HIV-1 and human genetic variation

Paul J. McLaren1,2 and Jacques Fellay3,4,5

Abstract | Over the past four decades, research on the natural history of HIV infection has described how HIV wreaks havoc on human immunity and causes AIDS. HIV host genomic research, which aims to understand how human genetic variation affects our response to HIV infection, has progressed from early candidate gene studies to recent multi-omic efforts, benefiting from spectacular advances in sequencing technology and data science. In addition to invading cells and co-opting the host machinery for replication, HIV also stably integrates into our own genome. The study of the complex interactions between the human and retroviral genomes has improved our understanding of pathogenic mechanisms and suggested novel preventive and therapeutic approaches against HIV infection.

HIV-1 is the human retrovirus responsible for the HIV/AIDS pandemic, which has claimed more than 30 million lives over the past four decades. HIV infection continues to be a major global public health issue, with currently around 40 million people living with HIV (PLWH). Lifelong antiretroviral therapy (ART) has transformed the disease into a manageable chronic health condition. When available, ART enables PLWH to lead long and healthy lives but there is still no effective vaccine and no cure.

Early in the pandemic, it became clear that the risk of HIV acquisition is highly variable across humans. Socioeconomic and behavioural factors played a central role in this variability with some risk groups, such as intravenous drug users and men who have sex with men (MSM), being disproportionately affected. Still, even among the most highly exposed individuals, a fraction remained HIV negative. Similarly, important differences in the natural course of HIV infection (such as time from infection to AIDS diagnosis and the occurrence of opportunistic infections or malignancies) were only partially explained by known variables such as age and comorbidities. Taken together, these clinical and epidemiological observations suggested a role for additional factors in the modulation of the individual response to HIV, including inherited variation in the genes and pathways involved in the retroviral life cycle and in innate or adaptive immunity against the infection.

HIV enters its main target cell, the CD4+ T lymphocyte, by binding to its receptor CD4 and to the co-receptor CCR5. This binding event triggers the fusion of the viral and human cell membranes, initiating a complex intracellular life cycle that will lead to the production of new viruses. The natural immune response against HIV infection relies mostly on CD8+ T cells, also called cytotoxic T lymphocytes (CTLs). Upon primary infection, intense HIV replication results in a very high plasma viral load, measured as copies of the HIV RNA genome per millilitre of plasma, which is then partly controlled by the specific CD8+ T cell response. The very diverse human leucocyte antigen (HLA) class I molecules play a central role in this immune response by presenting small viral fragments, called epitopes, at the surface of infected cells. The recognition of these epitopes by CTL leads to the elimination of HIV-infected cells. A more efficient immune response is linked to a lower viral load during the chronic phase of an untreated infection and to slower disease progression, though it is unable to eliminate the virus.

As a retrovirus, HIV can be described as a genomic pathogen. Indeed, it not only uses the molecular machinery of the infected cell for replication and dissemination but it also has the remarkable capacity to integrate a DNA copy of its RNA genome into a host cell chromosome. By becoming part of the human genome, HIV can persist in long-term cellular reservoirs for decades, making it extremely challenging to develop therapeutic strategies resulting in complete eradication.

To better fight HIV infection, we must once again consider the old Delphic maxim: ‘know thyself’. Because HIV is an expert at hijacking human cells and immunity, we have no choice but to improve our understanding of our inner machinery, starting with the most fundamental layer of biological information — the human genome. The exploration of human diversity at the DNA level, long hampered by technological limitations, has been fuelled by the development of new and more powerful tools over the past decades. Thanks to progress in our understanding of human genetic diversity, in genotyping and sequencing technology, as well as in bioinformatics and data science, it became possible to search for genetic factors that modulate the individual response to HIV, including the resistance and susceptibility to infection and the natural history of the disease in PLWH.

1National HIV and Retrovirology Laboratory at the JC Wilt Infectious Diseases Research Centre, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada.
2Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, MB, Canada.
3School of Life Sciences, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland.
4Swiss Institute of Bioinformatics, Lausanne, Switzerland.
5Precision Medicine Unit, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland.
6E-mail: jacques.fellay@epfl.ch
https://doi.org/10.1038/s41576-021-00378-0
In this Review, we first present an overview of the technological and conceptual developments that have fuelled HIV host genomic research. We then describe the major genetic factors modulating the natural history of HIV infection — in the HLA class I region and the CCR5 locus. Next, we highlight the recent convergence of human and HIV genomics, which allows longitudinal analyses of host–pathogen genetic interactions. Finally, we explain how genomic knowledge is poised to have a positive impact on PLWH, notably through pharmacogenomic interventions and stratification of care based on polygenic risk scores (PRS) before discussing the short-term and long-term perspectives for translational research and clinical applications of human genomics in the field of HIV.

Discovering host genetic variants
The search for human genetic differences that have an impact on HIV-related outcomes was first motivated by clinical observations, namely the striking variability in individual trajectories of patients in the absence of treatment. It was further propelled by a desire to uncover fundamental physiopathological mechanisms by the careful exploration of genomic variants and their impact on host and viral molecular processes.

Candidate gene studies. In the candidate gene approach, population-level associations are sought between HIV-related phenotypes and specific genetic variants in genes that have been selected based on previous biological knowledge or functional work. The selected variants are typically typed using targeted genotyping assays or Sanger sequencing of the region of interest. This framework was first applied to HIV host genetics in the early 1990s in analyses of allelic variants in genes known or suspected to play a role in HIV pathogenesis or in the antiretroviral immune response. Therefore, genetic associations were reported in two broad categories: genes coding for proteins involved in the HIV life cycle (such as PPIA and TSG101 [REF.11]) and immune-related genes encoding molecules implicated in innate and adaptive immune pathways (such as MBL2 [REF.12]).

Fig. 1 | Schematic representation of the HIV life cycle and the HLA-mediated host response. The viral envelope glycoprotein gp120 binds CD4 and CC-chemokine receptor 5 (CCR5) on the surface of target cells triggering the fusion of the viral and host cell membranes; host genomic studies have implicated genetic variants in CCR5 (purple background) listed as modifiers of infectivity. Reverse transcription of the single-stranded RNA genome into double-stranded DNA (dsDNA) occurs using proteins carried by the infecting virion. Viral dsDNA is trafficked to the nucleus, where it is integrated into the human genome. The transcription of viral dsDNA results in viral gene expression and genome replication. Viral mRNA is translated into polyproteins, which are cleaved by the viral protease. Functional proteins assemble with copies of the viral genome at the cell membrane and mature virions bud from the surface. In parallel, as part of the immune response, viral proteins are digested by the host proteasome and processed through tapasin I and II (orange rectangles) into the Golgi, where the epitopes are loaded in the human leucocyte antigen (HLA) class I molecules. The peptide-loaded HLA protein is trafficked to the cell surface and presented to cytotoxic T lymphocytes (CTLs). Variability in epitope presentation by HLA-B alleles, such as the protective allele B*57:01, and in expression levels of HLA-C and HLA-A modify the response to infection and the set point viral load (red background). TCR, T cell receptor.
Human leucocyte antigen (HLA). A protein, encoded by one of a group of HLA genes, that presents antigens that train the adaptive immune response. HLA genes are highly variable and allelic variants encode proteins that are differentially able to present antigens based on the amino acid sequences in the peptide-binding grooves.

