HHS Public Access

Author manuscript

AIDS. Author manuscript; available in PMC 2024 September 26.

Published in final edited form as:

AIDS. 2013 November 28; 27(18): 2831–2839. doi:10.1097/01.aids.0000432536.85335.c8.

On stand by: host genetics of HIV control

Mary Carrington[a,b], **Arman A. Bashirova**[a,b], **Paul J. McLaren**[c,d,e]

[a]Cancer and Inflammation Program, Laboratory of Experimental Immunology, SAIC Frederick Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland

[b]Ragon Institute of MGH, MIT and Harvard, Boston, Massachusetts, USA

[c]School of Life Sciences, École Polytechnique Fédérale de Lausanne

[d]Institute of Microbiology, University Hospital Center and University of Lausanne, Lausanne, Switzerland

[e]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

Abstract

The impact of host genetic variation on determining the differential outcomes after HIV infection has been studied by two approaches: targeting of candidate genes and genome-wide association studies (GWASs). The overlap in genetic variants that has been identified by these two means has essentially been restricted to variants near to the human leukocyte antigen (HLA) class I genes, although variation in the *CCR5* locus, which was first shown to have an effect on HIV outcomes using the candidate gene approach, does reach significance genome-wide when very large samples sizes (i.e. thousands) are used in GWAS. Overall, many of the variants identified by the candidate gene approach are likely to be spurious, as no additional variants apart from a novel variant near the *HLA-C* gene have been consistently identified by GWAS. Variants with low frequency and/or low impact on HIV outcomes are likely to exist in the genome and there could be many of them, but these are not identifiable, given current GWAS sample sizes. Several loci centrally involved in the immune response, including the immunoglobulin genes, T-cell receptor loci, or leukocyte receptor complex, are either poorly covered on the GWAS chips or difficult to interpret due to their repetitive nature and/or the presence of insertion/deletion polymorphisms in the region. These loci warrant further interrogation, but genetic characterization of these regions across a range of individuals will first be required. Finally, synergistic interactions between loci may affect outcome after infection, as suggested by associations of specific, functionally relevant HLA and killer cell immunoglobulin-like receptor variants with HIV disease outcomes, and these require further consideration as well.

Correspondence to Mary Carrington, PhD, SAIC-Frederick Inc., Frederick National Laboratory for Cancer Research, PO Box B, Bldg 560 Frederick, MD 21702, USA. fax: +1 301 846 6771; carringm@mail.nih.gov.

Conflicts of interest

There are no conflicts of interest.

Keywords

genome-wide association study; host genetics; human leukocyte antigen

The pre vs. postgenome-wide association study periods of host genetics in HIV disease

A large number of host genetic associations with HIV outcomes have been identified since the beginning of the epidemic, the first involving the human leukocyte antigen (*HLA*) loci [1]. One of the most compelling early studies illustrating a key role of *HLA* variation in HIV control was one in which HIV-positive hemophiliac sibling pairs who shared *HLA* haplotypes were found to be significantly concordant in $CD4^+$ T-cell decline and AIDS status over time, whereas there was no such concordance if the sibs were *HLA* mismatched [2]. Since that time, *HLA* associations with HIV disease outcome have been reported continually [3], and it is certainly the most extensively interrogated region of the human genome in genetic studies of HIV that employ the candidate gene approach. Multiple genome-wide association studies (GWASs) have identified the major histocompatibility complex (MHC) region as the dominant locus in terms of its influence on outcome to HIV infection. Table 1 lists studies published to date in which single nucleotide polymorphisms (SNPs) near *HLA* genes were the most significant genome-wide associations with various outcomes after HIV infection [4–10]. Like that for candidate gene studies, different outcomes were tested in the genome-wide studies, including set-point viral load, progression to $CD4^+$ T-cell count less than 350 cells/ml HIV-RNA and HIV-DNA levels during primary infection, ability to control viral load below 2000 copies\ml of plasma, and maintenance of disease-free status for more than 7–8 years (all in the absence of therapy).

The genome-wide studies consistently pointed to a SNP located within the *HCP5* gene, rs2395029, as the top or near top hit in every study performed in whites. This SNP marks the *HLA-B*57:01* allele nearly perfectly, which is a well documented protective *HLA* allele among this racial group [11–13]. The top genome-wide significant hit in a small cohort of African–Americans [10] was rs2523608, which marks *B*57:03*, an *HLA* allele well known to be protective in Africans. An analysis of individual amino acid residues within the *HLA* loci indicated that the major genetic associations between the *HLA* locus and HIV control were primarily due to polymorphisms in amino acids comprising the HLA-B peptide-binding groove (reviewed in [14]). The most significant amino acid positions associated with control were positions 67, 70, and 97 in HLA-B, which are all highly polymorphic, involved directly in binding peptides within the peptide-binding groove, and in strong linkage disequilibrium with the other variants involved in peptide binding (i.e. marking different *HLA-B* alleles). Thus, although these positions are important in peptide binding, there is no known distinct function of these amino acid positions in HIV control above and beyond their participation in peptide binding. One possible exception to this is a dimorphism at position 245, which showed a significant association with HIV control [15] and is important in the interaction with the CD8 molecule [16].

The rs9264942 SNP, located 35 kb upstream of the *HLA-C* gene (termed '−35'), showed consistent, independent effects in the Euro-CHAVI (Center for HIV/AIDS Immunology) cohort [5,6] and the International HIV Controller Study (IHCS, [10]), and in the IHCS, this SNP was more significant than the variant marking *B*57:01*. The rs9264942 SNP is absent on the earlier chip versions utilized by other studies, but the rs10484554 variant, which is located in close proximity to rs9264942, was identified as a top hit in the analysis of early HIV viral load in the PRIMO/ANRS (Agence Nationale de Recherche sur le Sida) 01 study (Table 1, [4]). It is likely that this SNP is marking the same effect as rs9264942 due to linkage disequilibrium between the two variants ($D' = 1$, $R^2 = 0.31$ based on the 1000 genome data). Many studies have interrogated the functional basis for B*57 protection against HIV (reviewed in [17]), indicating a multifaceted, complex scenario, but given the recent identification of the rs9264942 effect, only a few studies have considered the basis for its protection and those are discussed in the following section.

