial cells to help fight pathogens, including HIV-1. A variant located in the 5′ region of the β-defensin 1 gene (DEFB1) has been associated with higher level of HIV-1 RNA in breast milk [88] and with an increased risk of maternal–fetal transmission of the infection [89,90].

The mannose-binding lectin 2 protein, encoded by the MBL2 gene, is an important element of the innate immune system that is capable of activating the classical complement pathway. MBL2 deficiencies have been associated with susceptibility to autoimmune and infectious diseases. Several non-synonymous coding variants have been associated with lower serum levels of MBL2, as well as increased susceptibility to HIV-1 infection and/or accelerated disease progression [91–93].

DC-SIGN, or CD209, is a gene encoding a transmembrane C-type lectin with an extracellular portion composed of a tandem repeat region (‘neck region’) and a carbohydrate recognition domain, which is important for pathogen binding. Neck region alleles with <5 repeat units were recently shown to associate with HIV-1 resistance in individuals of Chinese ancestry [96]. Conflicting results have also been reported regarding associations between DC-SIGN promoter polymorphisms and HIV-1 outcomes [97,98].

Finally, recent evidence that HIV-1-derived products can activate immune cells through Toll-like receptor (TLR) signalling led to the description of genetic variants in TLR genes that associate with different rates of disease progression [99–101].

Conclusions
The mission of HIV-1 host genetic studies is to describe as comprehensively as possible human genetic influences on HIV-1 control. Although a number of genetic factors implicated in HIV-1 pathogenesis have been described over the past 15 years, progress is arguably slow mainly because of the number of studies with equivocal or controversial results. Technical limitations to genetic studies have now largely been overcome, but there are still too many flaws in HIV-1 host genetic studies that could and should be addressed [102]: clinical outcomes and biological phenotypes should be very precisely defined, the use of ancestry markers to correct for population stratification should now be the rule, and large cohorts of ethnically diverse origin should be analysed to decrease the likelihood of false-positive findings and to increase the coverage of human diversity. The goal is to provide the HIV-1 research community with complete, powered and definitive studies on clear end points.

To make genetic studies easier to design and quicker to run, a crucial step would be to include in the routine of all prospective cohorts a request for genetic consent, and also to obtain the permission to perform additional analyses as defined by local or national guidelines (for example, waivers for deceased patients or authorization to study cohorts under anonymity).

Genome-wide association studies have recently become the approach of choice to search for host factors involved in HIV-1-related outcomes and to assess their relative contribution [65–67]. They notably demonstrated the primary role of the MHC region in HIV-1 control and the intricacy of the genetic effects in the MHC. So far, studies have focused on viral load and progression phenotypes in infected individuals of European ancestry. A much larger effort is needed to address the question of genetic resistance to infection and, essentially, to target other ethnic groups. Still, genome-wide association studies face significant challenges and limitations [103,104]: they are only designed to detect common polymorphisms (generally >5% minor allele frequency) with effect sizes that are large enough to create significant associations after correction for multiple testing. More common variants with smaller population effect sizes could be identified through a continuous increase in the number of participants included in genome-wide studies. However, because rare genetic variants also play a role in the interindividual variability of HIV-1 phenotypes, targeted or whole-genome resequencing strategies will prove to be essential in order to better appreciate the global contribution of the human genome to HIV-1 control.

The ongoing historical transition into the genomic era brings new hope that host genetics will contribute substantially to understanding HIV-1 pathogenesis. This will in turn help identify new targets for drug and vaccine development, and potentially lead to the
creatin of predictive tools that would set the stage for personalized HIV-1 medicine.

Acknowledgements

I am grateful to Amalio Telenti and David Goldstein for their advice and careful reading of the manuscript.

Disclosure statement

The author declares no competing interests.

References