Polygenic risk scores (PRS). Statistics that are calculated by enumerating the number of risk alleles associated with a particular phenotype (often weighted by their population-level effect sizes) that are present in a single individual and comparing the individual’s score to the distribution of risk scores in the population.

Population stratification. Presence of systematic differences in allele frequencies between population subgroups owing to systematic differences in ancestry.

HIV progression. The natural disease course of HIV infection in untreated individuals, characterized by an acute phase, a chronic phase and the development of AIDS. The rate of HIV progression varies dramatically in the infected population.

HIV latency. The long-term persistence of HIV in an integrated but transcriptionally inactive form in the host genome. Because latent HIV resides in memory T cells, it persists indefinitely even in patients on suppressive antiretroviral therapy. This latent reservoir is a major barrier to curing HIV infection.

Set point viral load (spVL). Mean log viral load (HIV RNA copies per milliliter of plasma) measured in an individual with HIV infection during the chronic phase of infection. The spVL varies substantially within a population and correlates with disease progression.

and TLR9 (REF.15) as well as in specific antiretroviral defence mechanisms (such as APOBEC3G17 and TRIMS18). Dozens of genes were tested in multiple cohorts. Unfortunately, as has been the case in the broader field of human genetics, most reported associations turned out to be false positives, notably owing to the small sizes of the studied cohorts, population stratification and the lack of correction for multiple testing. Replication attempts in larger cohorts, where these factors could be better controlled, showed no association for the vast majority of variants16–18. In fact, only two major discoveries remain from the candidate gene era: the protective effect of a homozygous 32 bp deletion in CCR5 (CCR5Δ32) against HIV acquisition19–21 and the modulating effect of HLA alleles on HIV progression for which early studies, largely in MSM of European ancestry in the United States, noted a strong impact of the HLA-B alleles B*57 and B*27 on delaying time to AIDS onset22.

Genome-wide association studies. Advances in genotyping and sequencing technologies progressively transformed human genetic analyses during the first decade of this century. In particular, the commercial availability of genome-wide genotyping arrays marked the beginning of the era of genome-wide association studies (GWAS). The principle of a GWAS is to simultaneously test very large numbers of genetic variants throughout the genome for potential associations with a phenotype of interest15. This truly agnostic approach finally allowed for a more comprehensive exploration of the human genome. To date, most GWAS have been based on the genotyping of single nucleotide polymorphisms (SNPs) followed by imputation, a process that leverages the linkage disequilibrium property of the human genome to statistically infer the genotypes that are not directly measured. This approach allows near-comprehensive testing of common variants (that is, variants with a minor allele frequency of >1%) in most human populations41.

The first GWAS of any infectious disease focused on the level of detectable viral genetic material in the blood of untreated, chronically infected individuals during the period of HIV latency23. This phenotype, known as set point viral load (spVL), was selected because of its relative ease of measurement and its known correlation to the rate of progression to AIDS24 and transmission potential25. The spectrum of interrogated variants was limited by early DNA genotyping arrays, yet genome-wide significant associations were identified in the HLA class I region, the most polymorphic locus in the human genome, known to have a crucial role in the modulation of T cell immunity (see ‘HLA variation in HIV control’, below). These findings were soon validated and expanded by other GWAS performed in independent cohorts, which demonstrated that the genetic architecture of HIV spVL is comparable between the general population of PLWH16,26–30 and a particular group of individuals able to maintain low viral loads for prolonged periods of time in the absence of ART, the so-called HIV controllers17,24. The absence of specific genetic factors explaining the HIV controller phenotype was a disappointment in terms of potential therapeutic development. However, it is consistent with what has been found for many complex human traits and diseases — that individuals at the extremes of the phenotypic distribution are more likely to carry multiple common variants with weak effects rather than rare, high-impact variants32. Beyond genotyping, a single exome sequencing study has been published so far in the HIV field39 also indicating that rare coding variants with large effect sizes are unlikely to make a major contribution to host control of HIV infection.

GWAS were less successful in the search for determinants of HIV resistance, with no definitive evidence found of human genetic polymorphisms conferring an altered susceptibility to HIV apart from CCR5 variation16,34. However, recent genome sequencing studies of extreme exposure phenotypes35 have shown promising associations in CD101, a gene encoding an immunoglobulin superfamily member implicated in regulatory T cell function36, and in UBE2V1, which encodes a ubiquitin-conjugating enzyme involved in pro-inflammatory cytokine expression37,38 that associates with the HIV restriction factor TRIM5α39. Although both CD101 and UBE2V1 are plausible candidates, further functional studies are required to validate their role in HIV susceptibility. Finally, analyses of GWAS data provide evidence for residual heritability owing to additive genetic effects beyond CCR5 (REF.15) and genetic overlap with behavioural and socioeconomic traits39. These results suggest that larger genomic studies of HIV acquisition may identify additional loci that impact susceptibility and warrant further investigation, potentially in large biobanks.

Several intrinsic limitations make it difficult to investigate the genetic mechanisms potentially involved in HIV resistance. For example, sample sizes are usually small (in the tens or hundreds) because studies need to be performed on highly exposed yet uninfected individuals such as patients with haemophilia exposed to HIV through contaminated blood products40, sex workers in hyper-endemic areas41 or serodiscordant couples (stable heterosexual couples where one partner has HIV infection and the other is seronegative for HIV at enrolment)42. Frailty (or survival) bias is a limitation in cross-sectional studies of HIV cohorts with long-term follow-up, as these cohorts are enriched for genetic factors protecting against HIV disease progression. Another limitation is misclassification bias in studies comparing the genomes of patients with HIV infection to unselected controls from the general population, in which most individuals are in fact susceptible to HIV infection43. The identification of additional genetic determinants of individual susceptibility to HIV infection will require increased sample sizes (ideally in the thousands) as well as the use of sequencing approaches to characterize the rare functional variants that are not interrogated in studies based on genotyping arrays.

HLA variation in HIV control. The HLA locus in infectious diseases. The human major histocompatibility complex (MHC) located on chromosome 6 is one of the most genetically diverse loci in the genome4. The extended MHC occupies ~7.6 Mb of
Genetic architecture
Underlying genetic basis of a given trait, in terms of variant number, effect size, allele frequency and interactions.

HIV controllers
A group of people living with HIV whose plasma HIV RNA load is spontaneously maintained at very low levels for several years (usually at least 3–5 years) in the absence of antiretroviral therapy.