CCR5 was the second locus shown to have an effect on HIV outcome by the candidate gene approach, with *CCR5*Δ*32* heterozygosity and, subsequently, other variants within or near the locus associating significantly with HIV disease outcomes after infection [18]. It is striking that among the many genetic variants reported to have effects on HIV disease, the first two loci to be identified by the candidate gene approach remain the only two loci to be validated by GWAS [5]. Many of the associations previously reported to affect outcome to HIV infection through the candidate gene approach appear to be type I errors based on results from the agnostic GWAS approach (see a listing of these in the supplementary materials of [10]).

Homozygosity for *CCR5*Δ*32* is the only genetic effect on risk of HIV infection that has been consistently replicated across cohorts, and no other such variantswere identified in a GWAS of 431 high-risk seronegative hemophilia patients when compared with 765 HIV-positive controls [19]. These data indicate that any other genetic effects that protect against infection are either of low penetrance or very low frequency in the human genome.

Human leukocyte antigen-C expression: considering additional aspects of human leukocyte antigen beyond peptide binding

The first HIV GWAS identified a genome-wide significant effect of rs9264942 on HIV set-point viral load that was independent of all other significant hits in the region [6], a finding that was confirmed in subsequent studies comparing HIV controllers to noncontrollers [10,20]. This variant is located 35 kb upstream of the *HLA-C* locus, and it is in strong linkage disequilibrium with the *HLA-C* coding region in people of European descent, but not in those of African ancestry. Its identification was particularly interesting because prior to that time, variation within the *HLA-C* locus was not considered to be much of a driving force in HIV restriction. Compared to *HLA-B* and *HLA-A*, the *HLA-C* locus is less polymorphic [21], has lower expression on the cell surface [22,23], and more extensive ligand-receptor interactions with killer cell immunoglobulin-like receptor (*KIR*) [24]. Individual allelic effects of *HLA-C* are minimal in comparison to those at *HLA-B*, though the linkage disequilibrium between these closely neighboring loci (only 150 kb apart

from one another) muddies the assignment of causation to one locus vs. the other. Unlike HLA-B and HLA-A, HLA-C molecules are not downregulated by HIV Nef [25], which allows 'normal' recognition of infected cells by cytotoxic T lymphocytes (CTLs) restricted by HLA-C molecules [26]. Taken together, these data suggest that *HLA-C* variation has an impact on HIV, but that the mechanism by which this occurs may be distinct from that of *HLA-B* and *HLA-A*.

The rs9264942 variant was found to correlate with levels of *HLA-C* mRNA transcripts [27] and cell surface expression [20]. Additionally, high levels of HLA-C expression were associated with protection against HIV relative to those expressed at low levels [20]. The proposal that expression levels of HLA-C can impact HIV control was challenged by Corrah *et al.* [28], who concluded in their study that the association between rs9264942 and HLA-C expression was due completely to the low expression of the HLA-C*07 allotypes, and that associations between the rs9264942 variant and HIV outcomes reported previously [6,20] were due primarily to specific *HLA-B* alleles (and perhaps *C*07*) that are in linkage disequilibrium with rs9264942. Subsequent studies employing large sample sizes have attempted to determine more definitively whether HLA-C expression varies significantly across the gamut of HLA-C allotypes, the mechanism by which expression levels differ across allotypes, and whether there is a direct effect of HLA-C expression on HIV control that is independent of all individual HLA allelic effects.

The linkage disequilibrium between rs9264942 and the *HLA-C* coding region is very strong in people of European descent, but weak in those of African descent [29], a finding that is consistent with the differential ranges of linkage disequilibrium between variants throughout the rest of the human genome across these two ethnic groups [30]. Thus, the effect of rs9264942 on HIV control in people of European, but not African descent, indicates that rs9264942 is very unlikely to have a direct effect on HLA-C expression, as any potential effect of HLA expression levels would be expected to remain consistent across ethnic groups. A polymorphic microRNA (miRNA) binding site in the 3′UTR of the *HLA-C* gene was subsequently shown to account for some, but not all, of the variation in HLA-C expression across allotypes [31]. miRNAs are short (~22 nucleotides), nonprotein coding RNAs that bind to specific site in the 3′UTR of protein-coding mRNAs, resulting in post-transcriptional destabilization of target mRNAs or translational repression [32–35].

Variation in the miR-148a binding site of the *HLA-C* 3′UTR includes an insertion/deletion polymorphism 263 bp downstream of the HLA-C stop codon (rs67384697), and this variant is in significant linkage disequilibrium with rs9264942 in whites, but not in African–Americans. *HLA-C* alleles that have the intact miR-148a binding site, which includes an insertion at position 263 (263I), have significantly lower cell surface expression of HLA-C than those alleles that have a disrupted miR-148a binding site, which includes a deletion at position 263 (263D) [31]. Differential miR-148a regulation of *HLA-C* does not, however, completely explain the range in HLA-C expression levels across allotypes, which occurs as a continuous gradient [20] rather than three distinct expression groupings depending on the 3′UTR variant (i.e. 263D/D, 283D/I, or 263I/I). Indeed, there are *HLA-C* alleles that contain 263I and are actually expressed at relatively high levels. Thus, additional factors in linkage disequilibrium with the *HLA-C* coding region appear to affect HLA-C expression

in an allotype-specific manner, and factors unlinked to the *HLA-C* locus may also affect expression levels independently of HLA-C allotype, leading to some degree of variation in expression levels of a given HLA-C allotype. Nevertheless, variation in the miR-148a binding site, which is in strong linkage disequilibrium with rs9264942 in whites, accounts for part of the association between rs9264942 and HLA-C expression observed previously.

Variation in the miR-148a binding site of the *HLA-C* 3′UTR associates significantly with HIV control in European Americans, potentially because of its direct effect on HLA-C expression. However, eliminating the possibility that individual *HLA-B* or *HLA-C* alleles account for the association with the 3′UTR variant is difficult in a region where linkage disequilibrium confounds our ability to assign genetic causation. Recent data have investigated the effect of HLA-C expression as a continuous variable on HIV control [29], which is a more appropriate test for an effect of expression than is coarsely dividing individuals into three groups based on a genotype at the rs9264942 or the 3′UTR insertion/deletion variants. Cell surface expression of 16 common HLA-C allotypes was determined in 200 African–Americans and was shown to be very similar to those determined in a smaller group of European Americans [20]. The mean allotypic expression levels were then used to assign total HLA-C expression level for each of more than 5000 HIV-infected individuals. The continuous distribution of *HLA-C* allele expression correlated significantly with HIV outcomes in both whites and African–Americans, in spite of the extensive differences in *HLA* allelic frequencies and linkage disequilibrium patterns across these two ethnic groups. Importantly, these consistent effects were independent of all individual *HLA* allelic effects, including the strong protection conferred by *B**57*.

The genetic findings were then validated through an analysis of viral sequences from nearly 2000 patients in which greater selection pressure on the virus with increasing expression of HLA-C level was observed based on 22 viral escape mutations in 12 epitopes. A similar conclusion was drawn from a previous study in 99 Han Chinese patients, in which six *HLA-C*-associated viral mutations were identified that occurred more commonly among individuals homozygous for high *HLA-C* expression alleles as compared with those homozygous for low expression *HLA-C* alleles (based on the rs9264942 genotype) [36]. An analysis of CTL responses to overlapping peptides spanning the HIV proteome in 1010 African patients added further support to a protective effect conferred by high expression *HLA-C* alleles by demonstrating more efficient induction of responses with increasing expression of *HLA-C* alleles [28]. Taken together, these data strongly point to HLA-C expression as a determinant in HIV outcome after infection, and the possibility that the other *HLA* class I loci have an impact on HIV as a function of expression level is entirely plausible. Thus, GWAS identified for the first time an effect of HLA on HIV disease that does not directly involve the peptide-binding region, but rather another characteristic of HLA-C that affects its function [6]. This finding emphasizes the need to consider more fully different characteristics of HLA molecules that may affect disease outcome and are not a function of individual allelic effects.

Have we exhausted the human leukocyte antigen region in terms of its influence on HIV control?

The *HLA* class I genes are the most polymorphic loci in the human genome, with 2244 *HLA-A* (246 of which are designated as Common and Well Determined, cwd.immunogenomics.org), 2934 *HLA-B* (CWD = 367) and 1788 *HLA-C* (CWD = 146) alleles described in the IMGT-HLA database to date (www.imgt.org). The linkage disequilibrium between these genes is substantial, as mentioned above. *HLA* genes do not rely on specialized mechanisms for generating diversity, such as is found at the immunoglobulin or T-cell receptor (TCR) loci. Rather, their diversification appears to have arisen through point mutations, and interallelic and interlocus recombination (gene conversion) events [37–40], the latter involving very short stretches of nucleotide sequences that often encode multiple polymorphic amino acid positions, resulting in a patchwork pattern of diversity across HLA allotypes [41,42]. The distinct peptide-binding properties of HLA class I molecules depend on diversity throughout the peptide-binding groove and involve the specific combinations of many amino acid variants. Given the evolutionary history of the *HLA* loci in terms of how their polymorphism was generated, it is hardly surprising that most individual *HLA* class I and II alleles are poorly marked by a single SNP in a region that actually has very dense SNP coverage on commercial genotyping arrays compared with comparable 4 Mb regions elsewhere in the genome. *B*57:01* is one of the few exceptions (Table 2), existing in strong linkage disequilibrium with rs2395029 on the Illumina 1M chip (r^2=0.894, in a population of >1700 Europeans). However, no other *HLA* allele known to have a significant independent effect on HIV control is well marked by a SNP present on the 1M Illumina chip, including *B*27* (r^2≤0.481 for the SNP that best marks any B*27 subtype having frequency ≥0.01), *B*58:01* (r^2=0.229), and *B*35* (r^2≤0.406). We can, however, rule out sizeable effects of alleles that have a strong proxy SNP present on the chip along with a relatively high frequency, at least at an odds ratio (OR) of 1.5 or greater with 80% power in a sample size of 500 cases and 1000 controls (Table 2 for those with r^2 >0.09 and Suppl. Table 1, http://links.lww.com/QAD/A385 for r^2 <0.09). Overall, in spite of extensive coverage of *MHC* variation on chips, it cannot replace *HLA* allelic typing for disease association. Imputation of *HLA* alleles based on SNP data has been successful in validating previous *HLA* allelic associations with HIV disease [10], but this method has limitations in that low frequency alleles are poorly predicted.