Restriction factor
A host cellular protein that participates in antiviral defence by interfering with specific steps of the viral replication cycle.

the human genome and encodes more than 400 genes, many of which are key mediators of the innate and adaptive immune responses. Within this locus, alleles at the classical class I (HLA-A, HLA-B, HLA-C) and class II (HLA-DR, HLA-DQ, HLA-DP) genes have been associated with numerous autoimmune, inflammatory and infectious diseases (reviewed in Refs [41,42]) with recent preprints demonstrating extensive disease associations in large biobanks from multiple populations [43,44]. In the context of infectious disease, class I HLA proteins present endogenous peptides on the surface of infected cells for recognition by CTLs, triggering the development of an adaptive response. As discussed below, the variability in epitope specificity of HLA proteins and expression levels of HLA class I alleles has a dramatic impact on the progression of HIV disease.

Effects of amino acid variability encoded by HLA alleles.
An individual's genotype at class I HLA genes has been consistently demonstrated to be the major host genetic determinant of HIV spVL and rate of disease progression across geographic contexts and ancestries [25,26,47–50]. This observation was put in the genome-wide context by the first GWAS of HIV spVL [25] and HIV controllers [27] that exclusively identified SNPs in strong linkage disequilibrium with classical HLA-B alleles. Although array-based techniques for the genotyping of DNA samples do not allow for the direct resolution of classical HLA alleles, computational methods leveraging linkage disequilibrium structure between SNPs and sequence-based HLA types in reference populations allow for accurate imputation of classical HLA types from GWAS data [51]. The application of this technique to a sample of >6,000 PLWH of European ancestry underscored the dramatic effect of HLA-B*57:01 on reducing viral load, which was, on average, ~0.8 log₁₀ RNA copies/ml lower in individuals carrying this allele [28]. This study also demonstrated strong associations at multiple other classical class I HLA alleles that had a range of effects, from decreasing spVL (B*57:01, B*27:05, B*13:02, B*14:02, C*06:02, C*08:02, C*12:02) to increasing spVL (B*07:02, B*08:01, C*07:01, C*07:02, C*04:01).

To better understand how functional variation in HLA class I proteins can impact HIV spVL, recent studies have tested variable amino acid positions within these proteins to fine-map the classical allele associations. In a GWAS performed by the International HIV Controllers study, this technique was applied to demonstrate that previously identified associations between HIV control and classical HLA alleles such as B*57:01 could be explained by variability across a small number of amino acid positions within the HLA-B protein [29]. The strongest effect was observed at position 97 of the protein, which accommodates six alternative amino acids, including valine, which is unique to B*57 haplotypes. A recent preprint describing the comprehensive analysis of the impact of HLA amino acid polymorphisms on spVL in a multi-ethnic sample of >12,000 PLWH identified three amino acid positions in HLA-B (positions 67, 97 and 156) and one in HLA-A (position 77) as independently associating with spVL [30]. The positions within HLA-B map to classical HLA alleles known to impact spVL, whereas the HLA-A position suggests that HLA-A functions independently of HLA-B. Interestingly, all four positions are located in the peptide-binding groove of the respective HLA protein, supporting the hypothesis that epitope presentation is key for the natural suppression of HIV replication. Furthermore, there was no substantial evidence that the effects of these polymorphic positions differed across ancestry groups, suggesting biological relevance across global contexts.

Several mechanisms of action have been proposed to explain why different alleles of the same HLA gene have differential effects on HIV progression. Studies of epitope specificity have shown that certain protective alleles, including B*57:01 (which uniquely carries valine at position 97) and B*27:05 (which carries the protective cysteine and asparagine residues at positions 67 and 97, respectively), drive compensatory mutations in the HLA genome leading to reduced viral fitness [41–43]. In addition to differential epitope specificity, the CTL effector function induced by epitope presentation has been implicated in HIV control, with CTLs in carriers of some protective HLA alleles exhibiting an enhanced proliferative capacity and more polyfunctional responses [54–56].

Within-host diversity and epitope presentation.
In addition to the impact of specific class I HLA alleles on HIV progression at the population level, the within-host diversity of HLA alleles may be important at the individual level. An early study looking at the impact of allele combinations revealed that maximum heterozygosity at HLA class I genes (that is, individuals carrying two different alleles at all three class I genes) was associated with a reduced time to AIDS [31]. This observation was supported by a GWAS that showed that individuals carrying different HLA alleles at each class I gene had a significantly lower viral load than homozygous individuals, even after accounting for the additive effect at each allele [32]. This heterozygote advantage likely comes from the ability to present multiple HIV epitopes, supporting the hypothesis that the breadth of presentation is beneficial in preventing HIV progression. To further test this hypothesis, a recent in silico study used novel algorithms to predict the binding affinity of all possible 9-mer
peptides in the HIV proteome to HLA proteins encoded by the different class I alleles. Coupling these predicted affinities to clinical and genetic data demonstrated that spVL was negatively correlated with the breadth of the peptide repertoire bound by an individual’s HLA protein isoforms. Moreover, HLA-B isoforms had the largest predicted breadth of epitope recognition and conferred the strongest reduction of viral load (FIG. 2a). However, the quantity of epitopes alone is unlikely to fully explain the protective capacity of an individual’s HLA alleles, as

**HIV-suppressing immune response**

**Protective HLA-B alleles**

**HIV-non-suppressing immune response**

**Non-protective HLA-B alleles**

**miR-148a non-binding HLA-C alleles**

**miR-148a binding HLA-C alleles**

**Low HLA-A-expressing and HLA-E-expressing alleles**

**High HLA-A-expressing and HLA-E-expressing alleles**

- Polymorphic CTL recognizing more diverse, structurally conserved epitopes
- Monofunctional CTL recognizing less diverse, less structurally conserved epitopes
- HLA-B (protective)
- HLA-C
- HLA-B (non-protective)
- HLA-E
subsets of epitopes that are uniquely presented by protective HLA isoforms explained more of the observed variance in spVL than the entire predicted set. This observation is further supported by an in silico and functional study that demonstrated that HIV epitopes that encode structurally important residues are preferentially targeted by protective HLA isoforms and associate with elite control of replication64. Thus, the quantity and quality of HIV epitopes presented by combinations of HLA isoforms are the key drivers of spVL.