Like *HLA* class I, most class II loci are poorly marked by single SNPs on the 1M Illumina chip (Table 2, Suppl. Table 1, http://links.lww.com/QAD/A385). Very few consistent effects of *HLA* class II have been identified for the various HIV outcomes that have been tested. Although larger sample sizes are necessary to definitively rule out class II allelic effects, all GWASs to date show no support for a role of *HLA* class II alleles in HIV control. Interestingly, as neither the well described *HLA* class I alleles nor the CCR5 effects on HIV outcomes are easily detectable using single SNPs on GWAS chips, it would be useful to know how many other loci in the genome are also poorly interrogated due to the complexity of the locus and/or inadequate coverage on the chip. For instance, the TCR, immunoglobulin (*Ig*), and *KIR* regions all fall within this category, and all are reasonable contenders in controlling the efficacy of the immune response. However, investigation of SNP coverage

demonstrates that these loci are not nearly as well covered on the chips as the *MHC* (Table 3).

Synergistic interactions between genetic variants in HIV control

The possibility that variation at multiple, biologically related loci may act in concert to provide greater (or lesser) resistance to HIV seems obvious, although very few examples of genetic synergism against the virus have been reported. We define genetic synergism or epistasis as the phenomenon in which a particular combination of variants at two distinct loci confer an effect that is greater than the sum of their individual effects. Genetic synergism between pairs of functionally related loci is difficult to identify definitively because it requires very large cohorts, particularly if the combination of variants in question occurs at low frequency.

Genetic synergism has been proposed for specific combinations of variants at the *KIR* and *HLA* class I loci in HIV disease outcome. These data have been reviewed extensively [43] and will not be covered herein. *KIR* genes encode members of the Ig superfamily of receptors and consist of a group of regulatory molecules that are expressed on natural killer (NK) cells and a subset of T cells, and their ligands are encoded by the highly polymorphic *HLA* class I loci. The *KIR* locus displays extensive diversity in terms of gene content variation, allelic polymorphism, cell surface expression levels across allotypes of a given *KIR* gene, and their variegated expression pattern on NK cells. Any one or more of these characteristics may affect the regulatory capacity of a given *KIR* gene. Another complicating factor in identifying specific *KIR-HLA* genetic effects in human disease is the strong linkage disequilibrium across the *KIR* loci, which are situated in tandem across a distance of only about 150 kb. The same is true for *HLA-B* and *HLA-C*, which encode the greatest number of ligands for KIR. Thus, it is difficult from a genetic approach to determine the pair of *KIR/HLA* variants that are causing the effect being measured. Nevertheless, an interaction between the *KIR3DL1/3DS1* locus and *HLA-B* Bw4 alleles is almost certainly having an effect(s) on HIV disease [44–46], although the precise combinations of alleles at these loci that confer the effects are still in question.

Interactions between gene copy number at the *CCL3L1* locus and variants within the promoter region of its receptor, the *CCR5* gene, have also been proposed [47] and these make sense given the known functional relationship of the molecules encoded by these loci. However, copy number variation at the *CCL3L1* locus is difficult to assess and whether such interactions occur has been disputed [48–51].

The future of HIV host genetics

The extent to which additional host genetic variants, beyond those described above, contribute to HIV control remains an open question. In particular, the role of common variants with small effect sizes, rare functional polymorphisms, and copy number variants (including insertions, deletions and gene rearrangements) needs further study. Recent advances in sequencing technologies have made it possible to assess both common and rare variation in large patient samples. In particular, exome sequencing provides a cost-

effective way to identify host genetic markers in coding regions, many of which are not represented on genotyping arrays. As costs continue to fall, whole-genome sequencing will likely become the preferred methodology in host genetics as, in addition to coding regions assessed by exome sequencing, functional noncoding regions are also expected to influence complex traits [52]. Such an advance in technology would be most usefully applied in African populations that have, as yet, been underrepresented by GWAS and may carry causal variants not observed in European populations [15]. As a compliment to DNA sequencing, techniques such as RNA-Seq, which can provide both host genome variation information and transcript abundance, may become more widely utilized in future studies as functional inference (i.e. gene expression levels and allele-specific expression) can be made at associated loci. Indeed, measuring gene expression levels using microarray technologies has already permitted the genome-wide analysis of infected cells *in vitro* and *in vivo* in HIV-infected individuals [53]. Ultimately, bringing together genome-wide DNA sequence data, RNA transcription levels and proteome data (on both host and pathogen) in an integrated system biology framework should be a key goal for future investigations. In all cases, collaboration across individual groups currently acquiring such data is critical to maximize power for discovery. Future candidate gene studies, particularly those with small sample sizes, need to be interpreted in light of the much larger datasets that have been generated. Groups performing such studies should make every attempt to validate their findings in the currently available datasets to limit reporting of false-positive associations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

M.C. wrote initial draft and directed analyses. A.B. edited the entire review and compiled citations. P.J.M. performed all analyses and contributed to writing the text.

This project has been 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 U.S. Government. This research was supported in part by the Intramural Research Program of the NIH, Frederick National Lab, Center for Cancer Research.

References

1. Just JJ. Genetic predisposition to HIV-1 infection and acquired immune deficiency virus syndrome: a review of the literature examining associations with HLA [corrected]. Hum Immunol 1995; 44:156–169. [PubMed: 8666552]
2. Kroner BL, Goedert JJ, Blattner WA, Wilson SE, Carrington MN, Mann DL. Concordance of human leukocyte antigen haplotype-sharing, CD4 decline and AIDS in hemophilic siblings. Multicenter Hemophilia Cohort and Hemophilia Growth and Development Studies. AIDS 1995; 9:275–280. [PubMed: 7755916]
3. Carrington M, O'Brien SJ. The influence of HLA genotype on AIDS. Annu Rev Med 2003; 54:535–551. [PubMed: 12525683]
4. Dalmasso C, Carpentier W, Meyer L, Rouzioux C, Goujard C, Chaix ML, et al. Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. PLoS One 2008; 3:e3907. [PubMed: 19107206]

5. Fellay J, Ge D, Shianna KV, Colombo S, Ledergerber B, Cirulli ET, et al. Common genetic variation and the control of HIV-1 in humans. PLoS Genet 2009; 5:e1000791. [PubMed: 20041166]
6. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, et al. A whole-genome association study of major determinants for host control of HIV-1. Science 2007; 317:944–947. [PubMed: 17641165]
7. Guergnon J, Dalmasso C, Broet P, Meyer L, Westrop SJ, Imami N, et al. Single-nucleotide polymorphism-defined class I and class III major histocompatibility complex genetic subregions contribute to natural long-term nonprogression in HIV infection. J Infect Dis 2012; 205:718–724. [PubMed: 22238471]
8. Le Clerc S, Coulonges C, Delaneau O, Van Manen D, Herbeck JT, Limou S, et al. Screening low-frequency SNPS from genome-wide association study reveals a new risk allele for progression to AIDS. J Acquir Immune Defic Syndr 2011; 56:279–284. [PubMed: 21107268]
9. Limou S, Le Clerc S, Coulonges C, Carpentier W, Dina C, Delaneau O, et al. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). J Infect Dis 2009; 199:419–426. [PubMed: 19115949]
10. Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 2010; 330:1551–1557. [PubMed: 21051598]
11. Gao X, Bashirova A, Iversen AK, Phair J, Goedert JJ, Buchbinder S, et al. AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. Nat Med 2005; 11:1290–1292. [PubMed: 16288280]
12. Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ, et al. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nat Med 1996; 2:405–411. [PubMed: 8597949]
13. Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, 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 U S A 2000; 97:2709–2714. [PubMed: 10694578]
14. Carrington M, Walker BD. Immunogenetics of spontaneous control of HIV. Annu Rev Med 2012; 63:131–145. [PubMed: 22248321]
15. McLaren PJ, Ripke S, Pelak K, Weintrob AC, Patsopoulos NA, Jia X, et al. Fine-mapping classical HLA variation associated with durable host control of HIV-1 infection in African Americans. Hum Mol Genet 2012; 21:4334–4347. [PubMed: 22718199]
16. Martinez-Naves E, Barber LD, Madrigal JA, Vullo CM, Clayberger C, Lyu SC, et al. Interactions of HLA-B*4801 with peptide and CD8. Tissue Antigens 1997; 50:258–264. [PubMed: 9331948]
17. Goulder PJ, Walker BD. HIV and HLA class I: an evolving relationship. Immunity 2012; 37:426–440. [PubMed: 22999948]
18. Kaslow RA, Dorak T, Tang JJ. Influence of host genetic variation on susceptibility to HIV type 1 infection. J Infect Dis 2005; 191 (Suppl 1):S68–S77. [PubMed: 15630678]
19. Lane J, McLaren PJ, Dorrell L, Shianna KV, Stemke A, Pelak K, et al. A genome-wide association study of resistance to HIV infection in highly exposed uninfected individuals with hemophilia A. Hum Mol Genet 2013.
20. Thomas R, Apps R, Qi Y, Gao X, Male V, O'HUigin C, et al. HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. Nat Genet 2009; 41:1290–1294. [PubMed: 19935663]
21. Zemmour J, Parham P. Distinctive polymorphism at the HLA-C locus: implications for the expression of HLA-C. J Exp Med 1992; 176:937–950. [PubMed: 1383381]
22. Snary D, Barnstable CJ, Bodmer WF, Crumpton MJ. Molecular structure of human histocompatibility antigens: the HLA-C series. Eur J Immunol 1977; 7:580–585. [PubMed: 332508]
23. McCutcheon JA, Gumperz J, Smith KD, Lutz CT, Parham P. Low HLA-C expression at cell surfaces correlates with increased turnover of heavy chain mRNA. J Exp Med 1995; 181:2085–2095. [PubMed: 7760000]

24. Bashirova AA, Martin MP, McVicar DW, Carrington M. The killer immunoglobulin-like receptor gene cluster: tuning the genome for defense. Annu Rev Genomics Hum Genet 2006; 7:277–300. [PubMed: 16824023]
25. Cohen GB, Gandhi RT, Davis DM, Mandelboim O, Chen BK, Strominger JL, et al. The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. Immunity 1999; 10:661–671. [PubMed: 10403641]
26. Yang OO, Nguyen PT, Kalams SA, Dorfman T, Gottlinger HG, Stewart S, et al. Nef-mediated resistance of human immunodeficiency virus type 1 to antiviral cytotoxic T lymphocytes. J Virol 2002; 76:1626–1631. [PubMed: 11799157]
27. Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, et al. Genome-wide associations of gene expression variation in humans. PLoS Genet 2005; 1:e78. [PubMed: 16362079]
28. Corrah TW, Goonetilleke N, Kopycinski J, Deeks SG, Cohen MS, Borrow P, 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 2011; 85:3367–3374. [PubMed: 21248048]
29. Apps R, Qi Y, Carlson JM, Chen H, Gao X, Thomas R, et al. Influence of HLA-C expression level on HIV control. Science 2013; 340:87–91. [PubMed: 23559252]
30. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet 2008; 9:403–433. [PubMed: 18593304]
31. Kulkarni S, Savan R, Qi Y, Gao X, Yuki Y, Bass SE, et al. Differential microRNA regulation of HLA-C expression and its association with HIV control. Nature 2011; 472:495–498. [PubMed: 21499264]
32. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 2005; 433:769–773. [PubMed: 15685193]
33. Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, et al. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science 2006; 312:75–79. [PubMed: 16484454]
34. Wu L, Fan J, Belasco JG. MicroRNAs direct rapid deadenylation of mRNA. Proc Natl Acad Sci U S A 2006; 103:4034–4039. [PubMed: 16495412]
35. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 2010; 466:835–840. [PubMed: 20703300]
36. Blais ME, Zhang Y, Rostron T, Griffin H, Taylor S, Xu K, 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 2012; 188:4663–4670. [PubMed: 22474021]
37. Bender K The HLA-A:HLA-B crossovers and their contribution in analysing possible haplotype-specific recombination rates. Hum Genet 1982; 61:127–134. [PubMed: 6957374]
38. Rose M, Menzel GR, Geserick G, Doxiadis G, Grosse-Wilde H, Mukaida M, et al. Recombination within the class III region by a double cross over event. Tissue Antigens 1988; 31:183–190. [PubMed: 3406933]
39. Bouissou C, Pontarotti P, Crouau-Roy B. A precise meiotic map in the class I region of the human major histocompatibility complex. Genomics 1995; 30:486–492. [PubMed: 8825635]
40. Wu TC, Lichten M. Factors that affect the location and frequency of meiosis-induced double-strand breaks in Saccharomyces cerevisiae. Genetics 1995; 140:55–66. [PubMed: 7635308]
41. Parham P, Lawlor DA, Lomen CE, Ennis PD. Diversity and diversification of HLA-A,B,C alleles. J Immunol 1989; 142:3937–3950. [PubMed: 2715640]
42. Hughes AL, Hughes MK, Watkins DI. Contrasting roles of interallelic recombination at the HLA-A and HLA-B loci. Genetics 1993; 133:669–680. [PubMed: 8454208]
43. Carrington M, Martin MP, van Bergen J. KIR-HLA intercourse in HIV disease. Trends Microbiol 2008; 16:620–627. [PubMed: 18976921]