**Non-classical effects of HLA variation.** In addition to the classical effects of HLA genes on peptide presentation, several studies have suggested that non-classical effects may play a part in limiting HIV replication in vivo. In particular, the variable expression levels of classical HLA-C alleles have been linked to HIV control, with those expressed at high levels conferring protection against disease progression62. This effect has been observed across ancestries and has been linked to the absence of a variable microRNA-148a (miR-148a) binding site in the 3′ untranslated region of HLA-C64. The proposed model suggests that mRNA from alleles lacking the miR-148a binding site escape suppression by miR-148a; as a consequence, proteins encoded by these alleles and loaded with HIV epitopes are expressed at higher levels on infected cells, allowing for greater rates of detection by CTLs4. Similarly, proteins encoded by HLA-A alleles are also expressed at variable levels on the cell surface65. However, in contrast to HLA-C, HLA-A alleles expressed at high levels associate with poorer control of viral replication and with faster disease progression65. A combination of genetic and functional studies indicated that increased HLA-A expression levels correlated with higher viraemia in a combined cohort of more than 9,000 PLWH from sub-Saharan Africa and the United States. It was proposed that this effect may be the result of enhanced production of the HLA class I signal peptide that regulates HLA-E expression, a hypothesis that was supported by a correlation between HLA-A expression and HLA-E expression among 58 healthy donors tested66. HLA-E is a ligand for the natural killer group protein 2A (NKG2A) and their interaction results in strong inhibition of natural killer (NK) cell degranulation (Fig. 2c). Thus, the enhanced production of the HLA class I signal peptide in individuals carrying highly expressing HLA-A alleles may lead to enhanced inhibition of immune responses in infected individuals, resulting in poorer clinical outcomes.

Finally, it has also been observed that the combination of HLA genotype and the expression of particular killer cell immunoglobulin-like receptor (KIR) proteins variably modulated HIV disease course66. The KIR proteins are a highly variable set of cell-surface receptors expressed on NK cells (and some T cells) that, when engaged by their cognate receptors, either activate or inhibit NK cell-mediated killing (recently reviewed in REF66). In particular, the combination of the activating KIR3DS1 allele with a set of HLA-B alleles that carry isoleucine in the Bw4 epitope (Bw4-180) is highly associated with HIV control66. Taken together, these results demonstrate the complex interplay between epitope presentation, HLA protein expression and NK inhibition.

### CCR5 variation in HIV infection

**CCR5Δ32 and resistance against HIV infection.** Perhaps the most highly touted example of human genetic variability restricting infectious diseases is the observation that individuals carrying two copies of a loss-of-function variant in the gene encoding the cell receptor CCR5 are highly resistant to infection by HIV. CCR5 is a chemokine receptor expressed on the surface of multiple subsets of monocytes and lymphocytes, including CD4+ T cells, the major HIV target cells. At the earliest stages of infection, the HIV envelope protein gp120 binds CD4 and CCR5 on the cell surface, resulting in fusion of the viral and host cell membranes and in the release of the viral genome into the target cell. The discovery that individuals who carry homozygous loss-of-function alleles at CCR5 are resistant to infection was first made in a group of MSM that were multiply exposed to the virus but remained uninfected67. It was determined that these men all shared a 32-bp deletion in the CCR5 gene (the CCR5Δ32 allele) that leads to the production of a non-functional protein and the absence of functional CCR5 on the cell surface prevents HIV from entering target cells (Fig. 3a). The CCR5Δ32 allele is observed at ~10% frequency in individuals of European ancestry (homozygosity occurs at a frequency of 1%), at a reduced frequency in southern Europeans compared to those in the north and is not observed at an appreciable frequency in other continental populations. Compound heterozygotes (that is, individuals carrying one copy of CCR5Δ32 and a second loss-of-function CCR5 variant) are also resistant to infection, although these individuals are exceedingly rare67. The observation that individuals lacking CCR5 expression are resistant to HIV infection directly led to the development of the antiviral drug Maraviroc, a CCR5 antagonist68, as well as to the world’s first ethically fraught attempt at human embryo engineering69. Perhaps most interestingly, bone marrow transplants between CCR5Δ32 homozygous donors and recipients with HIV infection have resulted in the only two confirmed cases of long-term HIV cure70,71. Although promising, this effect has been difficult to replicate in engineered autologous stem cell models72 and is unlikely to be scalable to the level necessary to stem the pandemic. Additionally, the protection is not absolute, as several confirmed cases of infection in CCR5Δ32 homozygotes have been reported (reviewed in REF73), presumably by viruses that utilize the minor co-receptor CXCR4 or by dual-tropic viruses.

**Associations between CCR5 variation and spontaneous HIV control.** In addition to the impact of homozygosity on preventing infection, individuals with a single CCR5Δ32 copy exhibit lower spVL and delayed disease progression compared to those with two functional copies4,20,74, likely because the reduced levels of CCR5 protein on the cell surface lower the efficacy of HIV entry into target cells (Fig. 3a). The CCR5 locus was also identified in GWAS, first in a study of ~2,500 PLWH in Europe75 and then in an expanded set of 6,300 individuals from across the globe76. However, the CCR5Δ32 allele was not directly assayed on the genotyping
platforms used in these studies, thus only proxy SNPs were identified. In a combined analysis of GWAS data and direct CCR5Δ32 genotyping, it was observed that the CCR5Δ32 allele was not the most strongly associated variant in the region, suggesting that multiple independent genetic effects occur at this locus. Conditional analysis accounting for the effect of CCR5Δ32 showed that an additional marker, rs1015164, a single nucleotide polymorphism downstream of CCR5, affects cell surface levels of CCR5. In homozygous reference (A/A) and heterozygous (A/G) individuals, the surface expression of CCR5 is normal, whereas G/G homozygous individuals have a lower CCR5 surface expression and lower spVLs. b | rs1015164 regulates the expression of an antisense RNA termed CCR5-AS. When CCR5-AS (orange) is expressed at high or intermediate levels in homozygous reference (A/A) and heterozygous (A/G) individuals, respectively, CCR5 mRNA (blue) is protected from RALY-mediated degradation and results in normal levels of surface expression. In G/G homozygous individuals, CCR5-AS expression is diminished and CCR5 mRNA is degraded, resulting in lower surface expression at the cellular level and a lower spVL overall. IncRNA, long non-coding RNA.

**Fig. 3 | CCR5 expression modifies HIV progression.** a | A 32-bp deletion in CCR5 (CCR5Δ32) results in the reduced expression of CC-chemokine receptor 5 (CCR5) on the surface of target cells. Heterozygous individuals exhibit reduced CCR5 expression, lower set point viral loads (spVLs) and slower disease progression. Individuals carrying two defective copies of the CCR5 gene show no surface expression and are highly resistant to HIV infection. Additionally, rs1015164, a single nucleotide polymorphism downstream of CCR5, affects cell surface levels of CCR5. In homozygous reference (A/A) and heterozygous (A/G) individuals, the surface expression of CCR5 is normal, whereas G/G homozygous individuals have a lower CCR5 surface expression and lower spVLs. b | rs1015164 regulates the expression of an antisense RNA termed CCR5-AS. When CCR5-AS (orange) is expressed at high or intermediate levels in homozygous reference (A/A) and heterozygous (A/G) individuals, respectively, CCR5 mRNA (blue) is protected from RALY-mediated degradation and results in normal levels of surface expression. In G/G homozygous individuals, CCR5-AS expression is diminished and CCR5 mRNA is degraded, resulting in lower surface expression at the cellular level and a lower spVL overall. IncRNA, long non-coding RNA.

**RALY-mediated degradation** A mechanism in which the RALY protein binds to the 3′ untranslated region of an mRNA to promote its degradation.

HIV infection is directly influenced by the innate level of CCR5 expression within the infected individual. Whether additional functional polymorphisms in CCR5 have similar effects remains an open question.

**Host and pathogen genetic variation**

**Pathogen sequence variation as an indicator of host genetic pressure.** Most studies performed so far in the field of host genetics focused on clinically defined outcomes such as the susceptibility to infection or disease progression. However, intermediate phenotypes have been shown to be very valuable in identifying subtle genetic association signals that are not always detectable using more complex clinical outcomes. A particularly promising intermediate phenotype, unique by its nature to infectious diseases, is variation in the pathogen genome (Fig. 4). HIV is a highly variable virus that establishes a lifelong infection. Therefore, it represents
an ideal model to search for the potential effects of intra-host selective pressure on a human pathogen. While some of the variants observed in the HIV genomic sequence are present in the transmitted/founder virus, another fraction is acquired during the course of the disease resulting, at least partially, from the selective pressure exerted by the host response to infection. Signs of host-driven selection are clearly visible in the HIV genome. In particular, specific variants have been described in key viral epitopes presented by HLA class I molecules and targeted by CTL responses. Mutations have also been reported in regions targeted by KIR, suggesting the escape from immune pressure by NK cells. A non-negligible fraction of the HIV-1 genome (~12%) is under positive selection but only about half of the positively selected sites map to canonical CD8+ T cell epitopes, indicating that additional host factors could be driving evolution in non-epitope sites.

**Genome-to-genome studies.** Computational approaches developed over the past decade have allowed more comprehensive analyses of the reciprocal genetic signals resulting from the host–pathogen interaction. Joint analyses of human and HIV sequence variation start with the generation of large-scale genomic data from paired samples. The retroviral genome can be
Genome-to-genome (G2G) studies
Methods that test the hypothesis that host genetic variability causes pathogen genetic variability and can indicate novel host restriction factors.

Treatment as prevention
Strategy to prevent the sexual transmission of HIV through the prescription of antiretroviral drugs. It is a highly effective option for preventing HIV transmission. People living with HIV who take antiretroviral drugs and maintain an undetectable viral load have effectively no risk of sexually transmitting the virus to their HIV-negative partners.

Antiretroviral drugs target multiple stages of the HIV life cycle. Commonly used antiretroviral drugs target receptor binding, reverse transcription, integration and protease cleavage. Genetic variation in several human genes (in bold) have been shown to modify drug metabolism and contribute to adverse drug reactions (TABLE 1). CCR5, CC-chemokine receptor 5.

HIV precision medicine
Expanding antiretroviral therapy to eradicate HIV. With the now accepted knowledge that PLWH who do not have detectable plasma viral loads cannot transmit the virus to others, the Joint United Nations Programme on HIV/AIDS (UNAIDS) set an ambitious 90–90–90 target, where 90% of infected people know their status, 90% of those are on antiviral therapy and 90% of those are suppressing the virus below the level of detection. This aspirational treatment target would practically mean, given currently available technologies, that more than 34 million people would be on lifelong chemotherapy. Although this treatment as prevention approach would undoubtedly result in decreases in transmission and dramatic increases in life expectancy for the population with HIV infection, it also requires a deeper understanding of how human genetic variation relates to variability in drug toxicity and response to long-term therapy.

HIV pharmacogenetics. In addition to affecting HIV disease progression in untreated individuals, human genetic variability has also been implicated in modifying the response to treatment. A major achievement in the fight against HIV has been the development of multiple, effective therapeutics that target several stages of the viral life cycle. These include entry inhibitors, which prevent the binding of the viral spike protein gp120 to host cell receptors and fusion of the virus with host cell membranes; nucleoside and non-nucleoside reverse transcriptase inhibitors, which prevent the reverse transcription of the viral RNA genome into DNA; integrase inhibitors, which prevent the integration of the viral DNA product into the host genome; and protease inhibitors, which prevent the cleavage of viral polyproteins into their functional subunits (FIG. 5). For several classes of anti-HIV therapy, human genetic variability is known to influence response to the drug, which in some cases leads to severe adverse events and treatment discontinuation. Paradoxically, the HLA-B allele B*57:01, most notably associated with the control of infection, also predisposes carriers to a severe hypersensitivity reaction to the

---

**Fig. 5 | Antiretroviral drugs target multiple stages of the HIV life cycle.** Commonly used antiretroviral drugs target receptor binding, reverse transcription, integration and protease cleavage. Genetic variation in several human genes (in bold) have been shown to modify drug metabolism and contribute to adverse drug reactions (TABLE 1). CCR5, CC-chemokine receptor 5.
Reviews

Drugs Class Genes Variants Effect on patients

Abacavir NRTI HLA-B rs2395029 (*57:01) Hypersensitivity reaction
Atazanavir PI UGT1A1 rs8175347 (*28,*37) Predisposition to hyperbilirubinemia
Efavirenz NNRTI CYP2A6 rs1801272 (*2), rs5031016 (*7,*10,*19), rs2839943 (*9) Predisposition to high drug plasma levels and treatment discontinuation
Efavirenz and Nevirapine NNRTI CYP2B6 rs3745274 (*6), rs12721655 (*8,*13), rs35303485 (*11), rs36060847 (*12), rs35773040 (*14), rs35979566 (*15), rs28399499 (*16,*18) Predisposition to high drug plasma levels
Etravirine NNRTI CYP2C9, CYP2C19 rs1057910 (*3), rs4424285 (*2) Predisposition to high drug plasma levels
Lopinavir PI ABCC2, CYP3A rs717620 (1), rs6945984 (C) Predisposition to high drug plasma levels
Lopinavir PI SLCO1B1 rs4149056 (*5), rs17329885 (*4) Predisposition to low drug plasma levels

NNRTI, non-nucleoside reverse transcriptase inhibitor; NRTI, nucleoside reverse transcriptase inhibitor; PI, protease inhibitor.