44. Martin MP, Gao X, Lee JH, Nelson GW, Detels R, Goedert JJ, et al. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nat Genet 2002; 31:429–434. [PubMed: 12134147]
45. Martin MP, Qi Y, Gao X, Yamada E, Martin JN, Pereyra F, et al. Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. Nat Genet 2007; 39:733–740. [PubMed: 17496894]
46. Pelak K, Need AC, Fellay J, Shianna KV, Feng S, Urban TJ, et al. Copy number variation of KIR genes influences HIV-1 control. PLoS Biol 2011; 9:e1001208. [PubMed: 22140359]
47. Ahuja SK, Kulkarni H, Catano G, Agan BK, Camargo JF, He W, et al. CCL3L1-CCR5 genotype influences durability of immune recovery during antiretroviral therapy of HIV-1-infected individuals. Nat Med 2008; 14:413–420. [PubMed: 18376407]
48. Bhattacharya T, Stanton J, Kim EY, Kunstman KJ, Phair JP, Jacobson LP, et al. CCL3L1 and HIV/AIDS susceptibility. Nat Med 2009; 15:1112–1115. [PubMed: 19812561]
49. Field SF, Howson JM, Maier LM, Walker S, Walker NM, Smyth DJ, et al. Experimental aspects of copy number variant assays at CCL3L1. Nat Med 2009; 15:1115–1117. [PubMed: 19812562]
50. He W, Kulkarni H, Castiblanco J, Shimizu C, Aluyen U, Maldonado R, et al. Reply to: 'Experimental aspects of copy number variant assays at CCL3L1'. Nat Med 2009; 15:1117–1120. [PubMed: 19812563]
51. Urban TJ, Weintrob AC, Fellay J, Colombo S, Shianna KV, Gumbs C, et al. CCL3L1 and HIV/AIDS susceptibility. Nat Med 2009; 15:1110–1112. [PubMed: 19812560]
52. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012; 337: 1190–1195. [PubMed: 22955828]
53. Rotger M, Dang KK, Fellay J, Heinzen EL, Feng S, Descombes P, et al. Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. PLoS Pathog 2010; 6:e1000781. [PubMed: 20195503]

Table 1.

Genome-wide association studies on HIV infection outcomes with top signals detected in the *HLA* locus.

Cohort	Outcome	Sample size	Reference
Euro-CHAVI	Set-point VL	486	[6]
	progression to $CD4^+$ T-cell count <350 cells/μl	337	
PRIMO (ANRS 01)	Early plasma HIV-RNA and cellular HIV-DNA level	605 unselected seroconverters	[4]
GRIV (ANRS 02)	AIDS nonprogression	275 slow progressors, 1352 seronegative controls	[9]
Euro-CHAVI	Set-point VL	2362	[5]
	progression to $CD4^+$ T-cell count <350 cells/μl	1071	
International HIV Controller Study	HIV-1 control	974 controllers, 2648 progressors	[10]
GRIV, MACS, ACS (low frequency SNPs only)	AIDS nonprogression	365 slow progressors, 1394 seronegative controls	[8]
GISHEAL, PRIMO	AIDS nonprogression	144 slow progressors, 605 unselected seroconverters	[7]

SNP, single nucleotide polymorphism; VL, viral load.

Table 2.

Tagging single nucleotide polymorphisms for class I and class II *HLA* alleles on the Illumina 1M genotyping chip (r^2>0.090)[a].