Table 1 | Genetic variants known to affect the pharmacokinetics of anti-HIV drugs

- The high specificity binding of abacavir alters the binding pocket of HLA-B*57:01 and triggers reactivity to self-peptides. Similarly, variants in genes encoding the drug-metabolizing enzymes CYP2B6 (REFS 95–97), CYP2A6 (REF 98), CYP2C9 (REF 99), CYP2C19 (REF 99), CYP3A (REF 100) and ABCC2 (REF 101) have all been associated with slow metabolization kinetics of their cognate drugs (TABLE 1), in some cases leading to drug accumulation in the brain, psychiatric complications and treatment stoppage. The frequency of many of these polymorphisms varies depending on ancestral background, leading to reduced drug tolerance and therefore reduced efficacy in some populations. For example, the allele CYP2B6*6 (rs3745274), which results in the slow metabolism of efavirenz and nevirapine, two non-nucleoside reverse transcriptase inhibitors recommended for first-line use by WHO until recently, has an approximately twofold higher frequency in some African populations compared to Europeans. This increased frequency and the resulting adverse events led to thousands of cases of treatment discontinuation in Zimbabwe when the nation adopted a single-pill efavirenz-containing regimen. This example highlights the need to not only tailor the therapy to the individual but, in some cases, to the population as well. Newer generations of HIV therapies, such as integrase inhibitors and advanced nucleoside reverse transcriptase inhibitors, have more favourable pharmacokinetic and safety profiles. However, the effects of long-term treatment with these drugs and any potential interactions with human genetic variability remain to be understood.

Complex trait genomics in HIV medicine. In addition to the direct interactions between host genotype and drug metabolism, patients on long-term HIV therapy also experience early onset of several chronic diseases, including cardiovascular disease, metabolic syndrome, kidney disease and liver fibrosis. These conditions are all known to have high heritability in the HIV uninfected population and genetic risk factors for type 2 diabetes mellitus and cardiovascular disease have been shown to be enhanced in PLWH on therapy. Recently, there has been a push to develop PRS in the general population. These scores, built by summing the additive effects of dozens to thousands of genetic variants within an individual, have been shown to have a strong predictive ability for multiple metabolic, inflammatory, tumoural and cardiovascular conditions. Investigations of PRS in the specific context of HIV-infected individuals receiving long-term antiretroviral therapy have just begun, with the recent demonstrations that the prediction of chronic kidney disease can be improved through the addition of a PRS to the known clinical and pharmacological risk factors, and that a PRS can be useful to stratify PLWH at a high risk of cardiometabolic diseases who may benefit from preventive therapies. An important caveat is that PRS are not necessarily transferable across ancestral groups and, as in all areas of genomics, attention should be paid to enhancing diversity and ensuring equity in precision medicine approaches.

Conclusion and future perspectives

Host genomic studies have advanced our understanding of HIV biology in several important ways. Firstly, the demonstration of the dominant impact of HLA variation on HIV progression in the context of the whole genome reinforced the need to focus on T-cell responses in vaccine design. Moreover, the ability to accurately infer HLA allele types and protein-level variability from genotyping array data, an approach first piloted in HIV genomic studies, has greatly increased our understanding of how amino acid variability in HLA molecules contributes to multiple medically important traits. Secondly, dense genotyping and large sample sizes enabled the discovery of multiple, independent signals in the CCR5 locus, which provided a deeper understanding of how the expression of CCR5 is regulated and how it modulates HIV infection beyond the known impact of the CCR5Δ32 allele. Finally, amassing genome-wide data for large cohorts of PLWH has enabled the validity of previous candidate gene associations to be assessed, providing a new standard for identifying novel loci of HIV restriction.

In recent years, there have been several barriers to further advancing our understanding of how host genomics affects HIV susceptibility and progression. Firstly, current studies have predominantly included individuals of European ancestry, mirroring the lack of diversity in genomics in general, which is particularly problematic because the vast majority of PLWH are non-White. The example of the population-specific CCR5Δ32 allele further highlights the need to stretch beyond European cohorts to determine if other population-specific effects may exist. Attaining the large sample sizes required for genomic discovery in non-European populations will require a substantial investment of resources and building of capacity in low-income and middle-income
countries. Furthermore, understanding the potential function of genetic variants identified in diverse samples will require a shift towards inclusivity across genom-}
cicates databases. Secondly, with improvements in HIV care and broad adoption of test and treat strategies, the focus of host genomics studies has necessarily shifted away from the natural history of infection phenotypes to intermediate phenotypes, pharmacogenomics of long-term therapy, comorbidities or vaccine response. Thirdly, understanding other classes of genetic variation that are not well captured by genotyping arrays, for example, diversity of KIR alleles and T cell receptor usage, the other partner in the HIVA interaction, should be investigated to better understand how genetic variation in key innate and adaptive immune genes impact disease outcomes. However, capturing these types of variation requires in-depth sequencing to resolve genetic diversity and, in the case of T cell receptor var-}
ination, targeted immune assays to capture the relevant cells. Progress on computational methods for inferring variation at some complex loci from genotyping array data or next-generation sequencing data will greatly aid these efforts.

The full translational potential of host genomics dis-
covery in HIV has yet to be realised. Although the asso-
ciation between HLA allele type, epitope binding and