HLA	SNP	r^2	HLA_FRQ	SNP_FRQ (minor alleles)	OR_80%
*DQB1*03:02*	rs9275495	0.995	0.24	0.24	1.32
*DRB1*03:01*	rs2854275	0.994	0.23	0.23	1.33
*DRB1*13:02*	rs12660719	0.988	0.03	0.03	2.16
*A*01:01*	rs1611635	0.985	0.15	0.15	1.51
*C*01:02*	rs9366775	0.984	0.03	0.03	2.21
*C*07:02*	rs3132499	0.984	0.12	0.12	1.58
*A*33:01*	rs7749944	0.971	0.01	0.01	3.10
*C*04:01*	rs9461684	0.970	0.09	0.09	1.69
*A*32:01*	rs9259852	0.968	0.02	0.02	2.34
*DPA1*02:01*	rs987870	0.964	0.16	0.16	1.50
*A*03:01*	rs2499	0.961	0.12	0.12	1.60
*B*14:02*	rs2524229	0.957	0.02	0.02	2.82
*DQA1*05:01*	rs3129763	0.957	0.31	0.30	1.24
*DQB1*05:03*	rs3129727	0.957	0.02	0.02	2.56
*DQB1*02:01*	rs2854275	0.937	0.24	0.23	1.33
*B*07:02*	rs2596540	0.923	0.08	0.08	1.76
*C*07:01*	rs2524069	0.915	0.19	0.18	1.45
*B*08:01*	rs3131618	0.907	0.14	0.13	1.55
A 11:01*	rs16895757	0.895	0.06	0.06	1.93
*A*23:01*	rs4281004	0.894	0.02	0.02	2.60
*B*57:01*	rs2395029	0.894	0.02	0.02	2.67
*DQA1*01:01*	rs17533090	0.892	0.11	0.11	1.66
*DPB1*04:02*	rs1367728	0.890	0.09	0.09	1.74
*DQB1*06:04*	rs11752643	0.879	0.03	0.02	2.35
*C*12:02*	rs2270191	0.862	0.01	0.01	3.07
*A*02:01*	rs4947244	0.841	0.26	0.27	
*B*15:01*	rs3763288	0.835	0.08	0.09	
*DQB1*05:01*	rs3830135	0.810	0.09	0.08	

HLA	SNP	r^2	HLA_FRQ	SNP_FRQ (minor alleles)	OR_80%
*C*17:01*	rs9468942	0.794	0.01	0.01	
*C*08:02*	rs2524229	0.786	0.02	0.02	
*B*13:02*	rs12215963	0.785	0.02	0.03	
*DRB1*01:01*	rs3830135	0.784	0.07	0.08	
*DPB1*04:01*	rs3130215	0.781	0.37	0.35	
*DRB1*15:01*	rs3135388	0.777	0.05	0.05	
*A*29:02*	rs2743933	0.774	0.02	0.03	
*DQA1*03:01*	rs660895	0.772	0.31	0.33	
*B*50:01*	rs7742033	0.769	0.02	0.02	
*DQB1*02:02*	rs2858308	0.755	0.06	0.08	
*DPB1*11:01*	rs3130605	0.747	0.01	0.01	
*C*16:01*	rs7747738	0.737	0.02	0.03	
*DPB1*02:01*	rs2071349	0.718	0.14	0.16	
*B*52:01*	rs2270191	0.716	0.01	0.01	
*DQB1*06:02*	rs3135391	0.716	0.05	0.05	
*C*12:03*	rs2233982	0.713	0.05	0.06	
*A*24:02*	rs1150743	0.704	0.12	0.14	
*A*30:01*	rs11961941	0.698	0.01	0.02	
*C*05:01*	rs1049853	0.680	0.08	0.12	
*DPA1*01:03*	rs7905	0.643	0.23	0.16	
*C*06:02*	rs6931921	0.639	0.08	0.10	
DPB1 17:01*	rs16868943	0.633	0.02	0.02	
*DPB1*03:01*	rs3097671	0.629	0.12	0.17	
*DQA1*02:01*	rs2858308	0.610	0.08	0.08	
*B*44:03*	rs2524276	0.607	0.03	0.05	
*DRB1*07:01*	rs2858308	0.595	0.08	0.08	
*C*03:04*	rs1265110	0.586	0.10	0.14	
*B*18:01*	rs1063632	0.564	0.07	0.08	
*DQB1*03:01*	rs9357152	0.551	0.11	0.17	
*A*30:02*	rs6921314	0.548	0.02	0.03	
*B*55:01*	rs3819284	0.539	0.01	0.02	

HLA	SNP	r^2	HLA_FRQ	SNP_FRQ (minor alleles)	OR_80%
*A*26:01*	rs7741100	0.514	0.03	0.06	
*C*15:02*	rs2240066	0.485	0.02	0.04	
*B*27:05*	rs3819299	0.481	0.03	0.06	
*A*68:01*	rs2517922	0.477	0.03	0.05	
*C*02:02*	rs130075	0.436	0.04	0.05	
*A*31:01*	rs7760172	0.425	0.02	0.05	
*DRB1*04:04*	rs3763305	0.418	0.04	0.06	
*DPB1*02:02*	rs7738815	0.413	0.02	0.03	
*DRB1*08:01*	rs2395148	0.411	0.02	0.03	
*B*35:01*	rs130072	0.406	0.05	0.08	
*DQB1*06:01*	rs2859110	0.395	0.01	0.01	
*C*14:02*	rs6903896	0.385	0.01	0.03	
*DRB1*14:01*	rs3129727	0.383	0.01	0.02	
*DRB1*14:04*	rs3129727	0.376	0.01	0.02	
*DRB1*04:01*	rs660895	0.375	0.15	0.33	
*DPB1*14:01*	rs6937061	0.374	0.01	0.04	
*A*25:01*	rs3891157	0.370	0.02	0.05	
*B*56:01*	rs3819284	0.353	0.01	0.02	
*A*33:03*	rs362531	0.350	0.02	0.04	
*DQB1*03:03*	rs28732201	0.350	0.03	0.01	
*DRB1*15:02*	rs2859099	0.346	0.02	0.01	
*DRB1*01:02*	rs9469093	0.344	0.01	0.02	
*DQA1*04:01*	rs2395148	0.343	0.03	0.03	
*A*02:06*	rs7760545	0.339	0.01	0.02	
*DQB1*04:02*	rs2395148	0.330	0.03	0.03	
*B*44:02*	rs4713466	0.329	0.06	0.11	
*DQA1*01:02*	rs3135391	0.329	0.11	0.05	
*DPB1*01:01*	rs429916	0.326	0.08	0.09	
*DRB1*11:01*	rs17208888	0.320	0.03	0.07	
*B*49:01*	rs2524276	0.318	0.02	0.05	
*B*40:01*	rs2596503	0.304	0.05	0.15	

HLA	SNP	r^2	HLA_FRQ	SNP_FRQ (minor alleles)	OR_80%
*C*03:02*	rs2233954	0.301	0.02	0.05	
*DPB1*10:01*	rs6937061	0.291	0.01	0.04	
*C*07:04*	rs16899178	0.254	0.01	0.05	
*A*68:02*	rs11965797	0.253	0.01	0.03	
*B*38:01*	rs2233982	0.241	0.02	0.06	
*C*03:03*	rs2074488	0.236	0.04	0.16	
*B*58:01*	rs11757629	0.229	0.02	0.04	
*DPB1*09:01*	rs6937061	0.222	0.01	0.04	
*DRB1*13:01*	rs3763326	0.220	0.03	0.02	
*DQB1*06:03*	rs3763326	0.212	0.03	0.02	
*B*46:01*	rs2286654	0.202	0.004	0.01	
*B*51:01*	rs2256028	0.197	0.04	0.16	
*B*37:01*	rs13210132	0.194	0.01	0.03	
*B*14:01*	rs3093665	0.193	0.004	0.02	
*C*16:02*	rs7747738	0.176	0.005	0.03	
*DRB1*16:01*	rs17208888	0.169	0.01	0.07	
*B*42:01*	rs9468942	0.168	0.002	0.01	
*DRB1*12:02*	rs7748472	0.162	0.01	0.04	
*B*41:02*	rs9468942	0.155	0.002	0.01	
*B*47:01*	rs7738380	0.155	0.003	0.01	
*DQA1*01:03*	rs3763326	0.150	0.04	0.02	
*A*02:05*	rs6905408	0.141	0.02	0.05	
*B*07:05*	rs34636308	0.137	0.01	0.02	
*DRB1*13:03*	rs7356880	0.137	0.01	0.03	
*DQB1*06:09*	rs12660719	0.136	0.01	0.03	
*B*53:01*	rs1265180	0.133	0.01	0.02	
*DQA1*06:01*	rs2395148	0.133	0.01	0.03	
*DRB1*04:02*	rs483574	0.133	0.03	0.14	
*A*02:02*	rs9295820	0.131	0.003	0.02	
*DPB1*05:01*	rs13437000	0.131	0.03	0.15	
*DRB1*03:02*	rs2071288	0.126	0.002	0.01	

HLA	SNP	r^2	HLA_FRQ	SNP_FRQ (minor alleles)	OR_80%
*DQB1*05:02*	rs3104404	0.122	0.03	0.16	
*A*74:01*	rs9468559	0.121	0.004	0.02	
*DPB1*06:01*	rs3129270	0.120	0.02	0.12	
*DPB1*13:01*	rs2071354	0.118	0.03	0.16	
*B*13:01*	rs9468942	0.116	0.002	0.01	
*B*35:08*	rs3093665	0.115	0.003	0.02	
*A*02:07*	rs2286654	0.113	0.003	0.01	
*B*40:02*	rs3093671	0.112	0.01	0.03	
*B*41:01*	rs2071288	0.112	0.01	0.01	
*B*35:03*	rs130072	0.111	0.02	0.08	
*DPB1*23:01*	rs2229792	0.105	0.004	0.03	
*DRB1*01:03*	rs4151651	0.105	0.01	0.03	
*DRB1*04:05*	rs2621330	0.103	0.03	0.05	
*A*29:01*	rs7772240	0.102	0.003	0.01	
*C*15:05*	rs16899178	0.102	0.01	0.05	
*DRB1*10:01*	rs2227955	0.100	0.01	0.02	
*B*15:17*	rs11538264	0.098	0.005	0.03	
*8*27:02*	rs2516312	0.098	0.003	0.01	
*DR81*09:01*	rs2239688	0.098	0.02	0.01	
*8*39:06*	rs3132499	0.097	0.02	0.12	
*8*44:05*	rs3093561	0.097	0.002	0.01	
*8*40:06*	rs3819299	0.096	0.01	0.06	
*DR81*15:03*	rs6935715	0.091	0.005	0.01	

OR, odds ratio; SNP, single nucleotide polymorphism.

[a]For each human leukocyte antigen (*HLA*) allele, the correlation (r^2) between the classical allele dosage and the minor SNP allele was calculated in a cohort of 1100 individuals of European ancestry with both genome-wide SNP (Illumina 1M) and *HLA* types [10]. For the top 25 alleles, we list the minimum effect on HIV-1 control (odds ratio) necessary to detect the *HLA* allele through the tag SNP. Odds ratio was estimated using an 80% power threshold assuming a sample of 500 cases, 1000 controls, the observed SNP frequency, and genome-wide significance ($P<5 \times 10^{-8}$).

Table 3.

Density of single nucleotide polymorphism coverage on the Illumina 1M in genomic regions important for HIV control.

Region	Chromosome	Mb	SNPs total	SNPs per 1Mb
MHC	6	5.0	8959	1792
TCRb	7	0.5	398	796
TCRa	14	0.9	695	772
Ig	14	1.3	133	102
KIR	19	0.2	72	360

Ig, immunoglobulin; *KIR*, killer cell immunoglobulin-like receptor; *MHC*, major histocompatibility complex.