1. Friedland, G. H. & Klein, R. S. Transmission of the human immunodeficiency virus. *N. Engl. J. Med.* 317, 1135–1137 (1987).
2. Kulkarni, P. S., Butera, S. T. & Duer, A. C. Resistance to HIV-1 infection: lessons learned from studies of highly exposed persistently seronegative (HEPS) individuals. *AIDS Rev.* 5, 87–103 (2003).
3. Horton, R. E., McLaren, P. J., Fox, K., Kimanu, J. & Bail, T. B. Cohorts for the study of HIV-1-exposed but uninfected individuals: benefits and limitations. *J. Infect. Dis.* 202 (Suppl. 5), S577–S581 (2010).
4. Sabin, C. A. & Lundgren, J. D. The natural history of HIV infection. *Curr. Opin. HIV.* AIDS 8, S11–S17 (2013).
5. Chen, B. Molecular mechanism of HIV-1 entry. *Trends Microbiol.* 27, 878–891 (2019).
6. Letvin, N. L. & Walker, B. D. Immunopathogenesis and immunotherapy in AIDS virus infections. *Nat. Med.* 9, 861–866 (2003).
7. Dufour, C., Gantner, P., Fromentin, R. & Chomont, N. The multifaceted nature of HIV latency. *Nat. Rev. AIDS.* 130, 3581–3596 (2020).
8. Claussnitzer, M. et al. A brief history of human disease genetics. *Nature* 577, 179–189 (2020).
9. Narangbhav, V. & Carrington, M. Host genetic variation and HIV disease: from mapping to mechanism. *Immunogenetics* 69, 489–498 (2017).
10. An, P. et al. Regulatory polymorphisms in the cyclophilin A gene, PPIA, accelerate progression to AIDS. *PLoS Pathog.* 3, e88 (2007).
11. Bashirova, A. A. et al. Consistent effects of TSG101 genetic variability on multiple outcomes of exposure to human immunodeficiency virus type 1. *J. Virol.* 80, 6757–6763 (2006).
12. Bonisotto, M. et al. Polymorphisms in the MBL2 promoter correlated with risk of HIV-1 vertical transmission and AIDS progression. *Genes Immun.* 1, 346–348 (2000).
13. Bochud, P.-Y. et al. Polymorphisms in Toll-like receptor 9 influence the clinical course of HIV-1 infection. *AIDS* 21, 441–446 (2007).
14. An, P. et al. APOBEC3 genetic variants and their influence on the progression to AIDS. *J. Virol.* 78, 11070–11076 (2004).
15. Javardbakht, H. et al. Effects of human TRIM5alpha polymorphisms on antiretroviral function and susceptibility to human immunodeficiency virus infection. *Virology* 354, 15–27 (2006).
16. Fellay, J. et al. Common genetic variation and the control of HIV-1 in humans. *PLoS Genet.* 5, e1000791 (2009).
17. International HIV Controllers Study et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 330, 1551–1557 (2010).
18. McLaren, P. J. et al. Association study of common genetic variants and HIV-1 acquisition in 6,500 infected cases and 7,200 controls. *PLoS Pathog.* 9, e1003515 (2013).
19. This GWAS of HIV-1 acquisition showed no evidence for association outside of CCR5 including lack of replication of 22 loci previously claimed to impact acquisition.
20. Dean, M. et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CCRX structural gene. Hemophilia Growth and Development Study. Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort. *ALIVE.* Science 275, 1856–1862 (1996).
21. We believe this study to be the first description of host genetic resistance to an infectious disease observed in the clinic, the basis for a loss function mutation in CCR5 (previously known as CCRX).
22. Huang, Y. et al. The role of a mutant CCRX allele in HIV-1 transmission and disease progression. *Nat. Med.* 2, 1240–1245 (1996).
23. Liu, R. et al. Homozygous defect in HIV-1 co-receptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 86, 367–377 (1996).
24. Kashir, R. A. et al. Influence of combinations of human major histocompatibility complex genes on the course of HIV infection. *Nat. Med.* 2, 405–411 (1996).
25. Tam, V. et al. Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484 (2019).
26. Das, S., Abecasis, G. R. & Browning, B. L. Genotype imputation from large reference panels. *Annu. Rev. Genomics Hum. Genet.* 19, 73–96 (2018).
27. Fellay, J. et al. A whole-genome association study of major determinants for host control of HIV-1. *Science* 317, 946–947 (2007).
28. In what is considered to be the first GWAS of an infectious disease phenotype, the authors demonstrate the large impact of genetic variation in the HLA region on determining HIV spv.
29. Mellors, J. W. et al. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. *Ann. Intern. Med.* 122, 573–579 (1995).
30. Quinn, T. C. et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *N. Engl. J. Med.* 342, 921–929 (2000).
31. Pelik, R. et al. Host determinants of HIV-1 control in African Americans. *J. Infect. Dis.* 201, 1141–1149 (2010).
32. Lingappa, J. R. et al. Genome-wide association study for determinants of HIV-1 acquisition and viral set point in HIV-1 serodiscordant couples with quantified virus exposure. *PLoS ONE* 6, e28652 (2011).
33. Luo, M. et al. A genetic polymorphism of FREM1 is associated with resistance against HIV infection in the Pumani Sex Worker Cohort. *J. Virol.* 86, 11899–11905 (2012).
34. McLaren, P. J. et al. Fine-mapping classical HLA association with durable host control of HIV-1 infection in African Americans. *Hum. Mol. Genet.* 21, 4354–4367 (2012).
35. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* 50, 1219–1224 (2018).
36. McLaren, P. J. et al. Evaluating the impact of functional genetic variation in HIV-1 control. *J. Infect. Dis.* 216, 1065–1069 (2017).
37. Lane, J. et al. A genome-wide association study of resistance to HIV infection in highly exposed uninfected individuals with hemophilia. *A. Hum. Genot.* 22, 1903–1910 (2013).
38. Mackelprang, R. D. et al. Whole genome sequencing of extreme phenotypes identifies a CYP27B1 and CD101 and UBE2V1 associated with increased risk of sexually acquired HIV-1. *PLoS Pathog.* 13, e1006703 (2017).
39. Bouloc, A., Bagot, M., Delaire, S., Bensussan, A. & Boussemol, L. Triggering CD101 molecule on human cutaneous dendritic cells inhibits T cell proliferation via IL-10 production. *Eur. J. Immunol.* 30, 3132–3159 (2000).
40. Xia, Z.-P. et al. Direct activation of protein kinases by unanchored polyubiquitin chains. *Nature* 461, 111–119 (2009).
41. Pertel, T. et al. TRIMs is an innate immune sensor for the retrovirus capsid lattice. *Nature* 472, 361–365 (2011).
42. Powell, T. R. et al. The behavioral, cellular and immune mediators of HIV-1 acquisition: new insights from population genetics. *Sci. Rep.* 10, 3304 (2020).
43. Petrovski, S. et al. Common human genetic variants and HIV-1 susceptibility: a genome-wide study in a
homogeneous African population. AIDS 25, 513–518 (2011).
41. Shina, T., Hosomi, K., Inoko, H. & Kultski, J. K. The HLA genomic loci map: expression, interaction, diversity and disease. J. Hum. Genet. 54, 15–39 (2009).
42. Horton, R. et al. Gene map of the extended human genome. Nature 408, 889–899 (2000).
43. de Bakker, P. I. W. & Raychaudhuri, S. Interrogating the major histocompatibility complex with high-throughput genomics. Hum. Mol. Genet. 21, 1–8 (2012).
44. Kennedy, A. E., Ozbek, U. & Dorak, M. T. What has GWAS done for HIV and disease associations? Int. J. Infect. Dis. 2015; 23: e211 (2017).
45. Venkataraman, G. R. et al. Pervasive additive and gene-gene interactions in HIV disease. PLoS ONE 8, e58651 (2013).
46. Schneidewind, A. et al. Escape from the dominant restriction by variable HIV strains. Front. Immunol. 8, 283 (2017).
47. Arora, J. et al. HIV peptidome-wide association study demonstrates that it can be used to detect viral evolution owing to both innate and adaptive host immune pressure. J. Infect. Dis. 219, 4268–4275 (2015).
48. Ramsuran, V. et al. Epigenetic regulation of differential HLA-A allelic expression levels. Hum. Mol. Genet. 24, 461–478 (2015).
49. Ramsuran, V. et al. Elevated HLA-A expression impairs HIV control through inhibition of NKG2A-expressing cells. Science 359, 85–90 (2018).
50. Migueles, S. A. et al. HLA B*5701 is highly associated with sustained immunologic control of HIV-1 infection in a patient who received a stem cell transplantation. e3000224 (2019).
51. Chaturvedi, N. et al. Adaptation of hepatitis C virus to interferon lambda polymorphism across multiple viral genotypes. eLife 8, e42542 (2019).
52. Rungier, S. et al. The influence of human genetic variation on Epstein-Barr virus sequence diversity. Sci. Rep. 11, 4586 (2021).
53. Eisinger, R. W., Diefenbach, C. W. & Fauci, A. S. HIV viral load and transmission of HIV infection: undetectable equals untransmissible. JAMA 321, 651–652 (2019).
54. Joint United Nations Programme on HIV/AIDS. 90–90–90: an ambitious treatment target to help end AIDS epidemic (UNAIDS, 2016).
55. Lubomirov, R. et al. Association of pharmacogenetic markers with premature discontinuation of first-line anti-HIV therapy: an observational cohort study. Clin. Infect. Dis. 202, 256–263 (2016).
56. Malail, S. et al. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to the protease inhibitor abacavir. Lancet 359, 727–725 (2002).
57. Hetherton, S. et al. Genetic variations in HLA-B region and hypersensitivity to abacavir. Lancet 359, 1121–1122 (2002).
58. Norcross, M. A. et al. Abacavir induces loading of novel self-peptides into HLA-B*57. 01: an autoimmune model for HLA-associated drug hypersensitivities. AIDS 26, F21–F29 (2012).
59. Rutger, M. et al. Predictive value of known and novel alleles of CYP2B6 for efavirenz plasma concentrations in HIV-infected individuals. Clin. Pharmacol. Ther. 81, 557–566 (2007).
60. Aranibar-Amaedde, N. M. et al. Pharmacogenetics-based population pharmacokinetic analysis of efavirenz in HIV-infected individuals. Clin. Pharmacol. Ther. 85, 405–409 (2009).
61. Haas, D. W. et al. Pharmacogenetics of efavirenz and central nervous system side effects. An Adult AIDS Clinical Trials Group study. AIDS 18, 2391–2400 (2004).
62. di Iulo, J. et al. In vivo analysis of efavirenz metabolism in individuals with impaired CYP3A4 function. Pharmacogenet. Genomics 19, 350–359 (2009).
63. Lubomirov, R. et al. Pharmacogenetics-based population pharmacokinetic analysis of etravirine in HIV-infected individuals. Pharmacogenet. Genomics 23, 9–18 (2013).
64. Lubomirov, R. et al. ADME pharmacogenetics: investigation of the pharmacokinetics of the antiretroviral agent lopinavir corelated with ritonavir. Pharmacogenet. Genomics 20, 217–230 (2010).
65. Zanger, U. M. & Klein, K. Pharmacogenomics of cytochrome P450 2B6 (CYP2B6): advances on polymorphisms, mechanisms, and clinical relevance. Front. Genet. 4, 24 (2013).
66. Nordling, L. How the genomics revolution could finally help Africa. Nature 544, 20–22 (2017).
67. Kanter, S. et al. Comparative efficacy and safety of first-line antiretroviral therapy for the treatment of HIV infection: a systematic review and network meta-analysis. Lancet HIV 8, e510–e520 (2016).
68. Currier, J. S. et al. Coronary heart disease in HIV-infected individuals. J. Acquir. Immun. Defic. Syndr. 73, 506–512 (2013).
69. Obel, N. et al. Ischemic heart disease in HIV-infected and HIV-uninfected individuals: a population-based cohort study. Int. J. Cardiol. 44, 1625–1631 (2007).
70. Lang, S. et al. Increased risk of myocardial infarction in HIV-infected patients treated with the art of the general population. AIDS 24, 1228–1230 (2010).
71. Penard, D. et al. Atherogenic dyslipidemia in HIV-infected individuals treated with protease inhibitors. The Swiss HIV cohort study. Circulation 100, 700–705 (1999).
72. Naddeo, G. N. et al. The burden of dialysis-requiring acute kidney injury among hospitalized adults with HIV infection: a nationwide inpatient sample analysis. AIDS 29, 1061–1066 (2015).
110. Scherzer, R. et al. Association of tenofovir exposure with kidney disease risk in HIV infection. *AIDS* **26**, 867–875 (2012).

111. Li, Y. et al. Combination antiretroviral therapy is associated with reduction in liver fibrosis scores in HIV-1-infected subjects. *Medicine* **95**, e2660 (2016).

112. Rotger, M. et al. Impact of single nucleotide polymorphisms and of clinical risk factors on new-onset diabetes mellitus in HIV-infected individuals. *Clin. Infect. Dis.* **51**, 1090–1098 (2010).

113. Rotger, M. et al. Contribution of genetic background, traditional risk factors, and HIV-related factors to coronary artery disease events in HIV-positive persons. *Clin. Infect. Dis.* **57**, 112–121 (2013).

114. Dietrich, L. G. et al. Contribution of genetic background and data collection on adverse events of anti-human immunodeficiency virus (HIV) drugs (D:A:D) clinical risk score to chronic kidney disease in Swiss HIV-infected persons with normal baseline estimated glomerular filtration rate. *Clin. Infect. Dis.* **70**, 890–897 (2020).

115. Dietrich, L. G. et al. Rapid progression of kidney dysfunction in Swiss people living with HIV: contribution of polygenic risk score and D:A:D clinical risk score. *J. Infect. Dis.* **223**, 2145–2153 (2021).

116. Chang, H. et al. Genetic architecture of cardiometabolic risks in people living with HIV. *BMC Med.* **18**, 288 (2020).

117. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).

118. Rotman, D. DNA databases are too white: this man aims to fix that. MIT Technol. Rev. [http://www.technologyreview.com/2018/10/15/139472/dna-databases-are-too-white-this-man-aims-to-fix-that/](http://www.technologyreview.com/2018/10/15/139472/dna-databases-are-too-white-this-man-aims-to-fix-that/) (2018).

119. Vukcevic, D. et al. Imputation of KIR types from SNP variation data. *Am. J. Hum. Genet.* **97**, 593–607 (2015).

120. Kawaguchi, S., Higasa, K., Shimizu, M., Yamada, R. & Matsuda, F. HLA-HD: an accurate HLA typing algorithm for next-generation sequencing data. *Hum. Mutat.* **38**, 788–797 (2017).

121. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT2-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

122. Roe, D. & Kuang, R. Accurate and efficient KIR gene and haplotype inference from genome sequencing reads with novel K-mer signatures. *Front. Immunol.* **11**, 585015 (2020).

123. Sarkar, I., Hauber, I., Hauber, J. & Buchholz, F. HIV-1 proviral DNA excision using an evolved recombinase. *Science* **316**, 1912–1915 (2007).

124. Hu, W. et al. RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc. Natl Acad. Sci. USA* **111**, 11461–11466 (2014).

125. Ge, D. et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).

126. Thomas, D. L. et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* **461**, 798–801 (2009).

127. Thye, T. et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet.* **42**, 739–741 (2010).

128. Malaria Genomic Epidemiology Network. A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature* **526**, 253–257 (2015).

129. Zhang, Q. et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **570**, eabd4570 (2020).

130. Sul, J. H., Martin, L. S. & Eskin, E. Population structure in genetic studies: confounding factors and mixed models. *PLoS Genet.* **14**, e1007309 (2018).

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Publisher’s note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.